

NIST Special Publication 1206

**Proceedings of the 2015
International Symposium on
Forensic Science Error
Management**

John M. Butler (Ed.)

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.1206>

NIST Special Publication 1206

Proceedings of the 2015 International Symposium on Forensic Science Error Management

John M. Butler (Ed.)
Special Programs Office

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.SP.1206>

December 2016



U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Under Secretary of Commerce for Standards and Technology and Director

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

National Institute of Standards and Technology Special Publication 1206
Natl. Inst. Stand. Technol. Spec. Publ. 1206, 194 pages (December 2016)
CODEN: NSPUE2

This publication is available free of charge from: <https://doi.org/10.6028/NIST.SP.1206>

Proceedings of the International Symposium on Forensic Science Error Management

Held July 21-24, 2015 in Arlington, VA

**FORENSIC SCIENCE
ERROR MANAGEMENT**

**INTERNATIONAL
FORENSICS SYMPOSIUM**

JULY 20-24, 2015 • WASHINGTON, DC



http://www.nist.gov/director/international_forensics_home.cfm

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

*Proceedings prepared and compiled by
Donna Kimball, Grace Hao, Sabrina Springer, and John Butler*

Points of view presented at the symposium and in this document are those of the presenters and authors and do not necessarily represent the official position or policies of the National Institute of Standards and Technology. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

MEETING SUMMARY

A total of 432 registered participants from 11 countries contributed to a successful first-ever international symposium devoted exclusively to the topic of forensic science error management. Over the 3.5-day meeting and across 8 technical tracks and 42 sessions, there were 2 keynote and 10 plenary speakers, 106 oral presentations, 9 panel discussions, and 18 poster presentations. In their keynote address, Brandon Mayfield, a victim of a forensic science error, and Steven Wax, Mr. Mayfield's attorney, providing a gripping tale of the impact that an error in a fingerprint "match" caused Mr. Mayfield and his family. The symposium concluded with a moot court drama and Professor Bill Thompson serving as a "widget identification expert." *Science* magazine (31 July 2015 issue, 349: 462-463) even ran a short summary of the symposium. At a cost of \$178 per person, this conference was a bargain in part because it was subsidized by funds from the NIST Director's Office. Based on feedback received from participants where 97% (of the 126 who responded to the post-conference survey) felt that the meeting should be convened again, a second international symposium on error management is being planned for July 24-28, 2017.

Organizers

Mark Stolorow of the NIST Special Programs Office served as the Symposium Chair. The meeting was organized by **Donna Kimball** from the NIST Director's Office and **Mary Lou Norris** from NIST Conference Planning Services. Abstract submissions were reviewed and selected by a Technical Program Committee consisting of NIST scientists chaired by **Robert Thompson** and **Martin Herman**. Members of this committee included **John Butler**, **Michael Coble**, **Barbara Guttman**, **Marcela Najarro**, **John Roberts**, **Jessica Staymates**, and **Melissa Taylor**.

Session Moderators

Sessions were run by 26 moderators including members of the Technical Program Committee (**Robert Thompson**, **Martin Herman**, **John Butler**, **Mike Coble**, **Barbara Guttman**, **Marcela Najarro**, and **Melissa Taylor**) and other volunteers **Susan Ballou**, **Julie Bitter**, **Rebecca Bucht**, **James Darnell**, **Lynn Garcia**, **Melissa Gische**, **Will Guthrie**, **Peter Higgins**, **John Paul Jones**, **Alex Nelson**, **Nick Paulter**, **Erica Romsos**, **Toni Roberts**, **Meghan Shilling**, **Ed Sisco**, **Becky Steffen**, **Elizabeth Toomer**, **Elaine Wooton**, **Charlotte Word**.

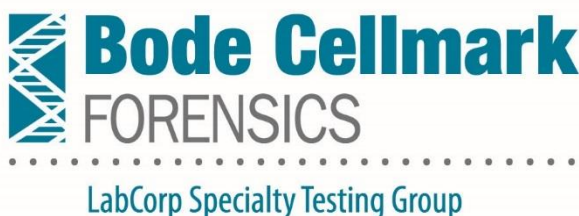
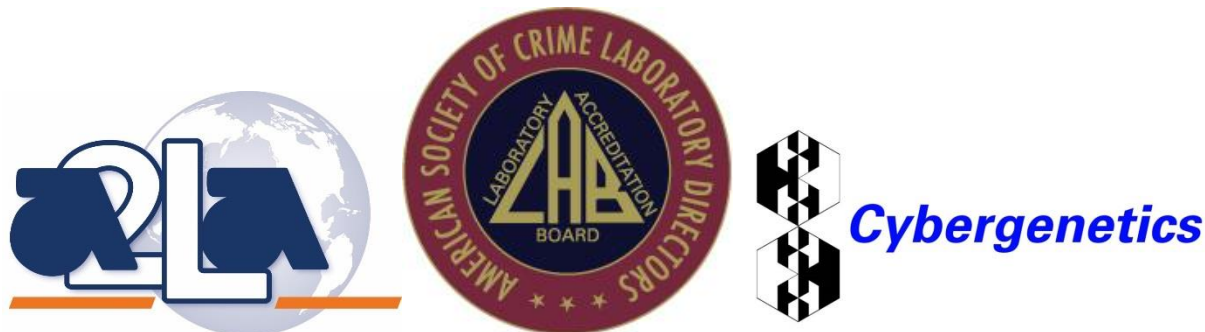
Website

The symposium website is <https://www.nist.gov/director/2015-international-forensics-symposium> and post-symposium information includes links to presentations or abstracts for oral presentations (<https://www.nist.gov/director/orals>) and archived video footage of the keynote and plenary sessions (<https://www.nist.gov/associate-director-laboratory-programs/recorded-sessions>). Hyperlinks on names connect to pdf copies of presentations for which permission was granted to share.

Symposium Proceedings

These proceedings include the presenters' abstracts, bios, photographs (when provided), and links to 85 available pdf presentations containing over 2500 slides. Only a few presenters took advantage of the opportunity to submit a more extensive manuscript after the meeting. These manuscripts are provided following the oral and poster presentation abstracts.

Thank you to our Exhibitors and Sponsors!



Examples of Errors in Forensic Science

<https://www.nist.gov/associate-director-laboratory-programs/technical-tracks-errors>

Analyst/Expert Error

- Errors due to human bias (i.e., cognitive bias, confirmation bias)
- Forensic examiner variability
- Errors due to improperly collected or improperly labeled evidence from crime scenes
- Errors due to break in the chain of custody
- Errors due to contamination and mislabeling of evidence
- Errors due to mishandling (i.e., losing samples, sample mix-ups, sample mislabeling and sample contamination)
- Errors due to misinterpretation of evidence
- Errors due to misinterpretation of data
- Errors in poorly following best practices, processes and methods
- Errors due to poor documentation and transcriptions
- Errors due to inadequately trained personnel
- Errors due to analyst incompetence
- Errors due to failure to review the analysis of the original analyst
- Errors due to misinterpretation of post-mortem artifacts (i.e., artifacts due to resuscitation, exhumation, decomposition, embalming, rigor mortis, toxicological, environmental)
- Measurement errors (i.e., systematic and random)

Fraud

- Errors due to examiner fraud
- Errors due to falsified reports
- Errors due to suppression of exculpatory evidence
- Errors due to exaggeration of test results
- Errors due to false testimony about test results

Methods/protocol error

- Errors due to unvalidated methods
- Errors due to methods without scientific underpinnings
- Errors due to inaccurate and misleading statistics
- Error rates in scientific techniques
- Measurement errors (i.e., systematic and random)

Instrumentation/Technology Limitations

- Errors in software packages
- Error rates in technology solutions
- Laboratory equipment errors (i.e., poor or no calibrations)
- Errors due to deficiencies in laboratory reference materials
- Measurement errors (i.e., systematic and random)

Technical Tracks

<https://www.nist.gov/associate-director-laboratory-programs/technical-tracks>



Death Investigation

Death investigators face the challenge of learning from the decedent only by detective work and scientific investigation. Identifying the cause, mechanism, and manner of death involves certain well-tested practices but is not free of controversy over some forensic pathology protocols, like shaken baby cases or uncertainty measurement of time of death. Learn causes and means to reduce error in death investigation.



Crime Scene

Detecting, recording, collecting, packaging, transporting, and preserving forensic science evidence to reconstruct criminal events is becoming ever more technologically advanced. Tools for video recording crime scenes, recording 3-D impression evidence, bloodspatter interpretation and collecting traces of biological and other evidence below the limits of detection by the human eye demand more training and expertise by crime scene scientists. Learn what challenges and potential errors occur with crime scene investigation and how to reduce them.



Human Factors

Forensic scientists have only recently come to recognize that their close relationships with prosecutors, defense attorneys and law enforcement officers that was traditionally regarded as vital to investigative teamwork might also provide unexpected bias from scientifically irrelevant but subconsciously influential information. What is the right balance of information necessary for forensic testing and sequential masking of unnecessary information? All human activities involve subconscious behaviors – learn how to identify and reduce cognitive, contextual and confirmation bias in actual casework.



Digital Evidence

Error management for digital and multimedia evidence – from mobile phones to digital images to social media analysis – presents some unique concerns. The digital world moves extremely rapidly and a digital forensic practitioner is likely to encounter novel situations. How do you address error management when you are extracting information from a device you have never seen before, or when the system is still running and information is changing while being extracted? How do you attribute data or actions to a particular user of a computer system when the system shows signs of compromise? How do you answer when asked about an error rate?



Legal Factors

Good lawyering and good scientific analysis of physical evidence share some things in common but objectivity is not one of them. Good scientist present conclusions in court based on objective facts they determine from scientifically valid protocols. Lawyers sum up those facts and their job is to influence the jury through advocacy. The intersection of science and the law presents challenges. Questions arise about admissibility, disclosure and defining limitations in reporting and expert testimony. Lawyers and scientists don't always agree how to proceed. Ramification of forensic science evidence in criminal proceedings – ethical and unethical expert testimony and lawyer conduct can impact appellate review. What can we learn from recent appellate rulings to improve forensic science practice and the law?



Quality Assurance

Forensic scientists can make errors. What can the quality assurance manager do in the laboratory to reduce the incidences of error, to measure the severity of the error and take appropriate action? How does the scientist get evaluated and corrective action or retraining get appropriately applied to reduce the chance of future incidences? How do clients get informed in a timely and effective manner? Are proficiency tests adequate? Random case reanalysis? Blind proficiency testing? Laboratory accreditation? Forensic science practitioner certification? In-service training requirements? Application of standards and guidelines to ensure consistent quality analysis, report writing and expert forensic science testimony are the tools of the laboratory QA manager. What challenges does the QA manager face every day?



Laboratory Management

Forensic science laboratory managers must be politically astute, effective negotiators, good accountants, good listeners, good detectives, good scientists, good media relations spokespersons, have squeaky clean integrity and be great leaders and role models for their staff, their superiors and their customers. The next unexpected laboratory incident looms just around the corner for the director of every crime laboratory, bar none. Managing to ensure quality, productivity and service to the criminal justice system – requires systems approaches to anticipate and proactively minimize error. What are those systems approaches?



Criminalistics

Forensic serology, DNA, bloodstain patterns, controlled substances, toxicology, hair and fiber, paint, glass, coatings, plastics, metals, soil and dust, arson debris and explosives residue, fingerprints and other friction ridges, footwear, tire treads, firearms and tool marks, gunshot residue, and question documents fall within the scope of criminalistics – that is, the forensic science disciplines traditionally practiced in crime laboratories. How do each of these disciplines detect, measure and reduce the likelihood of errors? What do the protocols include to signal the incidence of an error? What happens when an error is detected? What lessons can be learned? Learn causes and means to reduce error.

Keynote and Plenary Speakers

**Tuesday,
July 21**



**Mark Stolorow
Steven Wax
Brandon Mayfield
Willie May**

**Wednesday,
July 22**



**Mark Stolorow
Gillian Tully
Itiel Dror
Scott Shappell
Willie May**

**Thursday,
July 23**



**Rich Cavanagh
Alastair Ross
Ralph Kleuskens
Bryan Found
Mark Stolorow**

**Friday,
July 24**



**Rich Cavanagh
Lynn Garcia
William Thompson
Mark Stolorow**

Plenary Session Presentations

Presentations (click on name hyperlinks below): <https://www.nist.gov/director/orals>

Recorded Video: <https://www.nist.gov/associate-director-laboratory-programs/recorded-sessions>

Speaker Photos and Bios: <https://www.nist.gov/associate-director-laboratory-programs/plenary-speakers> and <https://www.nist.gov/associate-director-laboratory-programs/invited-keynote-speakers>

Tuesday, July 21, 2015

- [Willie May](#) (Welcoming Remarks)
- Jo Handelsman (A View of Forensic Science from the White House)
- Steven Wax and Brandon Mayfield (**Keynote:** Fingerprint Forensics and the Fourth Amendment in a Time of Terror)

Wednesday, July 22, 2015

- Gillian Tully (Learning from Errors)
- Scott Shappell (Got Sleep? How Human Factors Such as Fatigue Effects Work Performance)
- [Itiel Dror](#) (Cognitive Sources of Error and Ways to Minimize Them)

Thursday, July 23, 2015

- Alastair Ross (The Source of Errors: Systems, Policy and Practice)
- [Bryan Found](#) (To Err Was Forbidden: The Changing Culture of Error Exploration in Forensic Pattern Evidence)
- [Ralph Kleuskens](#) (Forensic Science Quality Management: Errors and Incidents)

Friday, July 24, 2015

- William Thompson (What is the Proper Evidentiary Basis for a Forensic Science Opinion?)
- [Lynn Robitaille Garcia](#) (The Importance of Trust and Collaboration in Tackling Forensic Problems)

Moot Court (*an hour-long session with examples of an unprepared and a prepared witness*) <https://www.nist.gov/associate-director-laboratory-programs/recorded-sessions> (See Day 4 Moot Court video file); Professor Bill Thompson (University of California-Irvine) as a “widget identification expert” and Jason Tulley (Washington DC Public Defender’s Office) as the cross-examining defense attorney

Breakout Sessions

<https://www.nist.gov/director/orals>

Tuesday, July 21

Crime Scene

- Crime Scene Errors and Issues - [H. Lee](#)
- The Trauma Patient in the Hospital Emergency Department: An Unrecognized Crime Scene - [J. Batts](#)
- Avoiding Non-Selective and Destructive Swabbing - [R. Bucht](#)
- 3D Laser Scanner Error Sources - [M. Shilling](#)
- **PANEL** - Organization of Scientific Area Committees (OSAC) Bloodstain Pattern to Address Error Management in the Field of Bloodstain Pattern Analysis - [E. Toomer](#)

Criminalistics I

- The Madrid Fingerprint Error: Root Cause and Procedures Implemented - [M. Gische](#)
- Quantifying False Individualization in Latent Identification - E. Tabassi
- Mending Justice and Learning from Error: NIJ's Sentinel Events Initiative - [T. Feucht](#)
- **PANEL** - Bridging the Gap Between Black Box/White Box Research and Implementation in Latent Print Casework - [T. Roberts](#)
- Policy Decisions in Latent Print Examination Affect Specificity - [C. Hall](#)
- Changes in Latent Print Examinations as a Result of Technical Review - [M. Montooth](#)
- A Forensic Latent Fingerprint Image Quality Metric for Preprocessing Quality Assurance - [H. Guan](#)
- Communicating Measurement Results in the Courtroom: A Matter of Error, Uncertainty & Inference - [T. Vosk](#)

Lab Management

- Guidelines for the Use of Root Cause Analysis to Reduce Error & Improve Quality in Forensic Science Lab - [J. Hollway](#)
- Corrective Measures Taken in Response to DNA Extraction Failures Using a Newly Validated Method - [J. McGrory](#)
- Cultural Aspects of Error Management - [V. Weedn](#)
- Errors in a DNA Testing Laboratory - [C. Word](#)
- Detection and Mitigation of Medical Errors - [T. Schenberg](#)
- The Challenge of Auditing Technical Records Generated During Complex DNA Casework Processes - [R. Askew](#)
- Laboratory Management Techniques and Practices for Creating A Culture Where Quality is Top Priority - [N. Morgan](#)
- Calculating the True Costs of Errors in Forensic Casework - [E. Mayo](#)
- Efficiency Efforts: An Examiner-Friendly Approach to Improving Laboratory Operations - [H. Maynard](#)
- Students in the Forensic Laboratory: Fostering Education While Maintaining Quality - [C. Baigent](#)
- Implementing the Duty to Correct Error in Forensics - [B. Scheck](#)

Quality Assurance

- Considerations on the Efficacy of Accreditation and Laboratory Performance - [E. Steel](#)
- Proficiency Testing Programme in Forensic GSR and Firearm Investigations - [L. Niewohner](#)
- Inaugurating Data Integrity in Forensic Science - [A. Mozayani](#)
- Error, Error Rates, and Uncertainty: What Are They, How Are They Determined, and Do They Facilitate the Discovery of Truth in the Courtroom? - J. Bono/T. Vosk
- Measuring Criminalistic Uncertainty - [J. Kadane](#)
- **PANEL** - The Use of Statistical Methods for Error Management in the Forensic Sciences - [W. Guthrie](#)

Breakout Sessions

<https://www.nist.gov/director/orals>

Wednesday, July 22

Crime Scene

- Forensic Science -The Quality Assurer? - [P. DeForest](#)
- Measurement Errors with Point Clouds - G. Walsh
- Pre-flashover Fire Pattern Repeatability on Gypsum Wallboard - [D. Madrzykowski](#)
- Error Management in Fire and Explosion Investigations - [C. Beyler](#)
- How Fire Investigation Organizations Get in Trouble with Preventable Errors - [C. Connealy/J. Lentini](#)
- **PANEL** - Improving Fire Investigation through Science, Best Practices, and Transparency - [L. Garcia](#)

Criminalistics I

- Arising Issues Implementing Advances in DNA Technology May Tarnish the Gold Standard - [A. Jeanguenat](#)
- NIST Concordance Evaluations to Assist in the Improvement of Commercial STR Multiplexes - [C. Steffen](#)
- Application of a DNA Quantitation Standard for Human Identity Testing - [E. Romsos](#)
- Effect of Drop-in on False Positive and Rank-Order Likelihood Ratios (LR) Calculated for a Mixture of Touch DNA - [C. Hughes](#)
- Error Rates in Probabilistic Genotyping Software for DNA Mixtures in Human Identification: How to Compare - [H. Miller Coyle](#)
- Objective DNA Mixture Information in the Courtroom: Relevance, Reliability and Acceptance - [M. Perlin](#)
- Validating Software for Probabilistic Genotyping - [J. Buckleton](#); Presented by [Simone Gittelson](#)
- Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S. (MIX13) - [M. Coble](#)
- A Large-scale Study of DNA Mixture Interpretation: Inter- and Intra-Laboratory Variability - [R. Aranda IV](#)
- Defining the Limits of Forensic DNA Profile Interpretation: An Assessment of the Information Content Inherent in Complex Mixtures - [K. Inman](#)
- Errors in Interpretation of DNA Profile Data - [C. Word](#)

Criminalistics II

- New Psychoactive Substances (NPS): How to Keep up with the Never Ending Need for New Reference Materials - [M. Collins](#)
- Error Analysis of a Forensic Controlled Substance Case - [H. Harris](#)
- Positive Identification Starting with a Skull Visualized by a 3D Scanner Using Image Superimposition and the 3D Max Software - [D. Humpire Molina](#)
- 3D Fingerprint Targets - [N. Paulter](#)
- Error Propagation in Shape Analyses with or without Landmarks - [S. Ferson](#)
- Strengthening Forensic Opinions through Objective Assessment of Evidentiary Value: A Prospective for Future Directions in Criminalistics - [M. Sigman](#)
- Mitigation of Fire Debris Compound Naming Error through Automation - [M. Sasser](#)
- Evaluation of the Degradation of Lotion Components Due to Age and Exposure - [J. Dake](#)
- Using Statistical Analysis to Assist with Writer Identification - [M. Wakshull](#)

Breakout Sessions

<https://www.nist.gov/director/orals>

Wednesday, July 22

Death Investigation

- Forensic Pathology as a Forensic Science: History, Current Challenges, Improving Quality and Understanding Cognitive Bias - [A. Baker](#)
- Mitigating DNA Identification Errors in Mass Fatality Response Operations Through Rapid DNA Technology - [C. Miles](#)
- Descriptive Frequency Analysis of 110 Bodies Found in Confined Water Spaces and Determination of the Manner of Death, Including Investigative Outcomes - [J. Richardson](#)
- Searching for a Standard: The Impact that Method Selection has on Evidence Recovery in Forensic Archaeological Investigations - [L. Evis](#)
- Forensic Science Will Greatly Improve Digital Evidence Error Management by Utilizing DICOM and SNOMED CT Standards in Dental and Visible Light Images - [A. Casertano](#)
- Medical Examiner Collection of Comprehensive, Objective Medical Evidence for Conducted Electrical Weapons and Their Temporal Relationship to Sudden Arrest - [M. Brave](#)
- Investigating Arrest-Related Deaths - [D. Ross](#)
- Confirmation Bias and Investigation of Arrest-Related Deaths - [S. Karch](#)
- Medical Examiner Mistakes in Tennessee: A Proposed Model for Formal, Comprehensive Death Investigation Peer Review - [A. Hawes](#)

Digital Evidence

- Error Management in Forensic Digital Imaging Should be a Resolution for Throughout the Entire Law Enforcement Community - [D. Witzke](#)
- Error Treatment in Forensic Authorship Attribution - [P. Juola](#)
- Speaker Detection in a Forensic Environment: Recognizing the Limitations, Improving the Science - [A. Martin](#)
- An Overview of Speaker Variability as a Source of Error in Forensic Automatic Speaker Recognition - [F. Kelly](#)
- **PANEL** - Speech Chain Analysis for Speaker Recognition Systems and Likelihood Ratios to Express Score Probabilities Under Competing Hypotheses - [P. Higgins](#)

Human Factors

- Cognitive Human Factors and the Use of Signature Features in Questioned/Known Signature Comparisons - [M. Merlino](#)
- Cognitive Human Factors and Call Accuracy in Limited-Information Signature Identification Tasks - [T. Freeman](#)
- Study on Methods of Quality Control and Evaluation of Fingerprint Identification - S. Liu
- **PANEL** - Human Factors in Pattern Evidence: How the OSAC Physics/Pattern SAC Subcommittees are Considering Measures to Mitigate the Effects of Bias - [M. Gische](#)

Breakout Sessions

<https://www.nist.gov/director/orals>

Thursday, July 23

Criminalistics I

- Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups - [S. Arora](#)
- A Novel Approach for Quantifying the Weight of Fingerprint Evidence - [H. Swofford](#)
- **PANEL** - You Want Us To Do What? Conducting a Retroactive Review of Hair Microscopy Cases as a Matter of Shared Ethical and Professional Duty - [L. Garcia](#)
- **PANEL** - Error Management and Forensic Document Examination - [E. Wooton](#)

Criminalistics II

- Congruent Matching - Theory & Application in Forensic Evidence Identification & Error Rate Estimation - [J. Song](#)
- Assessing Error Rates for Firearm Identification Based on the CMC Methods - D. Ott
- An Analysis of Correlation Scores Between Cartridge Cases - J. Yen
- Measured Error Rates in Cartridge Case Comparisons - [D. Baldwin](#)
- Gunshot Distance Determination, Crystal Ball Reading or Science? - [P. Paradis](#)
- Lead Density on a Target, A Significant Indicator of Firing Distance but is it Reliable? - [E. Gardner](#)
- Quantitative Metrics for Identifying Characteristic GSR Particles - [N. Ritchie](#)
- Uncertainty Associated with the Elemental Analysis and Forensic Comparison of Materials using Laser Ablation Inductively Coupled Plasma Mass Spectrometry - [J. Almirall](#)

Digital Evidence

- Improving Cyber Forensics & Cybersecurity through Block Chain Technology with Truth-based Systems - [K. Zatyko](#)
- Measuring and Mitigating Errors in a Digital Forensics as a Service Environment - [H. van Beek](#)
- Measuring Systematic and Random Error in Digital Forensics - [A. Nelson](#)
- A Logic Based Model for Error Management in Network Forensics Analysis - [A. Singhal](#)
- **PANEL** - Understanding the Relevance of Error Rates in a Digital World - [J. Darnell](#)

Human Factors

- Combating Confirmation Bias: Can Forensic Science Benefit from Importing Eyewitness Identification Procedures? - [J. Kukucka](#)
- The "Six Thinking Hats" Method of Removing Bias from Case Review - [R. Blackledge](#)
- Mitigation of Forensic Error Rates in Evaluating Pattern-Based Handwriting Evidence Using a Lineup Process - [L. Miller](#)
- Development of a Reasonable Minimum Documentation Standard in Latent Print Analysis and Comparison - [H. Eldridge](#)
- A Typology of Under-exploitation of Latent Print Evidence - [S. Cole](#)
- Objective Classification of Fingerprint Image Complexity - [H. Swofford](#)
- Minimizing and Leveraging Bias in Forensic Science - [N. Adams](#)
- High Resolution Heat Imaging Technology - A New Approach in Fingerprint Search & Securing - [N. Kreuter](#)
- Understanding and Detection, Preventing, Mitigating Errors in Stair Fall-Related Personal Injury Litigation Inspections and Other Legally Mandated Inspections - [J. Pauls](#)
- Visual Comparison of Complex Sets of Quantifiable Forensic Data - [D. Sorensen](#)

Breakout Sessions

<https://www.nist.gov/director/orals>

Thursday, July 23

Legal Factors

- Forensic Scientific Error and Omission Within the Legal System - Forensic Science in the Courtroom: Can we communicate better? - [C. Plourd](#)
- Statistical Issues and Reliability of Eyewitness Identification as a Forensic Tool - [K. Kafadar](#)
- Language Barriers - Creating Standard Terminology to Mitigate Forensic Errors in the Courtroom - [J. Gabel Cino](#)
- The Use of Blinded Eyewitness Identification as a Forensic Procedure - [B. Garrett](#)
- Uncertainty in Forensic Science: A Factor Deliberately Down-Played by Forensic Examiners - [C. Spiegelman](#)
- You're Doing a Great Job, But You're Doing It All Wrong - [R. Blackledge](#)
- Recognizing Sentinel Events in Forensic Science to Avoid Errors - [A. Young](#)
- To Err is Human, but How Might We Measure Error Rates in Forensic DNA Testing and What Would These Error Rates Really Mean? - [J. Butler](#)
- The Potential of Blind Collaborative Justice: Testing the Impact of Expert Blinding and Consensus Building on the Validity of Forensic Testimony - [C. Wong](#)
- Garbage In-Gospel Out? Moving Forensic Fire Investigations into the 21st Century - [V. Brannigan](#)

Quality Assurance

- Forensic DNA: The Gold Standard or Fools Gold? - [K. Reich](#)
- The Value of Regular Benchmarking Studies in Forensic Science to Understand Where Errors Can Occur: Lessons Learned from NIST DNA Interlaboratory Studies - [J. Butler](#)
- Automated DNA Examiner Assessment Tool (DEAT): A Software Program to be Utilized in DNA Benchmarking, Training, Proficiency, and Competency Testing - [R. Aranda IV](#)
- Quality Assurance and Quality Control in Breath Alcohol Measurement Programs Using Intoxilyzers Such as the Intoxilyzer 8000C - [S. Krishnan](#)

PANEL DISCUSSIONS

Tuesday, Crime Scene Session

Organization of Scientific Area Committees (OSAC) Bloodstain Pattern to Address Error Management in the Field of Bloodstain Pattern Analysis - [E. Toomer](#) (67 slides)

Bloodstain pattern analysis (BPA) is based on scientific analytical processes and critical reasoning. As with any field, the success of the practitioner is reliant on the quality of the data used for interpretation and the rigorous adherence to quality control mechanisms. This panel, comprised of experts in bloodstain pattern analysis practice, human factors/bias analysis, statistical evaluation and legal oversight, will address the challenges in determining error rates, sources of error, and current and suggested methods to mitigate errors in BPA. The panel will provide a brief status on the field of bloodstain pattern analysis and provide a multi-disciplinary discussion including potential sources of error in bloodstain pattern analysis.

Tuesday, Criminalistics I Session

Bridging the Gap between Black Box/White Box Research and Implementation in Latent Print Casework - [T. Roberts](#)

Since 2011, FBI and Noblis have published several papers related to Black Box/White Box latent print research. Implications or recommendations on latent print business processes can be drawn from research to include blind verification, standardization of minutia mark up, decision thresholds etc. How are the various latent print units dealing with these recommendations? Are these recommendations practical or intangible?

Tuesday, Quality Assurance Session

The Use of Statistical Methods for Error Management in the Forensic Sciences - [W. Guthrie](#)

The forensic sciences, like all other areas of science, must rely on imperfect analytical methods and metrics to assess, quantify, and interpret evidence to answer questions about important events about which the truth needs to be known. The methods used are potentially subject to both random and systematic sources of error that must be minimized through the use of best practices during analysis and whose typical residual magnitudes must be assessed and accurately presented to judges and juries in court. This panel session will present an overview of how statistical methods can be used for error management in the forensic sciences and invite audience discussion and feedback.

Wednesday, Crime Scene Session

Improving Fire Investigation through Science, Best Practices, and Transparency - [L. Garcia](#)

A major challenge for the community nationwide is to ensure current science and best practices are used consistently in fire investigations, regardless of jurisdictional resources. Attendees will learn about the tremendous transformation that occurred in the Texas State Fire Marshal's Office (SFMO) in an attempt to achieve consistency in applying best practices throughout the State.

Wednesday, Human Factors Session

Human Factors in Pattern Evidence: How the OSAC Physics/Pattern SAC Subcommittees are Considering Measures to Mitigate the Effects of Bias - [M. Gische](#)

Examinations in pattern evidence disciplines, such as Firearms and Toolmarks, Footwear and Tire, and Friction Ridge Analysis, can be influenced by various cognitive, contextual, and human factors. The interpretation of pattern evidence requires human judgment, which involves subjective decision-making. This panel will discuss practical considerations for implementing measures to mitigate the effects of various human factors during pattern evidence examinations.

PANEL DISCUSSIONS (cont.)

Wednesday, Digital Evidence Session

Speech Chain Analysis for Speaker Recognition Systems and Likelihood Ratios to Express Score Probabilities Under Competing Hypotheses - [P. Higgins](#) (17 slides)

Speech chain analysis could become an important systems engineering tool for assessing and managing the source of errors in the design of speaker recognition systems. Speaker recognition systems are designed to gather information opportunistically from sometimes uncontrollable environments via available technologies, which place a strong requirement on the analytic tools and personnel since the opportunity to recapture voice samples after the fact is non-existent. By mathematically modeling the speech energy/signal process of the entire speech chain, from the human source to the analysis of the resultant digital data, various speech capture and processing designs can be studied and analyzed to optimize the signal quality at each point and both understand the sensitivity of the comparisons in terms of error sources and work to reduce their impact and thus the total speaker recognition system error rates. This would be analogous to the image chain model for electro optical imaging systems. The application of speech chain analysis to quantify speaker recognition system error management will be discussed in this presentation. The presentation will also discuss modeling being proposed for Friction Ridge error predictions; followed by panelists presenting the pros and cons of Likelihood Ratios to express score probabilities under competing hypotheses.

Thursday, Criminalistics I Session

You Want Us To Do What? Conducting a Retroactive Review of Hair Microscopy Cases as a Matter of Shared Ethical and Professional Duty - [L. Garcia](#)

On July 18, 2013, DOJ announced it would review cases involving hair microscopy analysis, testimony & reports provided by FBI examiners before December 31, 1999. The DOJ recognized some FBI examiners had exceeded the limits of science by overstating conclusions that may be drawn from a positive association between evidentiary hair and a known hair sample. Attendees will learn how the Texas Association of Crime Laboratory Directors responded by publicly acknowledging their shared "ethical and professional duty, as scientists, to take appropriate action if there has been a miscarriage of justice."

Error Management and Forensic Document Examination - [E. Wooton](#)

A panel of members from the OSAC Document Examination subcommittee will provide a summary of research regarding the determinations of error rates for handwriting examination, a discussion of the usefulness of establishing error rates for non-handwriting examinations, as well as reporting/testimony strategies used to mitigate misunderstandings regarding the findings made by document examiners.

Thursday, Digital Evidence Session

Understanding the Relevance of Error Rates in a Digital World - [J. Darnell](#) (11 slides)

Like other forensic disciplines, digital evidence is prone to errors from several sources, including analyst errors, technique limitations, and imperfect software tool implementations, sometimes referred to as systematic errors. However, with proper quality assurance procedures in place, errors of this type can be recognized and potentially mitigated. Another type of error can be described as random in which a process produced error can be evaluated by a statistical rate. Unlike many other forensic disciplines, digital evidence is not purely seeking if two artifacts are from the same source, instead, digital evidence seeks to show or imply actions by an individual. As such, random errors are not necessarily appropriate as an evaluation tool in a digital evidence process. This panel brings together seasoned digital evidence professionals from government, academic, and private organizations in a discussion on how error rates should be addressed when evaluating the confidence of a digital evidence exhibit.

ORAL PRESENTATION ABSTRACTS

Session: Crime Scene / Tuesday, July 21 / 11:15am – 12:15pm

Crime Scene Errors and Issues

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/crime_scene_errors_and_issues-lee-crimescene-2.pdf (>200 slides)

Abstract

Scientific crime scene investigation is a process which not only includes the mechanical aspects of crime scene documentation and physical evidence collection, but also demands and expects more dynamic approaches to the crime scene such as scene survey, scene analysis and scene reconstruction. Physical evidence obtained from the crime scene or other segments of the investigation is often the cornerstone upon which the successful outcome of the case depends. Documentation of the crime scene and the collection of the physical evidence are important aspects of the crime scene investigation and, thus, must be properly performed. However, these activities are purely mechanical tasks. Any officer with minimal training and an open mind could perform these tasks easily. In addition, many police departments have standard protocols for crime scene procedure. As long as the investigator follows the protocols are often considered as met the legal and scientific requirements. In reality, history proved that many cases are never solved or prosecuted due to error in recognition and collection of evidence.. On the other hand, some innocent people are arrested and prosecuted because of misuse of crime scene evidence. Following are some of the important issues in crime scene investigation should be considered in the measurement of the crime scene errors:

- * Failure to secure and maintain the integrity of the crime scene.
- * Legal errors in collection and preservation of physical evidence.
- * Technical errors in crime scene documentation and crime scene security.
- * Technical errors in collection and preservation of physical evidence
- * Failure in recognition and collection of potential evidence
- * Over and under in collection/preservation of physical evidence.
- * Alteration, omission, modification and addition in handling evidence.
- * Failure or over interpretation of the crime scene pattern.
- * Miss leading or wrongful crime scene analysis.
- * Miss leading or wrongful crime scene Reconstruction.

Some of these errors due to careless of the investigator, lack of resources and/or lack of proper training. Other may be resulted from inherited bias, peer pressures and intentional motives. Detection and measurement of these errors are extremely critical but often are difficult. Case examples will be present to illustrate the detection, mitigation aspects of crime scene errors and the difficulties for the measurement of errors will also presented.

Speaker



Dr. Henry C. Lee is one of the world's foremost forensic scientists. Dr. Lee's work has made him a landmark in modern-day forensic sciences. He has been a prominent player in many of the most challenging cases of the last 50 years. Dr. Lee has worked with law enforcement agencies from 46 countries in helping to solve more than 8000 cases. In recent years, his travels have taken him to England, Bosnia, China, Germany, Singapore, Croatia, Brunei, Thailand, Middle East and other locations around the world.

Session: Crime Scene / Tuesday, July 21 / 1:30 – 1:50pm

The Trauma Patient in the Hospital Emergency Department: An Unrecognized Crime Scene

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/batts_abstract.pdf (abstract only)

■ Abstract

According to the FBI's Uniform Crime Reporting Program, in 2013 an estimated 1,163,146 violent crimes occurred nationwide.[1] When a victim of a violent crime survives their injury and is transported to the emergency department (ED) for treatment, it becomes another "primary" crime scene. Historically, the emphasis of emergency personnel has been to treat the patient's injuries without consideration of the associated forensic issues. In the United States and many other countries, the coroner or medical examiner is responsible for the investigation of unnatural or suspicious deaths. With the exception of specific guidelines relating to evidence collection in sexual assault cases, there is no physician or specialty responsible for addressing the forensic needs of living patients who survive their trauma. Because the majority of emergency providers do not receive training in forensic issues, they often make mistakes in regards to the detection, collection and packaging of physical evidence which is transported with the victim to the ED. Smialek[2] noted that due to a lack of standardized protocols for evidence collection, during the provision of patient care, critical evidence may be lost, discarded or inadvertently washed away. In addition, invasive procedures performed in the ED may alter wounds or cause injuries to be confused with events occurring during the resuscitation. Moreover, if wounds are not properly documented at the time of injury, they will heal over time and their appearance will change.[2] The lack of forensic training for physicians also leads to the misinterpretation of wounds. In one study [3], Randall, showed that physicians correctly identified entrance and exit wounds less than half of the time. Dr. Richard Carmona[4] also documented the problems of evidence collection in the ED. He reviewed the records for 100 admissions to the Trauma Service which had the potential for criminal or civil actions, and found poor, improper, or inadequate documentation in 70% of cases. In 38 of these cases, potential evidence was either not secured, not documented, or was discarded. Clinical Forensic Medicine involves the application of forensic medical techniques to living patients.[5] In the ED these techniques include the evaluation and documentation of traumatic injuries and the collection of evidentiary material for possible medico legal use.[5] In the majority of emergency departments, there are no standardized protocols for evidence collection and preservation. In addition, medical providers do not have the necessary training to understand the need to incorporate these protocols into patient care in cases with forensic implications. As a result, the mistakes we make may deny the criminal justice system access to accurate forensic information, and also deny the victim access to short-lived evidence of critical significance needed in subsequent criminal or civil proceedings. The lack of training in Clinical Forensic Medicine needs to be addressed. The well-trained emergency physician, when evaluating wounds in living patients, should apply the same forensic principals that the forensic pathologist applies to the dead. We also believe that the implementation of nationalized standardized clinical forensic protocols will clearly benefit the forensic needs of living victims of crime.

■ Speaker



Dr. Jayne Batts is a Professor of Emergency Medicine and has practiced for the past 25 years. During her tenure at Carolinas Medical Center in Charlotte, NC she was the Medical Director of the Clinical Forensic Medicine and Adult SANE programs. Jayne has co-authored a chapter and multiple articles in the area of Clinical Forensic Medicine and has also lectured extensively to medical personnel as well as law enforcement agencies on the topic of Evidence Collection and Preservation in the Health Care Setting. Due to her dedication and passion for education in the field of Clinical Forensic Medicine, Jayne was awarded the Chief's Award for Excellence in Policing by the Charlotte-Mecklenburg Police Department in 2000 as well as the Exceptional Service in the Public Interest award by the Federal Bureau of Investigation in 2001.

Session: Crime Scene / Tuesday, July 21 / 1:50 – 2:10pm

Avoiding Non-Selective and Destructive Swabbing

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/legislation/fy2015/EAC_2015_Transfer.pdf (20 slides)

■ Abstract

Swabbing is commonly used to sample suspected bloodstains, but is often not the best way and may complicate interpretation and destroy evidence. Careful thought should be given to alternate sampling modes on a case-by-case basis. All too often, areas suspected of containing bloodstains are swabbed recklessly in order to collect samples for DNA analysis. This can result in a more difficult DNA sample to analyze and, more importantly, a significant loss of potentially crucial information from the pattern which is being sampled.

This presentation aims to lobby for the proper visualization and documentation of blood stain evidence prior to sample collection as well as making an informed selection of which samples to target for DNA analysis. The benefits of alternative sample collection methods to swabbing will also be discussed.

Cases where senseless swabbing and improper sample selection have had adverse effects on the case will be presented, as well as some examples of where more appropriate sampling strategies have been successfully applied.

■ Speaker

Rebecca Bucht, Forensic Science Consultant, Finland

Session: Crime Scene / Tuesday, July 21 / 2:10 – 2:30pm

3D Laser Scanner Error Sources

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/shilling_abstract.pdf

(abstract only)

■ Abstract

3D laser scanners have great potential to quickly capture and preserve crime scene information. These systems collect a set of distance (range) measurements to objects within their field of view in equal increments of arc in both the horizontal (elevation) and vertical (azimuth) planes. The resulting data set, called a point cloud, can contain millions of individual data points that represent the surfaces in the scene. Typically, a laser scanner is placed in several locations within a crime scene the individual data sets are stitched together to give a more comprehensive view. NIJ's Forensic Science Technology Working Group, at the FY2014 meeting, recognized the potential of this technology: "many rapidly developing 3-D imaging technologies could enhance forensic science purposes (investigations, analyses and reconstruction)". Because they are capable of quickly capturing and preserving an entire scene, these systems hold real benefits for forensic science. The current method of using measuring tapes or other manual methods to record critical crime scene dimensions relies on the crime scene technician's ability to predict which measurements will be important in the investigation. As the investigation continues, different parts of the crime scene may gain significance. If the scene has been adequately captured using a 3D laser scanner, additional measurements can be made using the archived digital computer 3D representation of the scene. In fact, this technology has begun to be adopted by law enforcement agencies to capture crime scene information. This information is not only used for measurement but also for scene recreation and examination, and creating visuals to be presented in court. This presentation will give symposium participants an understanding of the sources of errors that may be present in data collected by a 3D laser scanner. We will begin by discussing the basic operating principles of 3D laser scanners. This will lead into a deeper discussion of both geometric error sources and errors that may result from the interaction of the laser with different surfaces within the scene. We will present simple test methods that can be used to detect some of the systematic errors in the 3D laser scanner. We will also briefly discuss the more comprehensive set of performance evaluation tests that is being developed with the ASTM E57.02 working group.

■ Speaker

Meghan Shilling, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Tuesday, July 21 / 11:15 – 11:35am

The Madrid Fingerprint Error: Root Cause and Procedures Implemented

Presentation Link:

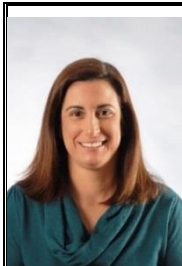
https://www.nist.gov/sites/default/files/documents/2016/11/22/madrid_fingerprint_error.gische.crim1_.pdf

(11 slides)

■ Abstract

This presentation will cover the changes made by the FBI's Latent Print Unit in the eleven years since the erroneous fingerprint identification in the Madrid train bombing case. Through both external and internal investigations, factors contributing to the error were analyzed. As a result, previous cases were reviewed, additional training was provided, and procedures were implemented to mitigate risk. In addition, numerous research studies have been conducted in recent years. The results of these studies provide a better understanding of the limitations associated with latent print examinations and have supported the changes in operations.

■ Speaker



Melissa R. Gische is a Physical Scientist/Forensic Examiner in the Latent Print Operations Unit at the Federal Bureau of Investigation Laboratory in Quantico, VA. She has qualified as an expert witness in the discipline of latent prints and testified in federal and state court. Melissa is Chair of the Organization of Scientific Area Committees (OSAC) Friction Ridge Subcommittee and the International Association for Identification (IAI) Latent Print Identification Subcommittee. Melissa received a Bachelor of Science degree in Psychobiology from the University of California, Los Angeles and a Master of Forensic Sciences degree in Forensic Science from The George Washington University.

Session: Criminalistics I / Tuesday, July 21 / 11:35 – 11:55am

Statistical Friction Ridge Analysis: Quantifying False Individualization in Latent Identification

Presentation Link: not available

■ Abstract

There are two types of error associated with latent print identification in particular and a forensic determination in general: false positive (or Type I or false individualization) and false negative (or Type II or false exclusion). False positive occurs when an individual is reported as the source of an impression when in fact he is not. False negative occurs when an individual is reported as not being the source of an impression when in fact he is. False positives results in false convictions, and false negatives result in letting guilty people go free. A better understanding of causes of false positives could reduce the number of false convictions. This work aims to quantify the probability of false individualizations and its associated uncertainty for latent identifications by investigating key factors that form the uniqueness of friction ridge patterns from general population. Theoretically, false positives can happen if areas of friction ridge skin of two different fingers are sufficiently similar. While the uniqueness of complete and high quality fingerprints are well understood, the evidential contribution of latent impressions that can be partial, distorted, and of poor quality, to our knowledge, has not been characterized or quantified. We examine the concept of 'sufficiently similar' by analyzing the occurrence of indistinguishable minutia configurations between two impressions of different fingers in terms of the quantity and quality of minutiae, the size and the spatial location of the configurations relative to the singular point(s) in a print. This will provide basis to quantify false individualization and to investigate the occurrence of natural confounding friction ridge patterns. We perform this study for three subpopulations: (i) unrelated individuals, (ii) twins, and (iii) same individuals. Our technical approach follows. We deployed a clique problem in graph theory to measure similarity of minutia configurations of two impressions in terms of Euclidean distance, ridge count, and minutiae angles of each minutia pair. A graph between two fingerprints is constructed as follows: take every pair of minutiae from the two fingerprints as a node; connect an edge between nodes if the minutiae pair properties (i.e., Euclidean distance, ridge count, and minutiae angles) are similar within empirical tolerance levels. All possible similar minutiae configurations are obtained by finding cliques in the graph. Two similar minutia configurations are considered indistinguishable, if their ridge structures are highly correlated at the image level. Moreover, we use a state of art commercial matcher to measure similarity of prints from different fingers (that is, impostor or non-mated) at various levels of completeness i.e., different size images with various number of minutia points. The preliminary results show that similar minutiae configurations with 4 to 15 minutiae are observed in fingerprints from different fingers or individuals. Additionally, fingers from the same individuals tend to exhibit more occurrences of similar ridge structures than fingers from unrelated individuals.

■ Speaker

Elham Tabassi, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Tuesday, July 21 / 11:55am – 12:15pm

Mending Justice and Learning from Error: NIJ's "Sentinel Events Initiative"

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/mending_justice_and_learning_from_error_NIJ_s_sentinel_events_initiative-feucht-crim1.pdf (12 slides)

■ Abstract

Industries like medicine and aviation have implemented system-level reviews to understand “sentinel events” – bad outcomes that signal underlying system weaknesses – and to learn from errors (as well as “near misses”) through a non-blaming, all-stakeholders approach. Is such an approach feasible in the justice system? How would it operate? Who would run it? What kinds of errors would it focus on? Most importantly: could a sentinel events approach help create a more coherent and effective justice system? These are a few of the questions the National Institute of Justice seeks to answer through its Sentinel Events Initiative. The presentation will cover NIJ's award-winning “Mending Justice” report; the recently completed 3-site “beta” project to conduct sentinel event reviews; and NIJ's other ongoing and future research on criminal justice sentinel events.

■ Speaker

Thomas Feucht, National Institute of Justice, USDOJ, United States

Session: Criminalistics I / Tuesday, July 21 / 3:30 – 3:50pm

Policy Decisions in Latent Print Examination Affect Specificity

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/policy_decisions_in_latent_print_exam_affect_specificity.hall_crim1.pdf (15 slides)

■ Abstract

Latent print examination has two main approaches to determine the utility of an impression as defined by the Scientific Working Group for Friction Ridge Analysis Study and Technology (SWGFAST). In one approach, impressions that cannot be identified will not be preserved, documented or retained. In the second approach, impressions that cannot be identified but can be excluded are preserved and compared. More recently, a third approach to define which marks cannot be excluded but can be identified has emerged. While none of these approaches are wrong, each has a different false negative rate and affects specificity; yet the policy changes are made with little consideration to the impact to the criminal justice system. Within the latent print community the value of exclusion decision has been minimized due to a perception of its limited importance to our traditional partners.

■ Speaker



Carey Hall is a forensic scientist working for the Minnesota Bureau of Criminal Apprehension, does consulting work for Elite Forensic Services and previously, she was employed for the Phoenix Police Department and now has a unique perspective on the variety of different office policies, workflow and technology within various agencies. Carey has worked in latent prints for 7 years and is also an I.A.I. Certified Latent Print Examiner. She obtained her Master's degree in Legal Studies from Sandra Day O'Connor College of Law, at Arizona State University, where she worked to better understand the criticisms of forensic science and how it might be improved. Standard setting and policy creation are Carey's biggest interests, specifically, how scientific and empirical research can lead to adopting better policies. To that end, Carey has been teaching for the last three years and has developed a several workshops related to error rates and how to respond to forensic errors for both managers and practitioners.

Session: Criminalistics I / Tuesday, July 21 / 3:50 – 4:10pm

Changes in Latent Print Examinations as a Result of Technical Review

Presentation Link:

[https://www.nist.gov/sites/default/files/documents/2016/11/22/changes_in_latent_print_exam_as_a_result_of_technical_review.montooth.crim1 .pdf](https://www.nist.gov/sites/default/files/documents/2016/11/22/changes_in_latent_print_exam_as_a_result_of_technical_review.montooth.crim1.pdf) (20 slides)

■ Abstract

Errors in latent print examinations are typically uncovered by defense review of a case. What types of errors are made by latent print examiners in all types of cases? Could these errors have been mitigated by technical review?

In 2012, the Indiana State Police latent print unit began 100% technical review of cases. The process involves the examiner completing the exam, verification by a second examiner, then a third examiner performing a technical review of the case. It became apparent that changes had been made in some decisions as indicated within the examination notes. No procedure had been in place to track these changes outside a review of the individual cases. A review of all cases from 2013 and 2014 was completed to determine what changes or errors were being made. Beginning in 2015, all changes are being tracked during the administrative review.

Changes were tracked for each individual examiner and categorized as non-significant or significant. Non-significant changes were tracked by case and include anything from a spelling error in the notes to a procedure that was not documented correctly. Significant changes in analysis, comparison, or evaluation decisions were tracked by case and individual instance. Furthermore, exclusions or identifications not verified and changes that resulted in additional individuals being identified in a case were also tracked.

Data was also collected to aid in determining trends related to the quality assurance concerns. This review produced knowledge that many changes, both insignificant and significant, had occurred in the examinations. Trends of the changes as related to examiner throughput of cases were determined. The results of the entire review and suggestions for improving the process will be provided in the presentation.

■ Speaker

Marcus Montooth, Indiana State Police, United States

Session: Criminalistics I / Tuesday, July 21 / 4:10 – 4:30pm

A Forensic Latent Fingerprint Image Quality Metric for Preprocessing Quality Assurance

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/a_forensic_latent_fingerprint_image_quality_metric_for_preprocessing_qa.guan_crim1.pdf (27 slides)

■ Abstract

Fingerprint detection, matching, and recognition problems are well studied in literature. In contrast, the study of forensic latent fingerprint preprocessing is in relative infancy. Forensic latent fingerprint preprocessing covers all image transformations performed on fingerprint images obtained at crime scene prior to submission of the image for identity analysis. While it is widely accepted that preprocessing is essential to render latent fingerprint images suitable for further analysis, standards and best practice guidelines for preprocessing are critically needed. In this talk we present preliminary analysis intended to inform the development of such guidelines.

Preprocessing is a complex procedure encompassing diverse activities including but not limited to: color filtration, contrast adjustment, edge enhancement, noise removal etc. We present a work flow analysis of the major steps and activities for preprocessing, and introduce general guidelines to preserve fingerprint information integrity and promote reproducibility. Furthermore we extend the SIVV (Spectral Image Validation and Verification) feature, previously used on flat or rolled fingerprint images, to make it suitable to investigate the effectiveness of preprocessing of latents. Preliminary results obtained with a newly created latent dataset demonstrate that this modified SIVV analysis may be useful as an aide to inform latent analysts in regards to the improvement of fingerprint information quality obtained through preprocessing. Finally, we outline our plans for a study underway in which a suite of computational metrics applied to latent fingerprint images before and after preprocessing, will be correlated with success rates in consistent identification of minutiae and other mark-up features obtained by a pool of forensic analysts working on the same set of images.

Through this research we hope to provide first steps toward creation of a science-based foundation for forensic latent fingerprint preprocessing. The long-term goal is that such foundations will serve to increase reproducibility of latent fingerprint analysis, and help to mitigate possibilities for error in this critical forensic tool.

■ Speaker

Haiying Guan, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Tuesday, July 21 / 4:30 – 5:00pm

Communicating Measurement Results in the Courtroom: A Matter of Error, Uncertainty & Inference

Presentation Link: not available

■ Abstract

Forensic measurements are relied upon to investigate and prove an array of civil and criminal charges, from speeding to murder. Unfortunately, measured results are often misunderstood and misapplied in the courtroom. Forensic scientists commonly present them in a simplistic fashion as 'accurate and reliable' representations of a quantity's value while failing to communicate the limitations of the inferences they support. Judges, lawyers and jurors often naively accept such results as conclusively establishing a quantity's true value. No measurement, though, no matter how sophisticated or carefully performed, can ever reveal a quantity's true value. If forensic measurements are to facilitate the discovery of truth in the courtroom, the relationship between measured and true quantity values must be determined and clearly communicated to fact-finders. There are two general approaches employed for this purpose: error analysis and measurement uncertainty. Both are valid and can be used to characterize the inferences supported by measured results. Although the terms error and uncertainty are often used interchangeably, though, they are distinct concepts, focused on different types of knowledge and conveying different information. This is not well understood in the forensic community. Thus, even when attempting to convey a result's limitations, forensic testimony is often incorrect and misleading. Error analysis (the traditional approach) relies upon frequentist methodology to characterize the inferences supported by measured results. Its object is to estimate a quantity's true value as closely as possible through the identification, minimization and quantification of error. The referent of this analysis is a measurand's unknowable 'true' value. Unfortunately, error analysis doesn't provide a method for generally combining systematic and random error which yields a statistically meaningful estimate of a measurement's total error. Hence, it provides no general measure of the likelihood that a particular result represents a measured quantity's value. Measurement uncertainty was developed to address this shortcoming and provide a standardized methodology by which a measurement's limitations can be determined and communicated. Grounded in Bayesian theory, the referent of this analysis is not a quantity's unknowable true value, but one's state of knowledge about that value. The objective is to determine the values that the information gained from measurement permits to be reasonably attributed to a quantity. Although one still cannot say what a quantity's value is, the uncertainty approach provides an unambiguous characterization of the quantity values our state of knowledge permits to be justifiably believed. This presentation will address the difficulties encountered when measurement results are presented in the courtroom and illustrate how even accurate and reliable results can lead to bad verdicts when presented incorrectly. Both error analysis and measurement uncertainty will be discussed with an emphasis on their distinctions and the types of knowledge each permits to be gained. A case will be made for the adoption of the uncertainty approach within the forensic sciences as well as the necessity of reporting a result's uncertainty whenever forensic measurements are presented as evidence. This should facilitate outcomes that hew more closely to scientific reality when measured results are relied upon during decision making.

■ Speaker

Ted Vosk, Attorney at Law, United States

Session: Laboratory Management / Tuesday, July 21 / 11:15 – 11:35am

Guidelines for the Use of Root Cause Analysis (RCA) to Reduce Error and Improve Quality in Forensic Science Laboratories

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/guidelines_for_the_use_of_root_cause_analysis_to_reduce_error_and_improve_quality_in_forensic_science_labs.hollway.labmgmt.pdf (16 slides)

■ Abstract

Forensic laboratories accredited under programs that adhere to the ISO/IEC 17025 (General requirements for the competence of testing and calibration laboratories) are required to "establish a policy and a procedure and . . . designate appropriate authorities for implementing corrective action when nonconforming work or departures from the policies and procedures in the management system or technical operations have been identified." A problem may be identified through a number of different techniques, including internal and external audits, reviews of the management system, customer feedback, or staff observations.

"Corrective actions" are potential solutions that eliminate or minimize the risk of repeating the nonconforming work or departure from policies and procedures. Corrective action is a requirement when any error or nonconformity is identified. To establish the best corrective actions, and as required by ISO 17025, an investigation is initiated to determine the root cause(s) of the situation or condition. Root Cause Analysis (RCA) is a critical step of determining corrective actions and may be the most important part of establishing proper corrective actions.

This presentation will review the principles of RCA and their applicability in a forensic science laboratory environment, including implementations of RCA in relevant environments, before providing recommendations drafted as part of the work of the National Commission on Forensic Science (NCFS) for the design, conduct, and optimization of RCA in a forensic laboratory setting.

The NCFS recommendations include information on when and how a RCA should be conducted; appropriate training to forensic laboratory personnel; construction of the investigative team conducting a RCA; determination of the extent of the adverse event and/or additional identical or similar adverse events; communication of the existence of the adverse event to individuals impacted by the adverse event; provision of Safe Harbor to employees who report adverse events or near misses, including use immunity for participation in an RCA and limitations on the disclosure of materials generated in the course of an RCA; implementation of interventions designed to minimize the chance of future similar adverse events and to appropriately redress injury caused by the adverse event; and documentation of both the adverse event(s) and the proposed interventions in a manner that does not reveal specific individuals or case information, and makes the learnings from the RCA available for the benefit of other laboratories.

■ Speaker

John Hollway, Quattrone Center for the Fair Administration of Justice, University of Pennsylvania Law School, United States

Session: Laboratory Management / Tuesday, July 21 / 11:35 – 11:55am

Corrective Measures Taken in Response to DNA Extraction Failures Using a Newly Validated Method

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/mcgrory_abstract.pdf (abstract only)

■ Abstract

The Centre of Forensic Sciences (CFS) recently moved from organic DNA extraction protocols to a semi-automated magnetic bead purification protocol. Following an internal validation (per ASCLD/ FBI QAS standards), methods were implemented using Qiagen Investigator extraction chemistry (buffer G2) in conjunction with the Qiagen EZ1 Advanced XL instruments. Shortly after that implementation, a subset of forensic substrates was identified where lower than expected yields of DNA were recovered. In addition to low yield, some of these samples also exhibited DNA profiles with peak height discordance or imbalance (using Applied Biosystems Identifier-Plus). The materials that were resistant to extraction were primarily dark coloured material and were not restricted to a particular body fluid. This led to a suspension of DNA extractions using the Qiagen chemistry and instruments pending the outcome of an investigation. A systematic review of approximately 1000 casework samples processed since implementation (some of which had already been reported to clients) was undertaken to determine those which may have been affected by the observed phenomenon. Based on specific criteria approximately 200 of these were re-purified or re-sampled from the original items in an attempt to recover more DNA and generate better quality profiles. The subsequent root cause analysis determined that the Qiagen G2 buffer was not removing the cellular material from the problematic substrates, while a re-extraction of the original substrate with the CFS organic extraction protocol did yield expected DNA quantities. A re-validation of the EZ1 instruments was undertaken using CFS in-house extraction chemistry. This approach was successful and the methodology was brought back online. However, shortly thereafter new issues related to the formation of precipitates with this buffer, thought to be due to the concentration of SDS, were experienced. Those samples that were affected presented high autosomal to male quantification ratios (using Promega Plexor-HY) as well as peak height imbalance (notably at D13S317). The methodology was suspended again, and further studies were undertaken to optimize the buffer composition as well as determine the impact on any affected samples. This presentation will focus on the investigative steps taken to determine the root cause of the incidents; the retrospective assessment of extraction results; the steps taken to mitigate the impact on individual cases; the communication of the issue to our clients; and the measures taken to re-validate and re-implement the silica bead technology purification. Our experience demonstrates the benefit for redundancy with multiple extraction platforms, and the value of a systematic approach to managing quality issues which impact a large sample set. Furthermore the need for continuous monitoring and critical assessment of newly implemented procedures, even those widely accepted within the forensic community, is highlighted.

■ Speaker



Joel McGrory has a background in molecular biology and originally worked in research in the biopharmaceutical field in the private sector then in the area of heritable connective tissue diseases in the public sector. He entered the forensic field in 1997 with the Centre of Forensic Sciences in Toronto where he has been ever since. During that time he has worked both as a DNA technologist, a reporting scientist and most recently as the manager of the DNA Unit. The Centre of Forensic Sciences provides service for the province of Ontario in Canada, covering Toxicology, Chemistry, Biology, Firearms and Toolmarks and Document examination. The Biology section processes approximately 5000 cases per year.

Session: Laboratory Management / Tuesday, July 21 / 11:55am – 12:15pm

Cultural Aspects of Error Management

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/cultura_aspects_of_error_measurement.wedn_labmgmt.pdf (35 slides)

■ Abstract

Geert Hofstede has studied cultural dimensions of nations; his theories are thought to set a framework for cross-cultural communication. Hofstede described five dimensions initially based upon a world-wide survey of employees for IBM in the 1960's and 1970's. A sixth dimension has since been added: power distance (PDI), individualism (IDV), masculinity (MAS), uncertainty avoidance (UAI), long term orientation (LTO), and indulgence versus restraint (IND). Different cultures cope with inequality, uncertainty, and relationships differently. The cultural clash between the various groups within and about the forensic science community can be seen from this perspective and nowhere is it more obvious than in error management. Law enforcement and prosecutors have a rule-based orientation, where defendants and defense attorneys are far more tolerant of infractions. Law enforcement rates high on power distance scales, but the scientific community has an egalitarian communal orientation. More fundamentally, the criminal justice system is primarily concerned with individual responsibility and blame, but scientists recognize uncertainty and error is fundamental. This can be seen as intolerance and tolerance for deviation from normal. Juries cope with tolerance based on words; scientists cope with tolerance based on numbers. A significant difference between these two models is intention, but this is eroding with the rise of statutory crimes. Management of error in forensic science laboratories calls for a root cause analysis, which precisely looks beyond the individual blame, on the other hand, their law enforcement administration may want to look for blame and prosecutors will have little tolerance for errors. In essence, this is why the air force has distinguishes Safety Investigation Boards which has a prevention focus and which specifically avoids blame from Accident Investigation Boards which has greater legal implications and may assess blame.

■ Speaker

Victor W. Weedn, MD, JD, George Washington University, United States

Session: Laboratory Management / Tuesday, July 21 / 1:30 – 1:50pm

Errors in a DNA Testing Laboratory

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/errors_in_a_dna_testing_laboratory.word_abmgt.pdf (23 slides)

■ Abstract

Over the ~17 year existence of a private DNA testing laboratory, some mistakes were made in various aspects of the DNA testing and reporting process. Some of the errors were minor and easily corrected, such as typographical errors in final reports and incorrect data in population databases. Other errors were more serious, including contamination in early proficiency tests and the mislabeling of individuals and their genotypes in a report leading to the incorrect conclusions being reported and presented in an admissibility hearing during which the mistake was recognized. An additional incident that will be discussed is the alteration of negative control data by a rogue scientist in the laboratory. Appropriate corrective actions to address and rectify each specific error as well as an assessment of how each error occurred and what actions could be taken to prevent a similar mistake from occurring in the future were implemented. The role of strategic planning by the laboratory in recognition that significant events could happen in the laboratory will be presented along with a discussion of several of the errors and the corrective actions applied. The presentation will also address the need for openness in discussing errors and the sharing of the information through laboratory audits, the discovery process and during testimony.

■ Speaker

Charlotte Word is currently a consultant in Human DNA Identification testing. She joined Cellmark Diagnostics in Germantown, MD (later called Orchid Cellmark) in 1990 and over the 15 years she worked there had several titles, including Deputy Director and Senior Manager, Forensics and Laboratory Director. She has performed technical reviews on many 1000's of cases from several public and private DNA testing laboratories and has participated in the validation of various DNA test systems. Charlotte has testified as an expert witness in over 300 trials and admissibility hearings around the country since 1990. She has published and given many presentations at meetings and workshops in the areas of DNA identification testing. In 1998-1999 she was a member of the Post-Conviction Issues Working Group of the National Commission on the Future of DNA Evidence. She has experience with ASCLD/LAB, NFSTC, DNA Quality Assurance Standards and ISO audits. Charlotte is a member of the editorial board of *The Journal of Forensic Sciences* and has been a guest reviewer for *Forensic Science International: Genetics*. For the past few years, she has been assisting with Mixture Interpretation training through a grant from NIJ awarded to Boston University. She is a member of the DNA Analysis 2 (DNA Reporting and Interpretation) Subcommittee of the Biology/DNA Scientific Area Committee of the Organization of Scientific Area Committees (OSAC) and of the Reporting and Testimony Subcommittee of the National Commission on Forensic Science. Charlotte received her B.S. in Biology from the College of William and Mary in Virginia and her Ph.D. in Microbiology, with specialties in Molecular Biology and Immunology from the University of Virginia. She did postdoctoral research as a Damon Runyon-Walter Winchell Cancer Fund Fellow at the University of Texas Southwestern Medical School in Dallas, TX and was a faculty member at the University of New Mexico School of Medicine prior to joining Cellmark Diagnostics.

Session: Laboratory Management / Tuesday, July 21 / 1:50 – 2:10pm

Detection and Mitigation of Medical Errors

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/detection_and_mitigation_of_medical_error_s.schenberg.labmgmt.pdf (23 slides)

■ Abstract

The medical biopsy testing process involves up to 18 different steps, each of which is an opportunity to introduce a specimen-switching error or contamination error (known as Specimen Provenance Complications "SPC's"). According to peer-reviewed literature, up to 3.5% of prostate biopsy samples are compromised due to occult (undetected) SPC's which can only be detected by use of DNA identity testing. Becoming aware of SPC's in 2008, Strand Diagnostics, an FBI-accredited forensic DNA testing lab, adapted its forensic DNA testing process for use in routine clinical medical practice to diagnose and prevent SPC's from adversely affecting prostate and breast patients. Known as KnowError, the system incorporates a novel tissue collection process (which has the effect of reducing errors) and incorporates Strand's DNA identity test (known as DNA Specimen Provenance Assay SPA's). The system includes the use of a bar-coded buccal swab for obtaining a sample from a patient for the development of a DNA profile which is then matched to the DNA profile developed from tissue samples putatively ascribed to that patient. TAT is usually less than 3 business days and a report is issued to the treating physician so as to confirm a patient's diagnosis. A patent has been issued by the USPTO for the KnowError system and it is now used in over 6% of all prostate biopsies in the US and 1% of breast biopsies. Since its commercial introduction in 2009, almost 200,000 prostate and breast biopsy patients have been protected by the KnowError system with 50,000 of those patients having had a DSPA test performed. Approximately 1% of those patients (500) had specimens which were either switched (single profile-to-single profile mixup) or were contaminated with another person's specimen. A Washington University article published in the Jan 2013 edition of the American Journal of Clinical Pathology confirms this rate of error and concludes that no clinical or laboratory setting is immune from occult SPC's. An article in the April 2015 issue of the Journal Of Urology concludes that 1 in 200 men suspected of having prostate cancer are affected by occult SPC's and that the wasted cost to the healthcare system is \$880,000,000 per year. The talk will expound upon why the medical biopsy process is fraught with errors; the documented rate of errors; anecdotal consequences of undetected errors; how the KnowError process was developed, its market acceptance, examples of how its use has caught errors/helped patients avoid unnecessary surgeries; as well as the application for use for DSPA in the growing world of personalized and precision medicine.

■ Speaker



Theodore "Ted" Schenberg is the co-founder of Strand Diagnostics LLC, a forensic and medical DNA laboratory located in Indianapolis, IN and has been its CEO since 2007. Mr. Schenberg is a successful, veteran businessman having founded and acquired several companies in the life-sciences, chemical, aerospace, construction and food processing industries. He has an extensive track-record of executive leadership as well as building value for his companies in partnership with his management teams. Three of Mr. Schenberg's companies have been acquired by publicly-traded firms. He has taught courses on business and entrepreneurship at assorted venues including Indiana University's Kelly School of Business from which he holds a degree in Accounting and a Masters in Business Administration.

Session: Laboratory Management / Tuesday, July 21 / 2:10 – 2:30pm

The Challenge of Auditing Technical Records Generated During Complex DNA Casework Processes

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/the_challenge_of_auditing_technical_records_generated_during_complex_dna_casework_processes.askew_labmgmt.pdf (36 slides)

■ Abstract

With an ever increasing demand from the legal system to provide scientific evidence to assist the Trier of Fact in the judicial process and the expanding palette of services to include suboptimal or reduced template DNA testing the trend of ever expanding case backlogs is not an entirely surprising phenomena. As backlogs grow stressing an already burdened lab system the potential for errors to be introduced into the workflow increases proportionally. Documenting this fact is finding #1 (Standard 4.13.2.1) in the 2013 ASCLD LAB top ten nonconformities report. Traditional LIMS systems utilize by laboratories focus on the end points rather than the branch points of the workflow and are well adapted to the linear workflow that is common to most forensic disciplines. This focus on the endpoints however falls short when attempts are made to adapt it to the dynamic workflow associated with complex DNA analysis. DNA analysis typically consists of a decision tree full of branch points that is dictated by the results of the individual steps of the process as well as a thorough review of the data generated from the process. This review process is designed to capture and correct errors before they introduced into the legal arena by creating layers of redundancy. This redundant approach is very efficient in instances where the process paradigm is straight forward, but begins to lose effectiveness as the decision tree associated with the sample becomes more complicated. Complicating the review process has a tremendous adverse affect on the overall efficiency of the entire system and leads to even more backlogged cases, resulting in negative cycle that can completely cripple a laboratory. We are proposing a better approach to DNA information management systems that would marry the strengths of traditional LIMS with the decision tree utilized by DNA analysts to create a more streamlined and error proof working paradigm. This will be accomplished by providing and recording a series of QC flags and rule firings to the analyst at each of the critical branch points. These flags will not only serve to assist the analyst in the real time workflow, but will be cataloged and provided to the reviewer to maximize the efficiency and effectiveness of the review time as well. Because this approach to information management provides a detailed map of the path a sample traverses through completion it can be associated with a timeline. The introduction of a timeline has the organic benefit of allowing the laboratory to data mine the operations using a layered metric engine to identify and resolve bottlenecks in the paradigm further optimizing sample turn around. The final step to maximizing laboratory efficiency while minimizing the potential for erroneous or omitted sample information would be to link the information management system to a computer assisted mixture interpretation tool. This holistic approach to information management will result in a dramatic decrease in the frequency of citations under Standard 4.13.2.1, providing excellent scientific support to the increased demands of the judicial system.

■ Speaker

Robert Askew has nearly 20 years of DNA analysis experience. He has worked as a bench scientist at various public and private laboratories. Recently Robert did a tour with the Joint Expeditionary Forensic Force (JEFF) under the direction of the United States Army Criminal Investigation Laboratory (USACIL). His primary assignment was the Reach Back Operations Center (RBOC) which provides forensic support for United States Military Assets in active theaters of combat. Robert now serves as the Senior Scientist for NicheVision Forensics. There he assists in developing innovative and effective solutions to some of the most challenging problems faced by the modern forensics laboratory.

Session: Laboratory Management / Tuesday, July 21 / 2:30 – 3:00pm

Laboratory Management Techniques and Practices for Creating a Culture Where Quality is Top Priority

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/lab_management_techniques_and_practices_for_creating_a_culture_where_quality_is_top_priority.morgan.labmgmt.pdf (20 slides)

■ Abstract

Laboratory managers have the difficult responsibility of balancing budgetary and staffing concerns while keeping the laboratory's culture focused on quality. At their core, DNA testing laboratories are targets of increased scrutiny due to the number of quality requirements already in place and the sensitive nature of the testing performed. However, with the formation of the National Commission on Forensic Science and Organization of Scientific Area Committee the entire forensic domain is at the cusp of establishing better practices. This presentation will focus on various aspects of Bode Cellmark Forensic's DNA quality control plan and how its evolution has enabled the successful completion of approximately 90,000 forensic cases and 1.5 million offender DNA profiles within the past 10 years. To combat the potential for errors in the DNA testing process, Bode Cellmark Forensics has created a transparent management system; one that enables all levels of personnel from managers to DNA analysts and technicians to visualize trends across the various operational teams. In this regard, the internal quality system is never static and it permeates every aspect of the laboratory starting with the training program and ending in the reporting of case results. To continue to meet the demands of the forensic DNA community, Bode has developed ways to constantly monitor quality in every step of the process. This includes ensuring that every individual understands their impact on the quality system and that even the smallest detail is not taken lightly. From an overall culture perspective, non-conformances and departures from procedure are collected not for punitive reasons, but to help identify trends and determine where process improvements are needed. Current management techniques have focused on efficiencies that increase the transparency among the operational teams and have proven vital. Quick status meetings each day, or team huddles, and the application of a social business enterprise system have been key to ensuring the steady flow of communication within the organization. These new tools only strengthen the value of previously implemented quality controls solutions at Bode Cellmark Forensics to enhance standard practices in the field of DNA processing, including: maintaining staff and visitor databases, using batched sample processing to visualize trends in reagent and substrate quality control, repurposing existing computer programs to detect sample to sample comparisons within a data tray, and engaging client and manufacturer relationships to optimize laboratory procedures. This look at the error detection practices and quality applications in a high throughput forensic DNA laboratory can be used as a model across the field. Bode Cellmark Forensic's team structure mimics some of the challenges faced among disciplines in any other forensic science laboratory, while at the same time is constantly able to overcome budgetary, staffing, and capacity difficulties and find ways to maintain the highest levels of quality.

■ Speaker

Natalie Morgan, Bode Cellmark Forensics, United States

Session: Laboratory Management / Tuesday, July 21 / 3:30 – 4:00pm

Calculating the True Costs of Errors in Forensic Casework

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/calculating_true_costs_of_errors_in_forensic_casework.mayo_labmgmt.pdf (52 slides)

■ Abstract

Implementation of robust quality management systems, best practices and accreditation have become industry standards leading to the global reduction of errors within the forensic community. More recently, increased awareness of human factors associated with bias in forensic casework will bring further improvements to quality practices. However, none of these changes will ever bring about a zero error environment. Errors will continue to occur and have an enormous impact on the delivery of accurate and timely results. Moreover, their costs and downstream effects may not be known, obvious or immediately calculable. In this presentation we will review some common forensic errors, identify factors contributing to their costs (monetary and otherwise) and review cost containment practices used by various industries in an effort to start a dialog for understanding and calculating the true costs of errors in forensic casework.

■ Speaker

Elissa Mayo, Assistant Bureau Chief, Bureau of Forensic Services, Office of the Attorney General, California Department of Justice, United States

Session: Laboratory Management / Tuesday, July 21 / 4:00 – 4:20pm

Efficiency Efforts: An Examiner-Friendly Approach to Improving Laboratory Operations

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/maynard_abstract.pdf (abstract only)

■ Abstract

This presentation will focus on the lab-wide "Efficiency Efforts" that were conducted at the Defense Forensic Science Center. This simple approach was used to understand and optimize case work by identifying and eliminating operational bottlenecks and minimizing unnecessary dwell times. The Efficiency Efforts utilized three straightforward sources of input (time tracking, process mapping and improvement boards) to help recognize and prioritize the best opportunities to save time. This information was used to write white papers that clearly defined the actual issues, potential solutions, and the impacts or benefits to the lab. The white papers were briefed by examiners to leadership so they would have the ability to further explain the issues and their solutions. Once leadership was aware of the examiners' issues they knew how to help and could begin to take action. This process resulted in the decrease of turn-around-times in one branch by more than 50%. Overall, the Efficiency Efforts helped the Defense Forensic Science Center streamline operations and save examiners time, which ultimately helped the lab obtain even faster turn-around times.

■ Speaker

Henry Maynard, Defense Forensic Science Center, United States

Session: Laboratory Management / Tuesday, July 21 / 4:20 – 4:40pm

Students in the Forensic Laboratory: Fostering Education While Maintaining Quality

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/students_in_the_forensics_lab.baigent.lab_mgmt.pdf (29 slides)

■ Abstract

The Metropolitan State University of Denver Human Identification Laboratory (MSUD-HIL) offers forensic anthropological search, recovery, and analysis services to medico-legal professionals throughout Colorado. The Rocky Mountain region is a geographically distinct area characterized by rapidly converging ecozones situated at elevations ranging from 1000 - 4400 meters. Outdoor search and recovery scenes are typically complex due to their magnitude and the changes to biological evidence affected by dynamic micro-climatic zones and attendant biotic and abiotic factors. In an effort to more effectively process vast outdoor scenes and provide students with the practical experience necessary for professional development, credited internships are offered to undergraduate students who have demonstrated exceptional academic performance and an interest in pursuing a career in forensic anthropology. Introducing an educational platform to scene processing may be beneficial to all involved, but presents a unique set of challenges necessitating foresight, planning, and a rigorous quality assurance (QA) program with the ultimate goal of maximizing the information recovered while maintaining the probative value of all lines of evidence. Toward that end, a multipronged training and competency testing system specifically designed for student participation (SQA) has been implemented in the MSUD-HIL and operates under the laboratory's primary QA protocols. The SQA is hierarchical in nature and devised of a series of benchmarks attained by the successful completion of internal and external training and testing. The level of student participation both in the field and in laboratory analysis is dictated by the certification level achieved. Certification levels are represented by a color-coded system so that qualified (and more importantly, unqualified) students may be easily identified in the field, allowing laboratory supervisors to easily manage students across broad topographic areas and rapidly delegate tasks to appropriately qualified individuals. The simple use of prominently displayed color-coded identification cards has demonstrated the added benefit of strict self-management and proactive training on the part of student interns. Because the tasks students are allowed to participate in are clearly outlined by the certification level achieved, students tend to proactively strive to attain higher qualifications; this is typically accomplished with the assistance of more highly certified interns. This has had the benefit of affording teaching and training experience to more skillful interns while reducing the burden placed on the laboratory director, engendering an ethos of active ongoing education, and practically demonstrating the importance of QA in the forensic laboratory. SQA is maintained through a series of standardized tests independently assessed by two members of laboratory management at each stage of certification. This ensures that standards are maintained while fostering a multidisciplinary, multi-perspective teaching environment. SQA in the MSUD-HIL required systematic and step-by-step processes. These included the development of the SQA framework within the primary QA framework, training and commitment amongst members of laboratory staff, additional certification manuals, methods for assessment, and integration of SQA programs into the laboratory's quarterly action plans. Once these controls are in place, students may make a valuable contribution to any laboratory while reaping the benefits of practical experience.

■ Speaker

Christiane Baigent, Metropolitan State University of Denver, United States

Session: Laboratory Management / Tuesday, July 21 / 4:40 – 5:00pm

Implementing the Duty to Correct Error in Forensics

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/scheck_abstract.pdf (abstract only)

■ Abstract

Scientific experts and scientific evidence are treated differently from other types of witnesses in the U.S. legal system. In order to serve the special needs of the 'customers' of the forensic science work product (prosecutors, defense counsel, judges, crime victims, criminal defendants, as well as civil plaintiffs and defendants) the OSAC must address their "special needs." These "special needs" include ensuring validated and reliable standards are established for each discipline covered by Scientific Area Committees. Establishing such standards are the best way to measure, detect, and mitigate error. Other key protections to be discussed include early detection of potential error through external blind proficiency testing on samples replicating case work, implementing a duty to correct and notify the all affected 'customers' of erroneous testimony or laboratory results, independent 'Coverdell' audits, all stakeholder reviews in conjunction with root cause analysis, and cognitive science 'fixes' to prevent testimony that exceeds the bounds of science or the relevant forensic data.

■ Speaker

Barry Scheck, Innocence Project, New York City, NY, United States

Session: Quality Assurance / Tuesday, July 21 / 11:15 – 11:35am

Considerations on the Efficacy of Accreditation and Laboratory Performance

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/Considerations_on_the_efficacy_of_accreditation_and_laboratory_performance-steel-qa.pdf (19 slides)

Abstract

It would seem logical that accreditation of a laboratory for conformance to a quality standard (e.g. ISO 17025) would be a good indicator of accurate output by an analytical laboratory and result in improved error management over time. This conclusion would be based on at least two assumptions: 1) the presence of a quality system sufficient for accreditation should assure quality in the laboratory measurements and 2) laboratories that do not produce quality results would not be granted accreditation. However, studies of accredited laboratories vs. nonaccredited laboratories in several fields over the past 20 years have in general shown limited impact of accreditation on laboratory results. Commonly, the studies are conducted by comparing the performance of accredited and nonaccredited laboratories on proficiency tests. Surprisingly, for many cases, laboratories that are accredited show no better performance compared to those that are not.(1,2,3) Other studies have shown a mixed or only a limited positive impact of accreditation on performance.(4,5) To our knowledge, only one program has shown that accredited laboratories performed substantially better than nonaccredited labs.(6) The reasons for the generally unexpected results of comparison of accredited and nonaccredited laboratories are not clear. Some speculations in the literature include: 1) accreditation has a greater effect in some fields than in others, 2) conduct of blind proficiency testing might give different results, and 3) quality systems improve repeatability within a laboratory but not reproducibility between laboratories. In this work some other factors possibly affecting the efficacy of accreditation are explored, including: 1) There may be human factors that act on accreditation bodies and serve to keep the majority of laboratories accredited despite variations in laboratory quality. 2) Well designed and executed proficiency testing and interlaboratory comparisons conducted over long time periods may be the best way to evaluate and improve the performance of laboratories. Laboratories are typically evaluated over short to intermediate time periods. 3) The assessments and proficiency testing may evolve to a very basic level reflecting the level at which almost all the laboratories can perform satisfactorily. 4) The measurement method used by the laboratories may not be developed sufficiently. 5) Quality systems may not be a significant contributor to the quality of analyses, as counter intuitive as that might seem. Other factors and possible means to evaluate the accreditation process will be discussed. Further consideration of the value of accreditation would seem to be an important area of study in the field of error management in forensics.

References:

1. Thompson, M., et al. (2009) *Accred. Qual. Assur.*, 14, 73-78.
2. R.G. Visser (1999) *Accred. Qual. Assur.*, 4, 108-110.
3. King, B., Boley, N. and G. Kannan (1999) *Accred. Qual. Assur.* 4, 280-291.
4. Siloaho, M., et al. (2006) *Accred. Qual. Assur.* 238-245.
5. Ehrmeyer, S.S. and R.H. Laessig (2004) *Clin. Chim. Acta* 346, 37-43.
6. Morris, A. and D. Macey (2004) *Accred. Qual. Assur.* 9, 52-54.

Speaker



Eric Steel is the Program Director for Forensic Science for the Material Measurement Laboratory at the National Institute of Standards & Technology (NIST) in Gaithersburg MD covering such areas as trace evidence, drugs and toxins, and genetic materials. He has served in a variety of positions at NIST including leader of the Microanalysis Research Group (1994-2003), Acting Deputy Director of the Chemical Science and Technology Laboratory (2003-2004), and Director of the NIST Program Office (2004-2009). Currently, as a senior microanalyst (2009 to present), Mr. Steel is developing standards and methods for characterization of particles, atomic resolution 3D chemical imaging, and quality systems for atom probe, microscopy and forensic analyses.

Session: Quality Assurance / Tuesday, July 21 / 11:35 – 11:55am

Proficiency Testing Programme in Forensic GSR and Firearm Investigation

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/pt_program_in_forensic_gsr_and_firearm_investigation.niewohner.qa_.pdf (39 slides)

Abstract

Within the framework of the ENFSI Working Group "Firearms/GSR" a proficiency testing programme has been set up and is performed over the last 20 years. One of them is the "Proficiency Test on the detection and identification of gunshot residues (GSR) by SEM/EDX". This test was created and performed for the first time in 1999 [1]. Since 2006 this test is scheduled by ASCLD-LAB as an approved GSR proficiency test [2]. Currently the GSR2015 test is under preparation and will be distributed in mid-2015. Sample description and evaluation: The test material is designed by an Advisory Board, consisting of two members of the ENFSI EWG Firearms/GSR, and manufactured on order by an external company. The participating laboratories are requested to determine the total number of PbSbBa containing particles on the test samples following their own laboratory specific methods of automated GSR particle search and detection by SEM/EDX. The test items for the proficiency test consist of a set of completely identical samples as it is demanded within the ISO 17043 standard for the performance of proficiency tests. Therefore the samples were produced using a special, patent protected method [3]. Synthetic "GSR particles" of a known composition of lead, barium, and antimony and a size distribution of 0.5 to 2.4 μm were precipitated onto a silicon substrate of a size of 8 x 8 mm, and mounted on a SEM stub. All participating laboratories are requested to analyse the received sample using their standard GSR examination routines on their SEM/EDX systems, and to report the results within 4 weeks to the organisation committee. Reports have to include the number of detected PbSbBa particles, their size and their exact position on the sample. Sample evaluation is performed by comparing the submitted data with the original dataset of the sample production. An evaluation of the laboratory's performance to detect GSR particles by SEM/EDX is carried out using z-scores according to relevant IUPAC and EURACHEM guidelines. Finally, a comprehensive report on the results of the proficiency test is prepared and sent to all participants. Further proficiency tests have been developed by the working group on the fields of firearms examination (FAID2009; FAID2012), and shooting distance determination (FDSD2015). The Proficiency Testing initiative of the EWG Firearms/GSR has been financially supported by the European Union within diverse OISIN, AGIS, and ISEC programmes.

References

- [1] Niewohner L; et al.; "GSR2005 - Continuity of the ENFSI Proficiency Test on Identification of GSR by SEM/EDX, J Forensic Sci., 53(1); (2008).
- [2] <http://www.asclcd-lab.org/approved-providers-and-tests/>
- [3] Niewohner L, Wenz, W., patent holder; "Verfahren zur Herstellung synthetischer Partikelproben"; Patent DE 199 32 357; 2001 Feb 8.

Speaker



Ludwigh Niewohner is a Senior Forensic Scientist at the Forensic Science Institute of the Bundeskriminalamt in Wiesbaden, GERMANY and is a member of the Steering Committee of the ENFSI Expert Working Group Firearms/GSR (chairman from 2011-2013).

Session: Quality Assurance / Tuesday, July 21 / 11:55am – 12:15pm

Inaugurating Data Integrity in Forensic Science

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/mozayani_abstract.pdf (abstract only)

■ Abstract

Data Integrity is generating the data that is ethical, analyzed by acceptable analytical techniques, traceable and defensible. While the vast majority of forensic analysts are all dedicated, hardworking seekers of the truth, the fact remains that forensic scientists and forensic science laboratories toil in an arena in which their actions frequently come under public scrutiny. This perception is fueled by news reports about individual analysts who are found falsifying results, contaminating evidence, producing improper testimony during legal proceedings or the laboratories that using the evidence by methods that are not validated or not accepted by scientific community. As a direct result of this continued public scrutiny, it is imperative that forensic laboratories should be proactive to develop, implement, and maintain data integrity policy for all the validated analysis and services that they provide to their customers and ultimately to the community. The objective of this presentation is to study the aspects and causes of data manipulation, and acquire steps to establish an effective and rigid policy and standard operating procedure that maintains data integrity across disciplines. To build the foundation of this procedure it is important to review causes of laboratory fraud. Unfortunately, these cases are bountiful, ranging from DNA contamination and dry labbing, to analysts selling drugs on the street that were smuggled from the laboratory in Florida and Texas. These cases have occurred in places such as Houston, Boston, and DNA contamination that suspended testing at the Washington D.C. laboratory. These cases include impropriety by analysts using instrumentation data incorrectly, as well as presenting improper or incorrect testimony in courtroom proceedings or many times not testing the evidence according to acceptable and validated methods. They also include examples of oversight by management, emphasizing quantity over quality as well as instances of unrealistic expectations for case acceptance and management. These cases can include negligence that may be caused by a need to increase productivity by limiting a thorough technical and administrative review. The addition of and strengthening of a quality control system is a key step in the development of a standard operating procedure for ensuring data integrity. The purpose of this procedure should at least be to describe the details of a laboratory's data integrity system, to point out the significance of ethics in the performance of all acceptable analytical work, to acquire the obligation of laboratory staff to the belief that all analyses shall be achieved in a controlled and documented manner, and to confirm that laboratory staff always convene the specific ethical requirements described in the organization data integrity plan. The data integrity policy will benefit the staff to understand the steps of monitoring data integrity through changing laboratory culture, emphasizing in non-conflict environment, training and monitoring of staff and assertive other managerial actions.

■ Speaker



Ashraf Mozayani is the Executive Director of Forensic Sciences & Professor at Texas Southern University. She has been recently appointed by the Governor as a Commissioner to the Texas Forensic Science Commission. Dr. Mozayani is a Board Certified Diplomat of the American Board of Forensic Toxicology (D-ABFT) and is an emeritus member of the American Crime Laboratory Director. Prior to her current position she was the Laboratory Director and Chief Toxicologist at the Harris County Institute of Forensic Sciences in Houston, Texas for more than fifteen years. She is also president and founder of the International Forensic Science Consultants, LLC.

Session: Quality Assurance / Tuesday, July 21 / 1:30 – 2:30pm

Error, Error Rates and Uncertainty: What Are They, How Are They Determined and Do They Facilitate the Discovery of Truth in the Courtroom?

Presentation Link: not available

■ Abstract

The National Academy of Sciences concluded in its 2009 Report that the results of all forensic science methods should include error-rates and/or uncertainties when reported in the courtroom. What does this mean, though, and is it a realistic expectation? The concepts of error, error-rates and uncertainty refer to distinct and typically ill-defined concerns both within and between forensic and legal communities. Before the goals set forth by the Academy can be realized, these concepts must be strictly defined. Error is commonly used to refer to a couple distinct situations. The first occurs when, regardless of whether the final results/opinions are correct or incorrect, there is a failure to perform a method or interpret data in a validated manner. The second occurs when, regardless of the manner in which a method was performed or data interpreted, the final result/opinion is incorrect. Much legal argument focuses on the presence or absence of the first type of error to deduce, respectively, the presence or absence of the second type of error. Whether an error of the first type has occurred may be important in as much as it tends to make an error of the second type more or less likely. It is not typically, though, the scientific focus when determining a method's error-rate or uncertainty. Scientifically, a method's error-rate or uncertainty is intended to convey the relationship between the result/opinion obtained, when a method is performed and data interpreted properly, and the correct result/opinion. Error-rates generally refer to a frequency with which results do not correspond to the "true" physical/factual state when obtained and interpreted in the validated manner. They cannot tell us about a unique event, rather, they convey the frequency with which one would expect to be wrong if repeated observations were carried out. There are a number of well-developed methods of determining a method's error-rate that are easily employed in the physical sciences. One reason they are easily employed in this context is because the methodology utilized can be distinguished from the researchers employing it. This permits the method itself to be isolated and characterized. This is not always the case in the identification sciences, though, where an analyst serves as an inseparable and unstandardized sensor or software component. This inseparability and lack of standardization makes the determination of a general error-rate for a method problematic. One way to address this is to follow the path taken by forensic DNA analysis. Eschew estimates of error-rates associated with a specific method of analysis and instead focus on general population characteristics associated with the data being examined. For example, instead of trying to characterize error-rates associated with forensic fingerprint analysis itself, one could characterize the likelihood that a given print will match another, randomly selected, fingerprint. This is analogous to the random match probability relied upon as an inferential tool in DNA analysis. While this still leaves to the fact-finder the burden of determining whether the method utilized to extract and match samples is reliable, it provides information concerning the uniqueness of the underlying physical entities. This leads to the final concept, uncertainty. Uncertainty relies upon Bayesian theory to characterize limitations in our state of knowledge about a physical/factual state rather than the state itself. This is what distinguishes uncertainty and error. The importance of this is that the characterization provided by uncertainty generally conveys something different about the inferences a result may support than does error. Uncertainty provides the limitations on what our current information, beliefs and models permit us to conclude about a unique event. This presentation will define and discuss various concepts which represent the limitations associated with scientific analysis, including specific types of error-rates and uncertainties. It will address several questions including: 1) whether the goal of forensic science in the courtroom should be to establish the presence or absence of actual error, or should it include a statement of the likelihood that errors may occur and the limitations of our state of knowledge; 2) whether it is always possible or even desirable to determine an "error rate" or uncertainty where such estimates may be not well-grounded and mislead or confuse fact-finders. The authors come from different professions and have different beliefs. The goal of this presentation is to explore and debate these and other questions from those differing points of view.

■ Speakers



Joseph P. Bono is a past-president of the American Academy of Forensic Sciences (AAFS). After four years on the adjunct faculty, in July 2011, he retired from a position as an instructor in the Forensic and Investigative Science Program at Indiana University Purdue University Indianapolis (IUPUI). Mr. Bono currently serves as an independent forensic science consultant to laboratories preparing for accreditation under the ASCLD/LAB International program. **Ted Vosk**, an attorney at law, gave part of the talk.

Session: Quality Assurance / Tuesday, July 21 / 2:30 – 3:00pm

Measuring Criminalistic Uncertainty

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/kadane_abstract.pdf (abstract only)

■ Abstract

There are many sources of uncertainty in criminalistics, including the integrity of the material at source, the chain of custody, the handling of the material in the lab, and the determination of the analyst. Some of these sources may be amendable to some degree of measurement and control, others may not. Reduction of total uncertainty rests on reduction of the most uncertain source. To the extent that laboratory uncertainty dominates, some experiments may shed light on, and permit reduction of, important sources of total uncertainty.

■ Speaker



Joseph B. (Jay) Kadane is the Leonard J. Savage University Professor of Statistics and Social Sciences, Emeritus at Carnegie Mellon University. One of the early proponents of Bayesian Statistics, Kadane has used the theory in both its decision-theoretic foundations and in problems of elicitation and computation to solve political science, law, physics, medicine and computer science problems. He is also well known for his contributions to the field of econometrics and for his work in applying statistics to decision-making. After joining Carnegie Mellon in 1971, he served as head of the Department of Statistics from 1972-1981 and instilled a balance within the department between theoretical and applied work. Kadane has authored more than 250 peer-review articles, and his latest book, "Principles of Uncertainty" won the 2011 International Society for Bayesian Analysis' DeGroot Prize. Kadane has been elected as Fellow of the American Academy of Arts and Sciences, the American Association for the Advancement of Science, American Statistical Association and Institute of Mathematical Sciences. He has served the statistical community in many capacities over the years; he is currently the vice chair of the American Statistical Association's Committee on Scientific Freedom and Human Rights and will become the committee's chair in January 2013. Kadane received his bachelor's degree in mathematics from Harvard University in 1962 and his Ph.D. in statistics from Stanford University in 1966. Prior to joining Carnegie Mellon's faculty, Kadane was assistant professor of statistics at Yale University (1966-1968) and a member of the professional staff at the Center for Analyses from 1968-1971.

Session: Crime Scene / Wednesday, July 22 / 11:00-11:30am

Forensic Science-the Quality Assurer?

Presentation Link:

http://www.nist.gov/director/upload/deforest_abstract.pdhttps://www.nist.gov/sites/default/files/documents/director/deforest_abstract.pdf (abstract only)

■ Abstract

Forensic science and the information obtained from a knowledgeable interpretation of physical evidence have a huge, relatively untapped, inherent ability to provide a quality assurance service to the criminal justice system which they serve.

The effects of cognitive bias on forensic science laboratory work and strategies to mitigate that bias continue to be discussed and researched, as well they should be. But laboratory testing is only a part of the physical evidence chain and the physical evidence chain is only a part of the investigative and adjudicative continuum. One should be at least equally concerned about the effects of cognitive bias on investigators, prosecutors, defense attorneys and the triers of fact. The information obtained from a knowledgeable interpretation of physical evidence has the ability to provide facts that are potentially far more reliable than other evidence such as eyewitnesses or confessions.

In order for the forensic sciences to properly provide this quality control function, several conditions need to be met. It is imperative that forensic scientists be present at the crime scene. They need to have authority over the items collected as well as which forensic examinations and analyses are requested. A forensic scientist's involvement is also needed to ensure a correct interpretation of the physical evidence record within the context of the case. Finally, strategies to ensure the efficient dissemination of this forensic science information to the investigators, prosecutors, defense attorneys, and trier(s) of fact need to be considered.

■ Speaker

Peter DeForest, John Jay College (retired), New York City, NY, United States

Session: Crime Scene / Wednesday, July 22 / 11:30-11:50am

Measurement Errors with Point Clouds

Presentation Link: not available

■ Abstract

Point clouds generated by laser scanners are composed of millions of points in 3 dimensional space, usually with attached intensities and sometimes with attached colors. Each scan point is represented by an idealized set of coordinates with no physical size, time, or uncertainty, giving the impression that the data was collected simultaneously with infinite precision and accuracy. For example, when zooming in on a point in a point cloud viewer, the point is usually displayed as a simple point occupying a single pixel, no matter how far the view is enlarged. Displaying point clouds in this manner ignores the physical reality that the point was measured using a laser beam of some diameter and pulse length with a given uncertainty in direction and range. The physical processes employed to collect the data may be removed from the presentation of the data, but the heritage of the points cannot be removed when considering the uncertainty of the measured points.

Using the example of computing the distance between any two points in a point cloud, one of the most basic operations and often the first step in using point clouds, we will explore the factors that contribute to the uncertainty of the distance. Two different cases are considered: first, the case where the points are both from the same scan position, and second, the case where the two points are from different scan positions. In the first case, the inherent properties of the instrument and how it is manufactured and calibrated, play a significant role in bounding the errors in the point cloud. Steps taken by Leica Geosystems in the production of their instruments to control these uncertainties are provided as an example, as are available in-field processes such as check and adjust, two-face target checks, and the inclusion of NIST traceable artifacts such as a twin-target pole. In the second case, the quality of the registration, that is, how point clouds from different positions, plays an additional role. How the laser scanner is deployed in the field and the data treated in the office now enters consideration. Some of these factors are explored in particular the use of tilt compensation, scanner placement, and proper tripods.

Standards in laser scanners are currently evolving and the breadth of applications for the point clouds generated by survey grade laser scanners is growing rapidly from its origins in commercial surveying. Until comprehensive standards are in place, a familiarity of the factors contributing to measurement errors is of interest. The processes developed by Leica for both the manufacture and deployment of survey grade laser scanners, as described in this talk, provide a path towards characterizing and controlling the measurement uncertainty. While the details of the talk are specific to Leica Geosystems, the physics and problems that are addressed are shared by every laser scanner point cloud no matter the origin. Certainly the question of determining the uncertainty of the distance between any two points in a point cloud is universally shared by anyone answering questions with point clouds.

■ Speaker

Dr. Gregory Walsh, Leica Geosystems AG, United States

Session: Crime Scene / Wednesday, July 22 / 1:30-1:50pm

Pre-flashover Fire Pattern Repeatability on Gypsum Wallboard

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/fire_pattern_repeatability.madrzykowski.crimescene.pdf (41 slides)

■ Abstract

In response to the National Academies' report, "Strengthening Forensic Science in the United States: A Path Forward", the National Institute of Standards and Technology (NIST) conducted a study to provide a technical starting point for understanding the repeatability of fire patterns. The primary objective of the study was assessing the repeatability of fire patterns on gypsum wallboard exposed to a limited range of source fires. The focus of this paper is an overview of the uncertainties of the measurements. The first step in determining the repeatability of the fire patterns was to obtain a measure of the repeatability of the fires that would generate the patterns. A variety of measurements were made to determine the repeatability of the source fires in terms of heat release rate, temperature, heat flux and flame height. Replicate source fire experiments were conducted in an oxygen consumption calorimeter in order to examine the repeatability of the fires in terms of heat release rate. The flame movement and height for each fire were recorded with photographs and videos for post-test analysis. The fire pattern experiments were conducted in a three-walled structure with a full floor and partial ceiling constructed from wood framing and lined with painted gypsum board. The source fires were positioned against the rear wall, midway along its length. Replicate experiments were conducted with each fuel. The fire patterns were documented and analyzed for repeatability. The fire pattern height results were then compared to the mean flame height results to examine the level of agreement. The end goal of this work was to guide the analysis of the fire patterns. The repeatability of the fire pattern cannot exceed the repeatability of the "fire that generated it". This research was conducted with the support of the National Institute of Justice (NIJ) and the NIST Law Enforcement Standards Office.

■ Speaker

Daniel Madrzykowski, National Institute of Standards and Technology (NIST), United States

Session: Crime Scene / Wednesday, July 22 / 1:50-2:10pm

Error Management in Fire and Explosion Investigations

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_management_in_fire_and_explosion_investigations.beyler.crimescene.pdf (14 slides)

■ Abstract

Fire and Explosion investigation has suffered over the years from investigative myths not borne out by science. This began to be recognized by the 1980's and by 1992, the National Fire Protection Association had published the first edition of NFPA 921, a Guide for Fire and Explosion Investigation with the goal of dispelling these myths. NFPA 921 introduced the scientific method as the backbone of the investigation process. It further, dispelled myths and introduced science-based methods into the investigation process. Now with its eighth edition published in 2014, one might imagine that the job is essentially done. Sadly, it is not.

The process of gaining acceptance of NFPA 921 in the fire and explosion investigation community was slow. It can be said that true acceptance within the profession was not achieved until 2000. Even so, that was 15 years ago and we still find serious quality issues in fire and explosion investigations. At the same time, by 2000 NFPA 921 can be said to have gained widespread acceptance in the courts as the standard of care in fire and explosion investigation.

The question must be asked, with a scientifically-based standard of care document that is well accepted in the profession and the courts, why and how could the quality of fire and explosion investigation continue to have serious quality issues? The answer to this question and the path toward error mitigation requires examination all of the elements that play a role in establishing and implementing investigation methods and managing errors. These include the following:

1. Scientific research basis underlying investigation methods
2. Standards and Guides Development
3. Education and Training
4. Certification and Accreditation
5. Report Writing and Review
6. Investigation Organization Management
7. Judicial Gatekeeping

The status of each of these areas will be reviewed. Required development will be identified. The activities by the OSAC Fire and Explosion Subcommittee will be reviewed.

Summary- There is a complex system of elements that contribute to high quality fire and explosion investigation that needs to be addressed to mitigate investigation errors. This goes well beyond the development of guides and standards, though these documents are foundational. A holistic approach seems to be indicated to reach the goal of high quality fire and explosion investigations with low and known error rates. A wide range of stakeholders must be involved in the process.

■ Speaker

Dr. Craig Beyler, Jensen Hughes, United States

Session: Crime Scene / Wednesday, July 22 / 2:10-2:30pm

How Fire Investigation Organizations Get in Trouble with Preventable Errors

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/how_fire_investigation_organizations_get_in_trouble.connealy_lentini.crimescene.pdf (15 slides)

Abstract

While organizations getting in trouble due to preventable errors is not unique to fire investigation groups, this presentation will focus on how these organizations can strategically minimize their risk exposure. The challenges faced by the Texas State Fire Marshal's Office (SFMO), Phoenix, Tucson, Jacksonville, and other fire investigation organizations around the United States share many common practices that yielded very bad outcomes. Errors can never be completely eliminated when organizations are comprised of human beings. It is difficult for any organization to avoid all preventable errors even when best practices are followed. Every organization needs to have a system in place for dealing with errors when the inevitable happens. As much as the public would like it to be so, the criminal justice system is not infallible. The number of overturned arson convictions in recent years demonstrates the need for a thorough, prevention oriented self-assessment. The stakes are too high for all concerned to ignore the issues we face. The importance of implementing strong leadership cannot be overstated. Fire investigators need to have leaders who are committed to continual improvement, and have the ability to acquire the resources for staff to comply with best practices. Fire investigation organizations must have progressive leaders who ensure that national standards and the latest scientific knowledge are incorporated into their training programs. Professional development plans must be in place for advancement within the organization that keep the unit continually going in the right direction. Personnel should be evaluated on knowledge of NFPA 921 annually. Mock Daubert hearings can be used to determine if staff can survive a reliability challenge. Education and training are at the core of continual improvement. Training programs should include participants passing an exam at the end of each training course, and organizations should encourage and support their line investigators improving their education. Unfortunately, the applicant pool for many fire investigation organizations includes many individuals with little or no scientific education. This needs to change. Retroactive reviews of cases should be incorporated to determine if new scientific advances or potential errors occurred that should be brought to the attention of the criminal justice system. Even more important, however, is putting in place independent technical reviews of important cases before they progress to criminal prosecution. The courtroom is not the place for quality assessment. There is also a critical need for public fire investigators to be able to utilize scientists and engineers to consult with on their cases. Too often, this is not available unless private fire investigators are brought in typically through insurance firms. There is much more work to be done in this area. This presentation will provide insight into real-world successes that have been implemented in Texas, and can be implemented in other organizations. The field of fire investigation has been under great scrutiny over the past ten years and the expectations of stakeholders continue to increase. The better path is to avoid negative attention through a commitment to excellence.

Speakers



Chris Connealy was appointed as the Texas State Fire Marshal on June 18, 2012. He previously served as fire chief of the Cedar Park Fire Department from 2004 to 2012. Chris has been in the fire service 37 years. He initially served 26 years with the Houston Fire Department and retired January 2004. He promoted through the ranks to fire chief. Connealy served as HFD's Accreditation Manager while an assistant chief and the department became the largest accredited fire department in the world in 2001. He has associate, bachelor, and master degrees. Chris completed the Harvard University State and Local Executive program in 2003. He is accredited as a Chief Fire Officer and is a National Fire Academy EFO graduate. Chris was named a 2014 Public Official of the Year by Governing Magazine and was a Dallas Morning News 2013 Texan of the Year Finalist. He was appointed by Governor Perry and served as presiding officer on the Texas Commission on Fire Protection, served on the Texas Fire Chiefs Association executive board, and received various additional awards during his career in the fire service. He is the first fire chief in the United States to have lead two cities (Houston and Cedar Park) to an ISO Class 1 rating. Chris serves as an adjunct instructor for TEEX. He loves his profession and serving others.



John Lentini is one of a handful of people certified to conduct both fire scene investigations and fire debris analysis. He has personally conducted more than 2,000 fire scene inspections and has appeared as an expert witness on more than 200 occasions. He is a frequent invited speaker on fire investigation science, and an active proponent of standards for fire and other forensic investigations. He is a member of the NFPA Technical Committee on Fire Investigations, and has served three terms as chair of ASTM Committee E30 on Forensic Science. John is the current Chairman of the AAFS Criminalistics Section. He also serves on the NIST/OSAC Subcommittee on Fire and Explosion Investigations. He is now an independent consultant living in the Florida Keys and doing business as Scientific Fire Analysis. His book, *Scientific Protocols for Fire Investigation*, now in its second edition was published by CRC Press in 2013. His website is www.firescientist.com.

Session: Criminalistics I / Wednesday, July 22 / 11:00-11:20am

Arising Issues Implementing Advances in DNA Technology May Tarnish the Gold Standard

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/jeanguenat_abstract.pdf

■ Abstract

Technological advancements in forensic DNA testing are continuously evolving and demonstrate promise to aid in criminal investigations. New DNA processes are being implemented that are more sensitive, software is available to interpret and perform statistics on complex mixtures that historically cannot be resolved, and an individual's phenotype can now be generated along with identity markers. As crime laboratories shift to expanded loci kits, probabilistic genotyping and massive parallel sequencing (MPS), the forensic community must carefully consider the implications and take steps now to responsibly implement these advances.

In 2017, the National DNA Database (CODIS) is expanding from 13 loci to 20 loci. In response, vendors have developed amplification kits that include 24 loci and are also generating full DNA profiles at a sensitivity of 100 pg (~17 cells). Kits now include 10 miniSTRs (loci <200bp) to improve sample recovery and overall achieve an 8 fold increase in the power of discrimination over current methods. Most laboratories do not have experience navigating data generated by miniSTRs or low copy number analysis. While guidelines were issued by SWGDAM in 2014 to address concerns about enhanced detection methods, crime laboratories should acknowledge that the new kits introduce enhanced abilities and be prepared to address this issue.

Mixture interpretation has proven to be difficult, subjective, and variable between humans. Probabilistic genotyping software is allowing the task of interpreting mixtures to be performed by algorithms that can predict outcomes and provide statistical power over current methods. However, if the DNA community does not understand the underlying statistical formulas and cannot come to a consensus on how to interpret mixtures, removing the human element completely creates a black box that may diminish the meaning of a match and the given statistical weight.

Massive Parallel Sequencing is moving the field beyond identity markers to include obtaining information on ancestry, DNA phenotyping, and genetic predispositions. MPS can be powerful and change the way crime is solved but will require careful coordination to properly implement. How the increase in knowledge provided by MPS is used and accepted by legislation and the Court system will steer the course for the future of DNA analysis. Ignoring possible ramification by using this technology for 'investigative leads' only, may be irresponsible considering all foundations used to identify a possible perpetrator may warrant an admissibility hearing.

This presentation will engage the DNA community to discuss the quality issues associated with implementing more sensitive amplification kits, probabilistic genotyping, and MPS. Weaknesses in DNA technological advancements will be identified in order to overcome and improve implementation strategies in the crime laboratory. While the DNA community has the 'power of more' information available now ensuring 'right' information will mitigate potential issues and maintain the gold standard status.

■ Speaker

Amy Jeanguenat, Bode Cellmark Forensics, United States

Session: Criminalistics I / Wednesday, July 22 / 11:20-11:40am

NIST Concordance Evaluations to Assist in the Improvement of Commercial STR Multiplexes

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/nist_concordance_eval_to_assist_inthe_improvement_of_commerical_str_multiplex_kits-steffen-crim1.pdf (36 slides)

■ Abstract

Concordance evaluations of prototype short tandem repeat (STR) multiplex kits are important to detect allelic dropout or null alleles present in a data set. These studies are performed due to a variety of available commercial STR kits utilizing different configurations of PCR primers. When multiple STR kits typing similar core loci are used within the community, there is concern that allele dropout may occur due to PCR primer binding site mutations that may affect one set of primers but not another. These null alleles become evident only when results from typing a constant and representative set of samples are compared. Null alleles are a concern because they could result in a false-negative or incorrect exclusion of two samples that come from a common source (only if different PCR primers are used).

Multiple concordance studies have been performed at NIST with a standard sample set of in-house U.S. population samples (~1450) and the six components of Standard Reference Material (SRM) 2391c: PCR-Based DNA Profiling Standard. These concordance studies have included STR multiplex kits from Life Technologies, Promega, and Qiagen, which include the new generation large STR multiplex kits. Several discordant results have been identified using concordance software developed at NIST, confirmed by DNA sequencing and reported to the forensic community on the null allele web page of STRBase: <http://www.cstl.nist.gov/strbase/NullAlleles.htm>. The results of concordance studies are reported to the commercial companies which allow them the opportunity to resolve the issue prior to official product release to the forensic community.

A summary of the results, including discordance and sequencing results, will be shown in order to help assess the benefits of performing concordance testing using a standard data set with STR multiplex kits that have different PCR primer sequences for the same markers.

■ Speaker

Carolyn Steffen, Applied Genetics Group, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Wednesday, July 22 / 11:40-12:00pm

Application of a DNA Quantitation Standard for Human Identity Testing

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/srm-2732_human_id_testing-romsos-crim1.pdf
(24 slides)

■ Abstract

Commercial short tandem repeat (STR) assays used by the forensic human identity community require tight control of the amount of sample DNA amplified in the polymerase chain reaction (PCR). This requires the ability to reproducibly measure the concentration of human DNA in a casework sample extract prior to input in the PCR reaction. Approximately 500-1000pg of input DNA will provide an interpretable profile. Commercially available quantitative PCR (qPCR) kits routinely are relied upon to determine the concentration of casework extract within forensic laboratories; however, assays employed rely upon commercial DNA standards for relative quantitation estimates. The NIST SRM 2372 Human DNA Quantitation Standard was produced to support the need for a human-specific DNA quantitation standard in forensic casework as a calibrant for commercially produced DNA standards. The use of SRM 2372 is intended to enable the comparison of DNA concentration measurements over time and laboratories. Manufacturers can use SRM 2372 to validate the values assigned to their own reference materials. Individual forensic laboratories can use SRM 2372 to validate DNA quantitation methods and to verify the assigned concentration of in-house or commercial DNA calibration standards.

■ Speaker

Erica Romsos, Applied Genetics Group, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Wednesday, July 22 / 1:30-1:50pm

Effect of Drop-in on False Positive and Rank-Order Likelihood Ratios (LR) Calculated for a Mixture of Touch DNA

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/effect_of_dropin_on_false_positive_and_rank_order_likelihood_ratios_lr_calculated_for_a_mixture_of_touch_dna.hughes.crim1_.pdf (27 slides and 11-page white paper)

■ Abstract

We have examined the LR statistical approach to analysis of low template DNA mixtures exhibiting drop-out and drop-in used in casework by the NYC Office of Chief Medical Examiner Department of Forensic Biology (OCME), called Forensic Statistical Tool (FST). We used Microsoft Excel to create a program to reproduce the results of FST on a three-person mock case mixture. We used the Excel add-on DigDB to create the roughly 200,000 possible combinations necessary to recreate the FST results of a suspect comparison to this mixture.

Allelic drop-in is defined by John Butler as contamination from an unknown source. In the FST program artifactual stutter alleles are lumped into the drop-in probability. There is general agreement in the field that drop-in should be a rare event. FST values for the probability of drop-in were derived from empirical studies using mixtures of pristine single-source buccal swab extracts. For 31-cycle amplification, FST expects the low probabilities of 0.035 single allele drop-ins per locus, and 0.005 drop-in of two or more alleles per locus.

We examined the effects of drop-in on the calculated LR using an OCME laboratory validation sample created by three known touch DNA contributors. We also examined one of the false positives for this sample found in a search of 1256 database profiles, including the NIST population database.

The mixture we examined contained 60pg of DNA in each of three amplifications, and was called a deducible 3-person mixture by the OCME analyst. This meant that FST used drop-out rates determined for the minor contributors to a 5:1:1 ratio mixture. The drop-in rates were as described above. The mixture generated false-positive associations with known non-contributors. The highest LR for database sample was 157, for sample from the NIST Caucasian database. By contrast, the two minor known contributors had LR of 4 and 10e-5.

This mixture provided a good example of drop-in. Out of 78 total alleles in the mixture, 20 were drop-in alleles not associated with any known contributor. 13 drop-in alleles were found in the 59 alleles in the consensus profile, being present in at least two of three amplifications. Out of the 45 individual loci, 14 loci showed one drop-in allele. An additional 13 loci showed two or more drop-in alleles. The FST drop-in rate predicts approximately 2 loci with one drop-in. FST expects no loci with drop-in of two or more alleles. Thus, the mixture had 10 times more foreign and stutter alleles than the FST program predicts.

Analysis of the mixture showed that drop-in affects LR in the same way that has been demonstrated for drop-out. Drop-in caused a strong false positive LR, which was higher than the LR for the known minor contributors.

■ Speaker

Clinton Hughes, The Legal Aid Society of New York City, United States

Session: Criminalistics I / Wednesday, July 22 / 1:50-2:10pm

Error Rates in Probabilistic Genotyping Software for DNA Mixtures in Human Identification - How to Compare?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_rates_in_probabilistic_genotyping_software_for_dna_mixtures_in_human_id.coyle_crim1.pdf (15 slides)

■ Abstract

DNA for human identification is and will remain a gold standard for criminal and civil casework. Although DNA is powerful and convincing technology, there is an inherent error rate associated with DNA mixture analysis methods whether computed manually or with software. The use of probabilistic genotyping software programs for the analysis of complex DNA mixtures is gaining momentum in certain regions of the United States and little comparative information exists in the published literature on error rates with comparable data sets using the same reference databases. While manufacturers claim they are able to powerfully extract the DNA profile of an individual from a complex mixture that was formally reported as 'inconclusive' using manual methods; the legal community has queried back as to how scientifically accurate this form of analysis could be with a request for the error rates to be reported along with the likelihood ratio for a fair and accurate representation of the science to the trier of fact. In reviewing validation studies, scientific articles and testimony about a novel analytical software program called Forensic Statistical Tool (FST), available only at the Office of the Chief Medical Examiner in New York City, sources of error were identified to explain the high false positive match rate for non-contributors. The false positive match (LR > 1 and as high as 10,000) can be as frequent as 1 in 1200 individuals or 0.08%. In comparison, True Allele software has an error rate of 1 in 20,000 or 0.005%. When investigating the reasons for these differences in scientific accuracy, they result from FST computer modeling based on drop-out rates, analytical thresholds, and use of local city database allele frequencies, along with confounding contamination events. Problematic software settings with FST include the difficulty in establishing the true number of contributors to the sample in duplicate stochastic amplifications of low template quantity and quality. Random DNA degradation and coincidental allele sharing cannot be effectively modeled with FST and influence the error rate. Individuals with low LR values for DNA inclusion have been acquitted from touch DNA cases due to the high false positive rates associated with the FST software package and resultant scientific uncertainty for the inclusion of a potential DNA donor to these randomly touched objects. Due to the differences in mathematical modeling and error rates with probabilistic software programs, it would be desirable to communicate these during testimony and include in written forensic reports.

■ Speaker

Dr. Heather Miller Coyle is an Associate Professor in the Forensic Science Department at University of New Haven, a small private University located in West Haven, CT. She obtained her B.S. in In Vitro Cell Biology from State University of New York – Plattsburgh in 1986 and her Ph.D. in Plant Biology from University of New Hampshire in 1994. Her work experience includes employment in the pharmaceutical industry (Boehringer Ingelheim Pharmaceuticals; Ridgefield, CT) and the DNA unit of a forensic science laboratory (Division of Scientific Services; Meriden, CT). She is also a private consultant for DNA review and independent quality control of forensic laboratory casework on request and is court qualified to testify in 5 states and federal court. Her research interests focus on touch DNA cases, trace biological evidence and DNA quality control issues.

Session: Criminalistics I / Wednesday, July 22 / 2:10-2:30pm

Objective DNA Mixture Information in the Courtroom: Relevance, Reliability and Acceptance

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/objective_dna_mixture_information_in_the_courtroom.perlin.crim1_.pdf (24 slides and 27-page white paper)

■ Abstract

DNA mixtures arise when two or more people contribute their DNA to a biological sample. Mixtures are seen in sexual assault kits, homicide evidence, handguns and other "touch DNA" surfaces. With advances in detection technology, they have become the predominant form of DNA evidence in many crime laboratories. While DNA from one person is easy to interpret, mixture data has complex patterns comprising many allele peaks of varying height. One person's DNA produces either one allele peak, or two of similar height, so a height "threshold" is meaningful. But data-simplifying thresholds fail to give accurate results when applied to complex mixture patterns. Ten years ago, NIST demonstrated a ten order-of-magnitude match statistic discrepancy between crime laboratories analyzing the same mixture data. Mixture "inclusion" analysis tests whether a subject's alleles are included in a set of (thresholded peak) alleles, but is inherently subjective to the analyst sees the subject's genotype during the analysis. An entirely objective (and potentially more informative) approach is to first separate out the genotypes of each mixture contributor without ever seeing the subject, and only afterwards make a comparison. This can be accomplished by sophisticated computing that considers many thousands of genotype alternatives, and how well their linear combinations explain the quantitative data. Multiple possibilities for a contributor genotype are assigned probabilities. Faithful modeling of the laboratory process can yield genotypes that accurately preserve DNA identification information. Comparison of a separated evidence genotype with a subject's reference genotype, relative to a population, gives a match statistic. This statistic is a simple ratio $\frac{\text{probability of genotype match}}{\text{the probability of random match}}$ divided by the random match probability. It is also a likelihood ratio (LR), or Bayes factor (BF), a standard measure of information change based on observed evidence. The LR is mathematically probative because it assesses how evidence data affects a hypothesis (i.e., whether the subject contributed their DNA to the mixture). And the LR's assessment is nonprejudicial, because (as a BF) the ratio factors out prior belief about the hypothesis. Thus genotype separation addresses FRE 403 relevancy balancing. The reliability of objective genotype separation has been extensively tested for at least one such system. Dozens of independent and developmental validation studies have been conducted, with seven peer-reviewed publications. These studies use the LR as an objective information measure to assess the method's sensitivity (true positives), specificity (false positives) and reproducibility (close numbers). This extensive testing, error rate determination, and scientific peer-review address FRE 702 and Daubert reliability factors. Courts have accepted this extensively validated computer method, which has withstood Daubert and Frye challenges in six states. Admissibility has been upheld at the appellate level. Separated genotypes provide easy to understand results. Objective DNA analysis elicits identification information from evidence. Validation establishes accuracy and error rates. Courts require solid science extensively tested and empirically proven to promote criminal justice, societal safety, and conviction integrity.

■ Speaker

Dr. Mark Perlin, Cybergenetics, United States

Session: Criminalistics I / Wednesday, July 22 / 2:30-3:00pm

Validating Software for Probabilistic Genotyping

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/validating_software_for_probabilistic_genotyping.buckleton_gittelson.crim1_.pdf (22 slides presented by Simone Gittelson)

■ Abstract

There is considerable interest in the validation of software for probabilistic genotyping. Initiatives are underway at SWGDAM and by the ISFG DNA commission. This has brought into focus questions such as what constitutes adequate validation, which parts can be done by the developers and which parts should be done by the laboratory implementing the software. We have asked ourselves what constitutes general acceptance within the US jurisdiction. More broadly we have been considering what responsibilities a developer carries to assist with competent usage, and what constitutes reasonable defense access. We argue for the limited usefulness of open source as an answer all to defense access.

These subjects are discussed in the light of recent experiences both positive and negative. The Frye rule states the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs which invites an analysis of what is the thing and what is the inference. This is dissected from a developer's perspective. We conclude that "the thing" must mean the principles of molecular biology and probabilistic inference. The inference must be the LR. Neither of these is the software.

■ Speaker

John Buckleton, ESR, New Zealand (he did not attend and his slides presented by Simone Gittelson, NIST Statistical Engineering Division, National Institute of Standards and Technology, United States)

Session: Criminalistics I / Wednesday, July 22 / 3:30-3:50pm

Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S. (MIX13)

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/interpretation_errors_detected_in_a_nist_interlab_study_on_dna_mixture_interpretation_in_the_us_mix13.coble_crim1.pdf (38 slides)

■ Abstract

Complex DNA mixtures from more than two individuals and/or profiles with small quantities of DNA, can be challenging for the analyst to interpret. Many laboratories in the U.S. have created their profile interpretation protocols based upon internal validation studies, best practices outlined in publications from the scientific literature, training workshops, and guidance from scientific groups such as SWGDAM and the DNA Commissions of the ISFG.

Interlaboratory studies allow a "big picture" view across different laboratories and can be used to measure consistency among analysts using the same protocol within the same laboratory. Since 1997, NIST has conducted a number of DNA mixture interlaboratory studies for the forensic community. In 2005 Butler and Kline [1] conducted an interlaboratory study (MIX05) with electronic data that showed a wide range of variation between and within the 69 participating laboratories.

After the publication of the 2010 SWGDAM Autosomal STR Interpretation Guidelines [2], many laboratories in the U.S. validated and established analytical and stochastic thresholds for their mixture interpretation protocols. The MIX13 interlaboratory study was designed to determine the 'lay of the land' within the U.S. in regards to STR mixture interpretation, subsequent to the publication of those guidelines. Examples of errors detected from the five profiles used in the MIX13 study from the 108 participating laboratories will be presented. Ideas for future training and research to improve mixture interpretation and reporting will be discussed.

1. Butler JM and Kline MC (2005) "NIST Mixture Interpretation Interlaboratory Study 2005 (MIX05)" poster at the 16th International Symposium on Human Identification (Grapevine, TX), September 27-28, 2005.
2. SWGDAM (2010) Autosomal STR Interpretation Guidelines:
http://www.swgdam.org/Interpretation_Guidelines_January_2010.pdf

■ Speaker

Mike Coble, Applied Genetics Group, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics I / Wednesday, July 22 / 3:50-4:10pm

A Large-scale Study of DNA Mixture Interpretation: Inter- and Intra-laboratory Variability

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/DNA_mixture_interpretation_study_DEAT-aranda-qa.pdf (24 slides)

■ Abstract

Forensic laboratories generally deconvolute DNA samples to generate genotype profiles present in a mixture sample. Sample complexity influences whether the resultant data is deconvoluted or deemed uninterpretable. As the complexity of a sample increases, so does the range of genotype interpretation generated by a DNA examiner. In general, the DNA forensic community recognizes that the variation in DNA mixture interpretation is dependent on many variables, including but not limited to: laboratory standard operating protocols, laboratory training, and the use of DNA mixture interpretation management tools and examiner subjectivity. Although variation is recognized to exist, the degree of mixture interpretation variation in the DNA forensic community is unknown. Thus, laboratories are unaware of the degree of mixture interpretation variation that exists within their laboratory, what is an acceptable range of variation compared to other DNA forensic laboratories, and how that variation can be reduced. In addition, DNA forensic laboratories are unaware of the error rate (mismatched genotypes) which is dependent on sample complexity. The Defense Forensic Science Center (DFSC) conducted a large-scale study to survey and quantify the variability that results from the DNA mixture deconvolution process using six (6) mixtures. The study involved over 180 DNA examiners from 50 DNA forensic laboratories and quantified the variation within and between DNA forensic laboratories using the automated DNA Examiner Assessment Tool (DEAT) developed at DFSC that calculates both the genotype interpretation metric (GIM) and the Allelic Truth/False (AT/AF) score. The GIM establishes a gradient to measure the multitude of potential genotype interpretations, converting the genotype interpretation precision into a quantifiable metric. The AT/AF score measures accuracy and error rates in calculating true and mismatched alleles. Six DNA mixtures were generated and varied in complexity level by altering the number of contributors, contributor ratios, frequency of allelic dropout, and level of stutter. Examiners analyzed the data and genotype interpretations of .fsa files provided to them. The techniques used to submit responses to DFSC included CPI/CPE, RMP, and Likelihood Ratio methods. Interpretations analyzed prior to technical review were collected, with GIM and AT/AF scores calculated for all six mixtures. Analysts were asked to identify the number of contributors, proportion of each contributor, and a final match statistic in each mixture sample. Results were individually compared within and outside their laboratory; the collective score of large labs were also compared. The results of this study may shed light on the sources, range, and acceptable range of mixture interpretation variation that currently exists in the DNA forensic community. These findings may inform and influence training programs for DNA examiners with the goal of reducing variation in the various interpretation methods for new and existing DNA examiners.

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense. Names of commercial manufacturers or products included are incidental only, and inclusion does not imply endorsement by the authors, DFSC, OPMG, DA or DoD.

■ Speaker

Roman Aranda IV, Defense Forensic Science Center, United States

Session: Criminalistics I / Wednesday, July 22 / 4:10-4:30pm

Defining the Limits of Forensic DNA Profile Interpretation: An Assessment of the Information Content Inherent in Complex Mixtures

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/defining_the_limits_of_forensic_dna_profile_interpretation.inman_crim1.pdf (16 slides)

■ Abstract

With the increasing sensitivity of DNA typing methodologies, as well as increasing awareness by law enforcement of the perceived capabilities of DNA typing, complex mixtures consisting of DNA from two or more contributors are increasingly being encountered. However, little research has been conducted to determine whether it is possible to distinguish a true contributor from a non-contributor in these complex samples, and under what specific conditions. In order to investigate this question, sets of six 15-locus Caucasian genotype profiles were simulated and used to create mixtures containing 2 to 5 contributors. Likelihood ratios were defined for various situations, including varying numbers of contributors and unknowns, as well as known non-contributors. This resulted in fourteen different sets of hypotheses, each of which was tested with both contributors and non-contributors; each was run 10,000 times through the experimental simulation. These experiments were intended to illustrate the best-case scenario, in which all alleles from the true contributors were detected in the simulated evidence samples. Therefore the possibility of drop-out was not modeled for this experiment. The computer program DNAMIX was then used to compute LR_s for all of the experimental conditions, i.e. true contributors (TC) and known non-contributors (KNC), assuming varying numbers of unknown individuals in the mixtures. This resulted in 140,000 LR_s for each of the two experimental conditions, which were assessed and compared to the known ground truth input profiles. These complex mixture simulations show that, even when all alleles are detected, (no drop-out), TCs can generate LR_s less than 1 across a 15-locus profile. However, this outcome was rare (7 of 140,000 replicates (0.005%)), and associated only with mixtures comprising 5 contributors in which the numerator hypothesis includes one or more unknown contributors. For KNCs, LR_s were found to be greater than 1 in a small number of replicates (75 of 140,000 replicates (0.05%)). These replicates were limited to 4 and 5 person mixtures with 1 or more unknowns in the numerator. Only 5 of these 75 replicates (0.004%) yielded an LR greater than 1,000. Thus, overall, these results imply that the strength of evidence that can be derived from complex mixtures containing up to 5 contributors, under a scenario in which no drop-out is required to explain any of the contributors, is remarkably high. This a useful benchmark result on top of which to layer the effects of additional variables, such as drop-out, contributor ratios, shared alleles, and other variables.

■ Speaker

Keith Inman holds a B.S. and M.Crim., both from the University of California at Berkeley. In his professional career he has been employed as a criminalist by the Orange County Sheriff's Department, the Los Angeles County Sheriff's Department, the Los Angeles County Chief Medical Examiner-Coroner, the Oakland Police Department, and most recently the California State Department of Justice DNA Laboratory. He also previously worked in private practice for six years, undertaking both prosecution and defense work. He has co-authored *An Introduction to Forensic DNA Analysis*, a book that has become the preeminent reference for both attorneys and crime laboratories, and *The Principles and Practice of Criminalistics*, a book aimed at practicing criminalists. His research and practice focuses on the interpretation of complex DNA samples, as well as crime scene investigation and reconstruction. He is a frequent speaker at forensic and legal conferences around the world, including the 2012 European Academy of Forensic Science meeting in The Hague, and The Royal Society in London in 2015. He teaches a variety of classes in the Criminal Justice Administration Department at California State University, East Bay, including Basic and Advanced Criminal Investigation, Criminal Identification, and Survey of Forensic Sciences.

Session: Criminalistics I / Wednesday, July 22 / 4:30-5:00pm

Errors in Interpretation of DNA Profile Data

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/errors_in_interpretation_of_dna_profile_data.word_crim1.pdf (20 slides)

Abstract

Throughout the history of forensic DNA identification testing, there has been a concerted focus on ensuring the accuracy of the DNA test results generated using various types of DNA testing procedures, commercially-available test kits and instrumentation. This has been done through carefully designed and implemented validation studies and standard operating procedures along with the requirement for the use of critical controls and other important quality assurance measures focused on the technological aspects of data generation. There has been minimal emphasis on the quality aspects of data interpretation. This is perhaps due in part to the majority of the DNA profiles generated during the first 20 years of forensic DNA testing consisted of results from only a single individual or two person DNA contributors, often where one of the contributors to the mixture was known. Even with a partial profile, a non-contributor to a single source or two person DNA profile could often be easily excluded as a possible contributor.

Today, however, the bulk of the samples tested in many crime laboratories worldwide contain small amounts of DNA (e.g., low template DNA) and often from several unknown contributors. This change in the types of evidence samples tested along with modifications to testing parameters has led to the generation of DNA profile data that may contain stochastic alterations which cannot be independently reproduced, and introduces a substantial increase in the risk of misinterpreting data. As a result, data and conclusions may be provided in reports and during testimony that are misleading or incorrect, including the reporting of false inclusions or false exclusions. The potential for misrepresenting the meaning of the data as it relates to a particular case has also significantly increased. During this presentation types and causes of DNA data misinterpretations will be discussed along with some suggestions for improving the quality of the data generated and interpretations provided. The need for and recommendations regarding relevant validation studies, proficiency testing and increased training will be discussed as well as the need for improved communications between DNA analysts and law enforcement, attorneys and the court.

Speaker

Dr. Charlotte Word is currently a consultant in Human DNA Identification testing. She joined Cellmark Diagnostics in Germantown, MD (later called Orchid Cellmark) in 1990 and over the 15 years she worked there had several titles, including Deputy Director and Senior Manager, Forensics and Laboratory Director. She has performed technical reviews on many 1000's of cases from several public and private DNA testing laboratories and has participated in the validation of various DNA test systems. Charlotte has testified as an expert witness in over 300 trials and admissibility hearings around the country since 1990. She has published and given many presentations at meetings and workshops in the areas of DNA identification testing. In 1998-1999 she was a member of the Post-Conviction Issues Working Group of the National Commission on the Future of DNA Evidence. She has experience with ASCLD/LAB, NFSTC, DNA Quality Assurance Standards and ISO audits. Charlotte is a member of the editorial board of *The Journal of Forensic Sciences* and has been a guest reviewer for *Forensic Science International: Genetics*. For the past few years, she has been assisting with Mixture Interpretation training through a grant from NIJ awarded to Boston University. She is a member of the DNA Analysis 2 (DNA Reporting and Interpretation) Subcommittee of the Biology/DNA Scientific Area Committee of the Organization of Scientific Area Committees (OSAC) and of the Reporting and Testimony Subcommittee of the National Commission on Forensic Science. Charlotte received her B.S. in Biology from the College of William and Mary in Virginia and her Ph.D. in Microbiology, with specialties in Molecular Biology and Immunology from the University of Virginia. She did postdoctoral research as a Damon Runyon-Walter Winchell Cancer Fund Fellow at the University of Texas Southwestern Medical School in Dallas, TX and was a faculty member at the University of New Mexico School of Medicine prior to joining Cellmark Diagnostics.

Session: Criminalistics II / Wednesday, July 22 / 11:00-11:20am

New Psychoactive Substances (NPS): How to Keep Up with the Never Ending Need for New Reference Materials?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/psychoactive_substances_need_for_reference_materials.collins.crim2_.pdf (23 slides)

■ Abstract

Lack of chemical reference materials is potentially a major cause of error in forensic chemistry. Since 2009 the number of new designer drugs including synthetic cannabinoids and cathinone derivatives has been increasing and reached a point where a new substance was appearing every week. From 1960 to 2008 approximately 200 substances were controlled. Between 2008 and 2015 that number had risen to almost 600! This dramatic increase has impacted law enforcement, health authorities, the justice system and forensic science. When completely new substances first appear their molecular structure is determined through a combination of nuclear magnetic resonance and mass spectroscopy usually at a university or a forensic centre of excellence. Ultimately, however, it is necessary to create certified reference materials (CRMs) for forensic chemists who typically rely on chromatographic techniques in hundreds of laboratories around the world. Keeping up with the demand for CRMs during this unprecedented explosion in NPS has been problematic. In Australia the National Measurement Institute (NMI) collaborated with the Australian Federal Police to address this situation through a program designed to prioritize the reference material needs of Australian forensic facilities. A massive program of synthesis and certification of NPS was begun to deliver CRMs to forensic laboratories.

■ Speaker

Dr. Michael Collins, Australian National Measurement Institute, Australia

Session: Criminalistics II / Wednesday, July 22 / 11:20-11:40am

Error Analysis of a Forensic Controlled Substance Case

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_analysis_of_a_forensic_controlled_substance_case.harris.crim2_.pdf (16 slides)

■ Abstract

This presentation will discuss sources of error in a typical forensic controlled substance case. This presentation will include some examples of errors in casework and some approaches to measuring and mitigating those errors.

Understanding and mitigating error is a critical component of good science, including forensic science. In *Daubert v. Merrill Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), the U.S. Supreme Court recognized the importance of measuring and reporting an error rate for a scientific analysis by including error rate in the list of factors to consider when evaluating the admissibility of scientific evidence. A quantitative error rate is a clear and easy measure of the reliability of a scientific analysis, especially for non-scientists such as judges and juries. The analysis of a controlled substance case is a unique set of tasks and decisions, each of which is susceptible to some type of error. However, error rates have not been determined for the individual tasks in a controlled substance case or for the analysis overall, primarily because of the qualitative nature of most controlled substance analyses. The tests utilized in a controlled substance case are designed to provide responses in a subjective binary manner rather than an objective quantitative manner. Personal interpretations and biases influence these tests and complicate the measurement of their error rates. Despite the challenge of measuring overall error rates due to the subjective nature of controlled substance casework, some tasks involved can be distilled into a measurable form. These include measurement of the weight of the evidence; proper generation, management and correlation of test aliquots; proper transcription of test aliquot information into data systems; and false positive/false negative identification rates. Approaches to error analysis in other forensic disciplines as well as other industries can be applied to these tasks to calculate an estimated error rate for a controlled substance case. Strategies for mitigation of errors can also be explored through an error analysis.

Upon its conclusion, this presentation will impact the forensic science community by educating administrators, managers, scientists and attorneys as to the potential of errors in forensic controlled substance casework and providing some suggestions as to the measurement and mitigation of these errors.

■ Speaker

Heather Harris is a certified independent forensic chemistry consultant and a PA licensed attorney. Ms. Harris holds a Master of Forensic Science from The George Washington University. From GWU, she went to the Bexar County Criminal Investigation Laboratory in San Antonio, Texas, where she was a Forensic Scientist working primarily in drug analysis and clandestine lab analysis. Upon relocating to Philadelphia, Ms. Harris began working at NMS Labs as a Forensic Chemist. It was also at this time that she attended Temple University's Beasley School of Law Evening Division. After successfully passing the PA bar exam, Ms. Harris began working independently as a forensic chemistry consultant for the Defender Association of Philadelphia, the Public Defender Service for the District of Columbia, and private attorneys in both criminal and civil cases. In addition to these duties, she teaches forensic chemistry and trace evidence as an adjunct professor at Arcadia University in the Master's of Science in Forensic Science program. Ms. Harris actively participates in professional organizations, such as the American Academy of Forensic Sciences, the Northeastern Association of Forensic Scientists, and the American Board of Criminalistics, by whom she is certified as a Diplomate in General Criminalistics and Drug Analysis.

Session: Criminalistics II / Wednesday, July 22 / 1:30-1:50pm

Positive identification starting with a skull visualized by a 3D scanner using image superimposition and the 3D Max software: Cases developed from 2004 to 2014 in the Forensic Science and Legal Medicine Institute of the Public Prosecutor's Office

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/molina_abstract.pdf

■ Abstract

The present article includes 50 cases notable for their international and national media impact, which occurred between 2004 and 2014. The study used digital analysis methods to compare photographs of skulls with photographs of possible victims. The method achieved 100% success as verified by established genetic (DNA), fingerprint, and odontogram analyses. The digital method has a significant advantage over the established methods because the identifications are completed more speedily, saving up to several months, with no loss of accuracy. The use of this method will make a significant contribution to the humanitarian impact of the work and possibly speed the processes of justice.

The present research was useful not only to identify the skulls, but also to study the facial morphology for identification of the fans or delinquents recorded by closed circuit cameras which would help in identifying the criminals by superimposing the images.

The objective of the research is to assess the effectiveness of two morphometric analyses in establishing the personal identities of skulls recovered from crime scenes.

The study compares the superposition of photographs of skulls on photographs of possible victims using software Adobe Photoshop with superimposition of photographs from different angles on images from the 3D scanner brand Polhemus model Fastcan Scorpion.

The study included images of 50 Peruvian cases of both sexes. The outcomes showed a margin for error of 3-2%, because of the aging of the photographs, for the 3D Studio Max software. Three solved cases by means of this equipment permit us to conclude that two morphometric analysis stages are necessary to produce the positive identification.

■ Speaker

Dr. Danny Jesus Humpire Molina, Forensic Science and Legal Medicine Institute of the Public Prosecutor's Office, Peru

Session: Criminalistics II / Wednesday, July 22 / 1:50-2:10pm

3D Fingerprint Targets

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/3d_fingerprint_targets.paulter.crim2_.pdf
(23 pages)

Abstract

Calibration of imaging systems typically involves the use of calibration targets [1] that are specially designed objects with properties similar to specimens that will be typically used by these systems. While calibration targets have been used for calibrating the imaging module of fingerprint readers [2] [3], these targets are not fabricated using materials similar in properties to the human finger skin. The goal of this research is to fabricate standard 3D targets with properties (hardness, reflectivity and conductance) similar to the human finger skin that will allow repeatable operational evaluation of fingerprint readers [4]. 2D calibration patterns with known characteristics (e.g. sinusoidal gratings of pre-specified orientation and frequency or synthetic fingerprints with known positions of core, delta and minutiae) are projected onto a generic 3D finger surface to create electronic 3D targets. A state-of-the-art 3D printer (X and Y resolution: 600 dpi and Z resolution: 1600 dpi) is used to fabricate the 3D targets with material similar in hardness and elasticity to the human finger skin. Our experimental results show that these 3D targets can be imaged by three popular optical fingerprint readers. Additionally, we conduct experiments to demonstrate (i) the fidelity of 2D pattern features after projection to synthesize electronic 3D targets, (ii) the fidelity of engraved features on the electronic 3D targets after fabricating physical 3D targets, and (iii) the end-to-end fidelity of 2D pattern features after fabricating physical 3D targets. We show that the similarity scores output by a fingerprint SDK on comparing the (i) projected 2D pattern to snapshot of the electronic 3D target, (ii) the snapshot of the electronic 3D target to the captured images of the physical 3D target using optical readers, and (iii) the projected 2D pattern to the captured images of the physical 3D target using optical readers, are significantly above the verification threshold score at FAR=0.01%. Furthermore, experiments are conducted to show that the intra-class variability between multiple images of the same 3D target captured using the optical fingerprint readers does not degrade the recognition accuracy of a fingerprint SDK at 0.01% FAR.

Our ongoing work investigates (i) alternative methods to fabricate the 3D targets with higher precision, (ii) devising ways to either fabricate or coat 3D targets so that they have optical and conductance properties similar to the human finger skin, and (iii) simulating the effects of dry and wet fingers, and cuts and abrasion in fingers to further study the imaging capabilities of different readers.

References:

- [1] D. A. Boas et al., Handbook of Biomedical Optics. Taylor & Francis, 2011.
- [2] N. B. Nill, Test procedures for verifying image quality requirements for personal identity verification (PIV) single finger capture devices, MITRE, Tech. Rep. MTR 060170, 2006.
- [3] N. B. Nill, Test procedures for verifying IAFIS image quality requirements for fingerprint scanners and printers v 1.4, MITRE, Tech. Rep. MTR 05B0016R7, 2013.
- [4] S. S. Arora et al., "3D Targets for Evaluating Fingerprint Readers", MSU Technical Report, MSU-CSE-15-3, February 23, 2015.

Speaker

Nicholas Paulter, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics II / Wednesday, July 22 / 2:10-2:30pm

Error Propagation in Shape Analyses with or without Landmarks

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_propagation_in_shape_analysis_with_or_without_landmarks.ferson.crim2_.pdf (39 slides)

■ Abstract

Shape analysis is often required in crime scene pattern matching. For instance, elliptic Fourier analysis can be used in analyzing outlines, tracks, signatures, silhouettes, symbols, ordered points, etc. This multivariate shape analysis decomposes geometric patterns into numerical coefficients that can be normalized to optionally remove effects such as rotation, size or magnification, translation, registration, orientation, resolution or pixel density, etc., which makes the coefficients suitable for statistical analysis. Interestingly, elliptic Fourier analysis can also be applied directly to landmark data if the points are endowed or assigned some order (which is always possible). The 'contour' in this case is simply a list of the landmark coordinates. The analysis does not require the point locations it uses as input to be contiguous or to form a connected or closed shape. Thus, the method can be applied to landmark sequences, arbitrarily complex outlines, closed contours, even self-intersecting tracks in two or three dimensions, and it can completely capture the shape if sufficiently many harmonics are used in the decomposition. This method is one of a family of techniques that reduce complex shape information into forms suitable for multivariate t-tests, outlier detection, classifications, statistical discriminations, and other quantitative characterizations. It induces a mathematical metric across the entire space of possible closed shapes that allows us to assess quantitatively how close one shape is to another. It would, for example, allow one to display an array of shapes that are as close to a test shape as a given shape is. Such an array can therefore be used to demonstrate the fidelity of a putative match in a purely visual and intuitive way that might be understood by jurors without appeal to numerical statistics. There are inescapable measurement uncertainties in the capture of contours and landmarks to be analyzed via elliptic Fourier analysis and similar techniques. If contour information is encoded manually by a human using a digitizer, the achievable measurement precision depends in large part on the care and skill of the operator. Even if done by machine, as is common in image analysis, the theoretical pixel-level precision cannot usually be obtained in practice. The tracing algorithm can often be calibrated for good, consistent performance, and its measurement precision can then be assessed empirically. There are other sources of uncertainty arising from errors of registration and deformation of non-rigid objects by stretching, twisting or drooping during the measurement. Current implementations of elliptic Fourier analysis do not characterize these measurement errors as uncertainty about the harmonic coefficients and thus cannot project them through subsequent statistical calculations. We show how such imprecision can be characterized using interval boxes or ellipsoids and how the uncertainty can be projected through intervalized elliptic Fourier analysis and intervalized statistical tests such as t-tests and discriminations. There are various ways to visualize the uncertainty. We explore how the resulting uncertainty about shape, and about the decisions that depend on matching shapes, can be visualized.

■ Speaker



Dr. Scott Ferson is a scientist at Applied Biomathematics in New York. He also teaches risk analysis at Stony Brook University. Dr. Ferson has over a hundred publications, mostly in risk analysis and uncertainty propagation, and is a fellow of the Society for Risk Analysis. His recent research, funded by NIH and NASA, focuses on reliable statistical tools when empirical information is very sparse, and distribution-free methods for risk analysis.

Session: Criminalistics II / Wednesday, July 22 / 3:30-3:50pm

Strengthening Forensic Opinions through Objective Assessment of Evidentiary Value: A Prospective for Future Directions in Criminalistics

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/sigman_abstract.pdf

■ Abstract

Shape analysis is often required in crime scene pattern matching. For instance, elliptic Fourier analysis can be used in analyzing outlines, tracks, signatures, silhouettes, symbols, ordered points, etc. This multivariate shape analysis decomposes geometric patterns into numerical coefficients that can be normalized to optionally remove effects such as rotation, size or magnification, translation, registration, orientation, resolution or pixel density, etc., which makes the coefficients suitable for statistical analysis. Interestingly, elliptic Fourier analysis can also be applied directly to landmark data if the points are endowed or assigned some order (which is always possible). The 'contour' in this case is simply a list of the landmark coordinates. The analysis does not require the point locations it uses as input to be contiguous or to form a connected or closed shape. Thus, the method can be applied to landmark sequences, arbitrarily complex outlines, closed contours, even self-intersecting tracks in two or three dimensions, and it can completely capture the shape if sufficiently many harmonics are used in the decomposition. This method is one of a family of techniques that reduce complex shape information into forms suitable for multivariate t-tests, outlier detection, classifications, statistical discriminations, and other quantitative characterizations. It induces a mathematical metric across the entire space of possible closed shapes that allows us to assess quantitatively how close one shape is to another. It would, for example, allow one to display an array of shapes that are as close to a test shape as a given shape is. Such an array can therefore be used to demonstrate the fidelity of a putative match in a purely visual and intuitive way that might be understood by jurors without appeal to numerical statistics. There are inescapable measurement uncertainties in the capture of contours and landmarks to be analyzed via elliptic Fourier analysis and similar techniques. If contour information is encoded manually by a human using a digitizer, the achievable measurement precision depends in large part on the care and skill of the operator. Even if done by machine, as is common in image analysis, the theoretical pixel-level precision cannot usually be obtained in practice. The tracing algorithm can often be calibrated for good, consistent performance, and its measurement precision can then be assessed empirically. There are other sources of uncertainty arising from errors of registration and deformation of non-rigid objects by stretching, twisting or drooping during the measurement. Current implementations of elliptic Fourier analysis do not characterize these measurement errors as uncertainty about the harmonic coefficients and thus cannot project them through subsequent statistical calculations. We show how such imprecision can be characterized using interval boxes or ellipsoids and how the uncertainty can be projected through intervalized elliptic Fourier analysis and intervalized statistical tests such as t-tests and discriminations. There are various ways to visualize the uncertainty. We explore how the resulting uncertainty about shape, and about the decisions that depend on matching shapes, can be visualized.

■ Speaker

Michael Sigman, National Center for Forensic Science, University of Central Florida, United States

Session: Criminalistics II / Wednesday, July 22 / 3:50-4:10pm

Mitigation of Fire Debris Compound Naming Error through Automation

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/sasser_abstract.pdf

■ Abstract

The state-of-the-art for analyzing fire debris samples involves a highly-trained technician evaluating each GC-MS peak to assure correct naming. This process is time-consuming and potentially error prone. The best way to mitigate errors is remove the opportunity to make the errors originally. We show results from a fully automated compound-identification package for GC-MS which accurately names chromatographic peaks to 0.1% response. The system names the broad array of hydrocarbons found in fire debris samples, such as toluene, trimethylbenzenes, alkanes, and naphthalenes; further it identifies a variety of common background compounds. The technique combines information from both the spectral match and the retention time to achieve consistent identifications. A visualization tool annotates the chromatogram with names of the compounds and displays associated Extracted Ion Chromatograms. Report generation gives full details of the location and size of each compound. As well as naming individual compounds, the method categorizes compounds into groups such as alkanes, alkenes, and aromatics, easing comparison to the ASTM E1618 standard. Using this objective method for analysis of fire debris samples significantly reduces the chance of error, saves time for the technician, and increases in-lab repeatability as well as improving reproducibility across laboratories.

■ Speaker

Myron Sasser, MIDI, Inc., United States

Session: Criminalistics II / Wednesday, July 22 / 4:10-4:30pm

Evaluation of the Degradation of Lotion Components Due to Age and Exposure

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/dake_abstract.pdf

■ Abstract

After attending this presentation, attendees will have a broad understanding of the forensically significant chemicals which comprise commercial lotions, and how the compositions of the lotions can vary due to age and exposure. This presentation will provide attendees with the knowledge base to examine and compare commercial lotions with residues collected from crime scenes, clothing, and other evidence. An evaluation of the variability between commercial lotions will be explored, as well as the effects of age and exposure on the chemical constituents of lotions, thus providing attendees with the knowledge to effectively interpret analytical results. This will assist in mitigating and minimizing errors associated with misinterpretation of evidence and data. The results of this project will also assist with providing scientific underpinnings for lotion analysis methods and providing more accurate statistical interpretation of data. Commercial lotions are virtually ubiquitous in modern society. Lotions and their residues may be encountered in forensic casework in a variety of fashions; by the transfer to items of evidence through contact in assaults, theft, breaking and entering, and sexual crimes. In this study a sample set of over 100 commercial lotion products was analyzed to assess the degree of variation between different products. The results of these analyses can provide numerous data points from which a comparison to a known source or questioned sample may be made. Factors such as the presence/absence of certain chemical constituents, the relative concentrations of various components, and the frequency of occurrence of the chemical components were evaluated. The frequency of occurrence was evaluated to determine the most discriminating components in lotion samples, and principal component analysis was performed to assess the discriminatory value of the components detected in lotion samples. Inter-lot and Intra-lot variability was also assessed. Several samples of the same product line, but of different lots, were collected and analyzed from several different manufacturers. Duplicates of each lot sample were also analyzed. Some minor variation was observed between different lots of the same product, which may be due to variations in raw materials or manufacturing location. Finally, the effects of aging and exposure were evaluated on the lotion samples from the inter/intra-lot variability examination. Noticeable changes in the composition were detected, particularly in the presence and concentration of light weight volatile compounds. An evaluation of the constituents which were lost due to evaporation and exposure will be discussed, providing valuable insight which can be utilized for data interpretation and sample characterization in the performance of forensic casework. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense. Names of commercial manufacturers or products included are incidental only and inclusion does not imply endorsement by the authors, Department of the Army or the Department of Defense.

■ Speaker



Jeffrey Dake is a Forensic Chemist employed by the Defense Forensic Science Center, working in the Chemistry and Firearms Branch at the Gillem Enclave outside of Atlanta, Georgia. He performs casework in the areas of: lubricants, explosives, ethanol quant, and miscellaneous chemical identification.

He is the President of the American Society of Trace Evidence Examiners (ASTEE), and served previously as the communications chair from January 2012 – July 2014 and President-Elect for 2014. He is also a member of the Midwestern Association of Forensic Scientists (MAFS). He graduated from Michigan State University with an M.S.F.S in 2005, and from Tulane University in 2003 with a B.S. in Chemistry.

Session: Criminalistics II / Wednesday, July 22 / 4:30-5:00pm

Using Statistical Analysis to Assist with Writer Identification

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/wakshull_abstract.pdf

■ Abstract

Identification of handwriting relies on the document examiner's ability to discover the variability of the subject's known writing. Although qualitative assessment is prone to judgment error by the document examiner, variability in forensic document examination is often presented in qualitative rather than quantitative terms. Qualitative analysis is typically neither repeatable nor reproducible. A quantitative approach to writer identification can and should be used to reduce bias and errors in document examiners' opinions. Literature shows that variability of the proportions of height of a person's written letters remains stable within common cause variability across writing sessions. This presentation applies aspects of statistical process control to determine whether a writer of a known document is a potential author of a questioned document, based on such variability. For example: The ratio of the height of lower case letters extending into the upper zone and height of lower case letters remaining in the middle zone is calculated. The ratios of the length of lower case descenders to the height of middle zone letters are calculated. For each calculated ratio a run chart is created for the ratios of the known writing. The same is performed for the slant angle of a given letter. The standard deviation of the ratios or angles are computed to one, two, and three sigma. These measurements are plotted to create a control chart. The measurement of the angle or ratio of the questioned writing is plotted on the control chart to determine how the questioned writing compares to the known writing with respect to the variability of the known writing. The number of standard deviations from the mean of the known writing is explored to determine writer identification. The null hypothesis is if the ratio of the given letters in the questioned writing falls beyond 2.5 standard deviations from the mean the questioned writing may have been written by a person other than the known writer. The result is an indicator of authorship rather than a conclusive determination.

■ Speaker

Michael Wakshull, Q9 Consulting, Inc., United States

Session: Death Investigation / Wednesday, July 22 / 11:00am-12:00pm

Forensic Pathology as a Forensic Science: History, Current Challenges, Improving Quality, and Understanding Cognitive Bias

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/forensic_pathology_as_a_forensic_science_baker_death.pdf (76 slides)

■ Abstract

In 2009, The National Research Council of the National Academy of Sciences published recommendations to strengthen forensic sciences. Though the report largely focused on broad issues spanning many disciplines, it is telling that medicolegal death investigation received its own chapter, recognizing the challenges unique to this community. This presentation will focus on the history of death investigation in the United States, address current challenges, and offer a number of proposals for improving quality and addressing concerns of cognitive bias.

Among the forensic disciplines, forensic pathology is (almost) unique because it is the practice of medicine. Its methods are unlike those in other forensic specialties, its goals are different, and the types of errors that may occur are unlike those in any other practice. Further confusing some of our legal and forensic colleagues is the patchwork of coroner and medical examiner systems, lack of standardization in laws and practices, and the sheer number of medicolegal death investigation jurisdictions in the United States. Though all jurisdictions ostensibly have the same mission to accurately identify decedents and determine their cause and manner of death systems run the gamut from ill-equipped offices that are little different from a 19th century coroner to nationally-accredited regional or statewide facilities employing board-certified forensic pathologists and state-of-the-art imaging equipment.

Forensic pathology supports many interests besides the criminal justice system. More than 90% of the sudden, unexpected, and suspicious deaths autopsied and certified by forensic pathologists are at the end of a thorough investigation found not to be criminal cases. The importance of these non-criminal cases should not be dismissed they provide closure for families and serve as powerful drivers for the allocation of public health and public safety dollars.

Some in the legal community have raised the specter of cognitive bias in forensic pathology, failing to recognize that legitimate medical and contextual history is not inappropriate cognitive bias, and further failing to separate the neutral term cognitive bias from pejorative accusations of incompetence, dishonesty, fraud, and corruption.

The optimal way to improve and ensure quality work in forensic pathology, and to minimize bias, is to safeguard the independence of medical examiners. This includes transitioning from elected coroner systems to appointed medical examiner systems, increasing the number of qualified forensic pathologists, creating model medical examiner legislation and providing incentive-based funding to assist states in converting antiquated and fragmented death investigation systems into modern medical examiner systems, ensuring that all forensic pathologists are trained and certified in their fields, ensuring that all forensic sciences laboratories including medical examiner and coroner offices are accredited to recognized national standards, and producing legislative or judicial guarantees that medical examiners are independent of law enforcement and prosecutors, and available to consult with all parties in the criminal justice system.

■ Speaker



Dr. Andrew M. Baker, MD is a board-certified anatomic, clinical, and forensic pathologist. He is a fellow of the College of American Pathologists. He is a fellow of the American Academy of Forensic Sciences and has previously served as the Chair of the Academy's Pathology/Biology section. He is a fellow of the National Association of Medical Examiners, and has previously served as the President and Chair of the Board of Directors of that organization. A veteran of the US Air Force, Dr. Baker began practicing as a full-time forensic pathologist with the Hennepin County Medical Examiner's Office in Minneapolis, Minnesota, in 2002. He was appointed to the position of Chief Medical Examiner for Hennepin County in 2004. In 2013, Dr. Baker's office expanded to provide all medical examiner services to Dakota and Scott Counties in the metropolitan Twin Cities area. In 2014, Dr. Baker was appointed to the Forensic Science Standards Board of the National Institute of Standards and Technology.

Session: Death Investigation / Wednesday, July 22 / 1:30-1:50pm

Mitigating DNA Identification Errors in Mass Fatality Response Operations through Rapid DNA Technology

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/miles-abstract.pdf>

■ Abstract

Inherently, a mass fatality DNA identification response operation is prone to errors, such as sample switches. This is because the operation is unprompted, involves the testing of a large number of samples in a short period of time and often requires people working extended hours performing tasks they do not routinely perform. In the initial stages of the World Trade Center DNA Identification effort, sample switches resulted in families receiving the body of an unrelated victim. Identifying and rectifying the errors was traumatic not only to the victim's family but also to the professionals involved in the DNA identification effort. Therefore, during the Louisiana Hurricane Katrina DNA identification effort, samples were collected in duplicate and processed independently by two different laboratories. Using this approach, sample switches were identified and appropriate actions were taken before identifications were reported and bodies released to the families. While this duplicate testing approach eliminated the reporting of DNA identification errors, it was expensive and time consuming. Rapid DNA technology has the potential to eliminate errors. Rapid DNA technology was developed through a joint research and development effort of the Departments of Homeland Security, Defense, and Justice and the Intelligence Community. Potential applications for Rapid DNA include verification of human identities, validation of claimed family relationships, and processing of crime scene DNA evidence. The Rapid DNA system integrates and automates the existing laboratory DNA processes into a single desktop system that generates and analyzes a DNA profile in the field. Because Rapid DNA is a sample in-profile out technology, the system eliminates the manual processing steps conducted by analysts; thereby reducing handling and transfer errors. Additionally, the Rapid DNA system is designed to be taken into the field where cases are being processed, mitigating potential chain of custody errors in the collection and reporting of results back to the operational entities. The Rapid DNA system also followed a Privacy by Design concept that serves to reduce operational errors. Operators verify their identity when they log onto the system with their credentials, a biometric, or user name and password. The operator's actions and system operations are tracked and any data transfers or deletions are logged. The samples are tracked through a RFID chip built into each swab or through barcode labels and these labels are tracked throughout the system and associated with all reports from the system. Data within the system, and when exported from the system, is encrypted and associated with the specific user who authorized the data transfer. While all of these concepts were adopted to ensure privacy protections, they will also help to reduce operational errors. This presentation will discuss the development of Rapid DNA and the potential of the technology to reduce errors in DNA testing in mass fatality response. This novel approach to minimizing errors benefits the families, and also benefits responding agencies and staff working to support the families following this most devastating type of disaster.

■ Speaker



Christopher Miles is a Biometrics Program Manager with the Resilient Systems Division (RSD) of the Department of Homeland Security (DHS) Directorate for Science and Technology (S&T) Homeland Security Advanced Research Projects Agency (HSARPA). He manages the development of rapid and low-cost DNA analysis for family relationship verification and past biometrics basic research portfolio efforts in multi-biometric research, standoff biometrics, and decision fusion research. Mr. Miles served as Co-Chair of the White House National Science and Technology Council (NSTC) Subcommittee on Biometrics & Identity Management (www.biometrics.gov) coordinating the biometrics research and development of 12 Federal Departments. He also served as the Chair of the Sponsor's Committee for the National Science Foundation (NSF) Center for Identification Technology Research (CITeR), a 10-year Industry/University Cooperative Research Center with 20 federal and industry sponsors and 13 Universities that have conducted over 200 research projects to date (<http://www.citer.wvu.edu>).

Session: Death Investigation / Wednesday, July 22 / 1:50-2:40pm

Descriptive Frequency Analysis of 110 Bodies found in Confined Water Spaces and Determination of the Manner of Death, Including Investigative Outcomes

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/110_bodies_found_in_confined_water_spaces_and_determination_of_the_manner_of_death.richardson.death.pdf (23 slides)

■ Abstract

The correlation between documented body position in bodies of water and manner of death showed that drownings are less likely to be considered suspicious and far more likely to be considered accidental.

The importance of interviewing techniques and reenactments can assist investigators in determining the plausibility of the purported circumstances.

To justify for conference inclusion includes:

1. Poor descriptors are used when describing position such as submerged, partially submerged, surface, floating, at bottom, submerged beneath surface. These are vague and these descriptors document neither depth of the water nor body position.
2. At risk population manner of death often overlooked as homicide, neglect versus accidental.
3. Risk of Co bathing. (cultural phenomena)
4. Statement analysis of family members and witnesses appear to be overlooked or not pursued further on red flag cases.
5. Assumption that drowning was 'tragic accident'.

■ Speaker

Joanne Richardson, Aspen Hope Center, National University, United States

Session: Death Investigation / Wednesday, July 22 / 2:10-2:30pm

Searching for a Standard: The Impact that Method Selection has on Evidence Recovery in Forensic Archaeological Investigations

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/method_selection_has_on_evidence_recovery_in_forensic_archaeological_investigations-evis-death-2.pdf (41 slides)

Abstract

Recent reviews of forensic science have found many of the methodological approaches used by practitioners in the various sub-fields of the discipline wanting. As a result, sub-fields such as forensic archaeology have been required to ensure that the methods that their practitioners use satisfy the standards set out in the legislative acts and admissibility regulations upheld by the courts in the country where the investigation is being conducted and/or tried. In general, these standards require methods used during the course of a forensic investigation to have been subjected to: empirical testing, peer review, have known error rates, have standards controlling their operation, and be widely accepted amongst the academic community from which the methodology originates. Until recently however, no substantial empirical testing had been conducted regarding archaeological excavation methods and recording systems, and no internationally accepted forensic archaeological investigatory protocol had been established. This is due to the fact that approaches to archaeological excavation and recording vary greatly from country to country, and the methods have evolved to their current state according to the practices advocated by practitioners and professional bodies in their country of origin, and the inherited traditions present in each. However, through adopting an accept all approach and attributing each excavation method and its associated recording system the status of a 'standardized' method, without any evidence to prove their suitability for use in forensic casework, the field of forensic archaeology was failing to meet the standards of the international courts. Nevertheless, this presentation will outline the recent research conducted by Evis and colleagues that has experimentally compared forensic archaeological excavation methods and recording systems using a controlled grave simulation, that will aid in satisfying the international court's requirements for empirical testing and the establishment of error rates. It will also discuss how the field of forensic archaeology now performs against each of the standards set out by the international courts, and examine the potential for the establishment of a standardized, tested, widely accepted forensic archaeological investigatory protocol.

Speaker



Dr. Laura Evis is a forensic bioarchaeologist. Her research specializes in the adaptation and application of archaeological and anthropological techniques to criminal investigations. She is currently based at the University of Exeter, Devon, in the United Kingdom and is the lead coordinator of forensic-based research and teaching in the humanities department. Prior to taking up this position, Laura was based at Bournemouth University, Dorset, in the United Kingdom, where she spent six years evaluating the impact that methodological approaches have on investigatory results.

Session: Death Investigation / Wednesday, July 22 / 2:30-3:00pm

Reducing Dental Forensic Errors by Using DICOM and SNOMED

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/forensic_science_will_greatly_improve_digital_evidence_error_management_by_utilizing_dicom_and_snomed_ct_standards_in_dental_and_visible_light_images.casertano.death_.pdf (31 slides)

Abstract

Data Format for Interchange of Fingerprint, Facial Information & Other Biometric Information, ANSI/NIST-ITL-2011 Special Publication 500-290, is currently being updated with dental and facial record information. This presentation summarizes ongoing work that supports digital evidence error management use of Digital Imaging Communications in Medicine (DICOM) and Systematized Nomenclature of Medicine - Clinical Terminology (SNOMED CT) by the American Dental Association (ADA) and the American Association of Orthodontists (AAO).

Forensic Use Cases include Recognition & Comparison of Dental and Facial Photographs, identification, documentation, collection, and preservation of evidence. Law enforcement requires a positive confirmation through the use of a visible light photographs and dental imaging. With the wide use of dental records management and imaging, dental photography and imaging is a vast resource for forensic science.

Ongoing coding work by standard development organizations are promoting interoperability by coding anatomic dental features such as Anatomic Regions and Modifiers for Intra-oral & Craniofacial Radiography, Primary Anatomic Structure for Intra-oral Radiography (Permanent, Deciduous & Supernumerary Dentition - Designation of Teeth. Additional DICOM work will standardize sets of images to support forensics including hanging protocols and structured displays.

Visible light photography as a biomedical image is much more valuable with associated metadata, such as identification of the patient, identification of the study performed, other images, and relevant clinical information (i.e., image acquisition context).

Image acquisition context includes the anatomic, chemical, functional, physical, spatial, and temporal conditions present at the time of image acquisition, such as anatomic region examined, contrast agent, vital stain, imaging modality, type of transducer, and position of the imaging subject during image acquisition. The attributes of image acquisition context have context-dependent value sets. Advanced forensic tools require the use of anatomic coded terminology, and controlled vocabularies to support interoperability to scan and mine big data.

Ongoing work by the ADA and the AAO on the Acquisition Context Sequence include the Laterality, Acquisition and Image Perspective, Subject state and Occlusion relationships. The Orthodontic Records include the Extra-oral Views frontal and side facial and oblique Profiles.

This DICOM and SNOMED standard metadata will be used to uniquely identify, store, display and retrieve the data quickly with the use of universally unique identifier (UUIDs) to minimize cost and errors that are used by all stakeholders from clinical acquisition to forensic data warehouses.

Speaker



Andrew Casertano, MS, has over 28 years of electrical engineering, healthcare information technology (HIT) and technical program management leadership experience. He has extensive experience in developing and managing large complex multi-site, multi-tiered healthcare information technology systems in support of the US Department of Veterans Affairs (VA), American Dental Association (ADA), American College of Cardiology (ACC), Blue Cross Blue Shield, and the US Department of Defense (DOD).

Session: Death Investigation / Wednesday, July 22 / 3:30-3:50pm

Medical Examiner Collection of Comprehensive, Objective Medical Evidence for Conducted Electrical Weapons and Their Temporal Relationship to Sudden Arrest

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/med_exam_collection_of_comprehensive_objec_med_evid_for_conducted_electrical_weapons_and_their_temporal_relationship_to_sudden_arrest-brave-death.pdf (36 slides)

■ Abstract

Background: The use of conducted electrical weapons (CEW) is now a common law enforcement force practice. Use of these devices is rarely temporally associated with the occurrence of sudden arrest-related deaths (ARD). Because such deaths are uncommon, few Medical Examiners (MEs) ever encounter one, and even fewer offices have established comprehensive investigative protocols. Without adequate data collection, the scientific assessment of the role, if any, played by a CEW in any given case is largely supplanted by conjecture. The purpose of this presentation is to provide MEs with a comprehensive evidence-based checklist to assist in the assessment of CEW-ARD cases.

Methods: PUBMED and Sociology/Criminology data bases were queried to find all medical, scientific, electrical, modeling, engineering, and sociology/criminology peer-reviewed literature for mentions of CEW or synonymous terms. Each paper was then individually reviewed to identify those that discussed possible bioelectrical mechanisms relating CEW to ARD. A Naranjo-type pharmacovigilance algorithm was also employed, when relevant, to identify and quantify possible direct CEW electrical myocardial stimulation. Additionally, CEW operational manuals and training materials were reviewed to allow incorporation of CEW-specific technical parameters.

Results: Relevant PUBMED mentions were surprisingly few (<250), and reports of death quite rare. Much relevant information was available from Sociology/Criminology data bases. Once the relevant published papers were identified, and reviewed, we compiled an annotated checklist of data that we consider critical to a thorough CEW-involved death investigation.

Conclusion: We have developed an evidenced-based checklist that can be used by MEs and their staffs to assist them in identifying, collecting, documenting, maintaining, and objectively analyzing the role, if any, played by a CEW in any specific case of sudden death temporally associated with the use of a CEW. Even in cases where the collected information is deemed by the ME as insufficient for formulating an opinion or diagnosis to a reasonable degree of medical certainty, information collected as per the checklist will often be adequate for other stakeholders to use as a basis for informed decisions.

■ Speaker



Mr. Michael Brave is Member/Manager of LAAW International, LLC. He is an attorney, sworn police officer and trainer. He has been retained as an expert in over 200 cases, and has been involved in reviewing 100s of law enforcement temporal death cases. He has presented on force option and other subjects 100s of times in the U.S., as well as Mexico, Canada, Panama, and the United Kingdom. He also serves as National/International Litigation Counsel for TASER International, Inc. He is a TASER Master Instructor, TASER's designated Person Most Knowledgeable, and Legal Advisor Member of the TASER Scientific and Medical Advisory Board and TASER Training Board.

Session: Death Investigation / Wednesday, July 22 / 3:50-4:10pm

Investigating Arrest-Related Deaths

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/ross-abstract.pdf>

■ Abstract

Annually a small number of individuals suddenly die after a physical restraint incident with police, correctional, or mental health personnel. The person exhibits bizarre and violent behaviors frequently related to chemical impairment symptoms associated with excited delirium, mental impairment, or both. Restraining officers use physical control techniques and less-lethal devices in order to control the combative person, as they have previously done without a death occurring, but in this incident an unexpected arrest-related death occurs. Some of these deaths have previously been blamed on positional or restraint asphyxiation. Recent prospective field research reveals that other factors are commonly associated with the death other than the restraint procedures applied by the restraining officers. A central problem often emerges in identifying the exact cause of death. As a result numerous questions emerge which are directed at the restraining officers, administrators of the arresting or restraining agency, and medico-legal investigators. Because these incidents often generate civil liability claims against the restraining personnel and agency investigators, it is paramount that a thorough and comprehensive investigation be completed. This presentation identifies protocols for conducting investigations into sudden arrest-related deaths and examines: the nature of these deaths and associated factors; the use of less-lethal force equipment; the use of physical control, restraint techniques and equipment; scene investigation; the body as evidence; evidence collection; decedent background; interviews of witnesses; terminal episode history; investigator at the autopsy, investigation checklists, maintaining the file, and composing the report. The presentation addresses key investigation components which every investigator should consider when tasked with an arrest-related death investigation. The presentation will assist the investigator in ensuring proper protocols are followed which work to mitigate potential errors of the investigation.

■ Speaker



Dr. Darrell Ross is a professor and the Department Head for the Department Sociology, Anthropology, and Criminal Justice and the Director for the Center of Applied Social Sciences at Valdosta State University, Valdosta, GA. He was 13 year practitioner for the Michigan Department of Corrections and taught in the police academy at Ferris State University for 7 years. His research and published works include: the outcomes of violent prone restraint and the use of force measures; excited delirium and restraint asphyxia death investigations in policing and detention; the impact of stress, human factors, and officer perception during lethal force encounters; the impact of stress and contextual cues on human performance and decision making during use of force encounters; patterns of subject resistance during arrest and during detention; use of force policy development; failure to train and use of force liability issues; liability issues and sudden in-custody deaths; liability issues and sudden deaths associated with ExDS, risk factors and unexpected in custody deaths; administrative liability issues; a risk management analysis of police and detention agencies; liability issues for emergency response teams; and liability and investigation issues of custodial deaths in jails.

Dr. Ross has made over 700 national and international conference presentations on these and other subjects. He served as an expert with DOJ/NIJ's scientific technology working group on Excited Delirium deaths (ExDS). He has provided consulting and training services for police & correctional agencies, private corporate and security agencies, and the military, nationally and internationally. Since 1988 Dr. Ross has been retained as an expert witness providing testimony in 30 states on behalf of police, correctional, and private security agencies.

Session: Death Investigation / Wednesday, July 22 / 4:10-4:30pm

Confirmational Bias and Investigation of Arrest-related Deaths

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/Confirmational_bias-karch-death.pdf (45 slides)

■ Abstract

Annually a small number of individuals suddenly die after a physical restraint incident with police,

A recent study confirms Sudden Cardiac Death (SCD) during psychological stress mostly occurs in young males who, after detailed cardiac examination, display no evidence of structural heart disease and have unremarkable autopsies. The same study shows that just over 10% die in police custody, and less than one third of the decedents had been restrained. Recent studies also show, police officer's risk of SCD is 34-69 times higher during restraints/altercations and 32-51 times higher during pursuits. When dealing with the small percentage who die suddenly temporal to police encounters many, if not most, Medical Examiners (MEs) feel compelled to qualify their diagnosis with terms such as "while during police restraint," "while in police custody," "while prone," or from "hypoxic effects of chest compression during police restraint." Others add "after being Tased," even if the TASER was applied long before the arrest. Given that all of the listed qualifiers have been found wanting in peer-reviewed controlled scientific as well as epidemiological studies, their persistent use raises serious questions about lack of foundational causation followed by confirmational bias; a politically acceptable way of saying that death investigators are interpreting and manipulating data in such a way that it confirms their own preconceptions. It also allows them to avoid misplaced public and media pressure to name a culprit, even a speculative one. This situation is not novel. In January 1692 in Salem, Massachusetts, the daughter and the young niece of Reverend Samuel Parris fell ill. William Griggs, the village doctor, was consulted. Griggs could not explain the symptoms the children displayed. For lack of a better diagnosis Griggs concluded they could only have been a consequence of bewitchment. Griggs' misdiagnosis led directly to the Salem witch trials and the death by hanging of 19 victims. Dr. Griggs demonstrated a lack of foundational causation and had a confirmational bias to make a diagnosis; using the evidence at hand, or lack thereof (which was insubstantial and untestable). We are more fortunate today because science has advanced and alternative diagnoses exist. Countless diseases are recognizable only at the molecular level, and methods of testing for them exist. If MEs were not constrained by their biases they would endeavor to identify these alternative causes. If resources for such a search were lacking, classifying the cause of death as undetermined would be most appropriate, without the speculative hyperbole. For that to happen, investigators need to overcome their biases; investigate all potential causes, and acquire a greater understanding of the underlying science. Investigators must recognize that many SCDs are undetermined; admit that coincidences do happen; acknowledge that many diseases of the heart are molecular, such as channelopathies; and accept such disease is detectable only by genomic testing. Even then confirmational bias will still exist among those without training in genomics. Time permitting, some of the bogus qualifiers will be debunked and the difficulties of genomic interpretation explained.

■ Speaker

Dr. Steven Karch, private practice, United States

Session: Death Investigation / Wednesday, July 22 / 4:30-5:00pm

Medical Examiner Mistakes in Tennessee: A Proposed Model for Formal, Comprehensive Death Investigation Peer Review

Presentation Link: <https://www.nist.gov/sites/default/files/documents/director/hawes-abstract.pdf>

■ Abstract

Forensic pathology is the practice of medicine, and as such is distinctly different from the majority of other forensic disciplines. Formal, standardized medical examiner/forensic pathology peer review is much less well developed than other medical specialties. Properly conducted standardized peer review would improve outcomes by reducing error, helping refine established standards of practice, encouraging objective defensible diagnoses and conclusions, and ultimately increasing public trust in the profession. Major barriers to forensic pathology peer review include: an extremely small pool of specialists, small specialist groups within each office, potential conflicts of interest, and concerns regarding immunity and privacy of those involved in the peer review process. Due to the structure of the medical examiner system in Tennessee, forensic pathology peer review is varied throughout the state. Tennessee is a county based medical examiner system with five regional forensic centers where autopsies are performed. The regional forensic centers are independently operated (two of the regional forensic centers are operated by county governments, one by a private professional corporation, and two by affiliation with universities) and each is responsible for establishing its own policies and procedures for peer review. Although Tennessee has statutory protections for records produced as part of a physician peer review (The Tennessee Patient and Safety Quality Improvement Act of 2011), discussion remains about whether medical examiners are covered as 'health care organizations' and 'health care providers' as defined in the statute. As yet, the protections have not been legally challenged. A Tennessee statutory requirement that all autopsies be performed at facilities accredited by the National Association of Medical Examiners (NAME) has improved the overall quality of death investigation throughout Tennessee. The National Association of Medical Examiners (NAME) accreditation standards only ensure a minimum standard for quality assurance, peer review, and performance monitoring activities. The NAME standards do not provide a framework or suggestion for how these baseline standards should be accomplished. The NAME Ad-Hoc Study of Quality Improvement by Peer Review in the Medical Examiner Office Committee is developing quality assurance/quality improvement programs for use in medical examiner offices, and the Committee is currently surveying members about participation in a centralized, inter-office peer review. Results of the survey will be presented. A standardized, comprehensive death investigation peer review process for forensic pathologists will be presented and discussed. This quality improvement plan is modeled after clinical peer review strategies. The proposed plan includes a 'template' style case evaluation, and has specific recommendations for the selection of cases for peer review, the suggested procedures for the review, and recommendations for reporting review results. The case selection process for review is tiered and includes both random selection of cases and specific trigger events. The proposed model reporting format is also tiered, allowing for diagnostic discrepancies that may simply be differences in professional judgement as opposed to an 'incorrect' diagnosis. The model could be easily adapted to assess the effectiveness and consistency of services provided by others such as death investigators and non-forensic pathologist medical examiners.

■ Speaker

Dr. Amy Hawes, Knox County Medical Examiner's Office, United States

Session: Digital Evidence / Wednesday, July 22 / 1:30-2:00pm

Error Management in Forensic Digital Imaging Should be a Resolution for Throughout the Entire Law Enforcement Community

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_management_in_forensic_digital_imaging_should_be_a_resolution_for_throughout_the_entire_law_enforcement_community.witzke.digevid.pdf
(43 slides)

Abstract

When judgments are made under uncertainty, two general types of errors are possible false positives and false negatives. A decision maker cannot simultaneously minimize both errors because decreasing the likelihood of one error necessarily increases the likelihood of the other (Green & Swets, 1966). Within the criminal justice community, there are a lot of judgments "conclusions" that are made under uncertainty. These judgments can result in the misidentification or missed identification of a fingerprint, a footwear impression or a questioned document as well as the inability to analyze characteristics in video images, such as the letters or numbers of a license plate. The reason for uncertainty in almost every one of these cases can be attributed to a lack of image resolution, improper use of image file formats, or the lack of standards. For example, when a footwear impression is captured with a resolution that is too low, it is impossible to enlarge the image sufficiently to make a reasonable conclusion about characteristics such as wear patterns or other unique characteristics. In addition, low resolution can blur the image with the background (substrate) upon which the impression was made. Failing to follow standards, such as including a scale, can prevent the image from being calibrated properly. Or, was the scale on the same plane as the impression? Was a tripod used or was the camera hand-held? Other artifacts can be introduced into the image using a compression algorithm, such as a JPG file format. While image compression should not be used for digital images that will be used for comparison, many agencies require their officers to compress images as much as possible so that they can get more images on the memory card in the camera. In far, far too many instances, the artifacts created by image compression significantly affect image quality, which in turn affects the ability to make a judgment with any degree of certainty. It is imperative that standards be in place to ensure not only reliable image quality, but to ensure consistency of clarity and detail. In digital imaging technologies, image resolution is the key factor in image quality. The level of detail can be increased with the use of color or, more specifically, the bit depth of color. Image quality together with image detail provides better clarity and thus, a higher degree of certainty. Most high-end digital SLR cameras today capture image detail with a bit depth of 12 or 14 bits-a dynamic range 4096 different contrast values (shades) or 16,384 different contrast values (shades). However, this dynamic range can be significantly reduced to only 256 different contrast values when using a file format, such as JPG or TIF when storing images on a digital camera. Once this detail is lost inside the camera, it can never be restored to the original values captured with the imaging sensor. While proper training will not solve all of the issues and remove all of the uncertainty, it is a very good place to start.

Speaker



Mr. David Witzke is Vice President, Program Management for Foray Technologies, is considered to be one of the foremost experts in forensic digital image processing technologies. His background includes software development, systems installation and integration, technical support, and training. Because of his established reputation in forensic digital image processing, Ski has been asked to speak at hundreds of regional, national and international conferences. Ski was also a guest instructor of the *Digital Imaging of Evidentiary Photography* training program at the FBI Academy in Quantico, Virginia for nearly eight years, and was the instructor for the *Forensic Digital Image Processing* program at the British Columbia Institute of Technology (BCIT) in Vancouver, BC for more than six years. Ski is a contributing writer for three well-known books: ***Crime Scene Photography*** (Second

Edition, Published 2010 by Academic Press), ***Introduction to Crime Scene Photography*** (Published 2012 by Academic Press) and ***Footwear, The Missed Evidence*** (Third Edition, Published 2013 by Staggs Publishing). (***Crime Scene Photography*** has been chosen by the IAI certification committees for the Certified Crime Scene Investigator (CCSI) Certification Test, the Certified Crime Scene Analyst (CCSA) Certification Test, the Certified Senior Crime Scene Analyst (CSCSA) Certification Test, and the Forensic Photography Certification Test.) Ski has also been the chairperson of the Digital Imaging Subcommittee of both the Chesapeake Bay and Florida Divisions of the IAI.

Session: Digital Evidence / Wednesday, July 22 / 2:00-2:20pm

Error Treatment in Forensic Authorship Attribution

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_treatment_in_forensic_authorship_in_attribution.juola_digevid.pdf (19 slides)

Abstract

Critics of forensic science, including the National Academy of Science, abound. A recent New York Times op-ed (March 30, 2015) suggested that "many forensic practitioners have offered either invalidated evidence or grossly exaggerated the value of the evidence" and have been doing so for years. However, much of the criticism is directed specifically at "pattern evidence" such as bite marks and shoe prints. By contrast, DNA and genetic evidence are often held up as models for the rest of the field (NAS 2009; Cheng 2013).

How does forensic linguistics, and specifically forensic authorship studies, stack up? This (sub) discipline focuses on the analysis of disputed writings when physical evidence such as handwriting or paper are unavailable. These writings could be web pages, text messages, email, or word-processed documents. As McMenamin (2011) has written, "At any given moment, a writer picks and chooses just those elements of language that will best communicate what he/she wants to say." By looking at these choices, one can identify the most probable author to have made those choices. We present several case studies drawn from the writings of McMenamin, Coulthard, Grant, Chaski, and Juola. These case studies range from murder investigations through contract disputes to the identification of J.K. Rowling as the real author of *The Cuckoo's Calling*.

Cheng (2013) has identified a "wish list" of three attributes one wants any forensic discipline to have: "a widely adopted, predefined [decision process]; a large, random sample of known exemplars [...]; and a well-understood theoretical underpinning." We claim that forensic authorship studies have, or at least are on a path to, all three of these characteristics, and provide examples.

One key practice among some forensic linguists that helps on this path is periodic TREC-style competitive evaluations. The Plagiarism Action Network (PAN) at the CLEF series of conferences, for example, has been running these evaluations for years (Juola, 2012; Juola & Stammatatos, 2013; Stammatatos et al. 2014 [and planned for 2015]). A typical problem setup would include fifty or more instances of questioned writings along with associated writings of known authorship. These are typically structured in what Koppel et al. (2012) have called "the fundamental problem of authorship attribution" --- is THIS document written by the same person as THIS (or THESE) documents? Results from 2014 show that high accuracy levels can be empirically established. These levels, in turn, can be used in formal protocols (Juola, 2014) to establish case-by-case error-likelihoods. The information gained from this kind of conference enables both error measurement and error mitigation (through the use of, e.g., ensemble methods).

We consider these events to be important practices in creating high quality forensic science. Not only does it establish and improve best practices generally, but it also provides a type of proficiency testing for individual scientists. It provides a third-party validation of expected error rates in a non-adversarial context, and creates a degree of transparency that is key for legal application.

Speaker



Dr. Patrick Juola is a professor of computer science at Duquesne University. He works in the field of computational linguistics and computer security currently serving as CEO of Juola & Associates and Principal of the Evaluating Variations in Language Laboratory. He is credited with co-creating the original biometric word list. Juola has also created a Java-based open source authorship attribution suite JGAAP, Java Graphical Authorship Attribution Program, with several students at Duquesne University. He has testified in numerous court cases, but his most high-profile analysis was unmasking JK Rowling as the true author of *The Cuckoo's Calling*.

Session: Digital Evidence / Wednesday, July 22 / 2:20-2:40pm

Speaker Detection in a Forensic Environment: Recognizing the Limitations, Improving the Science

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/speaker_detection_in_a_forensic_environment.martin.digevid.pdf (13 slides)

■ Abstract

Speaker recognition has a rather checkered history in terms of its use in a forensic context. Past claims of a capability to produce 'voiceprints' that could be regarded as comparable to fingerprints were vastly inflated if not downright false, and provided source material for key U.S. court rulings on what constituted acceptable bases for scientifically acceptable forensic evidence. It must be recognized that we currently generally cannot assert with certainty, based on automatic methods or human expertise, that a given person is the speaker in a particular recording. There are many factors that may affect the quality of available audio evidence in terms of being able to reliably make match/non-match decisions between voices, and what gets recorded at typical crime scenes is likely to be particularly challenging in terms of the channel qualities of the recording media, the durations of the utterances, and the cooperativeness and the physical and emotional states of the people involved. NIST and other organizations have been involved in studying the capabilities of both automatic and human based systems in performing successful speaker recognition with respect to varying types of speech utterances and to underlying channel and environmental conditions, but practical and ethical considerations make it exceedingly difficult to investigate the kinds of stressful conditions likely to prevail at crime scenes. Professional audio investigators need to recognize the limitations of what their methods can determine, and that in many circumstances it may be advisable to decline to pursue work on a specific case and punt. Often they might be more effective in investigatory roles to consider possible leads and rule out possible suspects than as primary expert witnesses in court. In its most recent evaluations of speaker recognition technology, NIST has conducted small scale evaluations, on a limited number of difficult trials, of systems encompassing human experts, showing limited success compared to the best performing automatic systems. In 2015 NIST and the Department of Justice created a Speaker Recognition Committee as part of its Organization of Scientific Area Committees (OSAC), to collaborate on creating consensus documentary standards and guidelines that we hope will put this field on a firmer scientific basis going forward.

■ Speaker



Alvin Martin has served as a mathematician in the Multi-Modal Information Group at the National Institute of Standards and Technology since 1991, and is currently a rehired annuitant. He coordinated NIST's series of evaluations in the areas of speaker recognition and of language and dialect recognition from 1996 until his 2011 retirement, and has contributed to its evaluations of large vocabulary continuous speech recognition. This work involved the collection, selection, and pre-processing of appropriate speech data, the writing of evaluation plans, the specification of metrics and charts for the scoring, presentation, and analysis of results, the implementation of statistical tests for determining the significance of performance differences, and the organization of workshops to review evaluation results. He currently serves on the Speaker Recognition Subcommittee of the Organization of Scientific Area Committees, the initiative of NIST and the Department of Justice to strengthen forensic science in the U.S.

He received a Ph.D. in mathematics from Yale University (1977), has taught mathematics and computer science at the college level, and worked on the development of automatic speech recognition and speech processing systems before coming to NIST.

Session: Digital Evidence / Wednesday, July 22 / 2:40-3:00pm

An Overview of Speaker Variability as a Source of Error in Forensic Automatic Speaker Recognition

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/an_overview_of_speaker_variability_as_a_source_of_error_in_forensic_automatic_speaker_recognition.kelly_.digevid.pdf (41 slides)

Abstract

The performance of automatic speaker recognition systems has seen consistent improvement over the last two decades. Progress has largely been driven by the regular NIST speaker recognition evaluations (SREs), which establish performance benchmarks on large and challenging datasets. Alongside this progress, the application of automatic speaker recognition to the forensic domain has become more prevalent. Forensic uses of automatic speaker recognition can be categorized as evaluative, in which the comparison of a small number of samples leads to a probability statement acceptable in court, or investigative, in which the comparison of a large number of samples leads to a "shortlist" of suspects. Both applications face the same challenge, in that the samples under comparison are completely uncontrolled. Current automatic speaker recognition systems demonstrate robustness to environmental sources of variability such as background noise, room acoustics, microphone type, and transmission channel. However, there are many additional sources of speaker variability, occurring frequently in forensic cases, which have not received the same level of research attention. Here we consider speaker variability, the change within a speaker's voice, as the combination of short-term fluctuations in emotional and physical state, and progressive long-term change due to aging. We present the effect of such variability on the speech signal, and demonstrate the associated impact on the error rates of current speaker recognition systems. Short-term variability due to stress may be induced by speaking in a noisy environment (the Lombard Effect), or as a result of physical exertion. A large change in the vocal effort of a speaker can result in whispered or shouted speech. Differences in the frequency content of speech in a neutral state and speech under stress, or at extremes of vocal effort, create challenges for speaker recognition. A speaker may also produce non-speech vocalizations, such as screams, whistles and coughs. It remains to be established whether enough speaker-discriminant information can be gathered from such vocalizations to enable recognition. Long-term variability due to aging is caused by gradual physiological change in the vocal apparatus, typically affecting the pitch, timbre, rate, and intensity of speech. Aging impacts negatively on current speaker recognition systems, causing error rates to increase progressively as the age difference between the samples under comparison increase. An additional challenge posed by the effect of aging is that its impact on the voice is both speaker and gender specific. Although we restrict our presentation to "normal" aging, additional vocal change may be observed in speakers with a physical or cognitive illness, such as Parkinson's disease, or those who relocate geographically. In real forensic comparison cases, one or more of the aforementioned sources of short-term variability are often present. If there is any significant time-lag in gathering recordings, the additional influence of aging must also be considered. We present an overview of recent research to address long- and short-term speaker variability in automatic speaker recognition, and highlight efforts for error mitigation in these cases. Finally, we mention the implications for forensic comparison and propose key areas for future study on this topic.

Speaker



Finnian Kelly is a research associate at the Center for Robust Speech Systems (CRSS) in The University of Texas at Dallas. Before joining CRSS, he was with the Sigmedia Research group at Trinity College Dublin, where he completed his PhD, 'automatic recognition of ageing speakers', in 2014. His main research interests are: speaker recognition in the presence of speaker variability, and the application of speaker recognition to forensics.

Session: Human Factors / Wednesday, July 22 / 11:00-11:20am

Cognitive Human Factors and the Use of Signature Features in Questioned/Known Signature Comparisons

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/cognitive_human_factors_and_the_use_of_signature_features_in_questioned_known_signature_comparisons.merlino.humanfact.pdf (30 slides)

■ Abstract

A substantial portion of Forensic Document Examiner (FDE) training is devoted to signature comparisons, handwriting, and hand printing. FDEs seek those features and characteristics which may represent the document's identifying attributes. Examiners first determine the presence or absence of features, and then qualitatively assign these features some degree of evidentiary weight to reach their decisions. Examiners are trained to look for both substantial similarities and differences among writing samples, and for repeated small characteristics which may sufficiently establish that writings are clearly the work of two individuals even though they may contain many general similarities. The number and quality of these features allow FDEs to determine the authorship of the specimen and the extent of their confidence in their decisions, yet not all FDEs reach the same conclusions given the same information. This paper reports findings from a national study of forensic document examiners (supported by Award No. 2010-DN-BX-K271) concerning the deployment of visual attention and the interpretation of signature features. Amos Tversky pointed out that most stimuli seem to be effectively described by the presence or absence of qualitative features. He and others argued that an object is represented by a set of features or attributes, and that judgments of similarity are achieved through a process of feature matching. Tversky's Contrast Model" systematizes this feature approach, and proposes that similarity depends on the proportion of features common to the two objects, and also on their unique features. Feature matching occurs by establishing differences in quality or quantity, such as differences in color or size, or the presence or absence of the features upon which the judgment is based, usually in terms of binary variables. This feature matching process, along with the deployment of attentional resources is a core process of forensic document examination.

The features available for forensic evaluation are determined in part by the nature of the writing specimens. Compared to stylized or mixed signatures, text-based signatures may offer a greater variety of features for evaluation. Additionally, signatures vary in terms of their complexity (e.g., the number of turning points and crossing lines), their semantic content, and any number of additional features commonly recognized within the profession as indicators of the authenticity of the writing.

Participant eye movements were recorded during a series of 66 questioned/known signature comparisons. FDEs approached the comparison aspect of the tasks differently from the Lay participants. The mean fixation count among FDEs was greater than that for Lay participants. The mean fixation count for mixed signatures and text-based signatures was significantly higher than that for stylized signatures, and among both FDEs and Lay participants the mean fixation count for stylized signatures was significantly lower than that for mixed signatures for both high complexity and low complexity signatures. These results demonstrate expertise effects in the deployment of attentional and cognitive resources, and the differences in accuracy indicate that the two groups weigh the available information differently.

■ Speaker

Mara Merlino, Kentucky State University, United States

Session: Human Factors / Wednesday, July 22 / 11:20-11:40am

Cognitive Human Factors and Call Accuracy in Limited-Information Signature Identification Tasks

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/cognitive_human_factors_and_call_accuracy_in_limited_info_signature_id_tasks.freeman.humanfact.pdf (22 slides)

■ **Abstract**

The development of expertise involves extensively greater proceduralization of problem-solving skills, tactics, and strategies. The cognitive advantages of perceiving and storing problems in terms of patterns, as well as the research demonstrating that experts in most domains are able to solve problems more quickly than are non-experts. This paper discusses findings from a national study of forensic document examiners (supported by NIJ Award No. 2010-DN-BX-K271) concerning the application of cognitive human factors to understanding the nature of attention, feature extraction and weighting, and decision-making in forensic document examination. Specifically, this paper will discuss the influence of top-down vs. bottom-up processing by comparing FDE and Lay participant decisions about whether text-based vs. stylized single signature specimens were genuine or simulated in two different experimental eye-tracking protocols.

The results of two protocols will be presented. In the single signature protocol, participants viewed one questioned signature without any known signatures for comparison for as long as they chose, and then made a decision about whether the signature was genuine or simulated. In the tachistoscope/extended view protocol, they viewed a single signature for one second, made a decision about whether the signature was genuine or simulated, then viewed the signature again for as long as they chose and again made a decision about whether the signature was genuine or simulated. Even when given a short period of time to view a signature, FDEs in most instances outperformed Lay participants. Compared to Lay participants, FDEs were able to reach a greater number of correct calls after viewing the signatures for only one second. This difference was even greater when participants were given the opportunity to view the signatures for an extended period of time. Lay participants were significantly more confident overall than were FDEs in their decisions. Eye-tracking results provide some support for the idea that different attentional and cognitive processes were deployed by FDEs and Lay participants. Semantic context provided by the text-based and mixed signatures allowed a greater extent of top-down processing and required fewer fixations, while the lack of semantic context increased the need to engage in bottom-up processing, requiring a greater number of fixations.

FDEs were more accurate overall than were Lay participants in both the tachistoscope view and the extended view of the signatures, although this varied among individual signatures. Given the limited information available overall, and the limited amount of time given to view the signatures in the tachistoscope view, these findings suggest that features such as line quality, speed and fluidity of execution, and other indicators of writing skill are valid and important indicators of signature authorship that are reliably used by FDEs to reach signature process decisions. These findings suggest that there is an expertise effect, such that FDEs are able to gather more information from minimal exposure to the signatures than are Lay participants, even though this effect is statistically unrelated to the amount of education, training, or experience among the FDEs.

■ **Speaker**

Dr. Tierra Freeman, Kentucky State University, United States

Session: Human Factors / Wednesday, July 22 / 11:40am-12:00pm

Study on Methods of Quality Control and Evaluation of Fingerprint Identification

Presentation Link: not available

■ Abstract

Fingerprint evidence plays an important role in criminal process because of its uniqueness and unchanged characteristics. But in recent years, the science of fingerprint identification has been questioned. The main issues are focused on opaque process, subjective judgments, lack of universal standard, no error rate expression and lack of scientific fundamental research data.

This paper is to set up a quality control and evaluation system for fingerprint identification process which is based on computer and internet technology, statistical tools and expert opinion and finally realize the fingerprint identification management model to increase the transparency and quality of fingerprint identification. We analysis the effect of expectation bias, background information deviation and character deviation to fingerprint identification quality and then suggest to establish recording system for whole identification process and finally point out effect of close no match fingerprints. This system includes four parts: mark quality evaluation system which quantify region and proportion of characteristics and then design a linear model for image quality evaluation, information recording system, data statistics and analysis system and quality evaluation system using R software.

106 identification agencies and about 200 fingerprint examiners around China were invited to participate experiments. We study accuracy and reliability of minutiae selection, cognitive ability of fingerprint examiners performance in analysis phase, reliability of fingerprint examiners judgment on comparison phase and accuracy of fingerprint identification conclusion in evaluation phase. The results show that there is a strong relationship between accuracy of fingerprint value judgment, stability of minutiae selection and accuracy of fingerprint identification conclusion. The interpretation of difference is an important factor affecting conclusion.

Basing on the findings above, we established scientific fingerprint identification procedures which includes an independent inspection procedures, a data recording and analysis system, a new fingerprint identification report review mechanism and a more detailed training model. Finally, we propose to future study on close non match fingerprint, distortion mechanism and scientific evaluation of fingerprint evidence.

■ Speaker



Dr. Shiquan Liu is a uniformed field technical investigator and also a latent fingerprint examiner in Jiangsu PD and has 8 years of experience in crime scene investigation (more than 1000 criminal cases) and fingerprint identification. Shiquan has finished his PhD thesis on setting up quality control and evaluation system to improve fingerprint identification quality in Chinese People's Public Security University (CPPSU). Shiquan is currently the post-doctoral researcher on physical evidence interpretation filed in China University of Political science and law (CUPL) and also the member of IAI, AAFS and ICSAI and is certified by The Ministry of Public Security of China as a senior latent print examiner.

Session: Criminalistics I / Thursday, July 23 / 11:40am-12:00pm

Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/crowd_powered_latent_fingerprint_ID-arora-crim1-2.pdf (20 slides)

■ Abstract

One of the most challenging problems in fingerprint recognition is comparing latent prints to rolled/slap (reference) fingerprints. Comparison of poor quality latent fingerprints to reference prints using an Automated Fingerprint Identification System (AFIS) does not typically yield satisfactory results. This is because many latents have only small friction ridge area with poor clarity and significant background noise. Therefore, it is a common practice to have a fingerprint examiner mark features on a latent for improving the AFIS hit rate. The increase in the hit rate, however, depends on the precision of the markup being input to the AFIS [1]. Imprecise markups can result in the mated reference print being returned at a lower rank amongst the retrieved candidates compared to when image alone is input to the AFIS [2]. Furthermore, markups for the same latent by different examiners can differ and, consequently, individual markups may lead to differences in the hit rate of the AFIS. To overcome the aforementioned limitations, we propose a synergistic crowd powered latent identification framework where multiple examiners and the AFIS work in conjunction with each other to boost the AFIS identification accuracy. Given a latent, its similarity scores to the top-N candidate prints output by the AFIS are used to determine the likelihood that a hit at rank-1 was found. A latent for which this likelihood is low is crowdsourced to a pool of examiners for additional markup. In this manner, the collective 'wisdom' of several examiners is utilized to obtain multiple markups for a latent when required. The manual markups are then used together with the AFIS to increase the likelihood of making a hit in the reference database. To evaluate the proposed framework, independent feature markups for the 500 ppi NIST SD27 latents were obtained from six certified latent print examiners affiliated to Michigan State Police (average feature markup time was about 5 min/latent). We also conduct experiments using two individual markups provided in the ELFT-EFS public challenge database. Experimental results show that the fusion of an AFIS with examiner markups improves the rank-1 identification accuracy by 7.75% on the NIST SD27, 11.37% on the ELFT-EFS public challenge database, against 250,000 reference prints. A greedy crowdsourcing strategy is also proposed where the latent is first crowdsourced to the best examiner to obtain a markup. The best examiner markup is then fused with the lights-out AFIS, and a decision whether additional markup is needed is then made. Experiments conducted using the six markups available for NIST SD27 indicate that after fusing three markups from the best three examiners, additional markups have negligible impact on the overall hit rate.

[1] Indovina et al. ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation# 1]. NISTIR, 7775, 2011.

[2] Dror et al. The Impact of Human Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. *Journal of Forensic Sciences*, 57(2):343-352, 2012.

■ Speaker

Mr. Sunpreet Arora, Michigan State University, United States

Session: Criminalistics I / Thursday, July 23 / 11:20-11:40am

A Novel Approach for Quantifying the Weight of Fingerprint Evidence

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/novel_approach_for_quantifying_the_weight_of_fingerprint_evidence.swofford.crim1.pdf (27 slides)

■ **Abstract**

Fingerprint analysts are faced with tremendous challenges when performing fingerprint comparisons and evaluating the significance of their findings. Not only are their analyses and comparisons typically performed visually without any tools capable of producing quantitative and statistically relevant data to assist in their interpretation of the evidence, but they must render and defend conclusions of source attribution based solely on their individual training and experience. Furthermore, these decisions are made without any formal or nationally accepted criterion or thresholds. Without tools capable of assisting the analysts with their interpretation of the evidence and standardized criterion by which decisions can be based, analysts have no internal quality assurance mechanism to protect them from making erroneous decisions other than the subjective examination of other analysts, which is valuable but not perfect having similar limitations. This presentation will discuss a novel, empirically derived approach for evaluating and quantifying the weight of fingerprint evidence based on the geospatial arrangement of friction skin features. The results of preliminary evaluations and policy guidelines which may be developed from these data will be presented along with the potential for transferring this technology into practice.

■ **Speaker**

Henry Swofford, Defense Forensic Science Center, United States

Session: Criminalistics II / Thursday, July 23 / 1:30-1:50pm

Congruent Matching – Theory & Application in Forensic Evidence Identification & Error Rate Estimation?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/song_abstract.pdf

■ Abstract

Reporting an error rate for forensic evidence identification is a fundamental challenge in forensic science. It is a national priority to establish a scientific procedure for quantitative error rate reports to support forensic evidence identifications in court proceedings. The Congruent Matching Cells (CMC) method was recently invented at NIST for accurate image-related forensic evidence identification and error rate estimation. The measured forensic images and topographies are divided into correlation cells. Four identification parameters are derived for identifying correlated cell pairs originating from the same source. This enables the estimation of error rates based on statistical analysis of the total number of correlation cells, the number of qualified CMC cell pairs, and the statistical distribution of the four identification parameters.

Initial validation tests on 780 topography image pairs of 40 cartridge cases ejected from guns with 10 consecutively manufactured pistol slides did not produce any false identifications or false exclusions. Validation tests have also been made using optical intensity images, using a different set of samples and three different correlation programs, all yielded clear-cut separation between known matching (KM) and known non-matching (KNM) image pairs. Validation tests using 946 image pairs of 44 fingerprints selected from the NIST's fingerprint database also showed correct identification and exclusion results. A statistical procedure using the validation testing results has been developed for estimating error rates. The false positive and false negative error rates for the validation tests are estimated; the Likelihood Ratio (LR) is above the "extremely strong" level as specified in the "2010 ENFSI Guideline in Evaluative Reporting in Forensic Science". The CMC method provides a statistical foundation to enable the estimation and reporting of error rates for court proceedings regarding ballistics identifications, thus emulating methods used in forensic identification of DNA evidence.

■ Speaker

John Song, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics II / Thursday, July 23 / 1:50-2:10pm

Assessing Error Rates for Firearm Identification Based on the CMC Methods

Presentation Link: not available

■ Abstract

Obtaining quantitative estimates for the error rate of firearm identification is a major challenge in forensic science. A new identification approach, the Congruent Matching Cells (CMC) method, provides an opportunity to develop a statistical approach to assess the uncertainty of the identification results. The CMC method is based on correlating patches (cells) of the correlated sample images or topography data. A match (identification) requires a sufficient number of CMCs, i.e., cell pairs that have a high degree of surface similarity and registration locations that are geometrically congruent. The CMC method yields a mathematically objective identification criterion, and addresses the challenge of comparing sample surfaces that contain areas with a poor transfer of the firearm surface topography. The method was applied to 780 pairwise comparisons of the breech face impression topography of 40 cartridge cases, ejected from 10 pistols with consecutively manufactured slides, and did not produce any false identifications or false exclusions.

Assessing the uncertainty of identification results requires models for the probability of obtaining a particular number of CMC cells in a correlation, both for samples originating from the same and different firearms. In principle, the respective probability distributions can be estimated from correlations of a sufficiently large collection of representative samples. However, the models have to accurately describe the tails of the distributions, as forensic tests with good sensitivity and specificity have a small overlap between the matching and non-matching distributions. Experimental data demonstrating false positive and false negative identifications may be limited, requiring a modeling approach that extrapolates the experimental data available.

Our initial approach to estimating the probability distributions was based on a binomial probability model for the number of CMCs in a correlation. This approach assumes that the cell identification trials are approximately independent and that all cells have the same false identification probability or the same false exclusion probability, which were estimated from the experimental data. For the known non-matching (KNM) sample correlations, this approach yielded an excellent match between the observed and modeled CMC distribution. For the known matching (KM) comparisons, however, the measured distribution was wider than the modeled distribution, indicating an estimated false negative error rate that is too low. We propose a new approach that provides better agreement by using a hierarchical binomial model. In this approach, the parameter for the false exclusion probability of a cell is also modeled by a probability distribution, currently the beta distribution, to account for differences in the firing conditions and domain of the samples. Using this same reasoning, we show that is possible to approximate the KM CMC distribution by a discretized and truncated Gaussian distribution, yielding comparable improvements to the fit between model and experiment.

Both models were applied to our experimental data and produced similar estimates for the false negative error rate. The variability of the estimated error rate was also evaluated through resampling of the experimental data (bootstrapping). These procedures, after further refinements of underlying assumptions, hold promise to enable the estimation of realistic error rates for CMC identification result.

■ Speaker

Daniel Ott, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics II / Thursday, July 23 / 2:10-2:30pm

An Analysis of Correlation Scores Between Cartridge Cases

Presentation Link: not available

■ Abstract

This presentation will contain some analyses of correlation scores between cartridge cases, focusing mostly only on data from NIST's National Ballistics Imaging Database Evaluation experiments.

The analysis will differ from previous treatments in focusing more on the differences between scores.

These differences can be seen as error and/or variation depending on the context, and may (or may not) be possibly be due to factors such as gun brand, ammunition brand, and measurement variability/error.

■ Speaker

James Yen, Statistical Engineering Division, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics II / Thursday, July 23 / 2:30-2:50pm

Measured Error Rates in Cartridge Case Comparisons

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/measured_error_rates_in_cartridge_case_comparisons.baldwin.crim2_.pdf (28 slides)

■ Abstract

This study was designed to measure examiner (not laboratory) error rates for false identifications and false eliminations when comparing an unknown to a collection of three known cartridge cases. Volunteer active examiners with Association of Firearm and Toolmark Examiners (AFTE) membership or working in laboratories that participate in ASCLD were provided with 15 sets of 3 known + 1 unknown cartridge cases fired from a collection of 25 new Ruger SR9 handguns. The ammunition was all Remington 9-mm Luger (manufacturer designation L9MM3) and sets were made up of cartridge cases fired within 100 cartridges of each other for each gun. Examiners were provided with a background survey, an answer sheet allowing for the AFTE range of conclusions, and return shipping materials. They were also asked to assess how many of the 3 knowns were suitable for comparison, providing a measured rate of how often each firearm used in the study produces useable, quality marks. The participating examiners were provided with known positives and known negatives from independent groups of samples, providing independent measurements of a false-positive rate and independent measurements of a false-negative rate, allowing the study to measure both rates and statistical uncertainties in those rates.

Responses were received from 218 participating examiners. The rate of false negatives (estimated as 0.367% from comparisons known to be from the same firearm but reported as eliminations) was quite low with the error distributed across examiners of various backgrounds (state, federal, local, private, etc. as determined from self-reported survey information). The overall rate of false positives (estimated as 1.01% from comparisons known to be from different firearms but reported as identifications) was significantly higher. However, most of the errors were reported by a small number of examiners; that is, individual examiners have varying error rates. For most examiners this rate is quite low while for some it is relatively high. Hence the overall rate is best interpreted as an average of widely varying individual rates. Inconclusive results were not recorded as errors. Rates of poor quality mark production for these handguns varied across the 25 sample handguns. Those rates were $2.3 (\pm 1.4) \%$.

False-positive and false-negative error rates for individual examiner performance on comparisons were measured. The rates are not uniform across the sample population with a few examiners providing most of the false-positive responses. False-negative rates are low and comparable to or lower than the rate of production of poor quality marks by the firearms used in this study. Laboratory reporting error rates may be significantly lower than these individual rates if quality assurance procedures are applied that can effectively manage to reduce or eliminate the propagation of false positives reported by individuals.

■ Speaker

Dr. David Baldwin, Midwest Forensics Resource Center, United States

Session: Criminalistics II / Thursday, July 23 / 3:30-4:00pm

Gunshot Distance Determination, Crystal Ball Reading or Science?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/firearm_distance_determination_science_or_crystal_ball_reading.paradis.crim2_.pdf (57 slides)

■ Abstract

When a firearm is fired besides the bullet other materials are expelled in the firing process. Smoke (soot), primer residues, partially burned and unburned gunpowder particles, metal fragments as well as any other residues that were left in the barrel from before the shot was fired. These particles can be used to determine the distance a muzzle was from the area of impact at the time the shot was fired.

‘Distance to target’ aka ‘muzzle to target’, and ‘distance determination’ tests have been used by forensic and crime laboratories for many years to determine the distance a gun was fired from. Factual evidence is often used to confirm a witness or suspect’s account of what happened. This presentation of distance testing science will focus on gun powder particles and nitrites. The firing process converts nitrates to nitrites and can deposit these nitrites on an item, commonly clothing. The detection of these particles is accomplished by visual, microscopic and chemical testing in combination as standard steps in the process. These particles can be as tiny as soot like material being very small to sub-microscopic to very large such as a partially burned gunpowder particle visible by the naked eye. Gunpowder is not totally efficient, many particles are discharged from the gun unburned acting as a mechanism to transport small particles of nitrite farther. This mechanism nitrite particles either staying on the gunpowder particle which adheres to the impact point or being transferred to the impact point by the impact. The item or clothing is subjected to a chemical process developing a color pattern. This pattern is compared to a known firing distance patterns created by the examiner to determine a distance. Generally speaking the larger and more diffuse a pattern the greater the distance

Over the years the author has found numerous mistakes with the ‘how’ the science is carried out in different laboratories. Often steps are cut out to save time and money. Generally speaking the process has validity. Not adhering to the standards as suggested by AFTE and SWGGUN as well as the departments conducting these tests own protocols often lead the examiners to faulty conclusions. There are other considerations that are not in many lab protocols as well. Clear violations of the protocols of these two organizations have led to numerous false reports and testimony as will be shown in this presentation. These will include examples of documented failure to follow protocols as well as examples of important information that is not considered in many lab protocols. It will be demonstrated to show that this ‘scientific’ procedure can easily give the illusion of science and be easily misrepresented by the examiner. This will be followed by suggestions to correct protocols and common errors.

Justification

Defendants lives rest on correct detection, testing and evaluation of evidence. This particular science is used often in the courtroom but has not undergone the scrutiny of other procedures. Acknowledgement here is needed.

■ Speaker

Paul Paradis, Paradise Firearms, United States

Session: Criminalistics II / Thursday, July 23 / 4:00-4:20pm

Lead Density on a Target, A Significant Indicator of Firing Distance but is it Reliable?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/lead_density_on_a_target_a_significant_indicator_of_firing_distance_but_is_it_reliable.gardner.crim2_.pdf (27 slides)

Abstract

There is an inverse relationship between the mass of lead on a target and distance. Quantifying the lead by atomic absorption spectroscopy (AAS) has been proposed as an objective method for determining firing distance. In the analysis, a series of test fires are made at known distances. The mass of lead contaminating the target is determined by extracting it from a specified area around the bullet hole and measuring the concentration via atomic absorption spectroscopy. A linear calibration curve is generated by plotting the natural log of the mass of lead against the firing distance. Previous studies have shown a strong linear relationship between lead density and distance, with coefficients of determination of $r^2 = 0.94$ or better.

The methods in the literature were reproduced using four firearm/ammunition combinations

- A. Hi Standard .22 Double-Nine revolver/ Aguila .22 Long Rifle High Velocity ammunition with copper plated bullets (40 gr)
- B. Hi Standard .22 Double-Nine revolver/Remington .22 Long Rifle High Velocity ammunition with round nose lead bullets (40 gr)
- C. Smith and Wesson Model 65 revolver/ Winchester .38 Special FMJ ammunition (130 gr)
- D. Smith and Wesson Model 65 revolver/ Freedom Munitions .38 Special RNFP ammunition (158 gr)

The results were mixed. Most notably, lead was detected and a calibration curve with an $r^2 = 0.96$ was generated for jacketed Winchester ammunition with a lead-free primer in the Smith and Wesson (C). Calibration curves for systems A and B were generated, with coefficients of determination $r^2 = 0.98$ (A) and 0.92 (B), respectively. However, when the calibration curves were tested with test fires from known distances, the rejection rate was nearly 28%, even though the distance was calculated with a 95% confidence interval. While the coefficient of determination is often used as a measure for the fit of a calibration curve, it is not an adequate method for validating models such as calibration curves. A better method would be graphical residual analysis.

In any calibration curve, the intensity of the response must depend exclusively on one variable. There are many factors that can affect the mass of lead deposited on a target. These include volume of powder, mass of the bullet, air currents, barrel temperature, or composition of the target. In fact, the mass of lead initially increased with distance when system D was used to shoot at a target of a cotton knit fabric. Further examination showed that the gunpowder residue was blown through the cotton fabric at close range and deposited on the cardboard support.

Finally, much of the literature on quantitative methods for muzzle-to-target distance determination focuses on method development. Few have tested the method with test fires from known distances or blind tests from unknown distances.

Even so, the results are quite promising. The solution to the lack of reliability may be as simple as increasing the number of replicate shots at each of the distances in the calibration curve. Further research into developing a reliable quantitative method for muzzle-to-barrel distance determination and future research will be presented.

Speaker

Elizabeth Gardner received her BS from Penn State and her PhD from Michigan State, both degrees are in chemistry. After doing her post-doc at Argonne National lab, she joined the Department of Chemistry at the University of Texas at El Paso as a faculty member. There she became interested in forensic science and trained the drug chemists in the El Paso Police Department Crime Lab in preparation for accreditation, which was attained in 2006. She then accepted a position at the University of Alabama at Birmingham Department of Justice Sciences, where she is currently an associate professor of forensic chemistry. Dr. Gardner research focuses on the identification of emerging drugs of abuse, illicit internet pharmacies, and pattern evidence. She is a member of the American Academy of Forensic Sciences and the American Chemical Society. She also gives workshops for the Girls in Science and Engineering Day and the Avondale Library Teen Book Club each year.

Session: Criminalistics II / Thursday, July 23 / 4:20-4:40pm

Quantitative Metrics for Identifying Characteristic GSR Particles

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/quantitative_metrics_for_identifying_characteristic_gsr_particles.ritchie.crim2_.pdf (24 slides)

■ Abstract

Most laboratories that perform gunshot residue (GSR) analysis use automated software to search for particles using a scanning electron microscopy with energy dispersive x-ray spectrometer (SEM/EDS). However, to be admissible as evidence in court, any particles discovered by the automated software must be relocated, reanalyzed and confirmed by a human operator. There is sense to this requirement. A trained human is able to view the SEM/EDS spectrum and identify the characteristic peaks indicative of the elements that must be present in a characteristic particle. A trained operator is also able to filter out false positives that may result from poor quality spectra or the mis-identification of elemental signatures. However, inevitably trained human analysts introduce subjectivity. It will mean different things to different trained analysts to say that evidence for an element is present in a spectrum. Some analysts will be satisfied with the mere hint of a characteristic peak while others may also require evidence of minor peaks. It could be argued that this subjectivity does not provide equal justice for all. It would be better to be able to make objective, statistically defensible statements about the presence or absence of an element in a spectrum. These metrics could be independent of laboratory and instrumentation. The metrics could weed out poor quality spectra and identify spectral artifacts. The SEM/EDS community knows how to do this but commercial tools and current standard protocols don't facilitate it. We will discuss techniques to answer whether evidence for an element is present in a spectrum and metrics to express the degree of confidence. We will also discuss briefly how implementing such a mechanism would simplify other aspects of a GSR standard (like instrument performance) which are otherwise somewhat arbitrary.

■ Speaker

Nicholas Ritchie, National Institute of Standards and Technology (NIST), United States

Session: Criminalistics II / Thursday, July 23 / 4:40-5:00pm

Uncertainty Associated with the Elemental Analysis and Forensic Comparison of Materials using Laser Ablation Inductively Coupled Plasma Mass Spectrometry

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/uncertainty_associated_with_the_elemental_analysis_and_forensic_comparison_of_materials_using_laser_ablation_inductively_coupled_plasma_ms.almirall.crim2_.pdf (57 slides)

Abstract

The chemical analysis and comparison of a number of solid matrices of interest to forensic scientists has progressed from fundamental research and method development to validation in several forensic laboratories and resulted in the publication of international method standards. Elemental analysis of forensic materials including glass, paint, soils, precious metals, diamonds, paper and ink on paper begins with the first reporting of analytical techniques in the scientific literature by a research laboratory and, eventually, forensic laboratories adopt the developed methods for use in casework. This presentation describes the many steps that need to take place in between the first research reports of a method and adoption of the methods for casework including a very careful evaluation of the 'errors' associated with taking the measurements and the overall uncertainty budget associated with Laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) for the analysis of these matrices. LA-ICP-MS has been called the 'gold' standard for solid-sampling and high-sensitivity elemental characterization of materials providing sub-ppm detection limits of elemental analytes encompassing more than 70% of the periodic table. In addition, LA-ICP-MS can provide true quantitative analysis data for use in numerical/statistical hypothesis testing to determine a 'match' and also to populate databases that are useful to estimate the frequencies of occurrence for a given elemental profile within a target population of the material. The efforts of several forensic laboratories collaborating on method development and optimization, as reported by the European Union-funded NITECRIME effort (2000-2005) [1] and continued by the NIJ-funded Elemental Analysis Working Group (EAWG) effort (2008-2012) [2] proved critical for an understanding of the uncertainty associated with this analysis. In the case of the glass example, carefully designed interlaboratory trials revealed the method performance and limitations of these methods for the analysis of glass in forensic laboratories [3] paving the way to the publication of international (ASTM) analytical consensus standards for the examination these materials [4]. More than 30 forensic laboratories around the world now routine employ the use of LA-ICP-MS for materials characterization on every continent and the experience, lessons learned and an understanding of the various sources of uncertainty elemental analysis provides a good model for how forensic method development should progress from basic research to routine use and acceptance in the courtroom.

1. C Latkoczy, et al, Evaluation of a standard method for the quantitative elemental analysis of float glass samples by LA-ICP-MS, J. of Forensic Sciences, 2005, 50 (6), 1327-1341.

2. T Trejos, et al, Cross-validation and evaluation of the performance of methods for the elemental analysis of forensic glass by μ -XRF, ICP-MS and LA-ICP-MS, Anal. Bioanal. Chem., 2013, 405, 5393-5409.

3. T Trejos et al, Forensic analysis of glass by μ -XRF, SN-ICP-MS, LA-ICP-MS and LA-ICP-OES: evaluation of the performance of different criteria for comparing elemental composition, J. Anal. At. Spectrom., 2013, 28, 1270-1282.

4. Standard Test Method for the Determination of Trace Elements in Soda-Lime Glass Samples Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry for Forensic Comparisons. ASTM E2926-13 (2013)

Speaker



José R. Almirall is a Professor in the Department of Chemistry and Biochemistry and Director of the International Forensic Research Institute (IFRI) at Florida International University. He was a practicing forensic scientist at the Miami-Dade Police Department Crime Laboratory for 12 years, where he testified in over 100 criminal cases in state and federal courts prior to his academic appointment at FIU in 1998. Professor Almirall has authored one book and ~120 peer-reviewed scientific publications in the field of analytical and forensic chemistry and presented ~600 papers and workshops in the U.S., Europe, Central and South America, Australia, New Zealand, Japan and South Africa. He was appointed to serve on the Forensic Science Standards Board in 2015.

Session: Digital Evidence / Thursday, July 23 / 1:30-1:50pm

Improving Cyber Forensics & Cybersecurity through Block Chain Technology with Truth-based Systems

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/improving_cyber_forensics_and_cybersecurity_through_block_chain_technology_with_truth_based_systems.zatyko.digevid.pdf (23 slides)

■ Abstract

Let's eliminate errors in chain of custody and evidence record keeping systems and add trust anchors and precise time to all entries. Could a chain of custody process be automated and verified through block chain technology similar to methods used by Bitcoin to verify truthfulness? Could this system scale for billions of entries?

The truth would be inside the block chain, which can be used to verify the status of infrastructure and provide complete chain of custody for all data generated and transmitted through that infrastructure. This system could even detect nefarious acts by insiders or hackers. Some changes may indicate a breach, which can be acted upon rapidly.

Keyless Signature Infrastructure (KSI) is an example of a blockchain technology optimized for the Internet. Trusted parties are eliminated for verifying the integrity and provenance of both infrastructure components and data generated from that infrastructure.

One of the most significant trends over the last few years, (spearheaded by Bitcoin and others) has been the move away from centralized trust authorities to decentralized "consensus" trust models where assertions about what is and what is not true can be verified independently using a public ledger built using consensus based decision making. Everyone can independently verify the status of that infrastructure and associated transactions. Some can be allowed to see content and some only a verification code of trustworthiness. Regulators get to audit all processes and everyone involved can verify what happened 'after the fact' and act in real time when things go wrong. It would transform our society from one that is trust based to one that is truth based, i.e. humans can choose to trust each other, but they can also verify; they can prove what happened without trusting anyone.

One way forensics practitioners can utilize this revolutionary technology is in their evidence systems to eliminate errors. It can also be used to save a considerable amount of time when conducting digital forensics investigations by providing a forensics ready environment for data integrity. Let us show you how.

■ Speaker

Ken Zatyko, Guardtime, United States

Session: Digital Evidence / Thursday, July 23 / 1:50-2:10pm

Measuring and Mitigating Errors in a Digital Forensics as a Service Environment

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/measuring_and_mitigating_errors_in_a_digital_forensics_as_a_service_environment.harmvanbeek.digevid.pdf (27 slides)

■ Abstract

In the Digital Forensics arena, in a simple case, the number of traces like files, pictures and log lines can easily reach the million mark. Because analyzing all those traces by hand is impossible, software needs to be developed to assist in that endeavor. However, developing software is not without risk, the number of errors in code can be as high as 15-50 in every 1000 lines of code [ref: Code Complete]. The type of errors in the forensic context can lead to the same type I and type II errors as in other forensic area's. Because low-level traces are always combined to obtain higher level traces, error propagation is very unpredictable. For example, events from a windows event log are combined to establish possible clock skew of the computer. An error in the event log parser may lead to missing events. The forensic impact of those missing events can range from missing where someone adjusted the clock, or 'only' adding more certainty to the already existing conclusion.

The presentation shows the initiatives taken by the department of Digital Technology & Biometrics of the Netherlands Forensics Institute to improve software and thus reduce the absolute number of type I and type II errors during the analysis:

More formal methods for describing data structures. A major source of errors in interpreting digital traces is that the interpretation is based on reverse-engineering effort of the binary data, without aid of documentation and/or original source code. By using formal notations for the interpretation, it is clearer when the interpretation is in error, thus improving awareness of the existence of errors.

Using frameworks that promote openness and transparency of the resulting traces, e.g. when was that trace generated by which version of which tool. By increasing the transparency, it is easier to re-apply improved tools only to those traces, and gauge the difference between the results.

Using a rigorous process for reviewing code as well as results. Literature and empirical studies have shown that the number of errors reduces when programming code is reviewed [ref: Code Complete]. By making code reviews required, we improve the overall error rate of code and resulting traces.

Create an environment where results, evidence and logging are centrally managed. By aggregating all logging generated when running the software, it will be easier to assess the impact of errors found.

The initiatives are the result of a quality improvement process at the NFI, involved in running Digital Forensics as a Service (DFaaS). When you run DFaaS, you run into issues an 'ordinary software vendor' does not have. One of the biggest issues is how to proceed when errors are found. Can you 'just rerun the analyses? Are you going to find all the users that accessed an incorrect result and tell them? This bears relevancy to other initiatives that work on 'Forensics as a Service' on a large scale as well.

■ Speaker

Harm van Beek, Netherlands Forensic Institute

Session: Digital Evidence / Thursday, July 23 / 2:10-2:30pm

Measuring Systematic and Random Error in Digital Forensics

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/measuring_systematic_and_random_error_in_digital_forensics.nelson.digevid.pdf (26 slides)

■ Abstract

Recognized sources of error in digital forensics include systematic errors arising from implementation errors, and random errors resulting from faulty equipment. But as digital forensic techniques expand to include statistical machine learning, another source of error will be statistical errors that arise because of chance disagreements between a statistical model and subject systems examined with that model. We consider two digital forensics systems with these different types of measurable error.

First, we show a mechanism for comparing the numerous and nuanced results of parsing a file system. Multiple storage system parsers were designed for or adapted to analyze a game console with a custom file system. However, it was initially unknown whether any of the parsers would produce a perspective of the storage system that was correct in reporting the files present and their characteristics. We adapted the parsers to produce an in-common, machine-differentiable format, and used a storage differencing algorithm to measure the relative incorrectness of each of the parsers. Discrepancies summarize errors in implementation or specification, an important report when any reverse-engineering is necessary. We discuss advantages and challenges in adopting this practice.

Second, we show how to construct a classifier using the hard drive from a multi-user computer that can determine the user responsible for creating a file. The classifier is constructed using allocated files and its accuracy determined with take-one-out cross-validation. Once created, the classifier can be used to predict the creator of files that can only be recovered with carving.

■ Speaker



Alex Nelson is a Ph.D. Candidate at the University of California, Santa Cruz, under Professor Darrell D. E. Long. His published research includes digital forensic storage analysis, analyzing and expressing the evolution of systems for forensic signatures. He is a member of the Computer Security Division at NIST.

Session: Digital Evidence / Thursday, July 23 / 2:30-2:50pm

A Logic Based Model for Error Management in Network Forensics Analysis

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/a_logic_based_model_for_error_management_in_network_forensics_analysis.singhal.digevid.pdf (16 slides)

Abstract

Network forensics is the science that deals with capture, recording and analysis of network events and traffic for detecting intrusions and investigating them. Network forensics involves post mortem investigation of the attack and is initiated after the attack has happened. Different stages of legal proceedings (obtaining a warrant or evidence to the jury) require reconstructing an attack scenario from an attacked system. In order to present the scenario that can be best supported by evidence, digital forensics investigators analyze all possible attack scenarios reconstructed from the available evidence.

There are a lot reasons for errors in this process:

- 1) The evidence can be a false positive as the Intrusion Detection System (IDS) did not generate the correct alert.
- 2) The number of security events is very large so finding an attack pattern is like finding a needle in a haystack given that IDS systems generate a large amount of duplicate events that need to be correlated prior to analyzing them for forensic analysis.
- 3) Some evidences cannot be found because the attacker used anti-forensic technique to destroy the evidence.
- 4) Some evidence may be missing due to storage limitations on the devices that collect them.

There needs to be a good method for detection, analysis and mitigation of such errors. In this paper we present a model and an accompanying software tool that systematically addresses these errors and how to resolve them to reconstruct the attack scenario. Firstly, we map the evidences to existing system vulnerabilities and use it to detect and mitigate errors due to false positive alerts. Secondly, we have developed an anti-forensic database that is used to detect and mitigate errors due to deletion of evidences by the attacker. Thirdly, we use primary, secondary and tertiary storage methods to continuously correlate and back-up the events through the memory hierarchy and provide an interface to recall 'useful' events that can be furnished as evidence. All three procedures need to be aware of the inherent errors in categorizing large quantities of evidence, using known anti-forensics databases and providing event records as potential evidence. In collecting a series of events as potential evidences there is a possibility for exacerbating the error margins to unacceptable levels and thereby exceeding the legally acceptable error margins. Also, the error margins may permit using the same series of evidence to support alternative explanations. Therefore there needs to be a uniform method that (1) assigns an overall error margin to the process from individual error margins and (2) provides an acceptable error margin separation between two potential explanation for the same set of evidence. We propose to use Bayesian networks to compute errors for the processes or scenarios and integrate them into our reasoning framework.

The goal of our work is to reduce the errors and the investigator's time/effort in reaching definite conclusions about how an attack occurred. Our experimental results show that such a system can be useful for error management in network forensics analysis.

Speaker



Dr. Anoop Singhal, is currently a Senior Computer Scientist in the Computer Security Division at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD. He received his Ph.D. in Computer Science from Ohio State University, Columbus, Ohio. His research interests are in network security, network forensics, cloud computing security, and data mining systems. He is a member of ACM, senior member of the IEEE and he has co-authored over 50 technical papers in leading conferences and journals. He has also co-edited a book on Secure Cloud Computing.

Session: Human Factors / Thursday, July 23 / 11:20-11:40am

Combating Confirmation Bias: Can Forensic Science Benefit from Importing Eyewitness Identification Procedures?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/combating_confirmation_bias_can_forensic_science_benefit_from_importing_eyewitness_id_procedures.kukucka.humanfact.pdf (18 slides)

■ Abstract

Forensic science errors have been uncovered in an alarming 47% of DNA exoneration cases (Innocence Project, 2015). In a recent target article (Kassin, Dror, & Kukucka, 2013), my colleagues and I argued that some such errors may be attributable to forensic confirmation bias i.e., the subconscious tendency for examiners to analyze evidence in ways that validate their pre-existing beliefs. Indeed, nearly a century of basic psychological research has demonstrated that confirmation biases are pervasive, automatic, and often problematic (Nickerson, 1998).

Two potential sources of bias have been identified. First, recent studies have shown that forensic examiners across a range of domains are influenced by a priori exposure to extraneous contextual information (see Kassin et al., 2013). As a countermeasure, many have espoused the adoption of sequential unmasking protocols, which shield examiners from potentially biasing information until after the critical stages of the analysis are complete (Krane et al., 2008).

Second, it has been argued that examiners may develop a base-rate expectation of guilt that predisposes them toward inculpatory judgments (Whitman & Koppl, 2010). To mitigate this expectation, some have proposed that examiners should make judgments from evidence lineups (e.g., Risinger, Saks, Rosenthal, & Thompson, 2002). That is to say, a questioned sample should be compared against an array of samples rather than against a single known sample.

In this presentation, I will discuss our ongoing program of research to test methods of counteracting these sources of bias.

First, Kukucka and Kassin (2014) found that individuals who knew that a suspect had confessed even though the confession was recanted saw greater similarity between handwriting samples from the suspect and perpetrator, and were more likely to incorrectly believe that the suspect had authored the questioned sample, relative to individuals who were unaware of the confession. In a follow-up study, we replicated these findings by showing that the same individuals adjusted their own judgments after learning of the confession. (Time permitting, I will also present a third study in which we tried to identify the conditions under which this bias effect was most pronounced.) Taken together, these findings underscore the value of sequential unmasking to protect against bias.

Second, I will present unpublished data from two studies which tested how evidence lineups impact the biasability and accuracy of handwriting identification judgments. In each of these studies, individuals who knew of a suspect's confession produced less biased judgments when using an evidence lineup. As such, these studies provide the first empirical support for the notion that evidence lineups might protect examiners against confirmation bias.

In these studies, evidence lineups produced systematic effects on judgment accuracy that were strikingly similar to those seen in studies of eyewitness identification (see Wells et al., 1998). These data thus support claims that evidence and eyewitness lineups function similarly in terms of reducing systematic error. These findings suggest that the forensic sciences may stand to benefit from designing their own procedures after best practices for eyewitness identification, which presents many fruitful avenues for future research.

■ Speaker

Jeff Kukucka, Towson University, United States

Session: Human Factors / Thursday, July 23 / 11:20-11:40am

The “Six Thinking Hats” Method of Removing Bias from Case Review

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/blackledge-abstract.pdf> (35 slides)

■ Abstract

A Case Review should occur prior to closing any major criminal investigation or taking a criminal case to trial. With no preconceived notions case review should consider all aspects of the investigation, all possible motives, subjects, interpretations of the evidence, and ask if there are any investigation avenues that haven't been pursued or pursued to a sufficient extent.

Unfortunately, we have all had the experience at meetings where one or two individuals by the force of their rank, overbearing personalities, loud voices, and strong, inflexible opinions dominate the discussion. These individuals remind me of my days in the military and the often-used entry in a fitness report: "This officer is often wrong, but is never in doubt." Such situations are totally counterproductive as far as the desired goal of a thorough and completely unbiased case review, and tend to skew the group's objective assessment of sentinel events in the criminal justice system.

There is a way of conducting meetings that prevent their domination by a few individuals, guarantee participation, and insure the topic under discussion is considered in an unbiased manner. This method was introduced by Dr. Edward deBono in his book, *The Six Thinking Hats*. The "Six Hats" are visualization tools that help sidestep the ego and provide a path to nonjudgmental decision making. The six "thinking hats" are different ways of looking at an issue that has to be decided. By giving each person a role (and each person eventually playing all of the roles), the method reduces the amount of personality-based conflict, encourages more participation, and gives validation to many different ways to present the question.

This presentation will provide a brief introduction of the visualization tools used in The Six Thinking Hats and how their incorporation into the Case Review process can minimize the likelihood of Sentinel Events.

■ Speaker

Robert Blackledge (he did not attend and his presentation was given by someone else), United States

Session: Human Factors / Thursday, July 23 / 11:40-12:00pm

Mitigation of Forensic Error Rates in Evaluating Pattern-Based Handwriting Evidence Using a Lineup Process

Presentation Link: https://www.nist.gov/sites/default/files/documents/director/miller_abstract.pdf

■ Abstract

It is documented that forensic science errors in pattern-based sciences can be mitigated through management of potentially biasing information. Biasing information can include contextual information about a case or a suspect, but errors can also be incurred through perceptual judgment (e.g., seeing similarities in patterns between samples while overlooking subtle differentiating details). Previously, through published research concerning bias in the forensic analysis of handwriting and human hair, forensic examiners error rate decreased when they were provided with a 'lineup' as a part of their decision-making process, helping to guard against contextual bias (Miller, 1984, 1987). When the testing subjects were provided with biased information concerning the case details and a single suspect, rather than a lineup, the subjects' error rate significantly increased.

To illustrate the problems associated with bias that occur in forensic examination and presentation of evidence in the courtroom, a forensic handwriting examination case study is presented by the authors involving a questioned block printscript sample that was 'matched' to a single suspect's handwriting samples and identified by a forensic document examiner. When the case was adjudicated, it was demonstrated, through a lineup process, how the simple, block printscript matched class or common characteristics of several writers, not only the suspect. The study also demonstrated how, through a lineup process, several evaluators mistakenly picked the wrong suspect out of the handprinting lineup. Through the case study, it is demonstrated how simple matches in pattern-based evidence are persuasive and potentially misleading to judges and jurors, especially since the sensationalism of forensic evidence in the media promotes perceptual error through making simple assumptions about matching samples.

Research has shown that the brain's experience with patterns can create perceptual expertise causing the brain to automatically process information and create holistic patterns rather than observe its featural characteristics (Curby & Gauthier, 2010). Assumptions about how handwriting is unique and identifiable also need critical examination as the accuracy of this assumption is dependent upon the unique qualities and quantity of the samples available. Forensic practitioners who evaluate pattern-based evidence need training that targets the automatic processing that occurs during perceptual expertise and challenge assumptions when working with comparative evidence. This training should also include how to counter the cognitive biases and perceptual expertise effects that judges and jurors are suspect to when observing comparative evidence. Creating checks and balances in the procedures and protocols in evaluating pattern-based evidence should be developed. Forensic practitioners can guard against and mitigate these errors by employing simple laboratory protocol and procedures. Methods and procedures for developing and using lineup processes accurately and effectively will be discussed and demonstrated.

Curby, K. M., & Gauthier, I. (2010). To the trained eye: perceptual expertise alters visual processing. *Topics in Cognitive Science*, 2(2), 189-201.

Miller, L. S. (1984). Bias among forensic document examiners: A need for procedural changes. *Journal of Police Science and Administration*, 12(4), 407-411.

Miller, L. S. (1987). Procedural bias in forensic science examinations of human hair. *Law and Human Behavior*, 11(2), 157.

■ Speaker

Larry Miller, East Tennessee State University, United States

Session: Human Factors / Thursday, July 23 / 11:40am-12:00pm

Development of a Reasonable Minimum Documentation Standard in Latent Print Analysis and Comparison

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/development_of_a_reasonable_minimum_documentation_standard_in_latent_print_analysis_and_comparison.eldridge.humanfact.pdf (35 slides)

■ Abstract

It is widely recognized that the comparison and identification of fingerprints is a process that incorporates subjective interpretation and decision-making at every step, thus it can come as no surprise to anybody that where human interpretation is involved, variability will inevitably follow. This variability may lead to disagreements between practitioners, errors, or simply confusion: how did two people observing the same data reach different conclusions?

Without documentation of the comparison process, none of these issues can be satisfactorily addressed.

Without a clear roadmap to how decisions were made, there is no way to reconstruct the decision-making process, no way to identify and mitigate the sources of error, and no way to transparently present the basis of the conclusion to a fact-finder.

When an error is discovered, it is critical to the root cause analysis and remediation process to understand how the error occurred. Was the failure one of interpretation, incompetence, bias, training, overconfidence, or deliberate fraud? The technical review process is intended to be a way for the reviewer to check that there is a sufficient basis for the conclusion reached, but without documentation of that basis, how is this to be achieved?

While it has increasingly been suggested by researchers, critics, and court decisions that contemporaneous bench notes for latent print examinations are desirable, many agencies do not currently have SOPs that define what notes are required in various circumstances, and many latent print examiners don't take any notes at all, take only minimal orientation- and anatomical source-indicating notes, or reserve note-taking for exceptionally difficult cases.

Part of the problem seems to be that agencies and practitioners fear that the note-taking process will be unnecessarily time-consuming, while for others, the will is there, but they just don't know where to begin:

Which marks require documentation? How much is necessary? How should we define 'complex'?

This lecture will first explore the philosophical reasoning behind providing contemporaneous documentation: why do we take notes? Who are we taking them for? What do we hope to accomplish or convey with our bench notes? How can they be used as a tool for mitigating error? The lecture will then describe the steps one agency has taken in the journey toward creating a reasonable minimum documentation standard that meets the needs and goals of documentation, while still maintaining an operationally productive environment.

■ Speaker

Heidi Eldridge, RTI International, United States

Session: Human Factors / Thursday, July 23 / 1:50-2:10pm

A Typology of Underexploitation of Latent Print Evidence

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/typy_latent_cole_human.pdf (25 slides)

■ Abstract

Previous discussions of error in latent print (fingerprint) analysis have tended to focus on erroneous individualization, at the expense of erroneous exclusions. In addition, such discussions have tended to apply a simple signal detection approach based on binary decision outcomes. These approaches are incomplete for two reasons. First, they oversimplify the full range of decision outcomes available to contemporary latent print analysts. Second, they fail to account for the full range of treatment of latent print evidence in the criminal justice system in actual criminal cases. This paper proposes a typology of fingerprint 'underutilizations' that seeks to account for the full range of possible analyst decisions and criminal justice system treatments of latent print evidence. The typology is based, in part, on a review of actual wrongful conviction cases that involved latent print evidence. It is hoped that the proposed typology can offer a coherent framework for discussing 'error' in latent print analysis and for measuring specific types of latent print underutilization. It is also hoped that the typology may serve as a model for similar typologies for other criminalistics disciplines.

■ Speaker

Simon Cole, University of California – Irvine, United States

Session: Human Factors / Thursday, July 23 / 11:20-11:40am

Objective Classification of Fingerprint Image Complexity

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/objective_classification_of_fingerprint_image_complexity.swofford.humanfact.pdf (25 slides)

■ Abstract

Fingerprint analysts are faced with tremendous challenges when performing fingerprint comparisons and evaluating the significance of their findings. Not only are their analyses and comparisons typically performed visually without any tools capable of producing quantitative and statistically relevant data to assist in their interpretation of the evidence, but they must render and defend conclusions of source attribution based solely on their individual training and experience. Furthermore, these decisions are made without any formal or nationally accepted criterion or thresholds. Without tools capable of assisting the analysts with their interpretation of the evidence and standardized criterion by which decisions can be based, analysts have no internal quality assurance mechanism to protect them from making erroneous decisions other than the subjective examination of other analysts, which is valuable but not perfect having similar limitations. This presentation will discuss a novel, empirically derived approach for evaluating and quantifying the weight of fingerprint evidence based on the geospatial arrangement of friction skin features. The results of preliminary evaluations and policy guidelines which may be developed from these data will be presented along with the potential for transferring this technology into practice.

■ Speaker

Henry Swofford, Defense Forensic Science Center, United States

Session: Human Factors / Thursday, July 23 / 2:30-2:50pm

Minimizing *and* Leveraging Bias in Forensic Science

Presentation Link:

http://www.nist.gov/director/upload/minimizing_and_leveraging_bias_in_forensic_science-adams-2.pdf

(16 slides)

■ Abstract

Addressing bias in forensic science must entail more than temporarily hiding information from a bench examiner though such information hiding is essential. We suggest that blinding should be complemented by measures to leverage remaining biases and that blinding measures should be embedded in a right mix of complementary measures, some of which we will describe as 'hierarchical' and others as 'distributed.' To make our case we explain that there are several very different ways to handle bias in forensic science. We offer a classification of strategies and comment on the strengths and weaknesses of each. It should be clear that the different strategies are complementary, but opinions may differ on which mix is best.

Hierarchical measures centralize some set of decisions in the hands of an elite, thus reducing the discretion of other actors in the system. Citing medicine as an example to be imitated, Simon Cole (2010) has called for more hierarchy in forensic science. A distributed system is the opposite of a hierarchical system. It disperses discretion and decision-making authority across a relatively large number of actors in the system. Research science is a distributed system in which any one lab may challenge the experimental results of any other lab.

Blinding measures are meant to reduce or eliminate bias. Measures that leverage bias turn a given bias into an instrument for truth revelation. Rather than attempting to somehow expunge the bias, such measures turn a bad thing, bias, to a good effect. The adversarial system of justice attempts to leverage the contrasting biases of the contesting parties to produce truth and justice.

Thus, we have four strategies for coping with bias in forensic science: (a) hierarchical measures that minimize bias, (b) hierarchical measures that leverage bias, (c) distributed measures that minimize bias, and (d) distributed measures that leverage bias. These strategies are not always in conflict. On the contrary, we advocate some use of each strategy as part of an integrated suite of measures to minimize the problem of bias in forensic science. We do suggest, however, that researchers have tended to overlook important obstacles to creating effective hierarchical measures to minimize bias.

Unfortunately, 'regulatory capture' may be an obstacle to effective hierarchical measures to minimize bias. Any proposal to erect a regulatory body, whether at the federal or state level, comes with the risk that the regulatory body will be 'captured' by special interests. The theory of regulatory capture does not tell us who will win in the contest of interests to capture a regulator. It is a continuous fight; victory may be partial and fleeting. Nevertheless, we can say that concentrated interests aid victory. Well-organized groups with relatively large and homogeneous interests have an advantage in the contest. For this reason, a coalition of law enforcement agencies may be in the best position to capture a federal regulator of forensic science. Forensic science should be a neutral control mechanism to reduce errors in the criminal justice system. Thus, it may be a suboptimal outcome if law enforcement interests capture a forensic-science regulator.

■ Speaker

Nathan Adams, Forensic Bioinformatics, United States

Session: Human Factors / Thursday, July 23 / 3:30-3:50pm

High Resolution Heat Imaging Technology – A New Approach in Fingerprint Search & Securing

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/hi_rez_heat_imaging_technology.kreuter.humanfact.pdf (46 slides)

■ Abstract

Since its first appearance at the end of World War II, heat imaging technology has helped to revolutionize working methods in many parts of industrial technology. It helps to improve energy efficiency, gives important contributions to working safety and health protection of work force and is used to develop the most recent environmentally sound technologies and processes.

Heat imaging cameras are capable of finding heat leaks in buildings or leakage points in petrochemical systems in a matter of seconds. Their ability to visualize smallest temperature differences makes them ideally suited for preventive maintenance, where temperature rise in motor bearings or generator housings may indicate system failure threads at an early stage and help to avoid system downtimes. They can monitor process temperatures in complex production facilities' contactless, from large distances (meters if needed) and at precision levels better than 0.1°F. In modern microelectronics, heat imaging is used to disclose local heat overload in PCBs and even in microchips.

Heat imaging devices have been in use by the police forces to find suspects and missing or disorientated people for the last 15 years.

However, applications of heat imaging devices in forensics are of more recent date. Here, heat imaging has proven its ability to find, uncover and electronically secure fingerprints without use of any kind of chemicals, contactless and within minutes. In this way, it has the potential to give a long-awaited-for contribution to lab work safety, trace DNA conservation and work efficiency. It is a significant contribution towards building a closed digital chain reaching from evidence analysis to identification in AFIS (or other future systems on IT base).

The heat imaging method allows examinations of surfaces difficult to the present approaches, like adhesive sides of packing- or duct tape or of finger prints under soot. In cases of high-value trace carriers, where chemicals cannot be used without doing significant loss in value, heat imaging system are ideally suited to find prints without doing any harm. The technology is not restricted to finger print search alone but can be of use in other cases. One example for this is to find smallest changes in textile textures typical for safety belt impressions.

The presentation gives a short introduction to the physical/technological background of evidence detection by heat imaging and illustrates the potential of this new technology in a variety of examples from daily practice including tests with German police forces.

■ Speaker

Normann Kreuter, German eForensics GmbH, Germany

Session: Human Factors / Thursday, July 23 / 3:50-4:10pm

Understanding and Detection, Preventing, Mitigating Errors in Stair Fall-Related Personal Injury Litigation Inspections and Other Legally Mandated Inspections

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/stair_fall_related_personal_injury_litigation_inspection_and_other_legally_mandated_inspections.pauls_humanfact.pdf (67 slides)

Abstract

Background. Stair-related falls are a leading cause of medically treated injuries, with on the order of three million such treatments each year in the USA and many millions more not receiving professional medical treatment. Relatively few are litigated relative to causation, however many more could be. Predictable, preventable injuries occur predominantly on stairways with design and construction flaws. The most potent category of such flaws 'systemic, random or both' are dimensional non-uniformities that are violations of reasonable care, construction standards and other legally imposed requirements. Societal costs of all stair-related injuries are on the order of 100 billion dollars annually in the USA with the vast majority of these being in homes.

Central Forensics Problems. Many Authorities Having Jurisdiction (AHJs) 'adopting and enforcing construction requirements' and their inspection staffs, make errors of commission or omission in the assessment of stairway dimension uniformity. These shortcomings are shared among many claiming to provide 'forensic' services in premises liability (personal injury) proceedings in, or intended for, court deliberation. Central problems include failures to understand what codes and standards actually require for stair dimensions and their uniformity, plus 'more fundamentally' the scientific bases (often in ergonomics) for such dimensions and their uniformity. As a result, many stairways 'from the time of their first uses' impose risks of injurious falls that are one, two or three orders of magnitude greater than expected for proper stairways.

Proper Methods for Original Construction Approvals and for Detecting Flawed Inspection Methods. US and Canadian building codes and safety standards mostly address correctly (if not in sufficiently clear or potent language) what the critical dimensions and uniformity criteria are. Practices however, too often, fall back on 'quick and dirty' methods for assessing dimensions and their uniformity, ignoring what the codes specify and, more fundamentally, the ergonomic bases for suitable dimensions. Over the last decade or so, in ergonomics, there has been much greater focus on both scientific studies of stair step dimensions, in terms of human performance on stairs, and on measurement methods that are based on science as well as adherence to what codes specify. To summarize this most succinctly, proper stair measurement cannot be made with only a length measurement device (e.g., a measuring tape or folding stick). Proper measurement requires, at a minimum, a spirit or digital level, along with a length measuring device, both used skillfully. Requiring even more skill are relatively sophisticated laser surveying systems that can produce great amounts of geometric data very quickly; these are only being discussed in a preliminary fashion in cutting-edge conferences of some of the world's top stairway safety researchers (e.g., at the International Ergonomics Association Triennial Congress in August 2015). It is hoped that, in terms of widespread practice, there will soon be a major reduction in faulty inspections in the various legal contexts where it occurs presently, both in governmental and private sector inspections under, respectively, administrative and litigation aspects of law.

Speaker

Jake Pauls, Jake Pauls Consulting Services, United States

Session: Human Factors / Thursday, July 23 / 4:10-4:30pm

Visual Comparison of Complex Sets of Quantifiable Forensic Data

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/visual_comparison_of_complex_sets_of_quantifiable_forensic_data.sorenson.humanfact.pdf (94 slides)

■ Abstract

The most important part of Forensic Science is accurately compiling data in order to make an absolutely correct decision. Making a decision involves making a choice between several options. Each option will have its own set of attributes. Attributes are generally quantifiable. Variation in the attributes distinguishes each option from all other options. For easier comprehension, attributes can be graphically represented. Since there are always a large number of objects and related variables, each decision maker or investigator must mentally merge a large number of graphics without correlation data to arrive at a decision. The conclusion made by each independent investigator could vary because of their personal knowledge and individual experience. Here is where the possibility of significant difference of opinion or errors could be introduced into the investigation. For example, if the various entities being investigated and/or compared had 15 quantifiable attributes, it would require 105 X-Y scatter plots to graphically compare each attribute against each other attribute. Having to mentally merge complex data is challenging and likely to introduce variations and error. Today, decision making is where data retrieval was before Google.

A unique technology has been developed at Brigham Young University called HyperSphere Graphics (HSG). HSG permits objects, each with several quantifiable attributes, to be displayed in a single graphical representation on a computer screen.

No conclusions or exclusion of data are required before HSG graphically represents the data. As a result, the entire data set is displayed objectively and the entire data set can be analyzed without segmentation.

Each attribute in the data set is displayed by HSG as an independent axis and all attributes are graphically included in a single HSG representation. The correlation between each attribute and all other attributes is also accurately maintained and graphically displayed. Each object is represented by an individual data point and all data points can be referenced to each and every axis.

The data input is a simple XLS spreadsheet format with a row for each object and a column for each attribute. Thereafter, the HSG software does all the data analysis including correlation analysis and positions all of the attributes as axis and all of the objects as data points in the single computerized graphic.

Once the graphic has been generated it can be explored and objects and/or attributes can be eliminated to simplify the graphic and further minimize any errors in the decision process.

Since HSG is an innovatively new approach to data analysis and decision support, it is important that the actual graphic be viewed in order to fully comprehend the extent to which the accuracy of decision making can be improved and errors eliminated. If selected for presentation at the Inaugural Symposium on Forensic Science Error Management, HSG will be visually displayed using related forensic data files so that the viewers can more fully comprehend its application to forensic science and specifically to elimination of errors.

■ Speaker

David Sorensen, WestCAMP, Inc., United States

Session: Legal Factors / Thursday, July 23 / 11:00-11:20am

Forensic Scientific Error and Omission within the Legal System – Forensic Science in the Courtroom: Can we communicate better?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/forensic_science_inthe_courtroom_can_we_communicate_better.plourd.legfact.pdf (42 slides)

■ Abstract

Few matters threaten the integrity of the legal system more than wrongful convictions. When flawed forensic science plays a role in convicting an innocent person, the damage to the forensic science community can be devastating. After such a case forensic scientists have to grapple with the challenge of preserving their credibility in the face of a growing and increasingly skeptical legal community. Instead of ignoring such errors the reasons for forensic mistakes need to be identified and corrective action should follow. This presentation will demonstrate some of the causes of diminished forensic science credibility, describe several forensic science errors in actual cases, and offer solutions.

The biggest blow to the credibility to forensic science is the increasingly contentious debate over wrongful convictions that were caused or contributed to by a forensic scientist. Public Attention on forensic science failures, although rare, is of such a tragic nature that they are able to gain high publicity and have the potential to completely erode the legal community's trust of forensic science.

The credibility gap between forensic scientists and the legal community is further exacerbated by the attitudes and understanding about the role and use of forensic science in the courtroom. Such attitudes and understandings can lead to a breakdown in communication between the forensic scientist and the legal consumers of the information, thus contributing to errors and misunderstandings. Forensic scientist should be willing to meet with members of the legal community and vice versa to open dialogues as to the nature and scope of each other's work. In other words there needs to be a conversation not a lecture by one to the other.

A core guiding principal for use of forensic science in the courtroom is that all players who have an interest in the outcome of forensic science results should work collaboratively to come to the right conclusion. This means that members of the legal community must learn to ask the right questions of the forensic scientist and the forensic scientist must speak clearly in response and/or provide communication tools for translating the nature scope and implications of their work.

Open communication is essential for identifying and rectifying the root causes of forensic science failures. For example, there is an emerging awareness of human factors, such as unconscious bias, in forensic science that causes errors to occur. Human factors can and do influence decision-making and can mask an otherwise reliable forensic result. Although applying human factors principles to minimize errors is difficult, it is nevertheless achievable and must be carefully considered in developing communication protocols between forensic scientists and the legal community.

Further, past errors should be studied irrespective of the hazard or difficulty such an investigation possess to the legal or forensic community. Developing a stockpile of strategies to evaluate forensic errors is an opportunity to improve forensic science practices. Existing efforts need to be reinforced and placed within a larger more cohesive operational framework.

■ Speaker

Christopher Plourd, Imperial County California Superior Court Judge, United States

Session: Legal Factors / Thursday, July 23 / 11:20-11:40am

Statistical Issues and Reliability of Eyewitness Identification as a Forensic Tool

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/statistical_issues_and_reliability_of_eyewitness_id_as_a_forensic_tool.kafadar.legalfact.pdf (28 slides)

■ Abstract

Among the 318 wrongful convictions identified by the Innocence Project that were later overturned by DNA evidence resurrected from the crime scene, 229 (72%) involved eyewitness testimony. Such courtroom identifications from an eyewitness can be tremendously powerful evidence in a trial. Yet memory is not a perfect video recording of events, and one's recollection of the events surrounding an incident is even less reliable. In October 2014, the National Academy of Sciences issued a landmark report evaluating the scientific research on memory and eyewitness identification. The Committee, comprised of researchers (psychologists, statisticians, sociologists) and representatives of the judicial system (judges, attorneys) reviewed published research on the factors that influence the accuracy and consistency of eyewitnesses' identifications, conducted via laboratory and field studies. I will describe the research on memory and recollection, focusing on the shortcomings in the statistical methods used in evaluating existing laboratory studies and the other scientific issues that led to the Committee's recommendations for standardizing procedures and informing judicial personnel of the factors that can have negative impacts on accuracy of eyewitness testimony. (The author was a member of the NAS Committee that issued the report.)

■ Speaker

Karen Kafadar, University of Virginia, United States

Session: Legal Factors / Thursday, July 23 / 11:40am-12:00pm

Language Barriers: Creating Standard Terminology to Mitigate Forensic Errors in the Courtroom

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/language_barriers_creating_standard_terminology_to_mitigate_forensic_errors_inthe_courtroom.gable_legfact.pdf (21 slides)

■ Abstract

Lawyers and forensic scientists should work together to create standardized terminology and quantification in forensic reports. This would allow for juries to hear consistent, reliable, and clear testimony on forensic evidence. Over more than one hundred years of jurisprudence, American courts have improperly sanctified various forensic disciplines without scrutinizing them under applicable standards related to the admissibility of novel scientific or technical evidence.

The 2009 NAS Report confirmed that the ultimate problem facing the use of forensics in an adversarial system is a lack of understanding as to:

(1) the extent to which a particular forensic discipline is founded on reasonable scientific methodology that gives it the capacity to accurately analyze evidence and report findings and

(2) the extent to which practitioners in a particular forensic discipline rely on human interpretation that could be tainted by error, threat of bias, or the absence of sound operational procedures and robust standards.

Often times, the legal system loses sight of these two necessities. We know that judges, lawyers, and juries have placed undue weight on forensic evidence that is subject to interpretation, rather than indicative of an objective answer. The presentation suggests that lawyers and forensic scientists need to work together to help develop standards that can be applied both within the science and inside the courtroom.

Very few national standards for terminology and reporting exist in the forensic sciences. Terminology plays a significant role in many court settings. A jury can hear that two samples are a 'match,' are 'consistent with,' are 'identical,' 'likely match,' or are of many other kinds of relation to each other, and take all these relations to mean the same thing even when they do not. On the other hand, a 'negative' fingerprint analysis could mean 'excluded,' 'inconclusive,' 'unable to locate,' or 'a poor sample' all of which are likely indistinguishable to the average lay person juror (and even the lawyers).

In addition, reporting standards also differ between laboratories. While some reports include detailed accounts of tests and protocols performed, others contain barely more than the scientist's brief conclusory statements regarding the test results (which, as stated, could have many different meanings). The decision over to how to use the report falls to the lawyer, who in many cases is uneducated on the underlying analysis.

Transparency in the comparison of evidence can be achieved by clearly defining what is observed and then interpreted. Examiners must document and be able to explain the analysis of what was originally observed and acknowledge the potential impact of biases. Without documenting these changes, the analysis misplaces what was originally observed and interpreted versus what an analyst now believes after comparing it with the record.

While adopting new and/or uniform standards would not necessarily solve all terminology and reporting issues, it would at the very least provide some reference points for uniform vocabulary and reporting protocols. It might also detect and prevent errors before they happen rather than leave their discovery to the fickleness of the post-conviction process.

■ Speaker

Jessica Gabel Cino, Georgia State University College of Law, United States

Session: Legal Factors / Thursday, July 23 / 1:30-1:50pm

The Use of Blinded Eyewitness Identification as a Forensic Procedure

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/garrett-abstract.pdf>

■ Abstract

Flawed scientific evidence has contributed to the most serious wrongful convictions that have come to light in the United States. Not only did over two-thirds of persons exonerated by post-conviction DNA involve eyewitnesses who misidentified innocence suspects, but over half of the over-300 cases involved unreliable and unscientific forensic analysis and testimony. First, I will discuss the eyewitnesses, who misidentified defendants using then-standard lineup procedures that were highly vulnerable to suggestion and contamination. Indeed, most of those eyewitness were unsure when they first saw the defendant in a photo array or other type of lineup. They became certain by the time of trial, although we now know they were wrong. The 2014 Report by a Committee of the National Academy of Sciences (of which the author was member) discussed a series of improvements to eyewitness identification procedures that can safeguard against errors. A central recommendation was that lineups be conducted blind, so that the administrator does not know which is the suspect, or is effectively blinded and cannot convey feedback to the eyewitness. Second, I will discuss how that same recommendation can be extended more broadly to prevent cognitive bias and error in a range of forensic disciplines. Scholars have recommended, and some crime labs have implemented, procedures to selectively blind the work that forensic scientists do, to help minimize everyday phenomenon of cognitive bias. The images that adorn our courthouses depict justice as blind for a reason; improvements to quality of scientific evidence that our criminal justice system routinely relies upon should include procedures to make the collection of that evidence blind.

■ Speaker

Brandon Garrett, University of Virginia School of Law, United States

Session: Legal Factors / Thursday, July 23 / 1:50pm-2:10pm

Uncertainty in Forensic Science: A Factor Deliberately Down-Played by Forensic Examiners

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/speigelman-abstract.pdf>

■ Abstract

Recent admissions by the FBI that it overstated testimony in hair analyses in judicial proceedings follows a related admission by the FBI in 2004 about comparative bullet lead analyses. In both cases, scientific facts and statistical analyses compelled the agency's changes in practice. Evidence of forensic testimony given beyond the limits of science in other areas of forensic practice have been more difficult to reveal as many forensic disciplines, such as firearms identification, are generally not amenable to DNA analyses. Testimony in impression and pattern evidence disciplines are subject to similar statistical issues that have been revealed by the FBI's review of microscopic hair comparison testimony. This presentation will discuss the current practices in forensic testimony and a satellite view of what science can support. This presentation will also suggest alternative testimony supported by the current state of research and statistical principles and offer firearm/toolmarks examiners tools they can implement today in testimony that are safely within the boundaries of the current state of the science as research to support the underlying validity and reliability of the discipline are underway.

■ Speaker

Clifford Spiegelman, Texas A&M University, United States

Session: Legal Factors / Thursday, July 23 / 2:10pm-2:30pm

You're Doing a Great Job, But You're Doing It All Wrong

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/blackledge-abstract.pdf>

■ Abstract

On November 22, 2005 Congress authorized the National Academy of Sciences to conduct a study on forensic science. A Forensic Science Committee was created and given a list of eight areas to look into. The efforts of the Forensic Science Committee culminated in a report published in 2009, 'Strengthening Forensic Science in the United States: A Path Forward'.

So that's what we have; not much point in going back and saying, 'would have, could have, should have.' This presentation will consider how well the forensic science community, and especially two large US Government organizations, the National Institute of Justice (NIJ) and the National Institute of Standards and Technology (NIST), have done in advancing the recommendations contained in that report.

NIST has done well as far as the efforts of their own scientists. However, NIST (this Symposium is an example) and especially the NIJ have failed woefully as far as inducing greater proactive involvement by the extended forensic science community and concerned citizens in general. Just as so often in the past, committees have been formed, blue ribbon panels named, and meetings and symposia hosted. Despite advances in technology and in our knowledge of human behavior, the approach so far is little different than in the past millennium.

It is now well known (as shown in the recent book, *The Wisdom of Crowds* by James Surowiecki) that the wisdom of crowds is not only superior to that of any individual no matter how well-qualified, but the wisdom of a large, diverse group is better than that of a panel of experts. In the past there were many technical and financial difficulties in getting a large crowd of people to look into a problem, but not today. A title for Surowiecki's book might well have been, *The Wisdom of the Cloud*! Today on the Internet there are an endless variety of interactive webinars. Research scientists are posing problems to the Cloud and anyone (whether a Nobel laureate or a child in the slums of Calcutta) with access to a smart phone may collect and transmit data or suggest solutions. Why are the NIJ, NIST, ASCLD, and AAFS not taking advantage of the wisdom of crowds and the cloud? Could it be that the Old Boy/Old Girl Forensic Establishment ('the usual suspects') has a vested interest in maintaining the status quo, or is their thinking just "boxed" in?

■ Speaker

Robert Blackledge (he did not attend and his presentation was given by someone else), United States

Session: Legal Factors / Thursday, July 23 / 2:30pm-2:50pm

Recognizing Sentinel Events in Forensic Science to Avoid Errors

Presentation Link:

<https://www.nist.gov/sites/default/files/documents/director/young-abstract.pdf>

■ Abstract

"Sentinel events" are a concept from the healthcare industry, whereby "any unanticipated event resulting in death or serious physical or psychological injury to a patient or patients, not related to the natural course of the patient's illness."

News headlines offer crime labs free opportunities to examine their own processes and to determine if they are vulnerable to the same flaws. This presentation will examine a number of headlines from the U.S. and the U.K., as a means of categorizing the underlying causes and the possible remedies thereof. These causes include misplaced trust, analyst incompetence, legal misconduct, and culture clashes between law and science, to name but only a few.

Another free opportunity comes from defense experts, whose comments and criticisms are not always self-serving and profit-driven; sometimes, they're right, which can point to deficiencies in the analyst or an underlying laboratory policy. In either case, however, the laboratory should go and retroactively review any and all affected cases to ferret out additional instances of the same error.

In the last three decades, forensic serologists have all but become extinct, rendered obsolete by forensic DNA analysts. Forensic DNA analysis has done more to free wrongfully-convicted people than any other technology, but it would be foolish to believe that it is fool-proof. But could likelihood ratios incorporating drop-out rates turn it into the new fertile garden of wrongful convictions, when a new and better technology emerges? Only by examining the minefield of mistakes of the past and the present can forensic science hope to move into the future.

■ Speaker

Arthur Young, Guardian Forensic Sciences, United States

Session: Legal Factors / Thursday, July 23 / 3:30pm-3:50pm

To Err is Human, but How Might We Measure Error Rates in Forensic DNA Testing and What Would These Error Rates Really Mean?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/error_management_dna_error.butler.legalfact.pdf (22 slides)

■ Abstract

The 1993 U.S. Supreme Court ruling in *Daubert vs. Merrell-Dow Pharmaceuticals* insists on knowledge regarding error rates and their application when evidence is introduced into court [1]. The 1996 National Research Council (NRC II) report *The Evaluation of Forensic DNA Evidence* discusses the topic of whether or not an error rate should be included in statistical calculations that accompany a DNA match [2]. This presentation will explore the meaning and potential calculations of error rates for DNA testing and will review the literature on this topic. Reported errors made during DNA analysis interlaboratory studies and proficiency tests will be considered and how useful this information may or may not be. A distinction will be made between error in a method and uncertainty in a measurement versus error in interpretation. Implications of open and subjective interpretation protocols compared to rigid method protocols and procedures will be discussed. Finally, the challenge of obtaining accurate and complete information to inform error rates while working in an adversarial environment will be considered.

[1] Butler, J.M. (2012). *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego: Elsevier Academic Press, pp. 534-535.

[2] National Research Council (NRCII) Committee on DNA Forensic Science. (1996). *The Evaluation of Forensic DNA Evidence*. Washington, DC: National Academy Press, pp. 85-87.

■ Speaker



John M. Butler has a B.S. in chemistry from Brigham Young University and a Ph.D. in analytical chemistry from the University of Virginia. His Ph.D. research, which was conducted in the FBI Laboratory's Forensic Science Research Unit with Bruce McCord, involved pioneering the techniques now used worldwide in modern forensic DNA testing. Dr. Butler is a NIST Fellow and Special Assistant to the Director for Forensic Science at the National Institute of Standards and Technology. He is author of the internationally acclaimed textbook *Forensic DNA Typing*—now in its third edition—as well as more than 150 scientific articles and invited book chapters. In August 2011, ScienceWatch.com announced that Dr. Butler was number one in the world as a high-impact author (number of citations per paper published) in legal medicine and forensic science for the decade of 2001-2011. He and his wife have six children, all of whom have been proven to be theirs through DNA testing.

Session: Legal Factors / Thursday, July 23 / 3:50pm-4:10pm

The Potential of Blind Collaborative Justice: Testing the Impact of Expert Blinding and Consensus Building on the Validity of Forensic Testimony

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/Blind_Collaborative_Justice-wong-legal.pdf (19 slides)

■ Abstract

Biased expert testimony is a leading cause of wrongful convictions, and new techniques are needed to reduce such biases. This study conducted an experimental investigation of two potential contributors to biased testimony within adversarial litigation involving forensic evidence: (1) experts' knowledge of their party representation (i.e., prosecution vs. defense counsel), and (2) lack of input from the relevant scientific community. A sample of 580 scientists was asked to read a vignette about a hypothetical criminal case and solve a statistical reasoning problem bearing on the case evidence. Participants were randomly assigned to one of three types of party representation (prosecution, defense, or blinded). Approximately half the participants in each representation were given the correct solution in advance of their responses. The correct solution was derived by an independent panel of 12 experts, and presented as 'consensus feedback.' The other half of participants in each representation received the consensus feedback after providing an initial response, and received an opportunity to change their initial response following that feedback. We found no evidence of an effect of blinding on accuracy. The results revealed a consistent, positive effect of expert consensus feedback on response accuracy. We conclude that expert consensus feedback could improve the validity of expert testimony, and discuss the importance of educating scientists about ways to reduce testimonial bias.

■ Speaker

Carolyn Wong, RAND Corporation, United States

Session: Legal Factors / Thursday, July 23 / 4:10pm-4:30pm

Garbage In - Gospel Out? Moving Forensic Fire Investigation into the 21st Century

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/garbage_in_gospel_out_moving_forensic_fire_investigation_into_the_21st_century.brannigan.legalfact.pdf (79 slides)

■ Abstract

Forensic fire investigation has been one of the last holdouts against the rejection of 'junk science' in the courts. Despite notable improvement in evidence gathering since the introduction of NFPA 921, the inferences and conclusions drawn in Forensic Fire Investigations are often little more than guesswork masquerading as science. The inferential process described in NFPA 921 is not the scientific method used by 'real' scientists. The goal of NFPA 921 is to tell a story of the fire 'consistent with' the physical evidence. But the process of logical inference of the best fitting answer from a set of facts is not deduction as defined in NFPA 921 but Abduction. Abduction, or inference to the best explanation, is a form of inference that best explains or accounts for the data. The important distinction is that abduction does not test a scientific hypothesis. Instead Abduction is routinely used to create a testable hypotheses. Abduction clearly describes the process of forensic fire investigation. But the 'cognitive test' of the hypothesis to show that it is the 'best fit' is not scientific proof.

Improvements in evidence gathering and fire science have allowed the rejections of many fire scenarios as being scientifically inconsistent with the physical evidence. But the fact that a particular scenario may be consistent with the physical evidence does nothing to rule out other scenarios which may be equally or more plausible. Pointing out that a scenario is plausible is simply not enough if there is no scientific reason at all to choose one or the other.

We believe that a key problem is the routine expectations of an overall forensic fire science conclusion without clearly differentiating those aspects of the problem which fall in the forensic science discipline from those issues of proof that fall outside it. E.g. fire experts have no special expertise in determining the credibility of witnesses. We propose to deconstruct the forensic problem into those modules which are supportable by scientific analysis that meets the Daubert/Kumho test and those which do not. Proof of these other modules must be supplied outside the forensic science process. The end result is a matrix in which the admissible reliable fire science evidence is properly connected to evidence which must be supported by other forms of evidence. Examples will be given from both cases and scientific analysis.

■ Speaker

Vincent Brannigan, University of Maryland-College Park, United States

Session: Quality Assurance / Thursday, July 23 / 1:30pm-1:50pm

Forensic DNA: The Gold Standard or Fools Gold?

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/forensic_dna_the_gold_standard_or_fools_gold.reich_qa_.pdf (15 slides)

■ Abstract

The lack of scientific foundation and rigor for many forensic fields has been evident to trained scientists for some time. There is now widespread acknowledgment of this fact; in contrast forensic DNA is held up as the 'gold standard' to which other forensic fields should aspire. While forensic DNA does have a real scientific foundation and indeed has some standards for controls and measurements, forensic DNA as currently practiced is riddled with errors, bias, inefficient and meaningless procedures, widely held dogmas that have no scientific support and other deficiencies. If these issues are not confronted and addressed, forensic DNA, despite its success as a tool for individual identification, will lose its imprimatur as an ideal model issues as the number of cases and samples and tests explodes across every country in the world. Here I will critically evaluate the conceptual and practical problems that bedevil the current practice of forensic DNA and provide possible solutions for these many issues.

First we must place forensic DNA in its proper context in the larger world of modern scientific research and acknowledge the current state of this field: compared to a modern, leading medical or biological research laboratory, forensic DNA is the high school athlete on the professional football team. Scientifically, forensic DNA is a forgotten backwater of molecular biology and as such receives no funding, support or attention from any of the agencies that fund scientific research, including the NIH, HHMI, or NSF.

■ Speaker

Karl Reich, Independent Forensics, United States

Session: Quality Assurance / Thursday, July 23 / 1:50pm-2:10pm

The Value of Regular Benchmarking Studies in Forensic Science to Understand Where Errors Can Occur: Lessons Learned from NIST DNA Interlaboratory Studies

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/nist_dna_interlab_studies.butler.qa_.pdf (24 slides)

■ Abstract

Over the past two decades, scientists from the National Institute of Standards and Technology (NIST) have conducted 13 exploratory interlaboratory studies addressing forensic DNA typing issues [1]. Five of these studies have dealt with DNA mixtures [2]. Interlaboratory studies involve multiple labs examining the same samples or data to look for consistencies or differences in results, trends in laboratory practice, and potential opportunities for further training and reference material development. These interlaboratory studies provide an interesting window into laboratory practices, protocols, and technology in use at the time of a study. In addition, results from these studies can expand the view of the community beyond each laboratory's focus on their own protocol and analyst performance. This presentation will discuss lessons learned and challenges faced in conducting the NIST DNA interlaboratory studies. Information from these studies and lessons learned can inform potential benchmarking studies conducted in the future for forensic DNA analysis and other forensic disciplines.

[1] See <http://www.cstl.nist.gov/strbase/interlab.htm>

[2] Butler, J.M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. San Diego: Elsevier Academic Press, pp. 150-151.

■ Speaker



John M. Butler has a B.S. in chemistry from Brigham Young University and a Ph.D. in analytical chemistry from the University of Virginia. His Ph.D. research, which was conducted in the FBI Laboratory's Forensic Science Research Unit with Bruce McCord, involved pioneering the techniques now used worldwide in modern forensic DNA testing. Dr. Butler is a NIST Fellow and Special Assistant to the Director for Forensic Science at the National Institute of Standards and Technology. He is author of the internationally acclaimed textbook *Forensic DNA Typing*—now in its third edition—as well as more than 150 scientific articles and invited book chapters. In August 2011, ScienceWatch.com announced that Dr. Butler was number one in the world as a high-impact author (number of citations per paper published) in legal medicine and forensic science for the decade of 2001-2011. He and his wife have six children, all of whom have been proven to be theirs through DNA testing.

Session: Quality Assurance / Thursday, July 23 / 2:10pm-2:30pm

Automated DNA Examiner Assessment Tool (DEAT): A Software Program to be Utilized in DNA Benchmarking, Training Proficiency, and Competency Testing

Presentation Link:

https://www.nist.gov/sites/default/files/documents/director/DNA_mixture_interpretation_study_DEAT-aranda-ga.pdf (29 slides)

■ Abstract

DNA forensic laboratories generally deconvolute DNA samples using data generated from single tandem repeats (STRs). The difficulty of interpreting STR data is amplified with increased sample complexity. This includes, but is not limited to: increasing the number of contributors in a sample, varying proportion ratios of each contributor, and the varying degradation state of each contributor. In complex mixture samples, the resultant genotypes generated vary by DNA examiner; this variation is the subject of the Defense Forensic Science Center's (DFSC) ongoing large-scale mixture interpretation study. To aid the study, a novel DNA Examiner Assessment Tool (DEAT) has been developed that assesses an examiner's generated genotypes for accuracy and quality. This automated tool generates two complementary metrics, a genotype interpretation metric (GIM) score and Allelic Truth/False (AT/AF) score. The GIM score measures the number of genotypes generated at the locus, contributor, and mixture level and the AT/AF score calculates the number of true and mismatched alleles generated at the locus, contributor, and mixture level when compared to prepared DNA mixture samples. We describe the calculation of these metrics, and demonstrate their use to: 1) highlight interpretation weaknesses at the examiner and laboratory level, 2) provide benchmarking scores to new examiners during training and experienced DNA examiners during retesting, 3) measure the effectiveness of any changes in standard operating procedures, 4) uncover the error rate and limitation of a laboratory's genotype interpretation accuracy depending on mixture complexity. In providing a tool that identifies and potentially reduces individual and systematic errors, DEAT is able to analyze data regardless of amplification kit used by a laboratory. In addition, DEAT's analysis can be applied not only to STR-generated data, but also to next generation sequencing (NGS) data. This tool will be made available to the DNA forensic community. The tool provides a standardized method of recording number and proportion of contributors, final match statistics generated, and genotypes/allele calls. These additional metrics are tabulated by DEAT for comparison to the known mixture profile and can also be used for examiner assessment.

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense. Names of commercial manufacturers or products included are incidental only, and inclusion does not imply endorsement by the authors, DFSC, OPMG, DA or DoD.

■ Speaker

Emily Rogers, Defense Forensic Science Center, United States

Session: Quality Assurance / Thursday, July 23 / 2:30pm-2:50pm

Quality Assurance and Quality Control in Breath Alcohol Measurement Programs Using Intoxilyzers Such as the Intoxilyzer 8000C

Presentation Link:

https://www.nist.gov/sites/default/files/documents/2016/11/22/qa_and_qc_in_breath_alcohol_measurement_programs_using_intoxilyzers_such_as_the_intoxilyzer_8000c.krishnan.qa_.pdf (10 slides and 4-page white paper)

■ Abstract

For the last few years Intoxilyzers, particularly Intoxilyzer 5000 and 8000 in the United States and corresponding Models 5000C and 8000C in Canada, are used in the determination of blood alcohol concentration (BAC) of persons accused of 'drinking and driving' or the DUI offenses. Although these instruments are reasonably accurate for the purposes for which they are used, one has to be careful and vigilant that errors do not occur in the field. This is because, these instruments are predominantly operated by police officers and other law enforcement officials who are not necessarily scientists and they operate the instruments under a protocol devised by the program managers.

While in Canada and the USA much thought have gone in to ensure that the BAC measured with an accused person is under good quality control and quality assurance protocols, we can take lessons from many other fields in which micro analytical methods are used such as in hospital laboratories, Government laboratories such as NIST and the industry. The following methods are suggested so that the quality of the BAC obtained is of the highest quality when it is presented in courts during the trial.

1. A composite sample of blood pool be developed using either actual or artificial blood containing common interferents that are possible such as acetone, isopropyl alcohol and other chemical which can be certified by laboratories such as NIST which can then be distributed to all units along with protocols to produce and analyze the headspace alcohol. This would be a good method to ensure that each unit can produce accurate BAC results.
2. A proficiency testing procedure is developed which is controlled by the Central quality control administrative unit which will sent 'blind' samples to all units periodically to check how each unit is performed. The not so well performing units can then be specially treated to identify the problems and rectify them.
3. A quality control unit is centrally and nationally established which gathers QC information and data about the Intoxilyzers and Simulators which are openly available to all interested parties and stake holders. This ensures that there is confidence in the system and reduce unnecessary litigation and court time.
4. A national unit is established to conduct research into why certain unexplained anomalies occur from time to time and provide, explanations, remedies and alterations in instrument components or protocols.

All the above will be discussed in details in this paper.

■ Speaker

Sivarama Krishnan, University of Toronto, Canada

POSTER PRESENTATION ABSTRACTS

Poster Presentations

1. **Regulating and Streamlining the Request and Validation Process of Programming Changes**
Michal Pierce, Harris County Institute of Forensic Sciences, United States; Cynthia Young [Laura Cook presented]
2. **Forensic Linguistics: Voice Biometric and Speaker Recognition using Gaussian Mixture Model and MFCC**
Nilu Singh, DIT, BBA University, Lucknow, India
3. **The Development of Certified Reference Materials of Drug Abuse (heroin, etc) for the Elimination of Measurement Error in Forensic Science**
SU Fuhai, National Institute of Metrology, P.R.China, China; LI Hongmei
4. **It is More Than Just Errors: A Quantitative and Qualitative Approach in Cognitive Contaminations Within ACE-V Methodology**
Francisco Valente Goncalves, INTREPID Forensics | University of Leicester, United Kingdom; Lisa Smith
5. **Learning-based Ballistic Breach Face Impression Image Matching**
Xiaoming Liu, Michigan State University, United States; Anil Jain
6. **A Model for Estimating Error Rates in Forensic Document Examination**
Steven A. Slyter, S. A. Slyter, LLC, United States
7. **Information Sampling and Weighting in Fingerprint Recognition**
Silke Jensen, University of Leicester, United Kingdom
8. **Real Science Admits its Uncertainty: DNA Evidence is Often Mischaracterized as Infallible**
Jason O'Rawe, Cold Spring Harbor Laboratory, United States; Scott Ferson
9. **Improved Law Enforcement Officer Reports with POV Cameras - Gathering Better Scene Information**
Donald Dawes, Lompoc Valley Medical Center, United States; Jeffrey Ho
10. **Conveying Technical Judgments Involving Uncertainty**
Leigh Cash, Los Alamos National Laboratory, United States; James Gattiker
11. **Potential Errors and Uncertainty in Microscopy Image Analysis**
Leigh Cash, Los Alamos National Laboratory, United States; Alison Tamasi
12. **The Qualification of Numerical Values Through Natural Language Expressions of Uncertainty in Expert Testimony**
Kari Sentz, Los Alamos National Laboratory, United States; Scott Ferson
13. **Minimizing Contextual Bias in Forensic Firearms Examinations**
Erwin Mattijssen, Netherlands Forensic Institute, Netherlands; Ralph Kleuskens
14. **Data Integrity Red Flags: Recognizing and Responding to Warning Signs; Lessons Learned from the Laboratory Industry**
Janine Arvizu, Independent Auditor, United States; Amy Sirignano
15. **Evaluation of Novel Technological Solutions for Detecting and Interpreting Presumptive Drug Tests**
Subrata Acharya, Towson University, United States; Kelly Elkins
16. **Considerations of Proficiency Testing Providers and Error Management**
Shirley Turner, NIST, United States; Stacy Doorn
17. **Validation of The Leica Geosystems ScanStation C10 for Crime Scene Investigations**
Thomas Wisner, Baltimore Police Department, United States; Amelia Faelyn
18. **Validation of TrueAllele Casework System with the PowerPlex Fusion Chemistry**
Thomas Hebert, Baltimore City Police Department, United States

Session: Poster 1

Regulating and Streamlining the Request and Validation Process of Programming Changes

Ms. Michal Pierce, Harris County Institute of Forensic Sciences, United States; Ms. Cynthia Young

■ Abstract

As documentation for laboratory functions becomes more automated and paperless, the need for computer programs and templates to be properly validated before use is unquestionable. Problems can arise if IT personnel do not successfully capture the scope of forensic validation. Even when computer-savvy lab analysts attempt to validate these programs, there can be a gap in ensuring the software is tested sufficiently to avert a series of errors. One error in a laboratory software program can cause multiple errors in released reports.

HCIFS Quality Management Division took on the challenge last year to standardize the way programming changes are requested, acknowledged, verified, and implemented. In order to prevent quality issues from arising, all steps in the process, from initial request to going live with the revision, are reviewed in advance, during validation, and prior to sign-off. Previously, programming requests were emailed to the IT department, where a member of IT would work on the project as time allowed. There was no pre-determined validation plan and no formal feedback on progress. Once a request was completed, section managers were often not aware of the programming scope or how it was tested. Under the new policy, the manner of communicating and reviewing programming is changed. HCIFS Quality Management (QM) is directly involved in the process.

The new procedure was created to ensure a structured process for requesting and reviewing program changes. It begins with a request and tracking form that requires a detailed description of the desired programming change. A lab manager approves the request and submits it electronically to the IT manager, who assigns it to a programmer and attaches a unique tracking number. The programmer outlines the steps required to fulfill the request, as well as the time needed for completion. After the programmer finishes the work, another IT member verifies that the new coding is correct and that it works properly. The form is returned to the requestor, who proceeds with validating the program in a systematic manner. Validation includes inserting a series of test data and comparing the revised program to the current program to identify differences in output. The form and all validation documentation are forwarded to a member of QM. QM checks that the process was completed according to the SOP and that the analyst's validation is sufficient. Signatures document each step; however, only after a member of QM signs off can the revised program be activated.

There are several benefits to this process. There are no longer informal requests sent to IT that can be misplaced, delayed, or misunderstood. Communication between the laboratory and IT is greatly increased because the procedure prompts each participant to answer key questions. Of great importance, QM is confident that new programs and changes to existing programs are now validated in accordance with accreditation requirements. This reduces the chance that errors will result from faulty programming.

Session: Poster 2

Forensic Linguistics: Voice Biometric and Speaker Recognition using Gaussian Mixture Model and MFCC

Mrs. Nilu Singh, DIT, BBA University, Lucknow, India

■ Abstract

To recognize a person's individuality through their voice is known as Automatic Speaker Recognition(ASR). Research in Automatic Speaker Recognition has completed approximately six decades but still research is in progress in this area due to challenges remains. In this paper make a contribution about the development of techniques and progresses have made in last six decades in automatic speaker recognition. Speaker recognition come in to the category of biometric security systems, Biometric related to human characteristics or individuality. Biometric verification or realistic authentication is used to recognize the human. Speaker biometric is the features of behavioral measurements or physiological of individual human. Behavioral Biometric can be done by Voice, Signature, Keystrokes, typing etc. and Physiological biometric for example Iris, face, retina, fingerprint, ear, DNA etc. Now days Voice biometric is broad research area. Speaker Recognition/voice biometric is a technology by using this human can be recognized through their voice. It is used in many security areas such as remote authentication or access control, voice message, surveillance etc. ASR has three common steps these are data acquisition, feature extraction and modeling techniques. To extract features from speech signal many feature extraction techniques are available such as MFCC, LPC, LPCC etc. and for modeling HMM, GMM, UBM etc. techniques are available.

Session: Poster 3

The Development of Certified Reference Materials of Drug Abuse (heroin, etc) for the Elimination of Measurement Error in Forensic Science

Dr. SU Fuhai, National Institute of Metrology, P.R.China, China; Mrs. LI Hongmei

■ Abstract

The production, characterization and certification of reference materials (RMs) is a key activity in improving and maintaining a worldwide coherent system of measurements. As detailed in ISO Guide 32 and ISO Guide 33, certified reference materials (CRMs) are used for calibration, quality control and method validation purposes. As part of quality control, the need for CRMs of illicit drug was emphasized by drug detection in the forensic science laboratories as a tool to improve comparability, ensuring accuracy and traceability of analytical results.

To develop the CRMs of illicit drugs, National Institute of Metrology (NIM) cooperated with Institute of Forensic Science Ministry of Public Security (IFS) since 2009. At present, 14 purity CRMs and 18 solution CRMs including heroin, methamphetamine, morphine, ketamine, cocaine, have developed. The certified values of these CRMs were assigned by two different methods. Q-NMR method was used for the value certification of drug abuse for the first time. Homogeneity of the CRMs was determined by an in-house validated liquid chromatographic methodology. Potential degradation during storage was also investigated and a shelf-life based on this value was established. The certified values for all the studied reference materials are traceable to the international system of units (SI). The uncertainties of these CRMs were evaluated based on certified methods, homogeneity testing, and stability study.

Nine calibration and measurement capabilities (CMC) of these drug abuse have been declared in KCDB in 2014. Ketamine and methamphetamine have been applied as CRM in the APMP comparison of APMP.QM-P20 Illicit Drugs in Human Hair in 2013, and gave the good results (Lab (7)) . These CRMs studied have been widely used as standards in routine inspection work for the labs affiliated Institute of Forensic Science.

The application of purity CRMs with accurate value could eliminated the measurement error in the forensic labs, which could give the accurate and reliable results for the forensic evidence. NIM will continue to carry out in-depth cooperation with IFS with raising the level of science to provide a greater level of rigor and confidence in forensic drugs evidence used in the Chinese criminal justice system.

Session: Poster 4

It is More Than Just Errors: A Quantitative and Qualitative Approach in Cognitive Contaminations Within ACE-V Methodology

*Mr. Francisco Valente Goncalves, INTREPID Forensics / University of Leicester,
United Kingdom; Dr. Lisa Smith*

■ Abstract

Currently forensic science has a considerable amount of research regarding cognitive contamination processes. Studies have been conducted on topics such as confirmation bias within different fields of expertise (DNA, face recognition or fingerprint comparison) leading researchers to some cases of miscarriages of justice. The present study focuses on the fingerprint examiners' performance.

ACE-V methodology suffered recent changes switching examiners' focus, where instead of looking to a number of minutiae, started to follow an independent verification phase after the three first phase [Analysis, Comparison and Evaluation]. Due to this, the principal aim in this study is to observe which variables can affect the examiners' performance during ACE-V verification phase.

In order to investigate this, the study divides the influences of cognitive contaminations into two categories, external and internal variables. External variables are associated with the features that examiners need to deal with from outside the examination environment, such as (1) knowing the type of crime from where the latent print originated, (2) having informations about the suspect's condition or (3) having pressures from investigators or prosecutors. Internal variables include contaminations that may occur inside the laboratory, namely (1) having case's information from colleagues, (2) dealing with stressful conditions time pressure or (3) the lack of guidelines to follow in phases as the verification stage.

The authors of the present study will employ a quantitative methodology in order to achieve the previous objectives. To do so computer-based experiments will be utilized, as well as an eye-tracker to obtain accurate and reliable data. However, the study also relies on what fingerprint examiners feel and think about their work regarding concepts such as motivation and confidence. To assess these concepts, individual interviews will be organized with fingerprint examiners in order to these professionals have the opportunity to give their own thoughts about the contamination phenomena.

The sample for the present study is composed of fingerprint examiners from different countries in the interest of assessing different perspectives and a wide range of data therefore allowing a better and more accurate generalization to the forensic science community.

Session: Poster 5

Learning-based Ballistic Breach Face Impression Image Matching

Dr. Xiaoming Liu, Michigan State University, United States; Dr. Anil Jain

■ Abstract

This work presents an algorithm or scientific technique to quantitatively evaluate whether a fired cartridge case has been fired from a specific firearm. Ballistic images of a cartridge case or bullet carry distinct "fingerprints" of the firearm, which is the foundation of widely used forensic examination in criminal investigations. In recent years, prior work has explored the effectiveness of correlation-based approaches in matching ballistic imagery. However, most of these studies focused on highly controlled situations and used relatively simple image processing techniques, without leveraging any supervised learning scheme.

Computer vision and machine learning have made great strides in both their individual capabilities and their confluence. It is generally agreed that a supervised learning-based approach is superior to a non-learning-based approach. Hence, to improve the matching accuracy, especially on operational data, we propose a learning-based approach to estimate the similarity (comparison score) between two ballistic images. We focus on the breech face impression (BFI) among the various ballistic toolmarks given its popularity in practice.

Given two 2D microscope images of cartridge cases, we first estimate the center of the BFI circular region of each image, followed by a global rotation estimation between the two images. In order to build a global feature correspondence among all available images a prerequisite for any learning-based method, we transform an image from the Cartesian coordinate space to the polar coordinate space, similar to the classic procedure of generating the iris code. Given the two unrolled images, one can use one or more local cells (local image region) with arbitrary size and location to compute the comparison score. To achieve the maximum discrimination between the genuine and false comparison scores, we employ a gentleBoost-based learning scheme to select a discriminative subset of local cells in the spatial domain, where each cell constitutes a weak classifier by using the classic cross-correlation function (CCF) score. Experimental results and comparison with the state of the art on the NIST database and a new operational database obtained from a forensics agency demonstrate the viability of our approach. We will also point out the limitation of the current study by analyzing the error rates.

Session: Poster 6

A Model for Estimating Error Rates in Forensic Document Examination

Mr. Steven A. Slyter, S.A. Slyter, LLC, United States

■ Abstract

This paper presents a theoretical model for estimating the risk of errors in the practice of forensic document examination. First developed in the mid-1980's prior to the development of any objective data on error rates, the author's model projects an uncertainty factor as high as .07 in "best case" scenarios. The author initially believed this rate to be far too high but Dr. Kam's research beginning in the mid-90's shows an estimate in the range of 7% to be correct.

In real-world practice, document examiners commonly encounter less than ideal cases. The model presented suggests that the risk of an error triples when any aspect of the evidence is compromised.

The author is a forensic document examiner with nearly fifty years of experience. The model is a series of graphs - a picture book, intended to be understood by document examiners and jurors who are unfamiliar with statistical analysis. The conclusions drawn from this model are presented in the author's book *Forensic Signature Examination* (Charles Thomas Publishing, 1995).

As accurately noted in NIST publications, as practiced today handwriting examination is a subjective field with a great deal of examiner variability. From his experience the author believes that the most common error made by document examiners is over-estimating the value of the evidence and then over-stating their level of confidence in an opinion. The model demonstrates the benefit of thoughtfully ranking the evidentiary value of the submitted materials before beginning any comparison of writing features.

The estimates of error rates developed in this model make it clear that in the great majority of cases handwriting identification should be seen as supportive evidence taken at an appropriate level of certainty. Identification can only rise to the level of dispositive proof in the rarest of circumstances when the depth of individualized detail is complex and equivalent to birthmarks, scars and tattoos.

In contrast with the popular use of five, six or even seven levels of opinion this model presents an objective argument for limiting the expression of a document examiner's opinion to one of four levels of certainty. Each level is clearly explained in relation to the available evidence and each level has a defined limit of certainty. Opinions are not semantically nuanced.

Quantifying the evidentiary value of the submitted materials is a first step toward quantifying the analytical comparison of writings. Just this first step allows a more objective review of the analyses. Acknowledging that evidence quality imposes a cap on the ultimate level of certainty can help minimize the influence of unconscious bias. Working from a platform with defined levels of risk helps the examiner express opinions within appropriate boundaries.

Session: Poster 7

Information Sampling and Weighting in Fingerprint Recognition

Silke Jensen, University of Leicester, United Kingdom

■ Abstract

Background: Fingerprint identification evidence has been used in a forensic context for over a century. Fingerprint examiners develop a high degree of expertise at complex pattern recognition. Understanding the perceptual processes underlying fingerprint examination can improve recruitment, training and examination guidelines, impact error mitigation and support expert witnesses in court. The present research describes a pilot investigation into a new task to examine information sampling and weighting across fingerprint regions. Method: We developed fingerprint stimuli consisting of a central core and a peripheral region. Both regions are of equal geometric area and can be independently manipulated, that is the core can be drawn from a different fingerprint identity than the periphery. The boundary of the regions is masked to avoid completion cues. Artificially generated fingerprints were used to control for further low-level cues of identity (e.g. luminance, contrast, size information). In a yes/no recognition task, a target and probe were presented serially, for two seconds each, with a one second inter-stimulus interval. Target-probe combinations can be summarised as follows: Only core regions match (Core-Match); only peripheral regions match (Periphery-Match); both regions match (Full-Match); no regions match (No-Match). Participants (8 fingerprint novices) were asked to indicate their recognition response (match/no match) and response confidence (6-point-scale). Results: While accuracy was above 90% in Full-Match and No-Match conditions, subjects made significantly more false identifications in Core-Match conditions compared to Periphery-Match and No-Match conditions. Conclusions: These findings imply that changes within the peripheral regions of a fingerprint are noticed or weighted less by novices than changes within central regions. Further studies involving fingerprint examiners, as well as a variety of stimulus and presentation types, are required to investigate information sampling and weighting strategies in fingerprint experts.

Session: Poster 8

Real Science Admits its Uncertainty: DNA Evidence is Often Mischaracterized as Infallible

Mr. Jason O'Rawe, Cold Spring Harbor Laboratory, United States; Dr. Scott Ferson

■ Abstract

Science is defined in part by an honest exposition of the uncertainties that arise in measurements and by careful propagation of those uncertainties through calculations and inferences, so that the reliabilities of its conclusions are made apparent. In 2009, the National Research Council criticized forensic science in the United States for failing to indicate the presence and implications of uncertainty in its measurements and calculations. The review suggested that DNA evidence is the only scientific component of forensics and called DNA evidence “real science at work”. The reality, however, is that technologies that measure the quantity and quality of DNA do not take full account of the uncertainties arising in these sequence data. Relatively standardized methods for analyzing DNA sequencing data have been developed. Their software implementations enrich applied and basic research programs and have allowed for data consumers to gauge, to some degree, how reliable their sequence data and inferences are. Despite this, none of the current methods allow for full-fledged uncertainty accounting. Analysts cannot propagate uncertainty through downstream calculations, which is made worse by the fact that software implementations for calculating with epistemically uncertain data are non-existent. Thus, even when uncertainties are assessed, they are not propagated through subsequent calculations, as though they are not relevant for making higher-order inferences.

The recent rapid development of high-throughput DNA sequencing technologies has dramatically increased the number of measurements made at the biochemical and molecular level. These data come from many different DNA-sequencing technologies, each with their own platform-specific errors and biases, which vary widely. Several statistical studies have tried to measure error rates for basic determinations, but there are no general schemes to project these uncertainties so as to assess the surety of the conclusions drawn about genetic, epigenetic, and more general biological and forensic questions.

There are numerous causes of potential error in DNA evidence across the wide range of sequencing technologies currently in use, including

- samples containing too little DNA,
- sample degradation,
- amplification biases that emphasize contaminating DNA,
- poor base detection due to overlapping signals,
- genetic mosaicism implying multiple genetic identities within a single organism,
- misalignments or errors in assembling short sequences into longer sequences,
- mistaken probability calculations (e.g., from inappropriately assuming independence).

Even slight mischaracterizations of uncertainties can lead to widespread misinformation. For instance, the headline-grabbing announcement by microbiologists in 2015 about finding the organism causing Black Death (*Yersinia pestis*) in New York City subways was made with significant uncertainty, yet this uncertainty was understated in the primary report and virtually ignored in most popular newscasts.

We review the state of uncertainty quantification in DNA sequencing applications. We describe the sources and conditions of errors that arise in high-throughput sequence data, and we describe components of these errors that are the consequence of manifestations of different types of uncertainties. We also propose methods that can be used to account for and propagate these uncertainties through subsequent calculations, and we advocate for the formal incorporation of uncertainties into the computational frameworks handling new DNA sequence data.

Session: Poster 9

Improved Law Enforcement Officer Reports with POV Cameras - Gathering Better Scene Information

Dr. Donald Dawes, Lompoc Valley Medical Center, United States; Dr. Jeffrey Ho

■ Abstract

The law enforcement officer (LEO) report can be a key element in a determination of a cause of death, particularly in deaths in which the gross anatomic findings at autopsy are not conclusive. In excited delirium deaths, for example, the physiology that contributed to the death (e.g., acidosis, hyperthermia, autonomic dysregulation, etc.) cannot be determined at autopsy and medical examiners must rely on other data, to include LEO reports, to make the determination of the cause of death. There is a robust literature on the re-constructive nature of memory and how errors are introduced. The common misconception is that memory is similar to a digital recorder, when, in fact, memory is a reconstructive process; a forming of complex connections in a neural network. Memory is not infallible, and stress, fatigue, and exertion, all factors commonly encountered by police during arrest, can further impair memory.

The purpose of our study was to examine the accuracy of the LEO report in the setting of a stressful encounter and to determine the improvement when officers used on-officer point-of-view cameras. The LEOs in our study came from 5 different law enforcement agencies in the Phoenix, AZ area. The average age was 37 and the average number of years on the job was 12. The average number of years on patrol was 8. The subjects were fitted with a TASER Axon POV camera and 'dispatched' to three back-to-back scenarios: 1) a domestic disturbance, 2) a vehicle stop, and 3) a theft report. After the three scenarios, the LEOs wrote their report on the first scenario, which required use of force to resolve the scenario, from their memory. After completing their report, the officers then amended their report as needed after watching their POV video. The LEOs had an average of 2.63 minor errors (range 0-7), 5.4 moderate errors (0-14), and 0.9 major errors (range 0-3) per report. The moderate errors included 3 errors related to the sequence of the use of force, 7 errors related to miscounting the uses of force, and 4 errors related to some description of the use of force (e.g., location of the force). The largest number of errors was with quotations and other statements important to the case, such as commands from the LEO. This accounted for 27 errors. Omitting important suspect behavior accounted for 10 errors. There were 2 major errors related to omitting a use of force in the report, and 4 errors with regard to verbal warnings. Of interest, there were many errors that persisted even after video review.

The relevance of the study to forensic examiners is that police use of point-of-view cameras will likely improve the accuracy of death investigations. Indeed, the conclusions from the study may be portable even to the consideration for the use of such cameras for scene investigators and for autopsies. Scene investigators or medical examiners may miss something in their initial observations that could later be caught on another review of the video.

Session: Poster 10

Conveying Technical Judgments Involving Uncertainty

Dr. Leigh Cash, Los Alamos National Laboratory, United States; Dr. James Gattiker

■ Abstract

Uncertainty information in communicating the results of technical analysis typically takes two different forms: the result of analysis of uncertainty in quantitative inference, and a qualitative assessment of additional factors outside the domain of the quantitative framework.

The methods and tools of Uncertainty Quantification (UQ) are used in forensics analysis for model-supported inference, producing statistically defensible posterior distributions on quantities of interest. A qualitative confidence in the analysis and/or conclusions is often appropriate and increasingly a requirement. The interpretation of a qualitative ‘high, medium, low’ judgment can be unsatisfactory both to the domain expert assigning the qualitative information, concerned about how this will be interpreted, and to the downstream consumer of the annotation, who may not be able to evaluate the issue behind the annotation.

This poster will support discussion on a proposal for how qualitative information can be understood by communicating parties, by considering how it would be integrated if it did represent quantitative probability. There are distinct categories of expressions that may be made about the analysis relating to expected bias or uncertainty. We also propose that qualitative results can be best communicated to technical consumers of information (i.e. collaborators or downstream technical analysts) in a systematic framework of graphical inference. On the other hand, the considerations of judicial review of evidence suggest that confidence should be expressed more succinctly.

This poster is inspired by experiences supporting the nuclear forensics analysis community. This forensics focus area relies heavily on both technical analysis from lab chemistry through computer models, as well as expert assessment with an emphasis on community collaboration.

Session: Poster 11

Potential Errors and Uncertainty in Microscopy Image Analysis

Dr. Leigh Cash, Los Alamos National Laboratory, United States; Ms. Alison Tamasi

■ Abstract

We use digital imaging in many areas of forensics. Microscopic characterization of samples of various materials is routine. Microscopy, and other imaging techniques, continues to advance our ability to characterize materials, and the quality and quantity of data is rapidly increasing. We are pushed to do more quantitative analysis of materials from these images. For example, the morphological and microstructural characteristics of particulate material (chemical, pharmaceutical, nuclear) are important forensic signatures that can be used to link materials to the processes by which they are made. However, as with all laboratory measurement data, there are many sources of bias, error, and uncertainty. These issues can result from typical laboratory sources (sampling error, instrumentation calibration) but also result from human biases (human-directed instrument settings that change image appearance, a tendency to select cool looking particles for further analysis), and are magnified due to the interim image analysis step (image processing and analysis that does not fall into the typical image enhancement methods) and the final quantification steps (image noise, digitization limitation, quantification algorithms.) As we try to become more quantitative, we require an error model that accounts for multiple layers of bias and uncertainty, and includes instrument, human, and algorithm factors.

As the technology and tools that generate chemical and microscopic image data advance we must be mindful of the introduction and propagation of errors in the analysis of these images. Many of these microscopy image analysis issues will even come into play when doing 'simple' comparative analysis of images of materials. This work will discuss some of the errors and uncertainties introduced into microscopy image analysis. We are particularly interested in quantitative analysis of particulate materials for nuclear forensics applications, and many examples in this presentation will come from the analysis of these materials.

Session: Poster 12

The Qualification of Numerical Values Through Natural Language Expressions of Uncertainty in Expert Testimony

Dr. Kari Sentz, Los Alamos National Laboratory, United States; Dr. Scott Ferson

■ Abstract

The use of specific words to qualify uncertain numerical values is ubiquitous in natural language and critical in the communication of scientific evidence and expert testimony. Examples from courtroom testimony include:

- the probability was about 1 in 73 million
- I was notified at approximately 10 o'clock
- you testified you would be able to find as many as thirty

We explore the qualification of numerical expressions through natural language constructs called linguistic hedges or approximators and the implications for the interpretation of scientific evidence in court. We present this through examples where hedged numbers are reinterpreted as precise values during cross-examination and examples where hedges are remarkably absent despite the obvious presence of scientific error. Both situations perpetuate the mistaken impression that scientific evidence is more precise than it actually is and this is often consequential for adjudication. We explore the linguistic hedging of numerical values in expert testimony and discuss recent experimental efforts to quantify the uncertainty associated with linguistic hedges in context. The aim is to socialize a better understanding of the uncertainty in scientific evidence through the disambiguation and use of linguistic hedges as numeric qualifiers.

Session: Poster 13

Minimizing Contextual Bias in Forensic Firearms Examinations

Mr. Erwin Mattijssen, Netherlands Forensic Institute, Netherlands; Mr. Ralph Kleuskens

■ Abstract

Solutions to cognitive and human factor challenges in forensic science are emerging [1,2], but have not yet been widely communicated or applied in practice. In this poster we will show the efforts that have been put into dealing with contextual information by the Firearms section of the Netherlands Forensic Institute [3]. A context management procedure [3-5] has been developed and implemented that mitigates the effects of domain-irrelevant information and maximizes the utilization of domain-relevant information. By going through the steps of the examination of a typical comparison it is illustrated how and when contextual information may influence the examiner's judgment and how this influence can be minimized. Guided by a categorization of different sources of contextual information [4], a flow-chart is constructed specifying the process of casework examinations and context management. Application of this context management system does not make a subjective examination objective, but can limit the risks of bias with a minimal investment of time and resources.

References:

- 1) Dror, I.E. (2014). Practical Solutions to Cognitive and Human Factor Challenges in Forensic Science, *Forensic Science Policy & Management* 4(3-4), 1-9.
- 2) Dror, I.E. & Stoel, R.D. (2014). Cognitive Forensics: Human Cognition, Contextual Information, and Bias, *Encyclopedia of Criminology and Criminal Justice*, 353-363, Springer Publishing.
- 3) Mattijssen, E.J.A.T., Stoel, R.D. & Kerkhoff, W. (in press). Minimizing Contextual Bias in Forensic Firearms Examinations, *Encyclopedia of Forensic Sciences*, Wiley.
- 4) Stoel, R.D, Berger, C., Kerkhoff, W., Mattijssen, E.J.A.T. & Dror, I.E (2014). Minimizing Contextual Bias in Forensic Casework, in *Forensic Science and the Administration of Justice: Critical Issues and Directions*, 67-86, SAGE: Thousand Oaks, California.
- 5) Found, B. & Ganas, J. (2013). The management of domain irrelevant context information in forensic handwriting examination casework, *Science and Justice* 53(2), 154-158.

Session: Poster 14

Data Integrity Red Flags: Recognizing and Responding to Warning Signs; Lessons Learned from the Laboratory Industry

Ms. Janine Arvizu, Independent Auditor, United States; Ms. Amy Sirignano

■ Abstract

Fraudulent and deficient testing has plagued the laboratory industry for decades. The consequences of unreliable lab work have been severe, for both the laboratories and their data users. Laboratory failures in the clinical and environmental industries have resulted in criminal charges and sizeable financial costs, and have motivated these industries to proactively address the issue of unreliable testing. The lessons learned through decades of dealing with data integrity problems can provide the forensic and legal communities with resources, tools, and insight for improving the quality of forensic laboratory work.

The discovery of poor science, deficient procedures, and forensic misconduct rocked the FBI Laboratory in the 1990s. These revelations were the first in a long line of forensic laboratory scandals. In recent years, instances of serious fraud, incompetence, and errors have occurred in forensic laboratories throughout the United States. Forensic laboratories have adopted quality measures in an attempt to prevent problems, but those using forensic results are woefully ignorant of warning signs that signal data integrity concerns and may result in erroneous findings and serious legal repercussions.

This presentation draws from the best practices of the clinical and environmental industries, and applies them to forensic testing. It provides forensic managers and users of forensic results with universally applicable “red flags” that serve as warning indicators. It also provides a practical process for identification and confirmation of the most common data integrity problems. In doing so, this presentation adopts an appropriately broad definition of the term data integrity. A forensic test result is considered to have integrity if its accuracy, representativeness, and validity are demonstrated objectively. Thus, this definition encompasses the full range of issues affecting forensic results, including intentional fraud, incompetence, and technical errors and deficiencies.

Session: Poster 15

Evaluation of Novel Technological Solutions for Detecting and Interpreting Presumptive Drug Tests

Dr. Subrata Acharya, Towson University, United States; Dr. Kelly Elkins

■ Abstract

The potential benefits of hand-held mobile and tablet technologies for crime scene investigation have clearly been demonstrated by proof of concept experiments with the functional applications, or apps, ColorAssist to aid in drug testing, the now widespread adoption of CrimePad for processing and documenting crime scenes and MagicPlan CSI for measuring and sketching crime scenes. These new technologies allow crime scene investigators to more quickly process crime scenes without sacrificing accuracy, precision or security. A recent survey by our group suggests that a wide range of apps are being used to assist investigators in processing crime scenes including collecting data, documenting evidence, and building reports. However, apps must meet hardware and software requirements, overcome security limitations with password protection, data encryption, and put limits on screen shot recording, and fulfill data ethics and compliance for legislative requirements for data storage. Despite the clear value of apps, there is limited software currently available. The technical challenges to further mobile tools and apps include integrating devices for data collection with the app that are portable, have sufficient battery power, rugged and are available at relatively low cost. The overall goals of this project is to more accurately interpret and record colors of presumptive drug tests and facilitate data submission to a remote server. To achieve these goals, we develop a framework and prototype application to run on Apple and Android smartphone devices and a Raspberry Pi device integrated with a commercial off-the-shelf USB UV-VIS Spectrometer that will decide the identity of drug evidence based upon the color recorded after performing an on-site chemical test based upon the best color match in a library. The color will be read using either the device camera in the Android and Apple applications or using the spectrometer with the Raspberry Pi. This represents a new approach to recording colors of presumptive drug tests that eliminates human variation and limitations with regards to recognizing and systematically describing different colors as two analysts may describe the same color differently, e.g. brilliant greenish blue vs. strong greenish blue. This approach also opens up these CSI tasks to less experienced CSIs and severely color-blind individuals. The resultant apps and device are compact the size of a smartphone or a little bigger and the captured data can be uploaded to a secure server and viewed by law enforcement or the laboratory while the investigators are still at the crime scene and facilitate real-time collaboration with the lab. The software will be provided at no cost to the end users. The software will all require secure login and the data will be encrypted. We will collaborate with active CSI expert users for on-site testing and evaluation to be compared to the traditional method.

Session: Poster 16

Considerations of Proficiency Testing Providers and Error Management

Dr. Shirley Turner, NIST, United States; Ms. Stacy Doorn

Abstract

Proficiency testing (PT) can be an important contributor to error management for laboratories. However, there may be a range in strength in the proficiency tests as far as error detection and mitigation. In some fields, there may be a natural evolution towards proficiency tests that are simple, have reduced costs and that are compatible with ease in satisfying requirements for accreditation (1). Perhaps a single value per analyte is evaluated in the proficiency test. The laboratory can gain insight as to its performance by comparison to other laboratories and/or a reference value but may have difficulty determining sources of error. Other proficiency tests have attempted a more in-depth analysis of errors by collecting experimental information and analyzing the performance relative to different experimental setups, instrumentation, etc. (2). The proficiency test can provide insight into weaknesses in the analytical field that can lead to improvement in analytical results. This work discusses development of a model whereby an even more active role for error detection and mitigation is undertaken by the PT provider. Though the PT program is not specific to forensics, the approach to proficiency testing described may provide a model for an alternative approach to PT in forensics.

The model has been developed and is being refined for a PT program for analysis of asbestos by transmission electron microscopy (TEM) that involves approximately 75 laboratories. It is associated with the National Voluntary Laboratory Accreditation Program (NVLAP) and is conducted by Research Triangle Institute (RTI) with technical oversight, direction and analysis from the National Institute of Standards and Technology (NIST). Some elements of this approach to proficiency testing include:

- 1) Investigation by the PT provider of the source of errors for outliers, where possible, by recalculation, reanalysis of returned samples, evaluation of submitted supporting data, etc.
- 2) Detection by the PT provider of errors resulting in near misses in addition to outliers. This allows for remediation of problems that could lead to outliers in future samples.
- 3) For some cases, development of approaches to error mitigation including development of training materials and reports for the laboratories.
- 4) Categorizing the major types of errors and monitoring performance of individual laboratories and laboratories as a group over time in each category. Where appropriate, assessing the effects of the approaches to error mitigation.
- 5) Adjusting the limits of acceptable values as proficiency increases and increasing the level of difficulty of the samples and/or required work to spur further advancement by the laboratories.
- 6) If a weakness is recognized in the laboratories as a group, development of a series of tests that along with the training reports gives laboratories the opportunity to gain expertise in the area.

Examples will be given for the different approaches listed above. The advantages and disadvantages of this type of proficiency testing will be discussed.

References:

1. Gardner, MJ (2007) *Accred. Qual. Assur.* 12, 653-657.
2. Zaninotto, M., et al. (2004) *Clin. Chem. Med.* 42, 1434-1441.

Session: Poster 17

Validation of The Leica Geosystems ScanStation C10 for Crime Scene Investigations

Thomas Wisner, Baltimore Police Department, United States; Amelia Faelyn

Abstract

The Baltimore City Crime Laboratory's Crime Scene Unit recently acquired a Leica ScanStation C10, complementary external camera, and Cyclone software versions 8.3 and eventually v.9. These additions will provide a futuristic documentation source resulting in a 3-D point cloud depiction of a crime scene as well as accompanying 360° panoramic photographs. Measurements can be taken from a point cloud rendering of scanned data with a stated accuracy of 0.002ft (6mm). Scenes from point clouds are frozen in time and can be measured at anytime following the incident. Data can be transferred to TruView software, a free software available to police and litigators. 3-D scanning is therefore an invaluable court presentation tool. An internal validation was completed to comply with ISO/IEC 17020:2012 accreditation using different traceable measuring devices including two Leica NIST certified Twin Target Poles.

Three scans from the Leica ScanStation C10 containing images of current measuring devices were compared alongside their theoretical values. The measuring devices scanned included a 1ft (0.305m) 2ft (0.610m) Sirchie Evidence Folding Ruler and a 50ft (15.240m) Lufkin/Sirchie Measuring Tape. Additionally, a Bosch Laser Distance Measurer also employed by the Mobile Unit was used for comparison with Leica scan data. Measurements from two NIST certified Twin Target Poles were utilized to verify the accuracy of the scanned data points.

A two-valued logic was used to determine whether the data points gathered from the Leica as compared to measurements from conventional methods were in same range given their uncertainties. Results showed overlap between corresponding values within the two ranges showing that the measurements were similar. Various T-Tests were done to compare differences between scans, between point cloud measurements and theoretical measurements, and to compare differences in values found between individual software operators. The averaged differences between paired scan values resulted in experimental T values of 1.079, 1.327, and 1.209 with a tcrit value of 2.776 at a 95% confidence level. The means of corresponding point cloud measurements from all three scans; in comparison to the theoretical values of current forensic devices; produced experimental T values of 0.143, 0.476, 0.018, 0.0002, and 0.355 with a tcrit of 2.132 at 95% confidence. Differences between corresponding values from separate point cloud operators yielded experimental T values of 1.772, 0.405, and 0.434 with a tcrit of 4.303 at a level of 95% confidence.

It was also found that a high resolution image provides 0.164ft (0.050m) with a horizontal and vertical spacing at 328ft (100m) while a medium resolution's spacing is 0.328ft (0.100m). A high resolution scan took approximately 28 minutes and had a file size of approximately 2.34GB compared to the medium resolution's approximately 8 min at nearly 700MB. Overall, these results show that the Leica C10, when used in tandem with the Cyclone 8.3/9 software, is an accurate and precise method of documenting and measuring a crime scene.

Session: Poster 18

Validation of TrueAllele Casework System with the PowerPlex Fusion Chemistry

Thomas Hebert, Baltimore City Police Department, United States

■ Abstract

Probabilistic genotyping is a powerful tool in analysis of complex and difficult DNA samples. The TrueAllele casework system is one such probabilistic system. It allows for direct analysis of raw data without the use of thresholds. Thresholds such as analytical, stochastic, and stutter and their uses often differ between labs and are points of contention among experts. The TrueAllele system was examined to determine the effectiveness of probabilistic genotyping and the elimination of thresholds. Samples from one to five contributors of differing amounts of total DNA were explored. Results were compared to human interpretation to determine the effectiveness and compatibility of the two methods.

White Papers As Submitted

Effect of Drop-in on False Positive and Rank-Order Likelihood Ratios Calculated by Forensic Statistical Tool for a Mixture of Touch DNA

Authors: Clinton Hughes, Arthur Speiser, Kevin Ramdass, Nicholas Corpuz, Khagay Nagdimov and Eli Shapiro (The Legal Aid Society of New York City)

Submitted to NIST: 9/30/15 (some reformatting performed by NIST)

Presentation (27 slides) available at

http://www.nist.gov/director/upload/effect_of_dropin_on_false_positive_and_rank_order_likelihood_ratios_LR_calculated_for_a_mixture_of_touch_DNA-hughes-crim1.pdf

ABSTRACT

We examined the effects of drop-in on Likelihood Ratios (LRs) obtained using a LR program called Forensic Statistical Tool (FST). FST uses pre-set drop-in and drop-out probabilities based on sample quantitation and modified from values determined by empirical experiments on mixtures created from pristine buccal swab extracts. We used the computational methods and parameter values described in the validation of FST to construct an Excel program and more recently a C++ program mimicking FST for calculation of the LR. The sample we analyzed in detail was a low-template touch DNA mixture of three known individuals created and analyzed as part of FST validation. The mixture contained 20 drop-in alleles, not associated with any of the known contributors, out of 78 total alleles. This drop-in rate is about 10-fold higher than the drop-in rate used by FST. The FST generated false-inclusions created by drop-in alleles, including one LR of 157 for a profile of "JB," a contributor to the NIST Caucasian database. This LR was higher than that obtained for the two known minor contributors to the mixture by FST (4 and $10e-5$). Drop-in also affected the rank-order of LR, creating many profiles with higher LR than those for known contributors. Drop-in created the highest-ranking LR, $10e14$, a non-contributor. FST strategy of using pre-set parameters not related to the actual evidence makes its LR calculations unreliable.

INTRODUCTION

Recent publications using two forensic statistics methods for calculating LR have reported data on false-inclusion rates when calculating a LR for complex DNA mixtures (1,2). Using different approaches, both studies found that non-contributor profiles could generate LR greater than 1 ('inclusionary'). Mitchell, et.al. (1) created artificial mixtures and measured LR values for known non-contributors using FST. Gill and Haned (2) used simulated non-contributor profiles to examine the LR they generated on casework mixtures.

Gill, et.al., (3) also studied rank ordering of random and evidentiary profiles, and examined false 'inclusionary' LR rates during database searches in a study that allowed for drop-out, but not drop-in. They found that non-contributor profiles could generate a higher LR than known contributors. They also found a higher rate of false inclusions in a mixture with more alleles. As noted by multiple authors, drop-in should always substantially weaken the strength of the evidence, and must be rare (e.g. 4, 5, 6, 7, 8).

We examined the effects of drop-in on the calculated LR using a laboratory validation sample created by three known touch DNA contributors. Allelic drop-in is defined by John

Butler (reference 9, Appendix 1, page 440) as “contamination from an unknown source.” FST also counts stutter artifacts as drop-in alleles. Therefore, drop-in means artifacts and contamination. With that in mind, we examined the effects of drop-in using one of the false-inclusionary LR calculated by FST (obtained from the FST validation, reference 10) for a known non-contributor profile from the NIST database. This false inclusion, found in a search of approx.1200 database profiles was caused by drop-in (10).

As described in Mitchell, et., al. (1) for the drop-in rates used by FST: “Our drop-in values (0.005–0.035) are consistent with Balding and Buckleton’s (2009) recommendation of less than 0.05 to reflect the relative rarity of this phenomenon.” In reality, drop-in rates for low template touch DNA mixtures amplified in triplicate can be quite high.

METHODS

The example we use, which we term “JB mixture,” is taken from the validation studies for the FST, conducted by the Department of Forensic Biology, NYC Office of Chief Medical Examiner (OCME) (10). The sample was referred to as PenB in the validation studies, and was one of the samples in validation study 3E, where three-person touched items were made by having three known individuals touch an item, after the item had been cleaned with bleach, water, and alcohol, to remove all DNA on the item prior to being touched by the known contributors. The item was then swabbed, and DNA and statistical analyses performed to determine the LR for the known contributors. In addition, LR calculations were performed on over 1200 profiles from known non-contributors (false-positive study). Non-contributor profiles were obtained in part from a public NIST allele frequency population database (1,10).

In 2012, The Legal Aid Society obtained the full FST validation study, as well as the OCME population database as part of its litigation in the Frye challenge against application of FST in two cases in Brooklyn, NY.

Obtaining a copy of the validation studies and OCME population database was critical because they included the pre-set drop-out rates used by FST, which are not listed in the OCME files, reports, protocol manual, or publications. The FST validation studies also contained the drop-in rates used by FST, which are reported incorrectly in reference 1. Finally, the Legal Aid Society was able to obtain the allele frequency tables used by the OCME for its casework statistics, after the OCME consented to provide the tables in response to an application for a judicial subpoena.

The FST Validation Studies included a binder called “Methods,” which laid out in some detail the formula that is the backbone of the FST calculations. It also included 24 spreadsheets of FST comparison at single loci that were used to check the FST code.

Among these 24 individual computations, there was one example following the JB scenario: a three-person mixture, where the numerator (prosecutor’s) hypothesis was a known profile (suspect) and two unknown unrelated individuals, while the denominator (defense) hypothesis was three unknown, unrelated individuals.

Using this information, we produced a spreadsheet designed to replicate the results of the FST. Our spreadsheet uses Microsoft Excel. We also used DigDB as an add-on to Excel, which created the combinations necessary to fill in the spreadsheets. More recently, we have created a program in C++ to perform these calculations. The most recent version of

the program, along with the source code and the data file for the JB comparison to the PenB mixture, can be accessed at https://github.com/mramdass/re-engineered_Quant-Based_Tool/releases. Tables listing the allele frequencies and drop-out rates we used, and which are ostensibly used by FST are presented in Supplementary Material 2.

Our programs follow the same basic format that Mitchell, et.al., laid out in her manual calculations (validation binder “Methods”) where the numerator calculation for the known contributor is physically separated on the spreadsheet from the denominator.

In order to examine the reliability of the FST and the effects of drop-in on LR, our FST spreadsheet was extended to report the LR of every potential contributor to a mixture, on a locus by locus basis. Creating LR for all potential genotypes, and not just that of the “suspect,” will give us an idea of the percentage of the general population that will fit into the mixture at any particular locus (see also reference 11).

The FST comparison reports that are part of the forensic case file issued by the OCME do not provide locus LR, or the drop-out rates, drop-in rates, or allele frequencies used by the program. This makes objective technical review of their results impossible.

RESULTS

The PenB/JB Mixture Comparison

Supplementary materials 1 includes the electropherograms used to calculate the LR. The quantitation value of the DNA extract (total volume 20 ul) from the mixture was 12pg/ul, and this sample extract was split into three aliquots of 5ul volume, each containing approximately 60pg total DNA. Each aliquot was amplified using high-sensitivity protocols, including 31-cycle amplification (12).

Using these electropherograms, OCME analysts called the JB mixture “deducible,” and programmed FST to analyzed the mixture as a “deducible” three-person mixture. This set the program to calculate LR based on numerator and denominator scenarios with three contributors, and drop-out rates empirically determined then modified for minor contributors to 3-person mixtures in a ratio of 5:1:1. FST uses drop-out rates that were measured using pristine mixtures made by pipetting buccal swab sample extracts, and then arbitrarily adjusted (10). Drop-in rates used by FST were arbitrarily set, and unrelated to any empirical data collected from touch DNA samples (1, 10).

Table 1 shows the alleles called in each run. It also shows the known profiles of the three expected contributors, and the DNA profile of “JB”, a known non-contributor whose profile is one of the NIST Caucasian database. As calculated by FST, the JB profile gave a high false-inclusionary LR of 157, the highest of the 9 false-inclusions generated by this mixture. Our Excel program and our C++ programs both also calculated a LR of 157 for JB. According to the protocol adopted by OCME, a LR of 157 provides “strong support” for the scenario that the DNA mixture is more likely if it originated from JB and two random unrelated contributors than if it originated from three random unrelated contributors. This qualitative scale is based on the one recommended by John Butler (9), but not supported by any FST validation studies.

The JB mixture provides a good example of drop-in (**Table 1**, alleles in red). Out of 78 total alleles in the mixture, 20 are drop-in alleles not associated with any known

contributor. 13 drop-in alleles are found in the consensus profile, being present in at least two of three amplifications.

Out of 45 individual loci (15 Identifiler loci in three separate amplifications) 14 loci showed one drop-in allele not associated with the known contributors. An additional 13 loci showed two or more drop-in alleles, not associated with any known contributor. By contrast, the FST program uses a drop-in rate of 0.035 single drop-in events per locus, which would predict approximately 2 loci with one drop-in. The FST expects no loci with drop-in of two or more alleles, using its programmed drop-in rate of 0.005 for two or more drop-ins per locus. But the JB mixture had over 10 times more artifactual and contaminating alleles than the FST program expects and uses to calculate LR.

Out of the 57 alleles of the three known contributors, three dropped out. One allele in the JB profile dropped-out. There was no drop out of alleles in the profile of the major contributor. Since the drop-out rates used by the FST in deducible mixtures is based on the rates measured for the minor contributors' alleles the LR calculated by the FST for the major contributor ($10e4$) is not accurate. There were two drop out alleles associated with each minor contributor. The low LR values obtained by the FST for these known contributors could also be the result of excessive drop-in events.

Table 2 shows the locus-by-locus calculation of LR for the expected contributors, Donor 21 and Donor 23, for known non-contributor JB, and for the profile with the highest LR, using our program. JB's profile has 7 alleles that dropped-in to the mixture. The profile with the highest LR calculated using the OCME Caucasian database, $10e14$, was created by 10 drop-in alleles.

Drop-in also affects the reliability of LR by adding to the total allele frequency of the alleles at a locus. For example, **Table 3** shows the results for the JB mixture at locus D3. Here, two drop-ins, alleles 17 and 19, created a situation in which the total allele frequency of the 6 alleles seen in the mixture is over 0.99. The alleles in the mixture almost comprise the allelic ladder at D3, and 98% of the population is included at D3, even without considering the possibility of drop-out. This locus has no probative value, yet by conditioning the LR on a suspect's genotype, over 30% of LRs will be above 1, and the highest LR is 15. This highest LR is for profile 18,19, and is caused by the drop-in of the 19 allele. Also, the drop-in of the common 17 allele results in increasing the false inclusionary LR of JB as well as lowering the LR for the known contributors.

Figure 1 shows one shocking example in a real criminal case report. The mixture contained 80 total alleles (not shown). The OCME reported FST statistics for a 3-person mixture. The FGA locus had 10 alleles, and since the defendant was homozygous, DNA from a minimum of 5 additional contributors was present at FGA in the Prosecution hypothesis. The combined allelic frequency for the 10 alleles in the OCME Caucasian database add up to 1.00 (due to minimum allele count). We could not find any methodology in the OCME validation, publications, or protocol manual for the FST to handle this calculation.

A drop-in allele, at the D7 locus of the PenB mixture, allele 9, created the highest LR at the locus, and the highest locus LR for the JB profile, 5.76 (**Table 2**). Since all three expected contributors were homozygous at this locus, as was JB, the presence of 4 alleles does not suggest the need for drop-in to explain the results. We calculated the LR at D7 without including the drop-in 9 allele. The locus now is a three-allele locus, with alleles 8, 10, 11, all expected alleles present in the known contributors. JB's profile would be considered

as a profile w,w. This removal of a single artifactual or contaminating allele caused the D7 locus LR for JB to drop to 0.0121, and the total LR for JB to drop below 1, to 0.327.

DISCUSSION

Analysis of the JB mixture showed that drop-in can affect LR in the same way that has been demonstrated for drop-out. Drop-in caused a “strong” false positive LR, which was higher than the LR for known minor contributors.

Drop-in created a highest-ranking false positive.

Drop-in alleles were repeating and in the consensus profile. This observation calls into question the use of consensus profiles and multiple replicates, since this approach does not eliminate spurious alleles. Rather, multiple amplifications provide an opportunity for additional drop-in.

The drop-in rates used by the FST were much lower than the actual drop-in rate seen in the JB mixture. Given this poor description of the actual “evidence,” the unreliable FST results are not surprising. Using an unrealistically low drop-in rate exaggerates the strength of the evidence by inflating confidence that any allele is not an artifact.

The reverse-engineered FST also allowed locus by locus analysis of the mixture. This showed how drop-in alleles caused the highest LR values. Examining the allele frequencies at loci with 5 or 6 alleles in the mixture, along with the calculation of the percentage of DNA profiles giving LR greater than 1, showed that a locus with too many alleles can lose its probative value, yet still provide very high LR solely conditioned on the genotype of the comparison sample.

OCME protocols use a variety of criteria related to the number of alleles at each locus to determine the number of contributors to a mixture (12,13). This methodology is flawed, since 6 alleles at a locus such as D3 can include over 95% of the population, while 6 alleles at a locus like D21 would include a much smaller fraction of the world.

The OCME methodology for determining the number of contributors to a mixture, as described in Perez, et.al., (13) also clearly shows a systematic bias towards under-estimation of drop-in. Perez, et.al., state, for example,

Two 5:1:1 260 pg three-person samples contained 45 or 38 different alleles, values below the average minus 2 standard deviations. Neither sample met the criteria for three-person mixtures regarding the number of loci with 4 or more repeating or different alleles. Moreover, for both of these samples, 3 or more alleles belonging to 1 or more of the sample donors were not labeled. Due to allelic drop-out these samples may be better described as two-person rather than as three-person mixtures.

In fact, the 25pg 5:1:1 mixture, which had the minimum number of alleles, 42, did not meet the **Table 2** criteria for a three-person mixture as it did not have any loci with 5 or more repeating alleles and only had 1 locus with 4 different alleles and 1 locus with 5 different alleles. Since 2 of the 3 contributors were missing at least 3 alleles each, this mixture may be better described as two-person than three-person. Similarly, 4 other purposeful 25 pg 5:1:1 mixtures that had either 43 or 46 different alleles did not meet the qualitative criteria for three-person mixtures, and at least 1 donor for each was missing 3 or more alleles.

In reality, this approach of conflating partial profiles to minimize the number of contributors automatically increases the number of drop-in alleles. When a contributor is “eliminated” due to drop out of three, four, or five alleles, the 10, 15, or 20 alleles from this contributor that do not drop out are now “drop-in” alleles!

FST is not validated for use with 4-person mixtures. OCME reported the maximum number of alleles seen in 105 3-person mixtures to be 66, with a mean of 57 and SD of 4. For 4-person mixtures, the maximum number of alleles seen was 75, with a mean of 67 and SD of 4 (Perez, et.al., reference 13, Table 1). Based on these findings the JB mixture is unlikely to be a 3-person mixture, unless the drop-in rate is adjusted to account for approximately 20 alleles (see Table 1 of this paper). However, it is common for OCME casework to analyze mixtures such as the JB mixture with over 70 total alleles as 3-person mixtures by using alternative criteria. In these casework situations, the drop-in rates used by FST are not changed to account for the additional alleles (see also Fig. 2). Our analysis of one such example from the FST validation shows the need for better ways to account for and control drop-in when evaluating a “touch DNA” mixture.

REFERENCES

1. Mitchell,A, Tamariz,J, O’Connell,K, Ducasse,N, Budimlja,Z, Prinz,M, Caragine,T. (2012) Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *FSI:Genet.* 6, 749-761.
2. Gill, P., Haned, H. (2012) A new methodological framework to interpret complex DNA profiles using likelihood ratios. *FSI:Genet.*
3. Gill,P, Bleka, O., Egeland, T. (2014) Does an English appeal court ruling increase the risks of miscarriages of justice when complex DNA profiles are searched against the national DNA database? *FSI:Genet.* 13, 167-175.
4. Balding, D., Buckleton, J. (2009) Interpreting low template DNA profiles, *FSI:Genet.* 4 (2009) 1–10.
5. Steele, C., Balding, D. (2014) Statistical evaluation of forensic DNA profile evidence. *Annu. Rev. Stat. Appl.* 1, 361-384.
6. Bright, J., Gill, P., Buckleton, J. (2012) Composite profiles in DNA analysis. *FSI:Genet.* 6, 317–321.
7. C.C.G. Benschop, C.P. van der Beek, H.C. Meiland, A.G.M. van Gorp, A.A. Westen, T.Sijen, (2010) Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results, *Forensic Sci. Int. Genet.*,5316–328.
8. Taylor, D., Bright, J., Buckleton, J. (2013) The interpretation of single source and mixed DNA profiles. *FSI:Genet.* 7, 516-528.
9. Butler, J. (2010) *Forensic DNA Typing*. Academic Press, UK
10. Transcript and Exhibits from Brooklyn, NY, Frye hearing, 2013.
11. Dorum, G., Bleka, O., Gill,P, Haned,H, Snipen,,L., Solve,S, Engeland, T. (2014) Exact computation of the distribution of likelihood ratios with forensic applications. *FSI:Genet.* 9, 93-101.
12. Office of Chief Medical Examiner of New York City Forensic Biology Protocols for STR Analysis. 2012
13. Perez, J., Mitchell, A., Ducasse, N., Tamariz, J., Caragine. T. (2011) Estimating the number of contributors to two-, three-,and four-person mixtures containing DNA in high template and low template amounts *Croat. Med. J* 2011:393-405.

BLE 1. Alleles called in each run of “JB” mixture, the DNA profiles of the three known contributors, Donor 21, Donor 13, and Donor 23, and the DNA profile of “JB,” who had the highest number of non-contributors to the mixture. Donor 13 was the major known contributor. See text for details.

SAMPLE # 388-36 3PCPer01 68mm 3x D 100000, where "36" from NIST Caucasian database has an FSI LR of 150

	D051179	D051511	D05420	CSF8PO	D051506	TH01	D055317	D055509	D051338	D055433	VWA	TPOR	D055591	AMEL	D055818	FGA
Run 39	8,10,13,14, 15	28,29,30,3 0,2	8,9,10,11	9,10,12	14,15,16,1 7,18	6,8,9,9,3	8,9,14	9,10,11,12 ,13	19,20,22	12,13,13,2,1 4,15,2	16,17,18	6,8,9,11 7,10	12,13,14,1 7,10	X	11,13	19,21,22, 3,24,25,26
Run 40	8,10,12,13, 14,15	28,29,29,2 ,30,30,2,31	8,9,11	10,12	14,15,16,1 7,18,19	6,8,9,9,3	8,9,11,12	10,11,12,1 3	19,20,22,2 5	12,13,13,2,1 4,15,2	16,17,18	8,9	12,14,17,1 8	X	11,13	19,21,22, 3,25
Run 43	8,13,14	29,30,30,2 31	8,9	9,10,12	14,15,16,1 7,18,19	6,8,9,3	8,9,11,13	10,12,13	19,27	12,13,13,2,1 4,15,2	16,17	8,9	12,13,14,1 6,17,18	X	10,11,12,1 3,14	19,21,22, 3,25,28
consensus	8,10,13,14, 15	28,29,30,3 0,2,31	8,9,11	9,10,11	14,15,16,1 7,18,19	6,8,9,9,3	8,9,11	10,11,12,1 3	19,20,22	12,13,13,2,1 4,15,2	16,17,18	8,9	12,13,14,1 7,10	X	11,13	19,21,22, 3,25,28
IB	12,34	28,30	9	10	16,17	6	11,14	11,13	22,23	12,14	17,18	8	14,16	X,Y	12,13	21,22
DONOR 21	13,15	19,29,2	11	12	15,16	9	11,12	9,13	20,24	14,15,2	14,18	8,9	14,17	X	11	19,23
DONOR 13	8,14	29	8	9,10	14,18	6,8	6,9	10,12	19,22	12,14	16,17	8,9	12,14	X	11,15	22,25
DONOR 23	10,15	28,29	10	12,14	15,16	6	12,13	11	19,25	14	14,16	8	13,14	X	11,12	23,24

TABLE 2. Locus-by-locus and full LR values calculated by our programs for Donor 21 (OCME Asian database), Donor 23 (OCME Hispanic database), and for known non-contributor JB (OCME Caucasian database). Also shown is the profile that gives the highest LR using the OCME Caucasian database, 10e14. The JB profile has 7 alleles that were drop-ins. The highest LR was created by a profile containing 10 drop-in alleles.

SUBJECT	LR	D051179	D051511	D05828	C01190	D051338	T0491	D133117	D163230	D251338	D195433	vWA	TPOX	D18551	D05818	IGA
JR profile	1.07E+02	1.7066 12.14	1.0320 20.30	3.7920 3	2.7388 18	0.6344 16.17	1.4478 5	3.1288 11.34	1.3278 11.13	1.1744 22.25	1.5788 52.14	1.8888 17.18	1.1823 8	0.9057 14.16	0.5400 12.13	8.9782 21.22
DONOR 21 profile	1.88E+08	1.4628 13.16	2.3433 20.29.2	3.8355 11	2.1388 12	0.2803 15.16	0.4886 9	3.8737 11.12	1.8388 9.15	0.3551 20.24	0.3673 54.15.2	3.2552 14.18	1.2518 8.9	1.9294 14.17	3.8334 11	2.6197 18.23
DONOR 23 profile	5.48E+05	0.3566 10.15	1.5389 25.25	3.1495 19	0.1115 12.14	0.3340 15.16	1.4805 5	3.4571 12.13	0.2338 11	3.7718 19.25	0.8236 54	3.1887 14.15	1.8018 8	2.3938 13.14	0.4519 11.12	5.5823 23.24
HH profile	1.25E+15	20.8164 0.12	15.8071 30.2.31	8.7483 3	6.7427 8.12	15.1106 18.19	3.5080 8	4.5427 2.14	6.2139 9.15	103.4748 22.27	83.4096 13.2.13.2	4.8399 19	4.2158 8	4.8190 17.15	5.9492 11	8.7488 25.25

TABLE 3. Six alleles are present at Locus D3, and the sum of their allele frequencies (from OCME Caucasian population database) totals 0.997. The highest LR for profile 18,19 is over 15, The “19” allele is a drop-in. Allele “17,” the most common allele at D3 is also a drop-in. See text for further details.

% Pop LR > 1		Allele Freq. (Cauc.)	
30.90%		(a) 14	0.132
Knowns		(b) 15	0.24
15,16	Donor 21	(c) 16	0.236
14,18	Donor 13	(d) 17	0.273
15,16	Donor 23	(e) 18	0.095
		(f) 19	0.021
		w	0.003
LRs Hi to Lo		Genotype Freq. (Pg)	
18,19	1.51E+01	14,14	0.0208613
14,19	1.09E+01	14,15	0.06336
16,19	6.08E+00	14,16	0.062304
15,19	5.98E+00	14,17	0.072072
17,19	5.26E+00	14,18	0.02508
14,18 (Donor 13)	2.71E+00	14,19	0.005544
16,18	1.52E+00	14,w	0.000792
15,18	1.50E+00	15,15	0.063072
17,18	1.32E+00	15,16	0.11328
14,16	1.10E+00	15,17	0.13104
14,15	1.08E+00	15,18	0.0456
14,17	9.51E-01	15,19	0.01008
15,16 (Donors 21 and 23)	6.07E-01	15,w	0.00144
16,17 (“JB”)	5.34E-01	16,16	0.0611051
15,17	5.26E-01	16,17	0.128856
18,18	1.26E-01	16,18	0.04484
14,14	9.12E-02	16,19	0.009912
16,16	5.14E-02	16,w	0.001416
15,15	5.05E-02	17,17	0.0804831
17,17	4.45E-02	17,18	0.05187
18,w	3.26E-03	17,19	0.011466
14,w	2.35E-03	17,w	0.001638
19,19	1.49E-03	18,18	0.0116043
16,w	1.33E-03	18,19	0.00399
15,w	1.30E-03	18,w	0.00057
17,w	1.15E-03	19,19	0.0010578
19,w	8.85E-05	19,w	0.000126
w,w	9.74E-07	w,w	9.873E-05

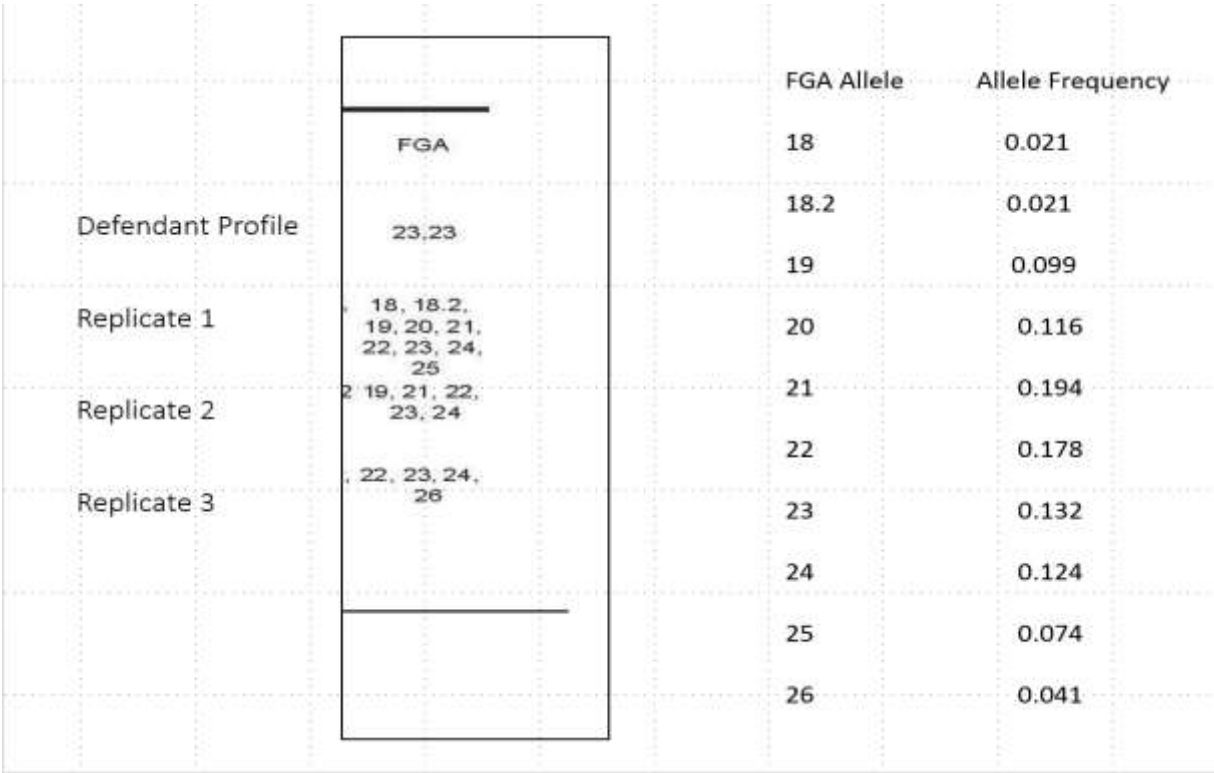


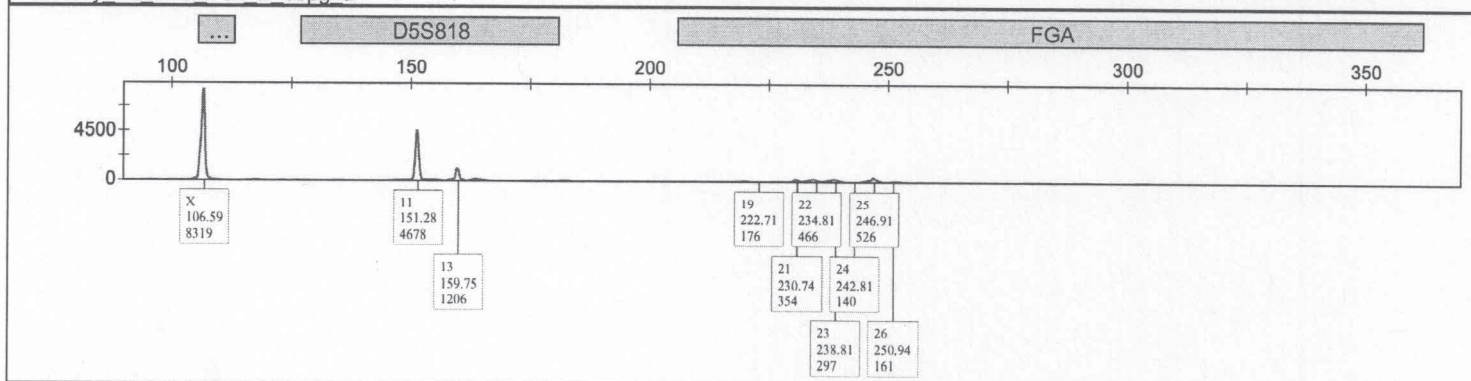
FIGURE 1. Casework example of Locus with 10 called alleles in a mixture analyzed by FST as a three-person mixture. Note that the defendant is homozygous at this locus. The entire mixture contained 80 alleles.

Supplementary Materials 1.

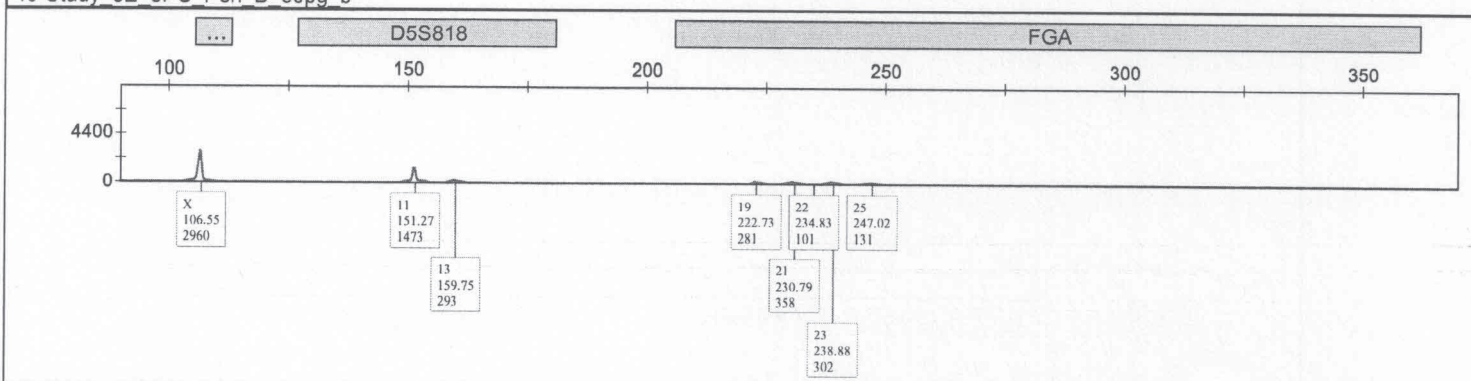
Electropherograms of the JB mixture. The sample was amplified in Identifiler in triplicate (electropherograms labeled "a" "b" "c"), and the results of the amplifications sorted to batch the color lanes containing the same loci next to each other. Based on these electropherograms the OCME analyst deduced the profile of the major contributor, and used FST to analyze the LR for minor contributors assuming a mixture ratio of 5:1:1.

Sample Name

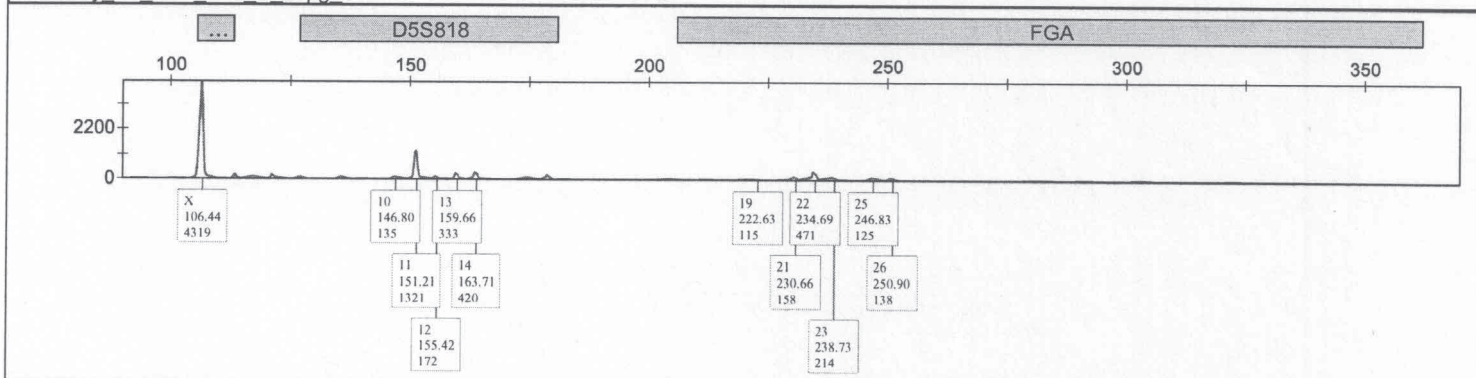
39-Study 3E 3PC Pen B 60pg a



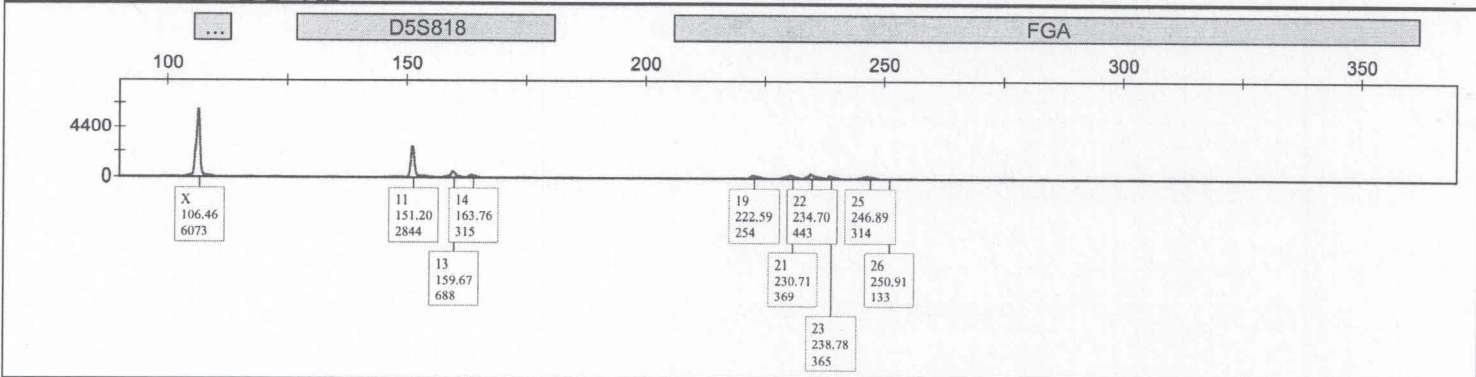
40-Study 3E 3PC Pen B 60pg b

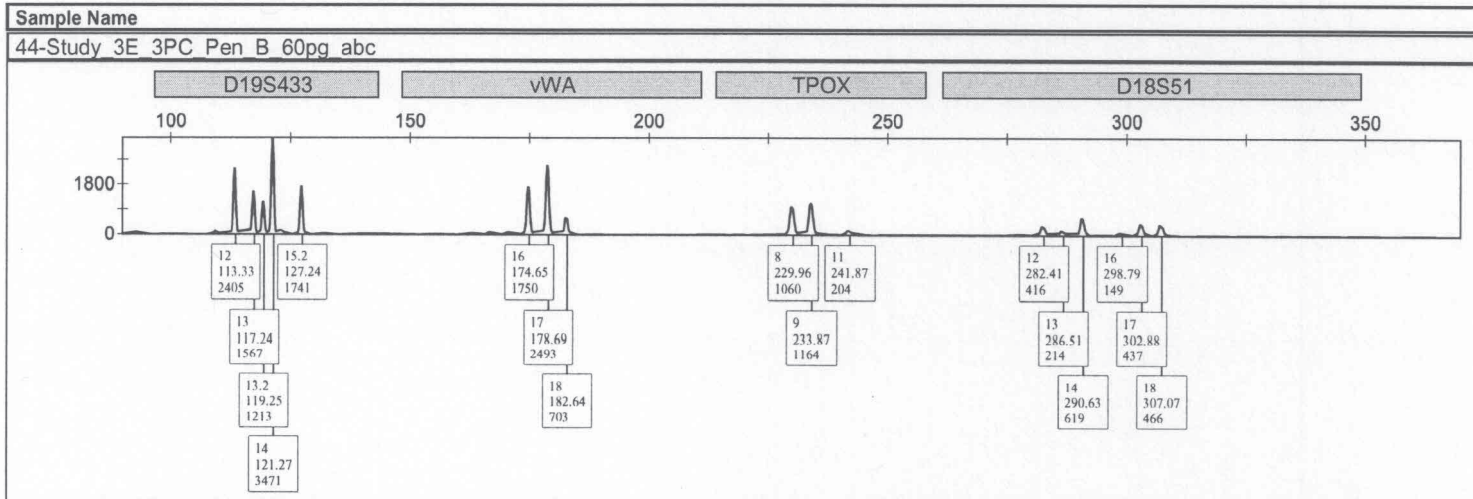


43-Study 3E 3PC Pen B 60pg c



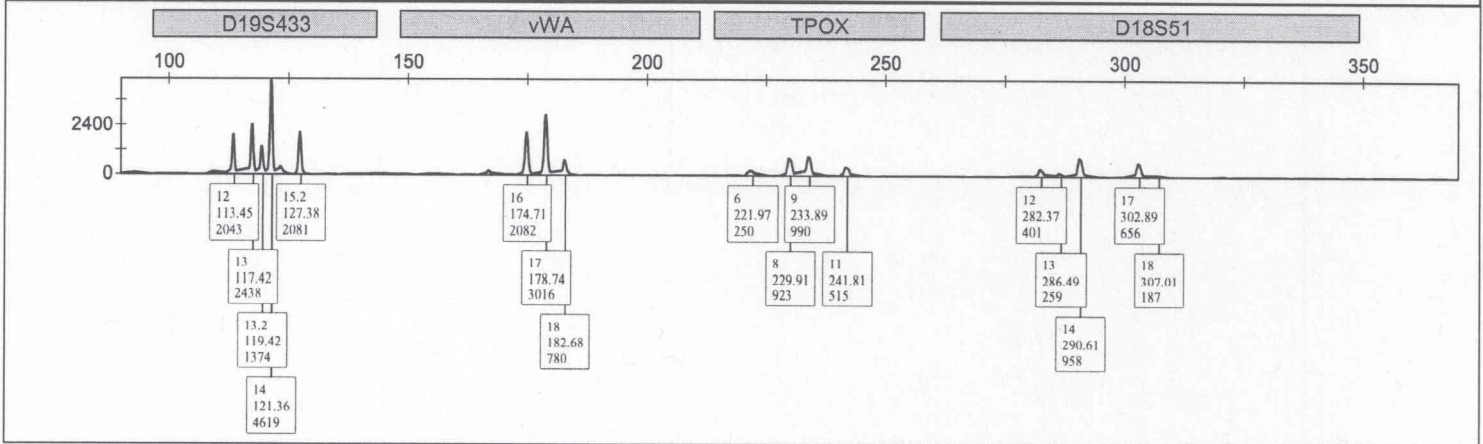
44-Study 3E 3PC Pen B 60pg abc



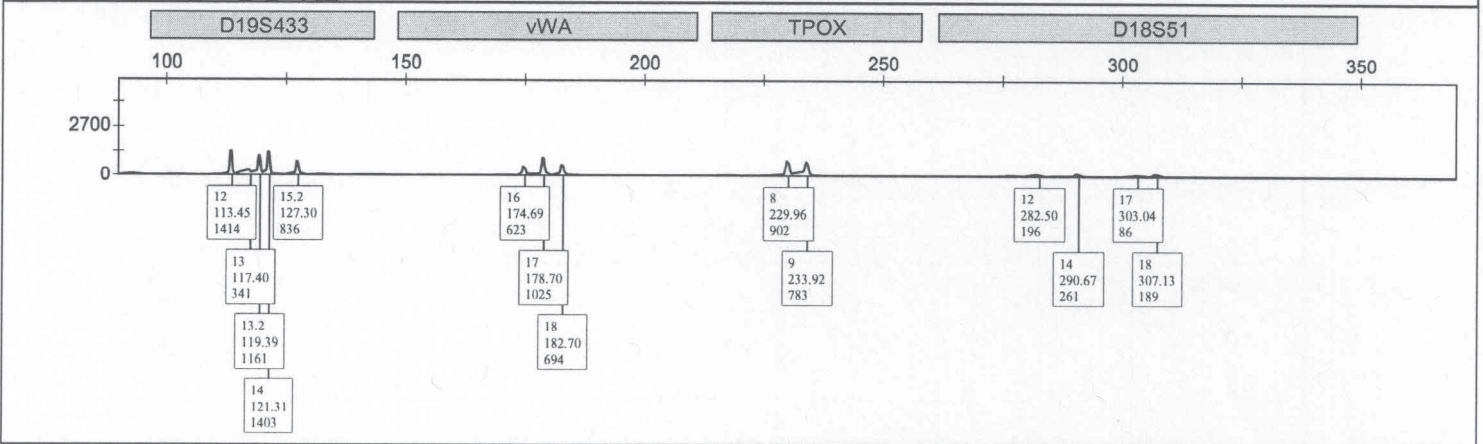


Sample Name

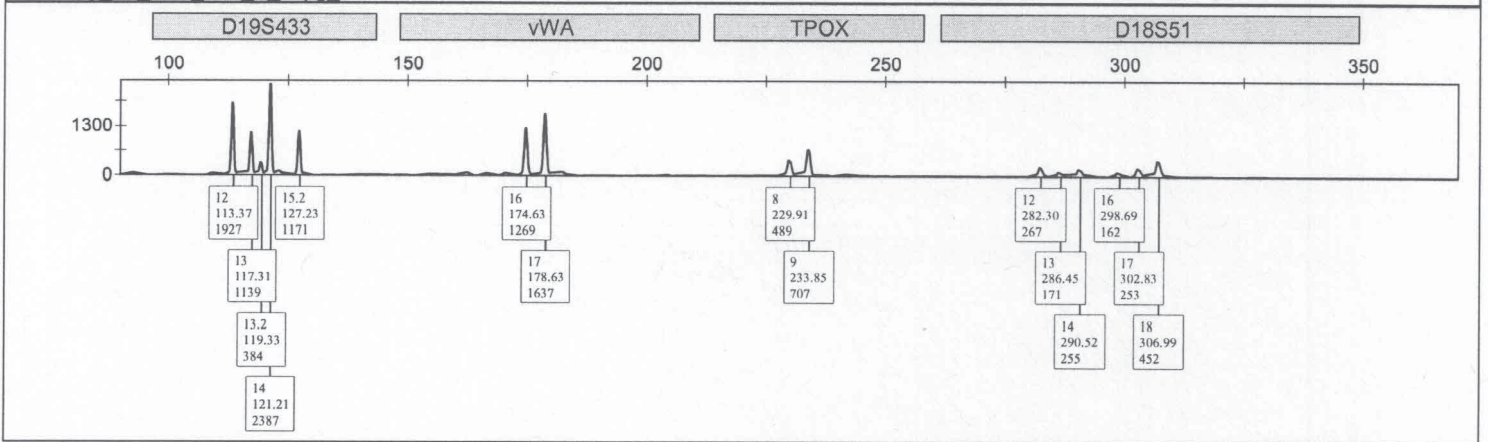
39-Study 3E 3PC Pen B 60pg a



40-Study 3E 3PC Pen B 60pg b



43-Study 3E 3PC Pen B 60pg c

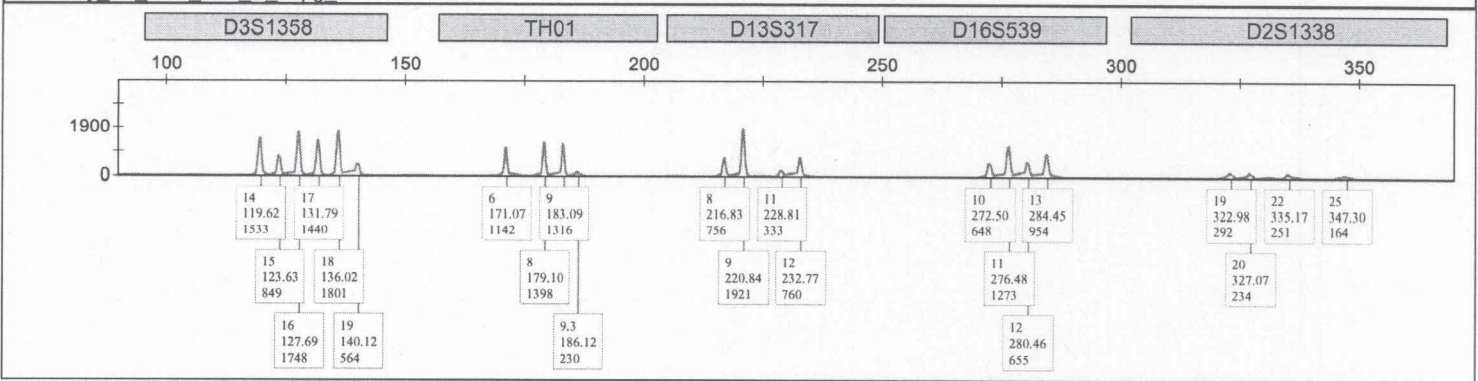


Sample Name

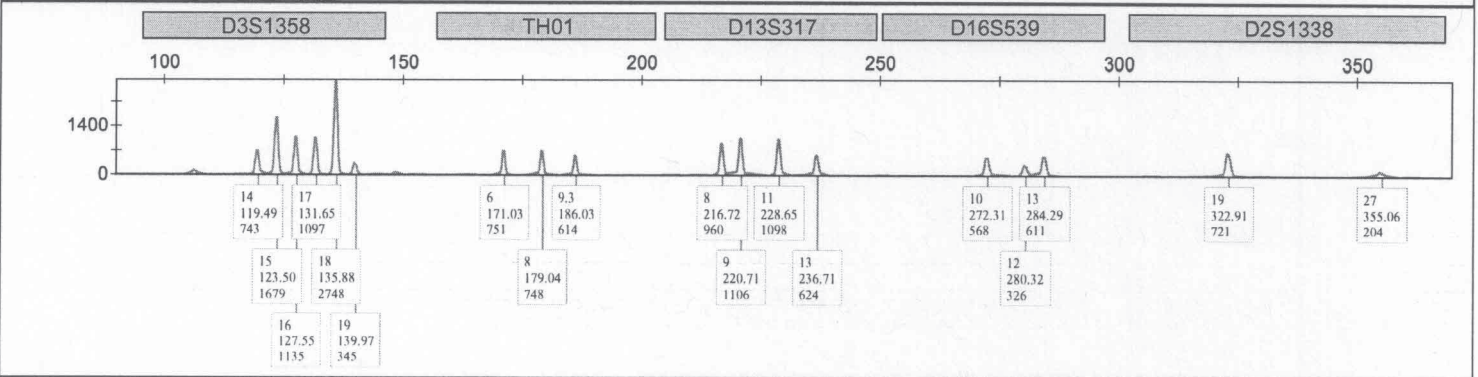
39-Study 3E 3PC Pen B 60pg a



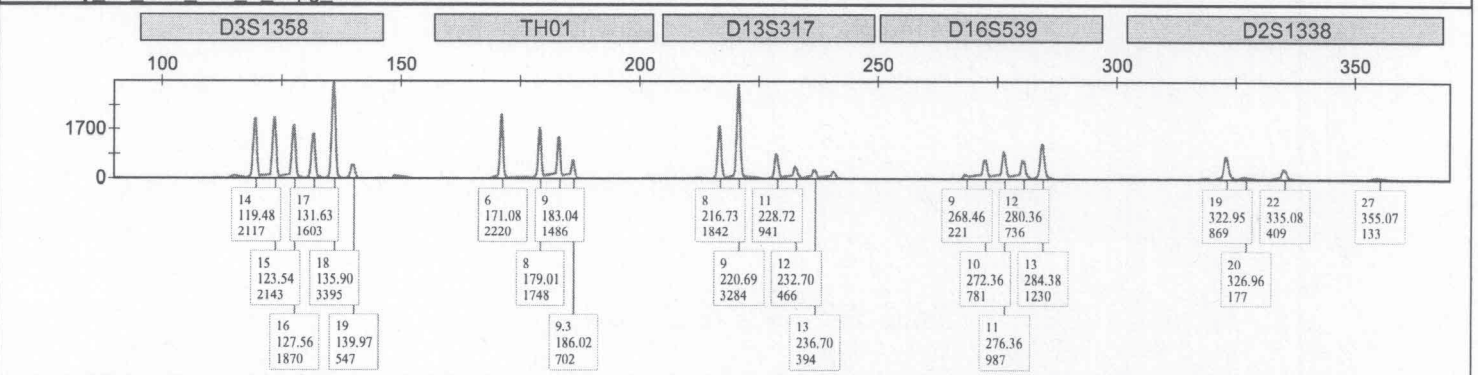
40-Study 3E 3PC Pen B 60pg b

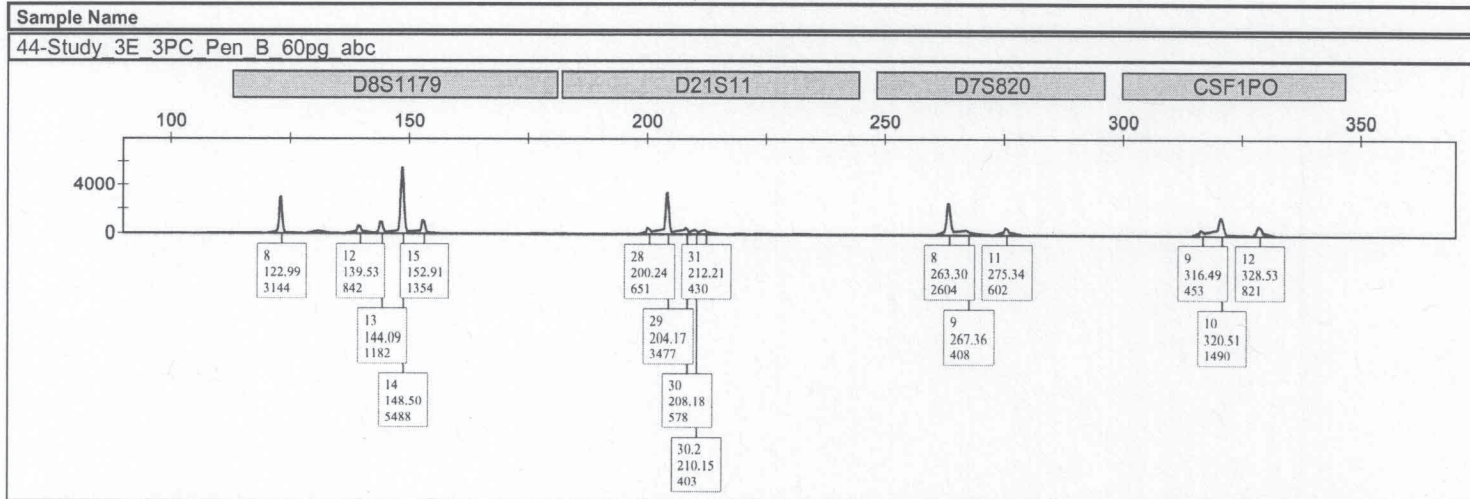


43-Study 3E 3PC Pen B 60pg c



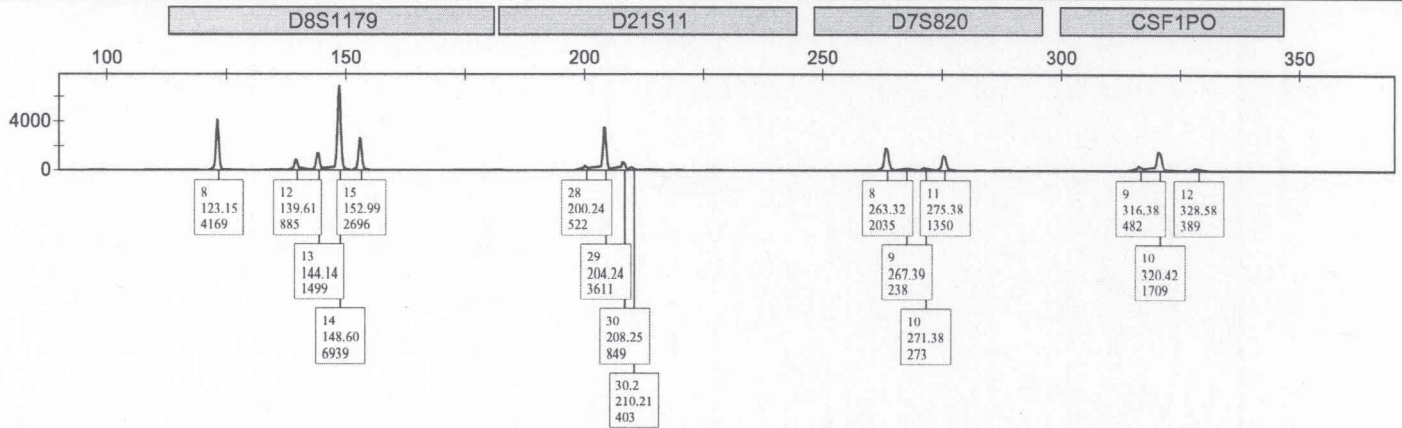
44-Study 3E 3PC Pen B 60pg abc



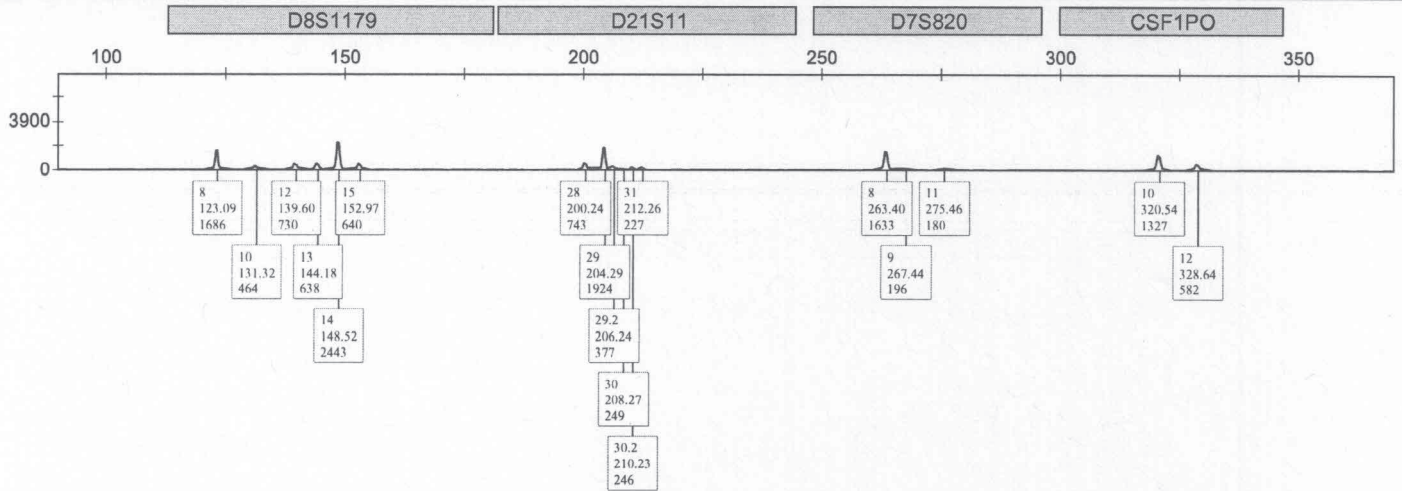


Sample Name

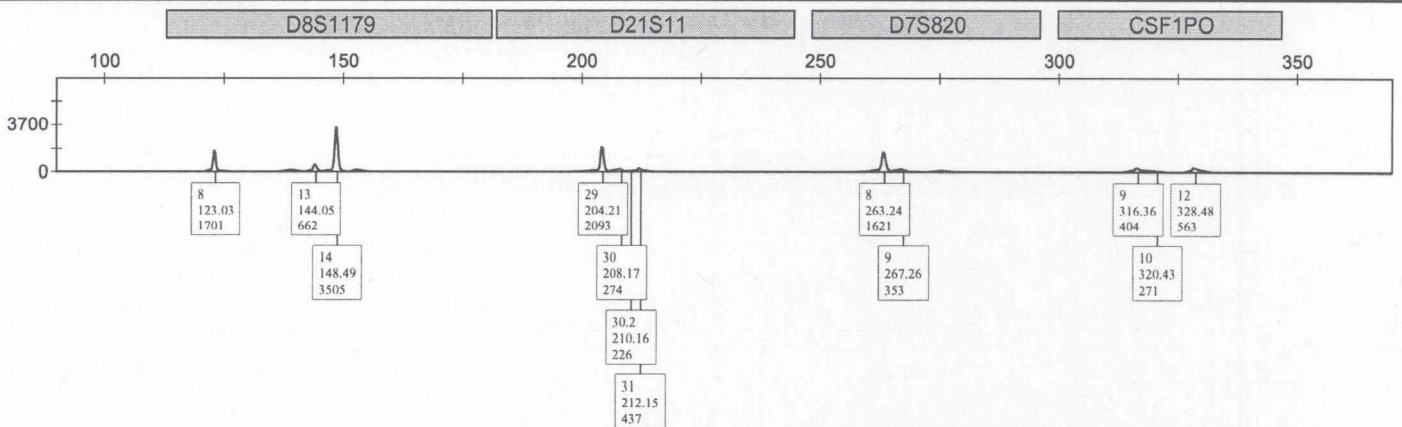
39-Study 3E 3PC Pen B 60pg a



40-Study 3E 3PC Pen B 60pg b



43-Study 3E 3PC Pen B 60pg c



Objective DNA Mixture Information in the Courtroom: Relevance, Reliability and Acceptance

Mark W. Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA

September 30, 2015

Presented in Arlington, Virginia on July 22, 2015 at
National Institute of Standards and Technology conference:

*International Symposium on Forensic Science Error Management:
Detection, Measurement and Mitigation*

Cybergenetics © 2015



Corresponding Author:

Mark W. Perlin, PhD, MD, PhD
Chief Scientific Officer
Cybergenetics
160 North Craig Street, Suite 210
Pittsburgh, PA 15213 USA
(412) 683-3004
(412) 683-3005 FAX
perlin@cybgen.com

ABSTRACT

DNA mixtures arise when two or more people contribute their DNA to a biological sample. Data-simplifying thresholds fail to give accurate results when applied to complex mixture patterns. An entirely objective interpretation approach is to first separate out the genotypes of each mixture contributor, without ever seeing the subject, and only afterwards make a comparison.

Comparison of a separated evidence genotype with a subject's reference genotype, relative to a population, yields a match statistic. This likelihood ratio is a standard measure of information change based on observed evidence that addresses FRE 403 relevancy balancing. The reliability of objective genotype separation has been extensively tested. Such extensive testing, error rate determination, and scientific peer-review address FRE 702 and Daubert reliability factors.

Courts have accepted this extensively validated computer approach, with admissibility upheld at the appellate level. Separated genotypes provide results that juries find easy to understand. Objective DNA analysis elicits identification information from evidence, while rigorous validation establishes accuracy and error rates. Courts require solid science – extensively tested and empirically proven – to promote criminal justice, societal safety, and conviction integrity.

Table of Contents

Introduction.....	4
Case Example	6
Bayes law	6
STR data	7
Genotype separation	8
Relevance and match	9
Match simplification	10
Exclusionary power	11
Case outcome	12
TrueAllele Validation	13
Specificity	14
Sensitivity	14
Reproducibility	15
Reliability	15
Acceptance.....	16
Conclusions	16
References	18
Figures.....	20

Introduction

Deoxyribonucleic acid (DNA) mixtures arise when two or more people contribute their DNA to a biological sample. Mixtures are seen in sexual assault kits, homicide evidence, handguns and other “touch DNA” surfaces. With advances in detection technology, they have become the predominant form of DNA evidence in many crime laboratories. While DNA from one person is easy to interpret, mixture data has complex patterns comprising many allele peaks of varying height.

One person’s DNA produces either one allele peak, or two of similar height, so a height “threshold” is meaningful. But data-simplifying thresholds fail to give accurate results when applied to complex mixture patterns. Ten years ago, scientists at the National Institute of Standards and Technology (NIST) demonstrated a ten order-of-magnitude match statistic discrepancy between crime laboratories analyzing the same mixture data [1]. Mixture “inclusion” analysis tests whether a subject’s alleles are included in a set of (thresholded peak) alleles, but it is inherently subjective – the analyst sees the subject’s genotype during the analysis.

An entirely objective (and potentially more informative) approach is to first separate out the genotypes of each mixture contributor without ever seeing the subject, and only afterwards make a comparison. This can be accomplished by sophisticated computing that considers many thousands of genotype alternatives, and how well their additive combinations explain the quantitative data [2]. Multiple possibilities for a contributor genotype are assigned probabilities. Faithful modeling of the laboratory process can yield genotypes that accurately preserve DNA identification information.

Comparison of a separated evidence genotype with a subject's reference genotype, relative to a population, gives a match statistic. This statistic is a simple ratio – the probability of genotype match divided by the random match probability. The statistic is also a likelihood ratio (LR), or Bayes factor (BF), which is a standard measure of information change based on observed evidence.

Mathematically, the LR is probative because it assesses how evidence data affects a hypothesis (i.e., whether the subject contributed their DNA to the mixture). The LR's assessment is also non-prejudicial, because (as a BF) the ratio factors out prior belief about the hypothesis. Thus genotype separation addresses Federal Rules of Evidence (FRE) 403 relevancy balancing.

The reliability of objective genotype separation has been extensively tested for at least one such system. Dozens of independent and developmental validation studies have been conducted, with seven peer-reviewed TrueAllele® publications. These studies use the LR as an objective information measure to assess the method's sensitivity (true positives), specificity (false positives) and reproducibility (close numbers). This extensive testing, error rate determination, and scientific peer-review address FRE 702 and Daubert reliability factors.

Courts have accepted this extensively validated computer method, which has withstood Daubert and Frye challenges in six states. Admissibility has been upheld at the appellate level. Separated genotypes provide results that are easy to understand.

Objective DNA analysis elicits identification information from evidence. Validation establishes accuracy and error rates. Courts require solid science – extensively tested and empirically proven – to promote criminal justice, societal safety, and conviction

integrity. This paper describes DNA mixtures, and how to objectively interpret them, focusing on relevance, reliability, and acceptance.

Case Example

We examine DNA mixture evidence in a Baltimore trial of Nelson Clifford; the author was an expert witness for the prosecution. Arguing consent, Clifford had been acquitted of sexual offenses on four previous occasions [3]. In this fifth case, mixtures were found on articles of clothing – a green shirt and a belt. The forensic question was: “Did suspect Nelson Clifford contribute his DNA to the victim’s clothing?”

A mixture sample contains DNA from two or more people. Figure 1 shows a relatively large amount of DNA from one person (blue) who has a 6,8 allele pair, a second person (orange) who is homozygous for allele 7, and a third person (green) with a 7,9 allele pair. The additive combination of these relative DNA amounts produces a data signature for this particular biological mixture.

Bayes law

Bayes law lets us reach meaningful conclusions from a small amount of data. Bayes uses this data to update belief. The probability law is 250 years old [4], but has gained considerable traction in the last 50 years with the advent of digital computing [5].

Bayes begins with a prior probability (brown, right side) of what we believe before we see data (Figure 2). We examine data through a likelihood function that describes

how well a hypothesis explains the data, giving a probability number (green, middle). All hypotheses are considered, determining how the data updates our belief (blue, left). The result is a posterior probability, our final belief after we have observed the data.

Genotype modeling is the application of Bayes law to genetic identification (Figure 3). We begin with a random genotype (brown, right) of probabilities for about 100 different allele pair possibilities at each locus. The quantitative data is then examined, usually for short tandem repeat (STR) data [6].

A computer considers all genotype possibilities, along with variables such as stutter, degraded DNA, variances, and other parameters. After examining the data, we derive a new genotype probability. This result represents our belief in the different genotype values for each contributor at every genetic locus.

STR data

Bayesian analysis starts with the data. We have STR genetic data comprised of quantitative peak heights, shown for the green shirt mixture at locus TH01 (Figure 4). There is a pattern of taller peaks at alleles 6 and 8, and lower peaks at 7 and 9.

It is important to use all of the data. Specifically:

- (a) The *amounts* of the DNA matter, expressed as peak heights that reflect the relative quantities of each allele in the biological sample.
- (b) The *pattern* of high and low peaks matter, as these patterns can be explained by different genotype hypotheses of allele pair quantities and their artifacts.
- (c) The peak *variation* is needed for modeling variance parameters; there can be

dozens of these parameters in a DNA mixture problem. For example, the 6 and 8 peaks here represent roughly the same amounts of DNA contributed by one person, but we see variation in their (unequal) peak heights.

Genotype separation

A likelihood function helps separate out the genotypes of each contributor to a mixture. The likelihood explains the genotyping data mathematically. Shown is one such explanation, out of many thousands that were considered (Figure 5). There is a major amount of a first 6,8 allele pair (blue), a minor amount of a second homozygote allele pair at allele 7 (orange), and a minor amount of DNA for third allele pair 7,9 (green).

Adding up these three different allele pairs forms a pattern, where the heights of those cumulative allele quantities are (to a first approximation) near the peak heights of the observed data. Since this pattern explains the data well, it has a relatively high likelihood and thus confers higher probability to each of the contributor genotypes.

A separated contributor genotype is shown in Figure 6. The locus vs. contributor table (center) lists 13 genetic loci, with TH01 in the first row, followed by another 12 loci. Each of the three assumed contributors has a separate column. There are thus 39 locus contributors ($13 \text{ loci} \times 3 \text{ contributors}$), each with its own separated genotype.

The bar graph (blue) shows one such genotype, here for a minor contributor at the TH01 locus. Out of a hundred or so possible TH01 allele pairs, the STR data has focused probability onto about a half dozen of these possibilities (x-axis). The probability scale is also shown (y-axis). Each bar gives the posterior probability of seeing an allele

pair (for this minor contributor at TH01), after having seen the STR mixture data.

This objective genotyping procedure is unbiased by the suspect's genotype; the computer is not given that reference information, only the mixture data. Moreover, the process is unbiased by a human analyst subjectively selecting data peaks. Data is entered into a machine, and then analyzed automatically. This mechanization facilitates workflow and productivity, but also ensures objectivity.

This process infers a separated genotype for each contributor at every locus. These objectively derived mixture genotypes are recorded on a computer's hard drive. We can now use these separated genotypes to calculate a DNA match statistic, relative to the suspect.

Relevance and match

Our forensic comparison goal is to assess the strength of match. We consider FRE 403, which governs the relevance of evidence. We want to assess the identification hypothesis "Did the suspect contribute his DNA to the mixture?" The legal role of relevance is to balance the probative force of DNA evidence against the danger of unfair prejudice to the defendant (Figure 7).

The likelihood ratio conducts this balancing mathematically. The LR is a form of Bayes theorem for a single hypothesis [7]. It quantifies the question "To what extent does the evidence increase or decrease strength in the identification hypothesis?"

The LR has a numerator (blue) that measures the extent to which the hypothesis is impacted upon by data. This numerator is inherently *probative*, since it centers on

how evidence affects the hypothesis. The denominator (brown) states the initial *prejudicial* odds of the identification hypothesis before seeing data. In dividing numerator by denominator, the LR factors out the prior prejudice from the evidentiary probative force.

After applying Bayes theorem and some algebra, we can calculate the likelihood ratio through genotype posterior probability [8]. At the defendant's genotype, we simply divide the probability *after* having seen data by the probability *before* seeing data. That is how genotypes give us match statistics. They provide a way of using DNA data to calculate a likelihood ratio for the identification hypothesis.

Match simplification

Separated genotypes are much easier to understand than unmixed STR data. With a separated genotype, mixture comparison is like random match probability (RMP), the standard DNA statistic involving just one genotype. We ask the question, "To what extent does the evidence match the suspect more (or less) than a random person?"

The graph in Figure 8 shows the same *posterior* genotype probability distribution (blue bars) as before – the separated contributor at the TH01 locus after the data has been seen. Now also shown (brown bars) are a half dozen (out of a hundred) allele pair possibilities having *prior* genotype probabilities for a random person in the population – the prior gives the chance that we are seeing a match purely by coincidence.

With these posterior (blue) and prior (brown) genotype probabilities, we can make a statistical comparison with anyone's genotype. In this case, the genotype of the

defendant happens to be a 7,9. We therefore focus our attention on that allele pair (red rectangle), looking at the ratio of posterior (blue bar) to prior (brown bar) probability at 7,9. This ratio of 47% to 13% equals 3.62, the value of the likelihood ratio at TH01.

The LR is the posterior genotype probability at the suspect's genotype, divided by the probability of a coincidence. We see that the numerator's 47% is less than the full 100%. A 100% numerator over a 13% denominator would be the simple RMP match statistic. But a DNA mixture introduces match uncertainty, so we must consider that reduced strength of match in the numerator, in addition to the usual genotype rarity in the denominator. Using separated genotypes, the LR is just the old RMP but with a reduced numerator; this idea is easy to understand and explain in court.

The match statistic is shown for each locus by a horizontal bar (Figure 9). The 13 loci are listed from top to bottom. Since STR genetic loci are independent, we can multiply these values together to calculate the joint LR. Stated in plain language, a match between the shirt and Nelson Clifford is 182,000 times more probable than coincidence.

Exclusionary power

Also of interest is the exclusionary power of a matching genotype. Comparing the contributor genotype (over all loci) with 10,000 random genotypes, we obtain a bell shaped curve of match statistics (Figure 10). This non-contributor distribution describes the match information (on a logarithmic scale) for someone who did *not* contribute their DNA to the mixture. The logarithmic mean is around -10 , for an average exclusionary

power of 1 over 10 billion; for a non-contributor, a coincidence is far more probable than an evidence-based match. The standard deviation (yellow bar) is around three log units.

From this non-contributor distribution (Figure 11), we can calculate an error rate for the match statistic (purple math). The LR is 182 thousand, which has a $\log_{10}(\text{LR})$ of 5.25. The normal distribution's z-score for this $\log(\text{LR})$ value is 5.02, or five standard deviations to the right (yellow bar). That deviation has a p-value tail probability of 2.53×10^{-7} . Therefore, the chance of observing a non-contributing individual with a LR of at least 182 thousand (i.e., a false inclusion) is 1 in 4 million.

Case outcome

Figure 12 shows a separated DNA mixture. TrueAllele separated the green shirt mixture into three genotypes: 11%, 82% and 7% contributors. These genotypes were objectively inferred, without examination bias from the suspect or some other reference. Following genotype separation, comparisons were made to three references (victim, elimination and Clifford), yielding match statistics to each of the three mixture contributors.

In this fifth Clifford case, the jury convicted him of third degree sex offense [9]. “Only DNA connected Clifford to the masked man who terrorized” his victims [10]. The defendant's prior sex offense was considered when he was sentenced to over 30 years in prison.

TrueAllele Validation

TrueAllele has been extensively validated in dozens of studies conducted by Cybergenetics and crime labs. Four peer-reviewed studies were performed on laboratory-synthesized data of known composition – mixtures that are made in the laboratory [11-14]. Three other peer-reviewed studies were done on casework samples, which have more realistic data complexity [15-17]. Both types of studies should be done when thoroughly validating a DNA mixture interpretation method.

A recent TrueAllele validation paper appeared in the *Journal of Forensic Sciences*. The study was conducted with co-author Kevin Miller in collaboration with the Kern County crime laboratory in Bakersfield, California. Entitled “TrueAllele genotype identification on DNA mixtures containing up to five unknown contributors,” the study employed a realistic randomized mixture design.

The Kern paper reported seven main results. The “contributor sufficiency” axis examined how changing the computer’s assumed number of contributors affects the match statistic. This axis showed that once there are a sufficient number of assumed unknown contributors, TrueAllele’s match statistic does not materially change.

For example, suppose there are actually three contributors in a DNA mixture. When the computer conducts separate runs assuming three, four, or five unknown contributors, the statistical match results will be essentially the same. Therefore, TrueAllele does not need to know the true number of contributors. Three other axes of interest were specificity, sensitivity, and reproducibility.

Specificity

Specificity validation studies are helpful in court. Figure 13 shows the distribution of log(LR) values for comparisons made between separated mixture genotypes and random genotypes. Millions of genotype comparisons were made, and the log(LR) values were recorded. The mixtures contained 2, 3, 4 or 5 unknown contributors.

The LR data are shown on a logarithmic scale. Zero log(LR) means there is no information (blue vertical line). As the number of contributors increases (from 2, to 3, 4 or 5), specificity (or exclusionary power) decreases. With five contributors in low-template DNA, the average is over one in a billion.

Specificity data can be used to develop a table of false positive events, as was done in this validation study. The table provides false inclusion error rate information. When asked in court, “What is the chance of seeing a false inclusion when the match statistic is a thousand?” the response can be an accurate numerical estimate. With log(LR) non-contributor data collected and tabulated, the error rate becomes a definite probability, whether one in a thousand or one in a trillion.

Sensitivity

Sensitivity examines to what extent a method can detect someone who actually contributed DNA to a mixture. As we increase mixture complexity from two to five contributors, the contributor distribution shifts leftwards towards less identification information (Figure 14). However, even with five contributors, and very low DNA

quantities, TrueAllele successfully made most of the identifications.

Reproducibility

We assessed TrueAllele reproducibility by running the program twice on the same data under the same conditions. Each point on the scatterplot in Figure 15 shows $\log(\text{LR})$ values from two independent computer runs on one mixture. The points line up nicely along a 45° angle, showing that the replicated numbers are essentially the same.

Reproducibility was measured using a within-group standard deviation statistic, and found to be well under a log unit, regardless of DNA quantity or contributor number.

Reliability

Reliability is important for the admissibility of scientific evidence. Expert evidence should be based on reliable methods that have been reliably applied to sufficient data. Daubert admissibility factors include whether a method is testable, has an associated error rate, has undergone peer-review, and is generally accepted in the relevant scientific community. The Frye standard considers only general acceptance.

TrueAllele has been admitted after Daubert challenge in Louisiana and Ohio. The system has withstood Frye challenges in California, New York, Pennsylvania and Virginia. Internationally, TrueAllele has successfully weathered “voir dire” challenges in Australia and the United Kingdom.

Acceptance

TrueAllele acceptance is widespread. Judicial acceptance has been facilitated by validation studies. The first TrueAllele case was tried six years ago in Pennsylvania, which led to an appellate precedent in that state [18].

TrueAllele has since been used in hundreds of criminal cases, and in over half of the states in the United States. TrueAllele experts appear mainly for the prosecution, but also testify for the defense. Five crime labs now regularly use TrueAllele in their criminal casework, with California having started in 2013 [19].

The main impact of TrueAllele is in bringing DNA evidence back into criminal cases. Past and current crime laboratory interpretation guidelines discard most mixtures as “inconclusive,” or assign weak statistics. This information loss precludes the evidence from being heard in court. TrueAllele restores mixtures as viable DNA evidence, with guilty pleas a common outcome.

Conclusions

Objective genotyping can help eliminate examination bias. When a calculator doesn't know the comparison profiles, interpretation can't be directed toward a desired answer. After separating out genotypes from a mixture, they can be compared against any number of people (one, two, ten, or a entire database).

Identification information (the likelihood ratio logarithm) is a standard information statistic. The $\log(\text{LR})$ quantifies DNA information in a case, as well as in a validation

study. The LR condenses the many aspects of genotype comparison into a single number. Scientific LR validation can help establish accuracy, applicability, and error rates. These assessments aid in understanding DNA mixture evidence, and how to use or explain it in court.

There are untested mixture interpretation methods. For example, the manual combined probability of inclusion (CPI) method has enjoyed widespread use for 15 years [20]. CPI is a probabilistic genotyping approach based on a very simple likelihood function, one that does not make much use of the data [21]. CPI accuracy has not been assessed, even though its reliability has been questioned [22,23]. Validation is needed to demonstrate CPI's relevance and reliability.

Courts need solid forensic science that has been empirically proven. Untested DNA mixture statistics should not be offered as reliable evidence. With objective and reliable science, better data interpretation achieves better criminal justice, helping to protect society and maintain conviction integrity.

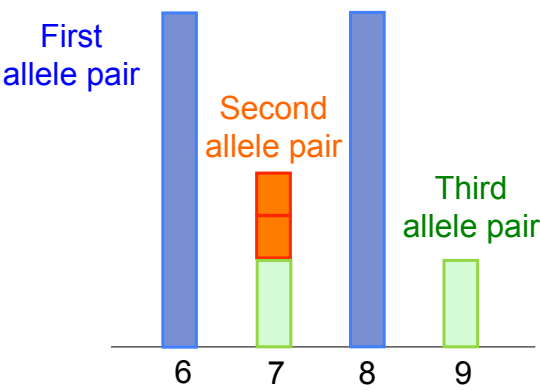
References

1. Butler JM, Kline MC (2005) NIST Mixture Interpretation Interlaboratory Study 2005 (MIX05), Poster #56. Promega's Sixteenth International Symposium on Human Identification. Grapevine, TX.
2. Perlin MW, Szabady B (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J Forensic Sci* 46: 1372-1377.
3. Anderson J, Duncan I (2013) Man charged in multiple sexual assault cases acquitted a fourth time. *The Baltimore Sun*.
4. Bayes T, Price R (1763) An essay towards solving a problem in the doctrine of chances. *Phil Trans* 53: 370-418.
5. O'Hagan A, Forster J (2004) Bayesian Inference. New York: John Wiley & Sons.
6. Butler JM (2005) Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. New York: Academic Press.
7. Good IJ (1950) Probability and the Weighing of Evidence. London: Griffin.
8. Perlin MW (2010) Explaining the likelihood ratio in DNA mixture interpretation. Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.
9. George J, Fenton J (2015) Jury convicts sex offender in fifth trial. *The Baltimore Sun*.
10. Fenton J (2015) Repeat sex offender apologizes, gets 30 years in prison. *The Baltimore Sun*.
11. Perlin MW, Sinelnikov A (2009) An information gap in DNA evidence interpretation. *PLoS ONE* 4: e8327.
12. Ballantyne J, Hanson EK, Perlin MW (2013) DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. *Sci Justice* 53: 103-114.
13. Perlin MW, Hornyak J, Dickover R, Sugimoto G, Miller K (2014) Assessing TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors (poster). AAFS 66th Annual Scientific Meeting. Seattle, WA: American Academy of Forensic Sciences.
14. Greenspoon SA, Schiermeier-Wood L, Jenkins BA (2015) Establishing the limits of TrueAllele® Casework: a validation study. *J Forensic Sci* 60: 1263-1276.
15. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, et al. (2011) Validating TrueAllele® DNA mixture interpretation. *J Forensic Sci* 56: 1430-1447.
16. Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® Casework validation study. *J Forensic Sci* 58: 1458-1466.
17. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE* 9: e92837.
18. (2011) Commonwealth of Pennsylvania v. Kevin James Foley. Superior Court of Pennsylvania.
19. Perlin MW, Miller KW (2014) Kern County resolves the DNA mixture crisis. *Forensic Magazine* 11: 8-12.

20. Scientific Working Group on DNA Analysis Methods (SWGDAM) (2000) Short Tandem Repeat (STR) interpretation guidelines. Forensic Science Communications 2.
21. Perlin MW (2010) Inclusion probability is a likelihood ratio: implications for DNA mixtures (poster #85). Promega's Twenty First International Symposium on Human Identification. San Antonio, TX.
22. Curran JM, Buckleton J (2010) Inclusion probabilities and dropout. J Forensic Sci 55: 1171–1173.
23. Brenner CH (2011) The mythical “exclusion” method for analyzing DNA mixtures – does it make any sense at all? (A111). AAFS 63rd Annual Scientific Meeting. Chicago, IL: American Academy of Forensic Sciences. pp. 79.

1. DNA mixture

Two or more people contribute their DNA to a sample



2. Bayes law

Use data to update belief (1762)

$\text{Prob}(\text{hypothesis} \mid \text{data})$ proportional to $\text{Prob}(\text{data} \mid \text{hypothesis}) \times \text{Prob}(\text{hypothesis})$

New belief, after seeing data	How well hypothesis explains data	Old belief, before seeing data
<i>posterior</i>	<i>likelihood</i>	<i>prior</i>

3. Genotype modeling

Apply Bayes law to genetic identification

$\text{Prob}(\text{genotype} \mid \text{data})$ proportional to $\text{Prob}(\text{data} \mid \text{genotype}) \times \text{Prob}(\text{genotype})$

New genotype
probability,
after seeing data
posterior

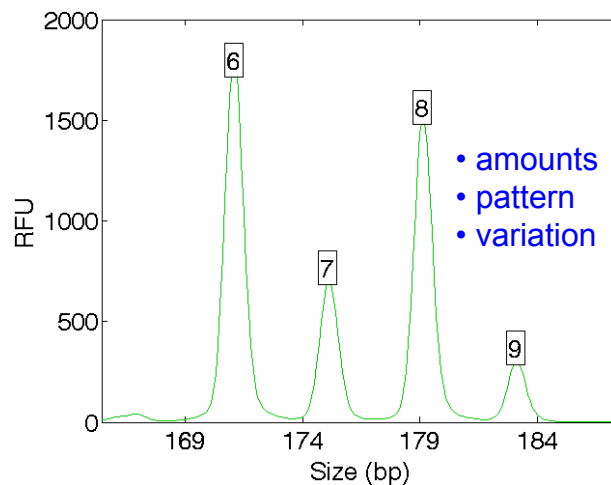
How well
genotype choice
explains data
likelihood

Old genotype
probability,
before seeing data
prior

Probabilistic genotyping

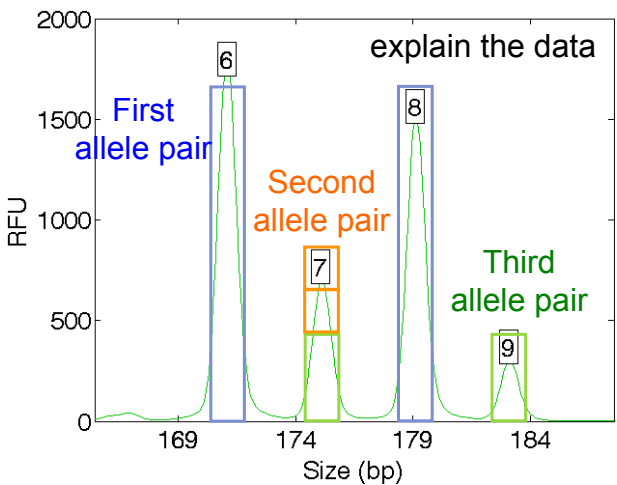
4. Genetic data

Quantitative peak heights at locus TH01



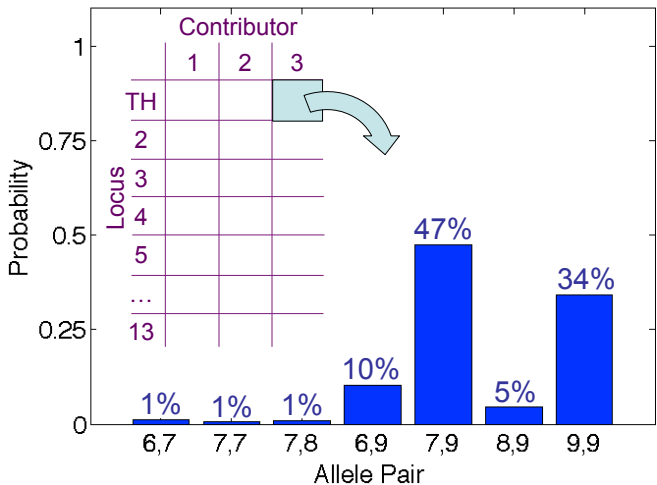
5. Separate genotypes

Consider every possible genotype (Bayes)



6. Separated genotype

Objective, unbiased – doesn't know suspect's genotype



7. Relevance (FRE 403)

Hypothesis = "suspect contributed his DNA"



likelihood ratio (LR)
is Bayes law
for a hypothesis

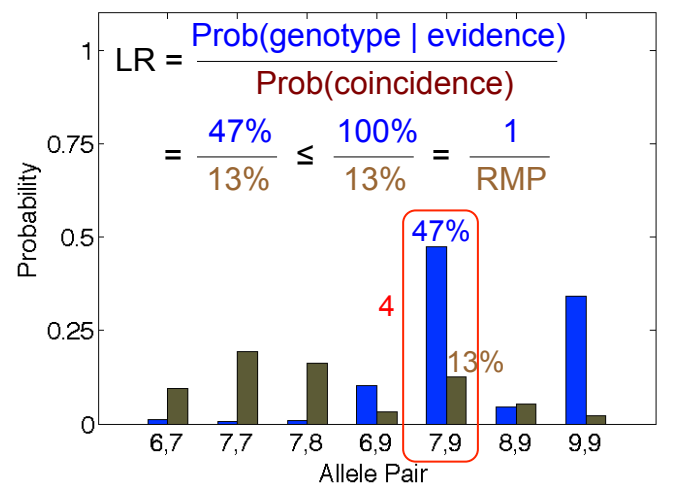
Probative

$$LR = \frac{\text{Odds(hypothesis | data)}}{\text{Odds(hypothesis)}} = \frac{\text{Prob(genotype | data)}}{\text{Prob(genotype)}}$$

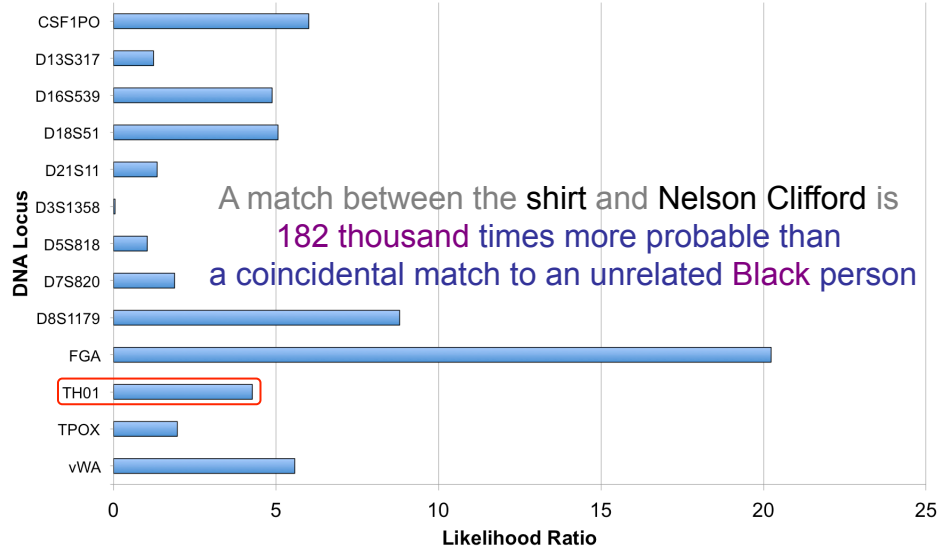
Non-prejudicial

8. Match statistic is simple

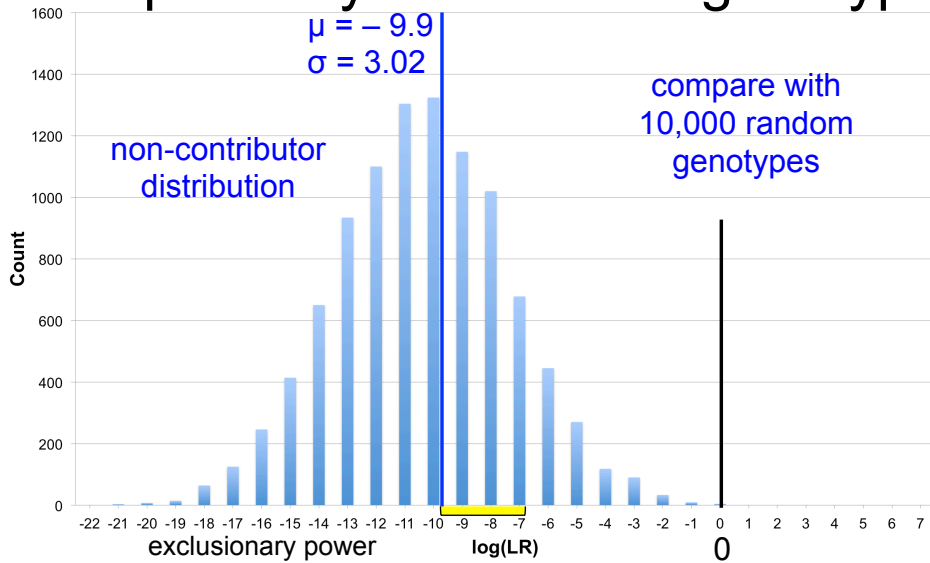
Suspect matches **evidence** more than **random** person



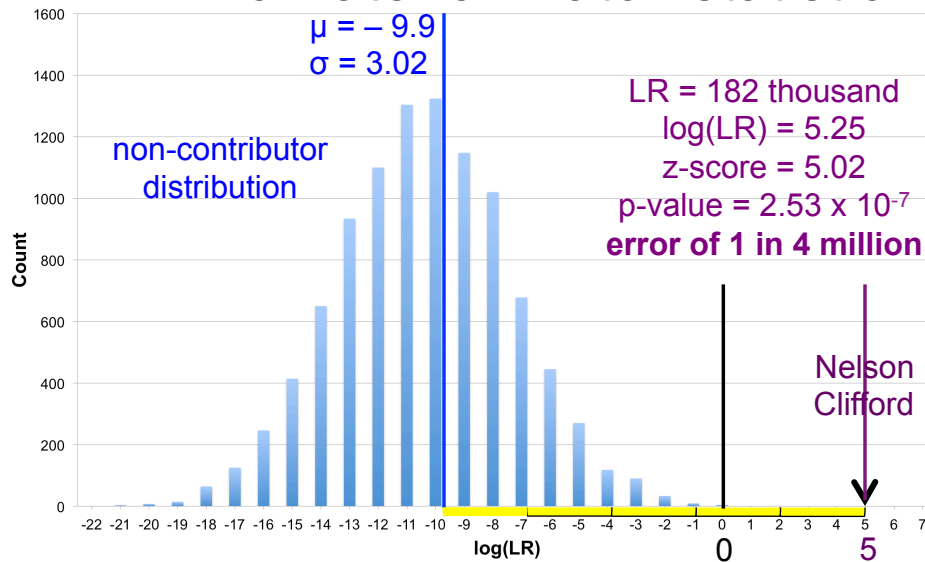
9. Match statistic at all loci



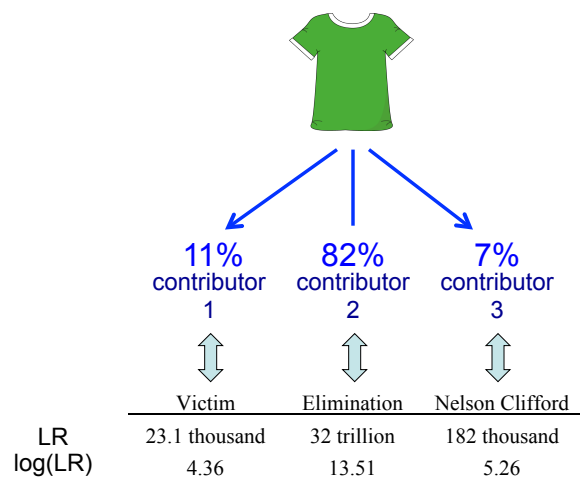
10. Specificity of evidence genotype



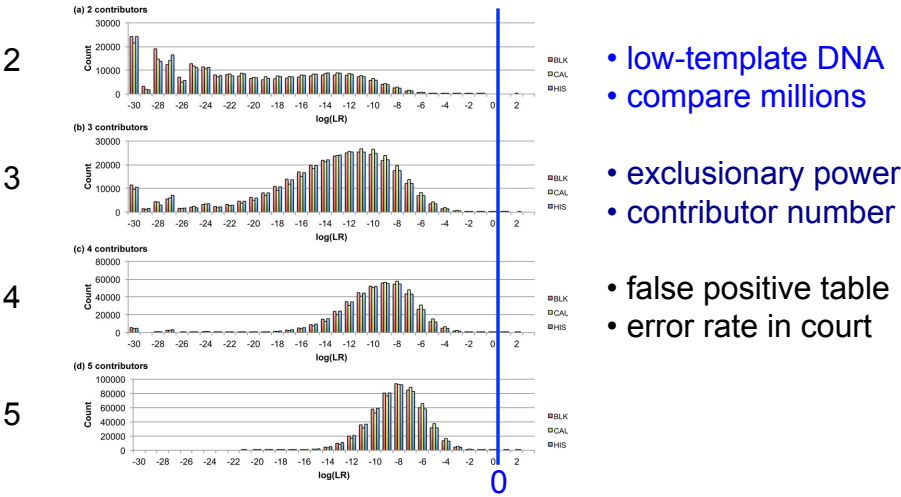
11. Error rate for match statistic



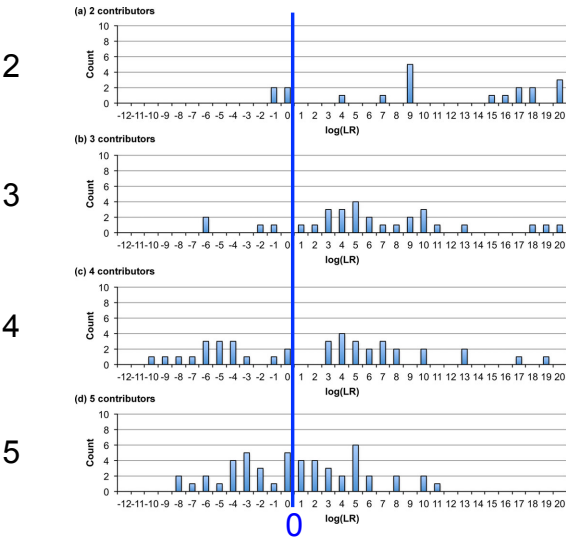
12. Separated DNA mixture



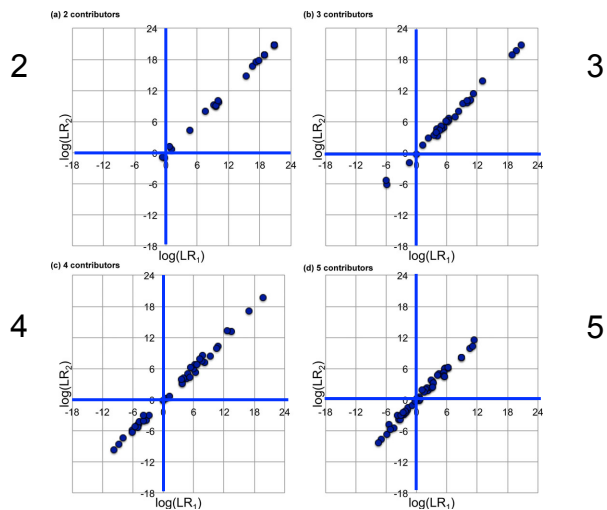
13. Specificity



14. Sensitivity



15. Reproducibility



QUALITY ASSURANCE AND QUALITY CONTROL IN BREATH ALCOHOL MEASUREMENT PROGRAMMES USING INTOXILYZERS SUCH AS THE INTOXILYZER 8000C.

By S. Krishnan, Ph.D., P.Eng, (Adjunct Professor, Retired, Faculty of Engineering, University of Toronto and Forensic Scientist Practitioner), Toronto, Canada.

INTRODUCTION

For the last several years Intoxilyzers, particularly Intoxilyzer 5000 and 8000 in the United States and corresponding Models 5000C and 8000C in Canada, are used in the determination of breath alcohol concentration (BAC) of persons accused of the Driving Under the Influence (DUI) offenses. Other similar instruments are also in use in some States of the USA and the Provinces of Canada. Although these instruments are reasonably accurate for the purposes for which they are used, one has to be careful and vigilant that errors do not occur in the field. This is because, these instruments are predominantly operated by police officers and other law enforcement officials who are not necessarily scientists. They operate the instruments, after a short training, according to a protocol issued by the program managers.

The issues discussed in this paper and the remedies suggested are applicable to all breath testing instruments and programs.

In Canada the offenses of driving an automobile while under the influence of alcohol (DUI) or while having a blood alcohol concentration (BAC) of over 80 milligrams of alcohol in 100 milliliters of blood (80 mg%) are criminal code offenses. The consequences of a conviction are quite serious.

Prior to the year 2008, Canadian law presumed that the BAC measured at the police station is accurate unless there is evidence to the contrary. Among the accepted evidence to the contrary, was the evidence from the accused person about the drinking pattern on the day in question. That drinking scenario should first be accepted as reliable by the Court. It should then be followed by expert evidence placing the accused with a calculated BAC of below 80 mg% at the time of the offense. This defense, known as the Carter defense was successfully used in many DUI cases involving 'over 80' charges. It was not required by the defense to prove where the instrument went wrong in providing a reading above the 80 mg% limit. Therefore, the defense seldom asked for extensive documentation to prove that the instrument malfunctioned or was operated improperly.

In 2008 a Canadian Law known as C-2 was passed which considers the BAC readings obtained at the police station as conclusive evidence of the BAC. The Carter evidence of the drinking pattern provided by the accused is no longer acceptable as a defense. That the BAC readings are accurate can be rebutted by the defense mainly by establishing that the instrument malfunctioned or was operated improperly (1). For this, the documentation has to be obtained from the prosecution which include information concerning the maintenance and operation of the instrument (2). These documents are usually generated and maintained under the Quality Assurance (QA) and Quality Control (QC) protocols of the breath testing program.

QUALITY ASSURANCE AND QUALITY CONTROL

Quality Assurance (QA) is a comprehensive ongoing program in which systematic procedures are put in place to ensure that the entire testing process is effective and the test results are accurate and reliable. Under the QA system, the Quality Control (QC) is the set of procedures and tests to assure that the measurement output viz. the accused's breath alcohol concentration is reliable and accurate (3).

In the United States, QA and QC issues are designed, handled and controlled by individual States. In Canada it is uniform across the country since the Criminal Code is applied nationally. Here, these procedures are designed by the Alcohol Test Committee (ATC) of the Canadian Society of Forensic Sciences which is the Advisory Body to the Canadian Government (4).

The components of the QA program include the rules of the testing procedures e.g. the requirements of the air blanks, duplicate tests, and fifteen minute deprivation period to ensure dissipation of mouth alcohol, the maintenance of the instruments and their accessories, the initial training, annual retraining of the Qualified Breath Technician and the overall management of the breath testing program.

In the QC procedure the blank tests, the diagnostics, the calibration checks and the duplicate subject tests are included.

PRESENT SITUATION

The C-2 legislation in Canada has given rise to lengthy litigation. The Crown and defense toxicologists disagree as to what constitutes sufficient and necessary information for defending DUI cases under the new legislation. The Crown toxicologists state that the Intoxilyzer printout of the subject test data is sufficient. The defense toxicologists believe that, additional disclosure documents, not just the aqueous standard calibration data, are necessary (2). The aqueous alcohol standard only checks the instrument's ability to measure pure uncontaminated alcohol. It does not check the instrument for its ability to correct for interferents such as acetone, isopropanol, methanol that may be present in human breath. Also, the calibration check does not use the entire instrument e.g. the pressure and breathe sampling systems, as does the human breath sample. These are the objections raised by the defense in using only the calibration check as the total check on the instrument's reliability.

The Crown toxicologists state that the previous status, earlier maintenance records and other earlier data regarding malfunction of the instrument is irrelevant to establish such difficulties at a later time when the BAC of accused persons are obtained. Yet nothing more than the Intoxilyzer printouts are available from the date of the offense, even after 2008. The system only has older records of repairs, malfunction and maintenance. These are sought by the defense to check that repairs and maintenance were properly carried out.

The procedure of using only an aqueous alcohol standard for a calibration check used in the Carter days has not been changed when the law was changed in 2008. Both in Canada and many regions of the United

States, this simple calibration check is used as confirmation of the proper working of the whole instrument.

RECOMMENDED QA and QC IMPROVEMENTS FOR BREATH ALCOHOL TESTING.

While in Canada and the USA much thought have gone in to ensure that the BAC measured with an accused person is under reasonably good quality assurance (QA) and quality control (QC) protocols, they now need updating and improvement. We can take lessons from many other fields in which similar chemical analysis procedures are used such as in hospital laboratories, government laboratories, the industry and the forensic laboratories. The following updated QA methods are suggested so that the quality of the BAC obtained is of the highest quality when it is presented in the courts. The improved and comprehensive quality control procedures recommended are also obtained at the same time as the accused breath sample is analyzed. This will provide direct and current evidence of the proper functioning of the entire instrument when human breath samples are analyzed for DUI prosecutions. The following are the recommendations:

1. A composite sample of blood pool be developed using either actual or artificial blood containing common interferences such as acetone, isopropyl alcohol and other chemicals which can be present in human breath. This can then be certified by laboratories such as NIST as Certified Reference Material (CRM). This should be made available to all units which conduct breath alcohol analysis. Along with this standard, protocols to produce and analyze the headspace alcohol should be available. Analyses runs of this CRM should be conducted before and after each subject test. If one does not want to use the CRM for every subject test due to expense, in-house secondary composite standards can be developed.

The routine procedure of using the blood CRMs containing both the ethyl alcohol and also the interferences cannot be run using the Intoxilyzer 8000C as is done with other instruments such as gas chromatographs. This is due to the fact that, when an interference is detected by the Intoxilyzer 8000/8000C, it would abort the analysis and will not proceed to determine the BAC. Therefore several CRMs with and without the interferences need to be run. The ones with interferences will indicate if the interference detectors are in proper working order. The ones without the interferences but only alcohol will be useful to check the accuracy of the BAC determination and produce calibration curves. It needs to be studied as to how to use these CRMs effectively to determine if the instrument and all its components are working properly at the time the subject is tested.

2. A proficiency testing procedure is developed which is controlled by the Central quality control administrative unit which will send 'blind' samples to all units periodically to check how each unit is performing. The not so well performing units can then be specially treated to identify the problems and rectify them.
3. A quality assurance unit is centrally and nationally established which gathers QA/QC information and data about the Intoxilyzers and Simulators used in the field. The data gathered should be

openly available to all interested parties and stake holders on a web site. This is to ensure and promote transparency and confidence in the system thereby reducing unnecessary litigation on disclosure issues.

4. A national unit is established to conduct research into why certain unexplained anomalies occur from time to time and provide, explanations, remedies and alterations in instrument components or protocols. For this purpose, the approved instrument should be freely available for purchase by academicians, researchers and private practitioners. Such availability in other fields of chemical analysis, has only helped in the improvement of the performance of the instruments, protocols and the quality of data obtained.

REFERENCES

1. R. v. St-Onge Lamoureux (2012) 3 S.C.R. 187.
2. R. v. Neville, 2013 ONCJ 500.
3. Kurt M. Dubowski, 'Quality Assurance in Breath-Alcohol Analysis', Journal of Analytical Toxicology, Vol. 18, October 1994, p. 306.
4. Recommended Standards and Procedures of the Canadian Society of Forensic Science Alcohol Test Committee. Can. Soc. Forensic Sci. J. 2013: 46(1): p. 1- 23.



2nd Int. Symposium on Forensic Science Error Management Being Planned for July 24-28, 2017