NIST Interagency Report
NIST IR 8510sup1

Communicating Forensic Findings: Current Practices and Future Directions

Workshop Proceedings

Sandra L. Koch

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8510sup1

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8510sup1

Communicating Forensic Findings: Current Practices and Future Directions

Workshop Proceedings

Sandra L. Koch
Special Programs Office
Laboratory Programs

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8510sup1

January 2025

U.S. Department of Commerce
Jeremy Pelter, Acting Secretary of Commerce

National Institute of Standards and Technology
Craig Burkhardt, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

NIST IR 8510sup1
January 2025

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

This publication is intended to capture external perspectives related to NIST standards, measurement, and testing-related efforts. These external perspectives can come from industry, academia, government, and other organizations. This report was prepared as an account of a workshop; it is intended to document external perspectives; and does not represent official NIST positions.

Presentations can be accessed at: https://www.nist.gov/news-events/events/2024/06/communicating-forensic-findings-current-practices-and-future-directions

Publication History
Approved by the NIST Editorial Review Board on 2025-01-15

How to Cite this NIST Technical Series Publication
Koch SL (2025) Communicating Forensic Findings: Current Practices and Future Directions: Workshop Proceedings, *Supplemental Document to Communicating Forensic Findings: A NIST Scientific Foundation Review* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8510sup1.
https://doi.org/10.6028/NIST.IR.8510sup1

Author ORCID iDs
Sandra L. Koch: 0000-0001-5872-2266

Contact Information
sandra.koch@nist.gov

Acknowledgments

The author gratefully acknowledges the workshop planning team (in alphabetical order): Sanne Aalbers, John Butler, Will Guthrie, Hari Iyer, Steve Lund, Melissa Taylor, and Sheila Willis. Special thanks to Donna Ramkissoon and Corrine Lloyd for their help with the planning and logistics of the workshop. This document summarizes the presentations and discussions that occurred during the workshop and would not have been possible without the detailed notes taken by a small group, whose attention to detail was greatly appreciated. Three current George Washington University Forensic Science Program graduate students served as notetakers: Darby Harris, Kelsey Johns, and Holly Zhao. Additional notes were provided by fellow NIST staff Kelly Sauerwein and John Butler, and workshop discussion facilitators Sanne Aalbers, Melissa Taylor, and Sheila Willis. Each presenter was given an opportunity to suggest revisions to clarify the summary of their presentations prior to publication

Abstract

This document supplements a forthcoming Scientific Foundation Review: NIST IR 8510 Communicating Forensic Findings and summarizes the presentations and discussions that occurred during a two-day NIST-hosted workshop held on June 25-26, 2024. The workshop was convened to assess the range of communication approaches currently in use among forensic disciplines, with a focus on how forensic findings are currently expressed in reports and testimony, and what the community can do to improve the communication and understandability of forensic findings. Presentations and discussions were centered on how forensic findings are communicated, the use of likelihood ratios (LRs) and verbal scales among multiple forensic disciplines, the strengths and clarity of these approaches to communicate the weight of forensic evidence, and what gaps there are in research or basic knowledge that may limit an end user's ability to apply or understand the findings. Attendees had an opportunity to provide input on how they thought the forensic science community could improve so that reports and testimony can more clearly convey the significance and limitations of forensic findings. This summary was developed with input from the workshop speakers and provides an overview of the different perspectives expressed by the stakeholders and discipline experts who were present at the workshop. Links to the presenters' slides are included below each summary. NIST plans to use the knowledge gathered during this workshop to inform the Scientific Foundation Review on this topic.

Keywords

forensic science; likelihood ratios; verbal scales; evaluative reporting; science communication

Table of Contents

1. Overview

On June 25 and 26, 2024 NIST convened a workshop to examine how forensic findings are communicated in reports and testimony: how they are communicated by different stakeholders in the criminal justice system, how the information being communicated is understood by recipients, and whether the information is understood in the way it was intended. The meeting was comprised of eight sessions, depicted in Figure 1.

The first day of the workshop explored how forensic findings are currently presented in the United States. The presentations and discussions examined the variety of approaches currently in use or under development, with specific attention paid to likelihood ratios (LRs) and verbal scales, how comprehensible forensic findings are for end users, and best practices for communicating scientific information to diverse audiences. On the second day, presentations and discussions focused on knowledge gaps that may impact an end-user's understanding of the findings, European colleagues shared how they approach forensic reporting and testimony, and attendees engaged in small group discussions sharing a range of perspectives on how the forensic science community can improve communication of forensic findings.

NIST plans to use the insights gained from the presentations and discussions to inform a Scientific Foundation Review on Communicating Forensic Findings. This event was intended for those actively engaged in communicating forensic findings, end users of the information being reported (e.g., legal community), and those who are working to improve the communication of scientific information with data analysis and research.

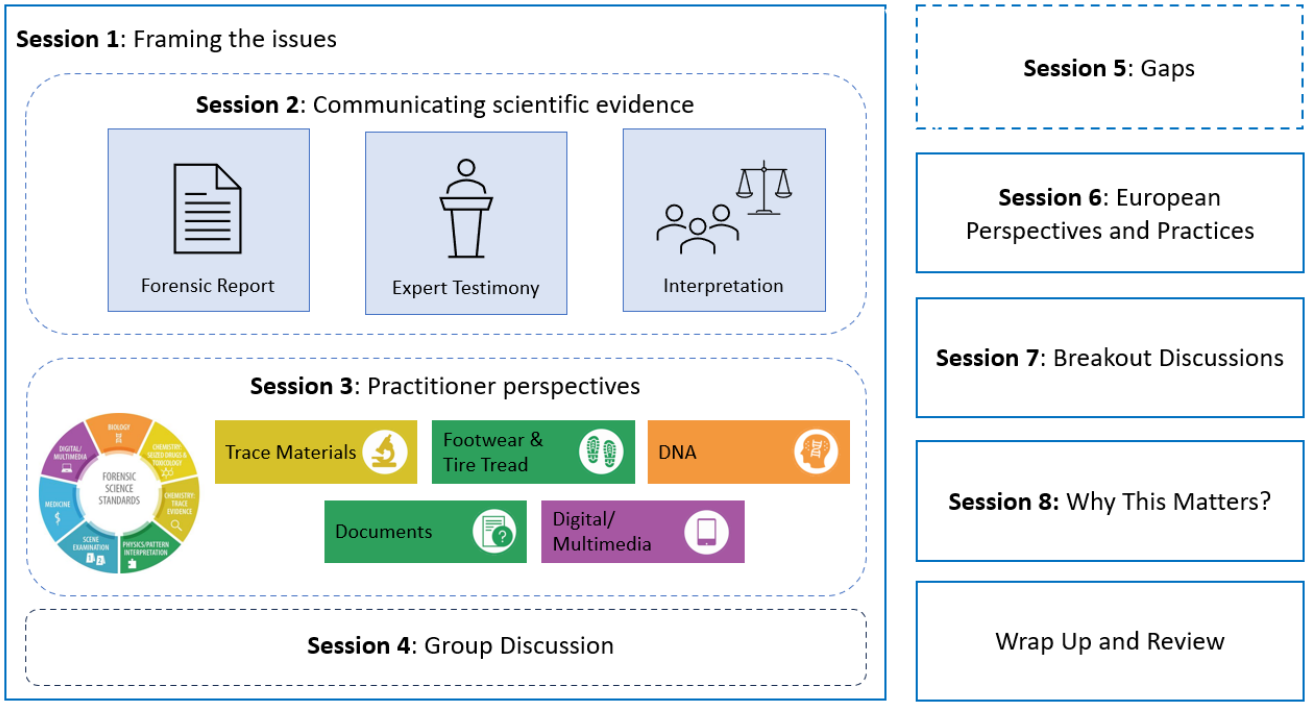

Figure 1: Meeting format (credit: Sanne Aalbers/NIST)

1.1 Agenda

Session 1: **Framing the Issues**

- **Sandy Koch** – NIST (SPO)
- **John Butler** – NIST (SPO)
- **Judge J. Michael Ryan** - DC Superior Court
- **Hal Stern** – UC Irvine/CSAFE
- **Steve Lund/Hari Iyer** – NIST (SED)

Session 2: **Communicating Results in Forensic Reports and Testimony**

Moderator: John Butler – NIST (SPO)

- **Julie Burrill** – Alan Alda Center for Science Communication
- **Heidi Eldridge** – George Washington University
- **Kristy Martire** – University of New South Wales (Australia)
- **Clinton Hughes** – Brooklyn Defender Services

Discussion

Session 3: **Practitioner Perspectives** (LR and OSAC Interpretation Scales)

Moderator: Sandy Koch – NIST (SPO)

- **Jarrah Kennedy** – Kansas City PD Crime Lab, DNA
- **Tatiana Trejos** – West Virginia University, Trace Evidence
- **David Kanaris** – Alaska State Crime Lab, Footwear
- **Lora Sims** – Ideal Innovations, Digital & Multimedia
- **Miriam Angel** – Los Angeles PD Laboratory, Documents

Discussion

Session 4: **Discussion Session part 1**

Moderator: **Melissa Taylor –** NIST (SPO)

Session 5: **Gaps and How to Fill Them**

Moderator: Will Guthrie – NIST (SED)

- **Bill Thompson** – UC Irvine/CSAFE
- **Kate Philpott**- Virginia Commonwealth University
- **Dan Rabinowitz** – Columbia University
- **JD Schmid** – 6th District Public Defender MN

Discussion

Session 6: **European Perspectives and Practices**

Moderator: Steve Lund – NIST (SED)

- **Alex Biederman** – University of Lausanne (Switzerland)
- **Anders Nordgaard** – National Forensic Centre (Sweden)
- **Marjan Sjerps** – Netherlands Forensic Institute (Netherlands)

Discussion

Session 7: **Discussion part 2 (Breakout groups)**

Moderators: **Shelia Willis** – Private Consultant**, Melissa Taylor** – NIST (SPO), **Sanne Aalbers** – NIST (MML)

Session 8: **Why this Matters?**

- **Lynn Garcia** – Texas Forensic Science Commission

Review of Group Findings

- **Sheila Willis** – Private Consultant**, Melissa Taylor** – NIST (SPO), **Sanne Aalbers** – NIST (MML)

Wrap up: **Where to Next? (plans for a NIST Scientific Foundation Review)**

- **Sandy Koch –** NIST (SPO)

1.2 Objectives

This workshop was held to better understand how the descriptive language and numerical analytical results used when communicating forensic findings in reports and testimony are conveyed and understood, the differences in presentation formats among disciplines, and how report information relayed by lawyers or other non-scientists could impact judicial outcomes. The workshop was structured around 8 sessions (Figure 1) to explore how forensic findings are communicated by different stakeholders in the criminal justice system, how the information being communicated is understood by recipients (the eventual users of the information), and whether the findings are understood in the way it was originally intended. Speakers were asked to consider how findings are explained in a report, what is presented in court, how the findings are understood by recipients (end users) (Figure 2) as well as what is missing, and what can be improved.

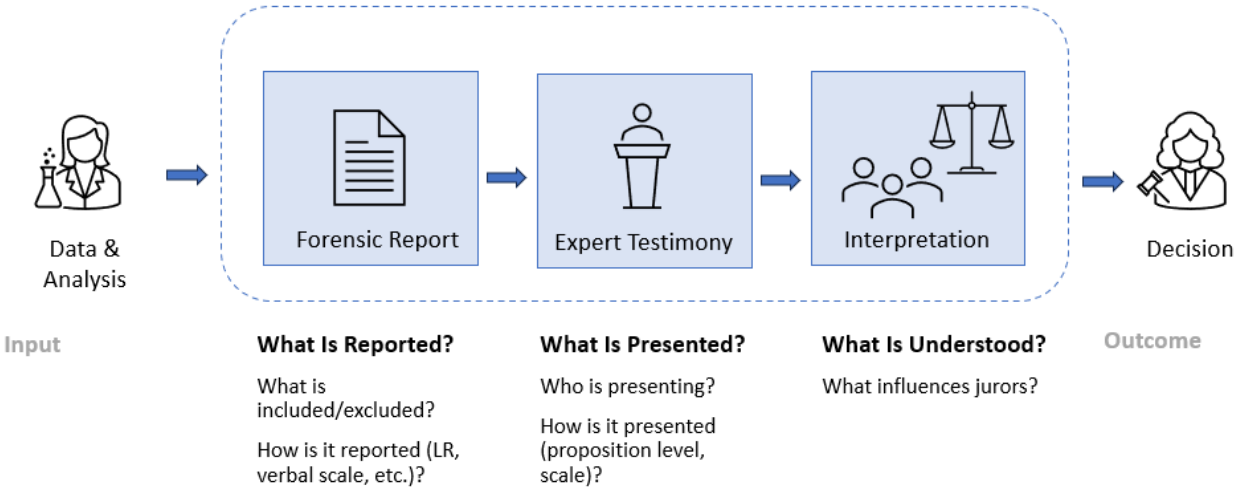

Figure 2: Communication process (credit: Sanne Aalbers/NIST)

From the personnel involved in collecting evidence at a crime scene to scientists conducting analyses in the laboratory and testifying in court, an individual's role as a communicator or recipient of information can change at different stages of a case (Figure 3).

Figure 3: Roles of communicators and recipients (credit: Sanne Aalbers/NIST)

Presentations and discussions were held on the impact different roles and how the mode of communication (written reports or verbal testimony) can impact how information understood and interpreted by different stakeholders (i.e., scientists, law enforcement, lawyers, judge, jury).

The sessions, speakers, and discussions were designed to capture this information to inform the scientific foundation review process. The goals for each session are described and summaries of the presentations and discussions are provided in the following sections.

2. Session 1: Framing the Issues

The aim of the first session was to provide context for later presentations and discussions and to ensure all workshop attendees had a similar basis for understanding what opinion testimony is and how it is admitted in court, what likelihood ratios do and do not mean, and the subjective nature of expert opinions in the interpretations of results. The session began with presentations on the goals for the workshop, the NIST Foundation Studies Program, and the process for conducting the reviews, as the presentations and discussions from the workshop are intended to inform a foundation review on communicating forensic findings followed by presentations on communication issues observed in the courtroom and interpretations of forensic evidence using likelihood ratios.

2.1. Welcome and Introductions – Sandy Koch, NIST

Communication is a process whereby information is conveyed from one person to another through verbal and non-verbal means. The goal for this workshop was to hear from a diverse group of speakers and stakeholders to better understand how descriptive language and numerical analytical results used in reports and testimony are conveyed, how this information is understood by recipients, the similarities and differences in reporting among disciplines, and how different ways of communicating information can impact judicial outcomes.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-current-practices-and-future-directions-koch

2.2. Foundation Reviews – John Butler, NIST

The goals for foundation reviews are to identify the scientific foundations that support and underpin forensic methods and to document and assess the empirical evidence for the reliability of these methods using publicly available data and peer-reviewed literature. The NIST IR 8351 DNA Mixture Interpretation Draft Report indicated a need for a separate review on how likelihood ratios (LRs) are calculated, understood, and communicated. The focus of this workshop was expanded to the communication of forensic findings by multiple disciplines, and not restricted to LRs and DNA. The foundation review will draw on the NIST human factors DNA report Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach [1] along with other resources related to communicating and quantifying the weight of forensic evidence.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-nist-scientific-foundation-reviews-butler.

2.3. View from the Bench – Judge J. Michael Ryan, District of Columbia Superior Court

Forensic scientists and lawyers operate within an adversarial system and if findings are miscommunicated or misunderstood, case outcomes can be impacted. Jurors take into consideration additional information beyond the findings being presented such as the witness's appearance, demeanor, and how they communicate.

- Trials are more of a search for proof than the truth: In the U.S. adversarial system it is about who is best at proving their case.
- Federal Rule of Evidence 702 (FRE 702) was rewritten because of the Daubert case. Initially it was about the process, not the testimony itself. Reliability is now the emphasis of Rule 702.

- A trial judge's responsibility is to rule on admissibility, whether that evidence can be considered by the factfinders (judge/jury) when deciding the case. For opinion testimony to be admitted, it must be demonstrated to the court that the testimony is the product of reliable principles and methods and that the opinion reflects a reliable application of the principles and methods to the facts of the case.
- Language is important! How witnesses are referred to can influence how the testimony is received. For example, if a witness is referred to as an "expert" the jury may add undue weight to the testimony of that individual. For this reason, some courts have moved away from referring to witnesses as "experts" in favor of "specialized opinion witness". This leaves jurors in a better position to evaluate opinion testimony admitted under FRE 702 and less apt to defer to a witness's expert status.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-rule-702-testimony-expert-witnesses-ryan

2.4. Communicating Uncertainty – Hal Stern, University of California at Irvine and the Center for Statistics and Applications in Forensic Evidence (CSAFE)

Van der Bles et al.[2] provide a useful framework to think about questions to be answered when communicating forensic findings: who is communicating and to whom, what is being communicated, in what form is the communication, and to what effect? For this workshop, the focus is on communicating the degree to which a known and unknown sample can be believed to come from the same source. When communicating forensic findings, the level of uncertainty in those findings is the most important aspect to communicate.

There are strengths and weaknesses associated with three common approaches to reporting forensic findings:

- Expert assessment based on experience, training, and use of accepted methods:
    - In the United States, expert opinion is the primary way that pattern evidence is currently assessed and reported. Examiners analyze evidence based on accepted methods, experience, and training then report conclusions as a categorization (identification, exclusion, inconclusive) or via multi-category scales.
    - Conclusions are easily understood by recipients, but they don't address uncertainty well. Black box studies can provide discipline-level performance data, but there are limitations to these studies as they are about the discipline rather than the individual expert or the individual case at hand.
- Two-stage procedure of analysis:
    - Stage 1: **similarity** involves a statistical test or procedure to determine whether or not two samples can be distinguished.
    - Stage 2: **discrimination** is an assessment of the probability that two samples from different sources could not be distinguished. Basically, how likely would it be to find two indistinguishable samples if they came from the same source or if they came from two different sources.
    - Strengths and Weaknesses: Stage 1 is a natural process for discrete or categorical variables like DNA alleles and blood types, but it is challenging when evidence is

summarized by quantitative measurements. Starting with an assumption that samples can't be distinguished is the wrong null hypothesis. Also, a binary decision can involve loss of information since one can only conclude distinguishable or not distinguishable. Stage 2 is difficult since one needs to carefully consider the relevant population. Results are not usually provided quantitatively for stage 2 at the present time.

- Likelihood ratios (LR): A statistical concept with the potential to serve as a unifying logic for evaluation and interpretation of forensic evidence.
    - Currently LRs are successfully used for single source DNA where underlying biology is understood and there is data. There have also been successful demonstrations for elemental analysis of glass. Europe has moved in this direction with the ENFSI Guidelines for Evaluative Reporting [3] and the work of several national organizations, especially the Netherlands Forensic Institute (NFI). DNA mixtures remain challenging. Pattern evidence is also challenging because it is not obvious how to best represent the evidence.
    - Strengths are that LRs explicitly compares the probability of the findings in light of two (or more) relevant propositions, avoids arbitrary match/non-match decisions through a continuous format, and can potentially accommodate a wide range of factors.
    - Weaknesses are that it is difficult to model for pattern evidence and challenging for people to understand and interpret.

Any approach for assessing the probative value of forensic evidence should: 1) account for two (or more) competing hypotheses about how the evidence (data) were generated, 2) be explicit about the reasoning and assumptions on which the assessment is based, 3) have relevant empirical support for the reasoning and assumptions, and 4) include an assessment of the level of uncertainty associated with the assessment.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-framing-issues-stern

2.5. Framing Some Issues – Steve Lund and Hari Iyer, NIST

In forensic science, communication is centered on experts imparting information to other parties in the criminal justice system. The types of information being communicated are observations about the evidence (e.g., factual, descriptive, directly observed with five senses) and opinions from expert(s) that are interpretive and personal, with some variability expected across opinions. There are challenges to observations in pattern disciplines, which can be complicated and difficult to interpret. Having an opinion scale can help recipients to understand expert interpretations but the main goal is to help recipients make informed decisions.

- Receiver operating characteristic (ROC) curves help to show the potential gain from an expert's ability to distinguish sources versus a recipient's ability to distinguish sources. The ROC data can show when an expert has meaningful information to provide or communicate to the recipients so the recipient can then use it in their decision making.
- Terminology is important and the terms "validated" and "error rates" can be problematic:
    - Validation is important for assessing the reliability of forensic methods and has a critical role in determining how much weight to give methods used in a specific case, recognizing where there are issues with overconfidence or unsupported claims, and helping recipients decide how much weight to give a method's result. But forensic

scientists don't talk about validation very well. The current focus is narrow with the impression being that once something is "validated" it is fine, that any uncertainty about method performance is inconsequential, and there is no need to continue to re-evaluate. The subjectivity of validation criteria is often hidden.

- Error rate studies bring attention to empirical performance studies but are oversimplifications, limited to correct or incorrect answers with little room for reliably interpreting inconclusive decisions.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-framing-some-issues-lund

3. Session 2: Communicating Results in Forensic Reports and Testimony

The goal of this session was to gather perspectives on how forensic findings are communicated in reports and testimony, specifically: 1) What is necessary to be included in reports and testimony, 2) How well are findings communicated, and 3) What can be improved?

The presenters were asked to discuss whether forensic findings are generally understandable to a layperson and what factors influence a juror's understanding of forensic testimony. They were invited to comment on what the community is doing right with respect to communicating forensic findings as well as where improvements can be made. Presenters were also asked how well the current reporting and testimony practices are supported by research data and whether there are discrepancies in how scientists and those reading their reports understand and use forensic findings (e.g., does the use of technical language impede understandability (scientific language and precision versus layman's terminology).

3.1. "It's Not You, It's Me" – Julie Burrill, Alan Alda Center for Science Communication

If people don't understand the findings in a report or testimony, it doesn't matter how correct the information is. Forensic scientists must be good communicators otherwise their results are not useful. To this end, more training is needed in this area.

Challenges in communication relate to:

- Content issues, specifically the complexity of the science, the need to elaborate on the subjective nature of certain types of forensic examinations, and the uncertainty that may be linked to an interpretation. The complexity of the language used in forensic reports (e.g., scientific jargon, use of passive voice, transition phrases, adjective chains, noun stacking), while precise, ends up being difficult to understand. It may be hard to say "I don't know" but it is important to be able say that and to use plain language that can be understood by those outside a discipline.
- Structural issues (related to reports and testimony) stem from the reality that the U.S. adversarial system is performative by nature, and information flows most often in a one-way communication format instead of a two-way communication model where there is greater interaction and feedback allows for clarification and improvements in understanding. Other structural issues include:
    - The last thing heard tends to have the most influence on recipients of information (primacy, recency, repetition biases)
    - The order in which information is presented impacts recipient understanding. Information provided under the scientific norm (with background information first, followed by supporting details, results, and conclusions last) differs from how human cognition prefers to receive information (where the main point is presented first followed by supporting details).

Communication skills can be improved and shouldn't be limited to training for moot courts or fitting information into template reports. Instead, forensic scientists should think about what is successful for delivery of information, be adaptable and flexible, and consider the use of improv during practice.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-its-not-you-its-me-burrill

3.2. You Keep Using That Word…I Do Not Think It Means What You Think It Means: Challenges in Communication Comprehension – Heidi Eldridge, George Washington University

The language used to convey results varies widely from discipline to discipline and this affects how recipients understand forensic findings. The focus should be on developing understandable language that is consistent across forensic disciplines. Cognitive psychologists and linguists should be engaged to better understand how recipients comprehend such information. Most of a forensic scientists' decision-making is documented in the report, but what is actually read and understood by recipients is not known. Even when forensic scientists think they are clear in their language (e.g., the use of the term "identified"), jurors do not necessarily know what the terms mean. Some of the differences in what is said versus what juries hear are related to language. Scientists have many phrases that are shorthand within a discipline or based on years of learning, but jurors do not have the same level of understanding, especially when it comes to conclusions and certainty. For example, to a forensic scientist, the phrase "100% certain" means "I looked at the evidence, and I am confident in my work." To a jury, this same phrase means an analyst didn't make a mistake.

Inconclusive findings may be perceived as not meaningful, but this can be misleading. A finding of "inconclusive" should only be used when there is no support for an opinion in either direction. Using latent prints as an example, the meaning of inconclusive results is unclear when comparing three-level scales to five-level scales. An inconclusive finding in a three-level scale can mean that there were some informative features where the examiner would lean in one direction; however, the evidence did not provide sufficient information to identify or exclude. In a five-level scale the support for different source, equal support, support for common source is clearer. There is discussion among the latent print community about moving away from a three-level categorical scale (identification, inconclusive, exclusion) toward a five-level scale using support for common or different source statements of findings [4] as well as for clarifying the meaning of inconclusive findings. It should be noted, however, that this does not yet have the full support of the latent print community.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-challenges-communication-comprehension

3.3. What is Successful Communication of Scientific Findings? – Kristy Martire, University of New South Wales

Results from a study on how statistical evidence is understood by jury-eligible lay people showed variable results. The five indicators of comprehension were consistency in evidence evaluations, ability to infer new information, sensitivity to evidence strength, orthodox belief-updating, and coherent reasoning [5]:

- Consistency in evidence evaluation indicates one can give equal weight to evidence of equal strength; e.g., a layperson understands that "1 in 1 million" means the same thing as "0.0001%." About 59% of laypeople can make these comparisons of mathematical equivalence; however, mathematical equivalence does not guarantee psychological equivalence and lay people may not treat these equally.
- The ability to infer new information from evidence signifies that one could correctly apply information from findings and use it to solve an evidentiary problem (e.g., how many random matches in a population could there be?). Sixty-nine percent of lay people could do this.

- Sensitivity to strength of evidence suggests that in the face of strong evidence versus weaker evidence, people can differentiate the weight of the information they are receiving. It was shown that 69% of people are able to differentiate the weight of information but this is done more at a general level as people understand smaller versus larger, but not how much smaller or how much larger, so it is more categorical instead of nuanced comprehension.
- Orthodoxy represents the ability to update beliefs based on (Bayesian) normative expectations. Evidence for this is mixed as only 24% of people update beliefs in line with Bayes. It is difficult to get the right priors and to test this aspect as many people are not Bayesian reasoners and there can be a tendency to underestimate the value of certain types of evidence.
- Coherence includes the logical and rational interpretation of evidence (ideally, reasoning without error) or the ability of people to avoid fallacies. Comprehension was found to be 11% under this criterion indicating clear evidence of aggregation errors and fallacious reasoning in the literature (e.g., prosecutor fallacy).

Research may look at some of these indicators, but when using all five of these components to assess findings presented using categorical conclusions, random match probability, verbal labels, and likelihood ratios, the study authors found a normal distribution of comprehension, with some people not understanding at all, some using all five components to understand the findings, and most using two to three of the comprehension indicators. The only exception was coherence for LR, which performed worse than the other indicators, possibly because the prosecutor's fallacy is more likely to occur here. When forensic scientists give more detailed reports to recipients there is an impact on lay people's understanding but there was not much of a change in the understanding of legal practitioners (the latter are already more discriminating between disciplines). More work is needed to understand what it would look like for a lay person to genuinely comprehend scientific findings.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-what-successful-communication-scientific-findings

3.4. Communicating Results in Forensic Reports and Testimony (It's not me, it's you) – Clinton Hughes, Brooklyn Defender Services

From the defense perspective, (with over 10 years focusing on forensic DNA) analysts should not just read from their reports when testifying in court but instead they should explain what their findings mean (i.e., offer a scientific opinion). There is a general lack of knowledge of validation procedures among DNA analysts who testify to results in court. There are knowledge gaps on both sides, scientists as well as lawyers, so the right questions may not be asked during a trial, or misstatements may not be caught. A prominent issue of the moment is the analysis, interpretation, and reporting of DNA mixtures. Some specific issues related to DNA mixture analysis include:

- Issues where likelihood ratios varied depending on how the software was set up, the number of related individuals in mixtures, and situations where interpretations can exceed the limits of validation studies.
- No "interpretation method can claim that the rate of false support is zero", as "there will always be uncertainty about the source of the DNA, as we cannot know who left the DNA trace."
- No one type of evidence should be relied upon in a case. Instead, evidence should be considered in combination with other elements of the case.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-workshop-current-practices-and-future-directions-hughes

3.5. Session 2 Discussion

The speakers were asked about challenges in communicating findings (e.g., information overload, short attention spans of receivers), and how forensic scientists can improve their communication of findings. Panelists noted the importance of practicing communication regardless of the experience level of the practitioner and the need to define technical terms repeatedly while presenting information to juries. One panelist suggested improv as a way to practice explaining information in the moment and stressed the importance of obtaining feedback from people who are in the target audience (e.g., family and friends without a scientific background). Defining technical terms and repeating what those words mean helps to take the burden away from the jury to remember meanings and helps them to understand better what is being communicated.

With respect to reports, the discussion participants shared that 1) the most relevant information should be presented first and not at the end of a report, 2) tables may be helpful as a neutral way of relaying the results, and 3) appendices would be useful, especially for visuals and supporting data. As for how reports are understood, that is difficult to research as it is not always clear what is meant by "well understood" and proxies are often used in studies (such as law students) who may have a different base knowledge than the typical jury. There was a suggestion to look at interactions between jurors to understand more of the dynamics that go into jury decision-making processes. No matter how witnesses prepare and practice good communication skills, the jury may not have the ability to understand the scientific details and interpretations. The burden is on the expert to use plain language to communicate in reports and testimony, using visual aids and analogies if appropriate and allowed in court. Preparation should not solely be the responsibility of the expert witness. The prosecution and defense should also work to ensure clear communications with the forensic scientist to ensure juries understand the meaning of results and any limitations to the findings.

When testimony monitoring and mechanisms for improvement was discussed, attendees acknowledged that while laboratories often obtain transcripts from trials in which their employees testify for purposes of testimony monitoring, this happens after the trial is over. Beyond internal laboratory QA procedures there is not a current practice to deal with external communication issues that may arise in situations where recipient stakeholders misunderstand or miscommunicate results (e.g., attorney summary arguments). Overall, the discussion centered on the need for more feedback for scientists and the courts and interdisciplinary communication to ensure clarity when information is presented and improve understanding of forensic findings, especially when complex scientific or statistical concepts are being presented.

4. Session 3: Practitioner Perspectives (LR and OSAC Interpretation Scales)

The goal of this session was to gain perspectives from forensic practitioners involved in the Organization of Scientific Area Committees for Forensic Science (OSAC) on the topic of likelihood ratios (LRs) and conclusion scales currently in use or under development among a selection of forensic disciplines. It was important to hear how different disciplines currently report forensic findings and the direction each is moving toward. The discipline representatives were asked to speak on:

- What background information is necessary to be communicated in reports and testimony (e.g., how best to educate factfinders about a discipline, types of data included to support results, phrasing required to clarify the strength or limitations of results)?
- How consistent is their disciplines in reporting and testifying to findings?
- How well are results supported with data in the report?
- How much of the underlying data is included in a report versus summarized?
- Are the forensic findings as reported understandable to a layperson?
- Have the conclusion scales been tested?
- If the conclusion scales have been tested, who tested them (e.g., juries – for clarity, other forensic practitioners – for consistency, law enforcement – for relevance and clarity)?
- What can be improved for broader comprehension of forensic results?
- What gaps are there related to the communication of forensic findings that should be studied?

4.1. Communicating Forensic Biology Findings – Jarrah Kennedy, Kansas City PD Crime Laboratory

Forensic biologists analyze evidence for the presence of biological materials and evaluate the evidence under a hierarchy of propositions (offense, activity, source, sub-source) to determine whether they can help answer questions relevant to the case (e.g., who the source of the blood may be). "Help" is the key word as DNA analysts are not directly answering these questions. For example, if the issue is whether a particular activity (e.g., an assault) took place that led to the transfer of the biological material, then the DNA comparison in isolation is insufficient to help address those activity questions. DNA testing is currently most concerned with the sub-source level (i.e., whether someone is or is not the source of the DNA). Technology has improved to the point where trace amounts of DNA are now more easily detected so the focus is typically on who the source of the DNA is. Such low-level detection creates additional complications as it is often not clear how the results relate specifically to the case in question (i.e., questions regarding how the DNA ended up where it was found gain in relevance).

- It is difficult to attribute generated DNA profiles taken from biological samples to their cell sources, especially in complex mixtures.
- There is a greater need to focus on task-relevant case information prior to assessing evidence and to formulate propositions prior to testing.

The current state of communicating biological results in the U.S. involves both screening test results and probabilistic genotyping:

- Screening test results are often stated as facts; however, they are not confirmatory, and it is often left to the factfinder to correlate these findings with DNA results.
- Not every laboratory is using probabilistic genotyping. But those that have adopted probabilistic genotyping tend to adopt likelihood ratios (LRs) as well.
    - Software is used to discern profiles (mixtures) and assign LR values for DNA comparison, but this often focuses on the results from the analysis not the interpretation of the results in relation to the propositions.
    - Manual methods of complex mixture interpretation express similarity between profiles from evidence and rarity may be expressed by combined probability of inclusion (CPI)/ random match probability (RMP) if similar.

The DNA field has the same issues as other disciplines when assessing similarity. Numbers and words can be hard for recipients to understand, and complex scientific principles and statistics are difficult to communicate!

- It is important to clearly state limitations of comparisons and define what words do and do not mean (e.g., rarity does not mean unique; large LR values do not automatically convey attribution).
    - DNA is one part of the overall case, and an evaluation of the DNA comparison cannot conclusively identify an individual as the source of the DNA.
    - DNA results *alone* cannot provide information on how or when the DNA was deposited in a particular case.
- One possible consideration for improvement is the placement of a cap on the LR to prevent some common misconceptions and potential cognitive fallacies that may result due to numbers greater than the world's population being presented.
- Verbal equivalents can be difficult to understand as well, but there should not be a special DNA-only scale. Instead, verbal qualifiers should reference both propositions by conveying the level of support provided by results with respect to one proposition versus the other and only after the LR number is assigned (e.g., SWGDAM verbal scale[1]).

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-communicating-forensic-biology-findings

4.2. Trace Evidence Perspective on Interpretation Scales – Tatiana Trejos, West Virginia University

Trace evidence examiners analyze evidence for the presence of small particulates (e.g., fibers, hairs, glass, paint) that may have transferred between people or a person and an object or location as a result of contact. Using a variety of analytical methods trace evidence examiners characterize materials and compare the items to known samples in order to determine whether they could share a common source. Findings from these examinations are used by investigators to answer activity or source questions about what may have occurred, how the materials may have transferred, or where the sample may have come from to link people or scenes to each other or to a potential source material. The trace evidence community has focused on developing an interpretation scale that harmonizes reporting

[1] https://www.swgdam.org/_files/ugd/4344b0_dd5221694d1448588dcd0937738c9e46.pdf

language for multiple materials (fibers, hairs, glass, paint, tape) to aid in communicating and understanding results. A review paper on the scientific foundations of trace evidence was published [6] and an interlaboratory study [7] was conducted to evaluate use of the scale, and the proposed scale has been submitted to ASTM.

The current guide details systematic approaches and criteria that work for evaluating results on a source level using a conclusion-based scale that is compatible with LRs. Analyses of class characteristics, i.e., the physical, optical, or chemical properties that either establish membership in a group or do not establish that the items came from the same source, only that they share characteristics and could have come from the same source. Class associations can have varying degrees of significance based on the relevant populations the items correspond with.

- Physical fits represent the highest degree of correspondence. A physical fit is reached when the items that have been broken, torn, or separated exhibit physical features that correspond or re-align in a manner that is not expected to be replicated. A physical fit is not currently based upon a statistical evaluation of data; it is also not based upon exhaustive comparisons to all potential sources.
- There are three association categories for class evidence:
    - Associations with highly discriminating characteristics (e.g., glass fragments characterized by elemental analysis using inductively coupled plasma spectrometry (ICP)-based methods; these may be based on statistics such as RMP or LR.
    - Associations with discriminating characteristics (e.g., four-layered OEM automotive paint)
    - Associations with limitations (e.g., blue denim cotton that is ubiquitous in the environment)
- Inconclusive
- Two levels for different sources:
    - Exclusion (Elimination) is used when the items exhibit differences indicating that the items originated from different sources.
    - Exclusion with limitations is used when an item exhibits differences from the comparison sample; however, limiting factors prevented an Exclusion (Elimination) from being reached (such as source variability).

An interlaboratory study performed to evaluate this interpretation scale showed a high level of agreement among practitioners analyzing paint.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-trace-evidence-perspective-interpretation-scales-trejos

4.3. Perspectives from the Footwear Practitioner Community – David Kanaris, Alaska State Crime Laboratory

Footwear exams start with an evaluation for design, size, wear, and randomly acquired characteristics to reduce the population of possible sources for comparison. In the past, footwear examiners narrowed down to one source only, which is now not considered to be the best approach. Uniqueness is only falsifiable and hard to express.

- The current conclusion scale published in 2013 ranges from exclusion to identification with indications of non-association, limited association of class characteristics, association of class characteristics, high degree of association in between but no inconclusive.
- The goal was to move away from categorical conclusions, and to have more transparent and justifiable opinions that were clear to stakeholders. To this end, the field moved to a weight-of-evidence scale (without underlying statistics) where a practitioner considers two competing propositions during their evaluation of the evidence.
- The result is a new scale developed under OSAC with input and contributions from the Center for Statistics and Applications in Forensic Evidence (CSAFE) that ranges from decreasing dissimilarity through neutral to increasing similarity with source exclusion split out and the strongest association being strong support for a known source.
    - Indeterminate is the chosen term to describe a neutral outcome instead of inconclusive due to the shift away from the use of "conclusions."

The presentation can be found here https://www.nist.gov/document/communicating-forensic-findings-perspectives-footwear-practitioner-community

4.4. Digital and Multimedia SAC Standard Guide for Image Comparison Opinions – Lora Sims, Ideal Innovations

Digital and Multimedia examinations involve comparisons of images of people, objects, or scenes and the information is used for intelligence gathering, identity management, and forensic comparison purposes. Some of the limitations for this discipline include image quality/resolution, distance of camera to subject, and the fact that images are static.

Examiners in this discipline have used multiple scales for different purposes (intelligence gathering and forensic comparison). The proposed SAC Standard Guide ultimately chose an evidence-centric scale in which the forensic scientist offers an opinion about the evidence instead of a conclusion-centric scale which is more the decision-makers role.

- Their scale includes 5 levels:
    - Strong support for different source
    - Support for different source
    - Inconclusive
    - Support for common source
    - Strong support for same source

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-dm-sac-standard-guide-image-comparison-opinions-sims

4.5. Forensic Document Examination – Miriam Angel, Los Angeles Police Department Laboratory

Document examiners are often requested to address whether the writer of a known handwriting sample also wrote a questioned writing sample or not. Handwriting is the result of a behavior, so (1) a person can purposefully change their writing through disguise or simulating another person's writing and (2) a person's writing can change due to unusual writing conditions, such as writing while intoxicated or in an

uncomfortable position. The traditional basis for the examiner's ability to distinguish between different writers is that "no two people write alike." It is more appropriate to condition that statement: "Given a sufficient amount of natural writing, no two people are likely to produce the same combination of handwriting characteristics." Lack of "sufficiency" of natural writing is a limitation in the examination that can result in a weak or inconclusive/indeterminate opinion. The assessment of the impact of limitations on the strength of a handwriting opinion is subjective, meaning it is based on the examiner's judgement and experience.

- The current scale in use is a nine-point linear scale, with elimination (exclusion) on one end and identification on the other. Inconclusive is in the middle. In between inconclusive and the ends of the scale are opinions stated in terms of posterior probability, such as "probably wrote" or "probably did not write." Aside from the criticism that an opinion of identification/individualization in the form of posterior probability is not scientifically supportable, these opinions can also be problematic in terms of logical reasoning when there are multiple known writing samples submitted for comparison.
- A new scale under development in OSAC requires that opinions be stated in terms of the degree of support the findings provide for one proposition over another. The propositions represent each side of what is typically argued in court, such as "The known writer wrote the questioned document" and "Someone other than the known writer wrote the questioned document." The new scale has nine levels to ease the transition from use of the traditional to the new scale.
- "Support" statements adhere to the following logic: If the probability of observing the handwriting features if proposition X is true is larger than the probability of observing these features if proposition Y is true, then the findings support proposition X over proposition Y. If the probability of observing the handwriting features is about the same under both propositions, then the findings do not support one proposition over the other.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-forensic-document-examination

4.6. Session 3 Discussion

Each of the session's speakers shared how their respective discipline currently presents findings in reports and the language and scales that are currently in use or under development. They were asked to determine whether there are any similarities between the discipline-specific scales and whether they could envision adopting a universal scale for use in the United States. While the panel thought a universal scale would be very useful for end users and that such a scale may be closer to implementation than they would have previously thought, there is more work that needs to be done at the foundational level to discuss forensic evidential reasoning and inference before such a scale could be adopted in the United States. Some factors that would need to be considered include agreement on terminology to use, and statistical underpinning from all disciplines to ensure the scale could apply similarly to all disciplines. The same terminology may not have the same strength in context for different evidence types and it may be difficult to gain buy-in if such a scale would be a significant departure from current practice. Regarding terminology, the OSAC Lexicon[2] was mentioned as evidence that scientists, lawyers, jurors etc. don't all define words similarly so work would need to be done within disciplines to come to consensus on how the terms are used, which would help everyone to understand the scales.

[2] https://www.nist.gov/organization-scientific-area-committees-forensic-science/osac-lexicon

Another issue that was brought up was that the use of words changes over time as people learn how what was intended is understood by others. For example, forensic scientists no longer use the term "match" as it was taken as a strong identification and there are concerns with using the term "consistent with," so the terminology for similarity may also change. Attendees agreed that there is some overlap in the broader terminology currently in use or under development so the comparative disciplines should work to use words in a similar way.

Another topic of discussion involved the numbers related to scales versus vague words like "rare". While having data to support the numbers is the goal, not all disciplines are there yet; therefore, it is difficult to convey what "highly discriminating" means. It was also pointed out that end users are even worse at interpreting statistics than forensic scientists are at explaining them. Ultimately conclusion scales that are flexible to accommodate quantitative and qualitative opinions was suggested as a goal, but that the opinions should not be isolated without the rationale for that opinion being included in the report. The speakers also pointed out the necessity of validation of the methods used, having black box studies to understand error rates within the disciplines, and providing guidance to the community to evaluate the consistency of interpretation and use of scales.

Communicating the limitations per discipline standards and other guidance documents being put out by OSAC is important, but how clearly the limitations are communicated in casework was a concern expressed from the audience. There was variability among the disciplines represented by the attendees in how clearly limitations are expressed in reports. Some believed their limitations were plainly expressed in the text of a report while others relied on use of phone conversations and pre-trial conferences to provide more detailed explanations. Appendices were brought up as a good way to explain the various limitations in more depth and to have that information available to all who receive the report. Many felt that additional work needs to be done to improve the clarity of forensic findings and if there are changes to the way disciplines are reporting, then more black box studies will need to be conducted. They felt that this research should be started while the language is being standardized so there will be performance data to support new standards, though attendees also acknowledged that balancing the need to do this was a challenge given the competing demands of casework and limited number of practitioners potentially able to participate.

5. Session 4: Group Discussion – Facilitated by Melissa Taylor, NIST

5.1. What questions would you expect to be addressed in a foundation review on communicating forensic science?

The focus of this discussion centered on broadly understanding successful communication, how limitations should be expressed, the language used among disciplines, and training and education of all stakeholders. The workshop attendees suggested NIST focus on understanding what successful communication looks like generally as a first step: how information is transferred (e.g., reports, testimony, non-verbal cues), how the information is understood by recipients (e.g., understood as intended, misinterpreted, prosecutor fallacy), and how the information is used especially in regard to the probative value of evidence. In order to communicate forensic science results effectively, one first needs to know how to communicate science. A survey of cognitive psychology research would be useful to learn about communication modalities and to see how they could be applied to forensic science.

Beyond that, attendees felt that the forensic community should work toward consensus for how LRs are assigned, the reporting of interpretations as findings, and the data that supports the findings to ensure that recipients of forensic reports and testimony correctly understand the meaning of the findings and how they fit within a range of potential conclusions. Furthermore:

- Forensic scientists need to clearly communicate limitations to their findings, the context the evidence is being evaluated under, any performance data or absence of such data that is relied upon to evaluate evidence in a case (especially in relation to the rarity or commonness of a material), and the empirical support for conclusions so the receiver knows how the opinion is supported or not supported by data.
- Interdisciplinary guidance on appropriate language to provide clarity on strength of evidence is necessary so recipients can effectively integrate information from multiple disciplines. Such guidance would include distinctions based on the strengths of evidence, contextual information relied upon for interpretations, and discipline- and laboratory-specific information (e.g., validation, error rates) that contributes to the meaning of the evidence. This was discussed as needing to be universal because all evidence is evaluated under the same legal framework. A universal interpretation scale would be ideal as the disparate forensic evidence found in a case is subject to the same rules of evidence and ultimately ends up with a small group of recipients who are charged with using the assorted information within the same legal framework.
- Training and education for different stakeholders is needed to limit potential misunderstandings by those who are not experts in the field. Training needs to be robust, so stakeholders have the knowledge to communicate and understand findings. Cross-disciplinary training is important to allow scientists, lawyers, and law enforcement to better understand each other's role and how information is used and communicated by different professions.
- Effective communicators often consider the audience and situation (e.g., verbal testimony to a jury, written reports for law enforcement and lawyers, non-verbal cues). The forensic community needs to better recognize how people learn and understand science in the general sense before focusing on discipline-specific recommendations to clarify results. Is there value in using the same scale for all evidence types? Would your audience be more apt to understand the proper weight of evidence if the same scale were used?

- Attendees thought that having a standardized scale (and one based on LRs) was a good goal. While it is not known how often multiple reports on different types of evidence are received and evaluated in a case or to know whether there is confusion among recipients when different scales are used to communicate the interpretations of the evidence within the context of a case, attendees perceived that variability in conclusion language can cause confusion and the goal should be to mitigate that.
- The workshop attendees were struck by the similarities among the presented scales for different disciplines and believed it was important that there be consistency in the language used to help recipients understand the weight they should give to different findings. For recipients to evaluate and combine conclusions about different types of evidence in their deliberations, a standardized scale based on LR would be beneficial. It was noted that the NRC 2009 report emphasized the importance of using universal language and terminology, and attendees noted that the forensic community has made a great deal of progress [8].
- While most attendees were in favor of evaluating evidence using a likelihood ratio framework, the reality of its potential use in the U.S. adversarial system was questioned. Attendees acknowledged there is a lack of statistical knowledge among forensic practitioners and recipients of their findings and felt that the lack of broad statistical expertise would make adoption of LRs difficult in the US. Consistency in language used among disciplines and having scales calibrated to empirical data for each discipline will help move forensic science forward. If terminology and a uniform scale could be agreed upon, each discipline could provide examples showing where their evidence fits and what portions of the scale would not be applicable or would be an overreach for that discipline.
- For disciplines that do not yet have data to calculate LRs, extrapolation or interpolation of the performance data that are available was suggested; however, that process adds another level of subjectivity and uncertainty. One suggestion was to learn from agencies and countries that have used LRs for multiple disciplines and find out what their experience in communicating to end users has been. By studying the experiences of others who have incorporated LRs into their systems, forensic scientists and the legal community would be better positioned to know the factors necessary to incorporate them under the U.S. adversarial system prior to implementation, as well as the limitations and training necessary before moving in that direction.

5.2. What are the properties of a suitable scale? And how do we check whether a scale has those properties? What makes one scale better than another?

- Discussions focused on terminology, the number of levels for conclusion scales, strength of the evidence, consistency, accuracy, repeatability, and reproducibility. A good scale is one that measures what it claims to measure, is understandable by a jury, is coherent, and ordered. Ordinal categories were favored over a symmetrical scale as there may be more gradations in one direction over another and the bins for each category may not all be the same size.
- Both ends of a scale were discussed. “Identification” was not supported as a category while exclusion had more support. “Exclusion” was discussed as a finding that is based on an observation of completely different characteristics (with a probability of 0). A two-stage approach was offered where stage 1 would be an evaluation for gross differences, where

exclusion would be appropriate, and the stage 2 comparison would use an ordinal scale that did not include a categorical exclusion or identification.

- "Inconclusive" was found to have different meanings among workshop attendees (e.g., not suitable, uninformative, indeterminant, similarities and differences observed, insufficient evidence to move away from neutral) with some favoring a move to neutral support statements for H_1 and H_2 instead of inconclusive.
- Discussions on statements of similarity focused on the degree something is similar or dissimilar, how rare or common the potential comparison source is, and how to provide useful information while ensuring that findings are appropriate and not overstated or understated.
- Understandability is an issue as some recipients may not comprehend a verbal scale and others may not comprehend a probability scale. Valuable information could be lost, so findings in reports should include a visual representation of the scale used in order to enable more people to understand where the findings lie along a continuum of potential conclusions.

5.3. Are there other questions that NIST should be exploring?

- What is the impact of hypotheses and numbers used in court?
    - How do forensic scientists arrive at LRs for disciplines where LRs have not yet been implemented and what is the impact of using subjective numbers on juries?
        - The LR numbers can vary based on the alternative hypothesis under which the evidence is evaluated. In the U.S., the defense does not have to disclose its alternative hypothesis, which can lead to challenges if the hypothesis is unknown.
    - There was a concern that courts might not fully comprehend or properly address LR-based approaches.
- What is the minimum requirement for information provided in reports?
    - An explanation of limitations on what can be determined and the methods used.
    - An explanation of assumptions used in evidence evaluations and the hypotheses upon which the LRs are based.
- What can we learn from science communication in other disciplines:
    - How do recipients understand information? Consult learning specialists on how to present information in a way that factfinders can more easily comprehend.
    - Cognitive science and communication expertise should be integrated into forensic training to improve clarity in report writing and testimony.
- What is known about decision making?
    - Psychology research on signal detection theory can help us recognize how people make decisions. Knowing more about the decision thresholds of experts can help us to gauge the effects of expert decisions on error rates.

        - Include a survey of forensic science providers and recipients of information to learn how receivers process the information being provided by the laboratories.
    - What is the forensic scientist's role in court? Are forensic scientists the ones who should be making decisions or should they be providing information to the factfinders who then make the decisions? And how is that difference best communicated?
    - How do the different purposes of reports (investigative versus evaluative) influence decision making, and should the purpose of a report be explicitly stated as investigative or evaluative?

The overall focus for the forensic community moving forward (and to be considered in a NIST foundation study) should be on improving the clarity and effectiveness of forensic science communication in court and reports, and the way information is perceived and processed by the audience.

6. Session 5: Gaps and How to Fill Them

The goal of this session was to gain diverse perspectives on 1) potential gaps in research to improve communication and understanding of forensic findings, 2) training on the use of likelihood ratios, and 3) how effective the different ways forensic scientists currently communicate their findings in reports and testimony might be. These perspectives were addressed through discussions relating to the types of information that should be shared to support conclusions including: How can others assess the reliability of forensic findings being presented today? What can be improved? What knowledge or data is missing in the forensic and legal community that can help improve clarity in presentation and better support accurate, understandable, and reliable assessments of forensic findings?

6.1. Gaps and How to Fill Them – Bill Thompson, UC Irvine

Forensic science findings should be reported in a way that is justifiable, complete (with limitations, assumptions, and uncertainties disclosed), and understandable, so that factfinders can give the right weight to evidence. However, for many disciplines there is disagreement on the probative value of evidence, the basis for opinions on such evidence, and whether the forensic community can agree on what is needed to meet the goals of justifiable, complete, and understandable reporting.

- Among many disciplines there are disagreements about the strength of evidence, especially in complex cases, and this is often when forensic expertise is needed the most.
- Task-irrelevant information may shift the decision threshold [9]
- Biedermann et al. (2011) [10] and Kalafut et al. (2024) [11] articles offer insights on how to disclose assumptions and uncertainties.
- LRs can vary based on different reference populations or number of contributors. Providing reports for each possible assumption is most informative for factfinders.
- What “conservative” means for analysis, reporting, and the effect on cases is not well known so it should not be a goal until that is better studied.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-gaps-and-how-fill-them

6.2. Identifying Gaps and Limitations via SOPs – Kate Philpott, Virginia Commonwealth University

Validation studies should test the limits of analysis so the scope of the method and its interpretation reliability can be defined. Laboratories should then incorporate the limitations identified by the validation studies in their standard operating procedures (SOPs) both for transparency of a method’s scope in casework but also to ensure analysts themselves understand the limitations and tested factor space.

- There is a gap in validation studies as they should address all Venn diagram options of a factor space and laboratories should incorporate this information into SOPs.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-identifying-gaps-limitations-sops

6.3. Statistical Theory for Likelihood Ratios in Forensic Analysis – Dan Rabinowitz, Columbia University

There is always a distribution underlying a measure, whether it is provided as a likelihood ratio or a verbal/categorical scale, though there may not be a lot known about that distribution.

- Someone must explain the LR calculation to the jury. A number should not simply be provided by the forensic scientist without sharing how the number was arrived at, what it means and does not mean, and how it can be used by a recipient to determine their posterior probability.
- The math is right because it is math, but does the model and/or your assumption really reflect the case? Reporting the uncertainty is necessary, including uncertainty on assessments of uncertainty.
- How to assess the prior probability needs more attention.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-statistical-theory-likelihood-ratios-forensic-analysis

6.4. Using a Validation Study Model to Minimize Wrongful Convictions – JD Schmid, 6th District Public Defender's Office, MN

The increased sensitivity of DNA testing methods has led to an increase in the detection of DNA that is unrelated to the crime being investigated. Prosecutions against innocent sources of "irrelevant" DNA has, in turn, led to wrongful convictions and significant deprivations of liberty. The ability of the American court system to adequately protect against these miscarriages of justice using existing procedural safeguards is not well understood.

The path forward suggested during the presentation included conducting mock juror studies modeled after laboratory validation studies, using fact-based scenarios derived from exoneration cases (or near-miss cases). Such fact-based scenarios are analogous to samples used in laboratory validations because they:

- approximate ground truth,
- replicate conditions that are expected to be encountered in casework, and
- test the limits of the system.

Using such samples would also allow researchers to assess whether procedural safeguards like cross-examination, expert evaluations of evidence given activity-level propositions, limiting jury instructions, and/or defense expert testimony can adequately prevent wrongful convictions.

Modelling mock juror studies after laboratory validations recognizes that the goals, methods, and values of forensic science laboratory systems are in many ways analogous to the goals, methods, and values of the criminal justice system in a trial. Both systems:

- seek to distinguish true positives from true negatives in circumstances where ground truth is unknown,
- use a vast array of procedures and rules designed to accurately make this distinction, and
- incorporate a value that a false positive outcome is worse than a false negative outcome.

The main difference between forensic science laboratories and the court system is that laboratories must validate their methods before they put them into practice. The court system has never been subjected to such a requirement and, to the presenter's knowledge, there has not been any serious effort to validate the court system in the same way that is required of laboratories. The failure to conduct validation studies on the criminal justice system significantly undermines confidence that the system is accurately distinguishing the guilty from the innocent, especially in difficult cases. The proposed research seeks to fill this gap by providing empirical data about whether, and to what extent, procedural safeguards available to American courts can prevent wrongful convictions when irrelevant DNA is recovered during a criminal investigation.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-using-validation-study-model-minimize-wrongful-convictions

6.5. Session 5 Discussion

Questions for the session 5 speakers revolved around gaps in data and how improvements could be made through research. Black box studies are not broadly representative so they should not be used as population data because of their study design. Some have been deliberately designed to be more complex to test the methods with close non-mates. Black box studies should not be used for frequencies, calculating LRs, or error rates. The data may be misleading if used as error rates for a particular discipline, as error rates will vary based on how easy or difficult the samples are and the ability of the analyst who examines it. Sample distributions within studies may not be reflective of random comparisons found in casework. It was agreed that you do get insights into the variability of sample complexity and performance within a discipline that can be useful to gauge whether forensic scientists should be saying extremely strong support, or if such statements should be limited.

Proficiency tests have room for improvement and stress-testing the system was discussed as a way to look at a different baseline for casework. Blind proficiency testing (PT) is hard to do but blind PTs that include samples that deliberately stress-test the system can help examiners reset their baseline. If examiners see more difficult samples, it may affect how they interpret casework. Routinely seeing challenging samples may be helpful to ensure there is not an overconfidence in how evidence is interpreted. The consensus among the attendees was that validation is necessary to challenge methods and to determine the limits of what the method can and cannot do. There was broad support for the proposal that limitations established from validation testing need to be more explicitly incorporated into casework or clearly communicated to recipients of reports.

7. Session 6: European Perspectives and Practices

The goal of this session was to gain an understanding from the European perspective on the use of likelihood ratios and how forensic scientists should communicate their findings in reports and testimony. These perspectives were obtained by addressing the following questions:

- What data is needed to calculate LRs?
- If that data is not available, how should results be stated?
- What types of training and how much do forensic scientists and the legal community need in order to be able to present and understand findings clearly and accurately?
- What issues exist in other legal systems that may differ from those in the United States?
- How much training do lawyers and judges receive to understand and reliably use forensic findings?

7.1. Guidance Documents for Evaluative Reporting in Forensic Science: European Developments – Alex Biederman, University of Lausanne (Switzerland)

The Fundamentals of Probability and Statistical Evidence in Criminal Proceedings[12], ENFSI Guideline for Evaluative Reporting in Forensic Science [3], RSS Guide to Statistics and Probability for Advocates [13]and the RSS Science and Law Statistics Primer [14] were all written for different audiences, but share the message that measurements of uncertainty use probability and likelihood ratios as the way to measure the value of evidence. Reporting results of analyses without interpreting them in the context of the case can be misleading. The ENFSI Guideline for Evaluative Reporting in Forensic Science aims to promote logical, balanced, transparent, and robust evaluation and reporting of results. To support the implementation of the ENFSI Guideline, it is important to provide training for practitioners. One strategy is to train selected personnel who can act as ambassadors for evaluative reporting in their laboratories.

- The ENFSI Guideline applies only to evaluative reports for use in court, not to investigative, intelligence, or technical reports.
- Findings are evaluated with respect to competing propositions set by case circumstances or by a mandating authority.
- For transparency, the probability assignment is based on body of knowledge available for auditing and disclosure.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-guidance-documents-evaluative-reporting-forensic-science

7.2. When, How, and for Whom? – Anders Nordgaard, National Forensic Centre (Sweden)

The National Forensic Centre (NFC) classifies uncertainty into categories based on the sources of uncertainty in forensic investigations. For uncertainty related to measurement, contamination issues, human factors, or situations in which a limited sample of material is analyzed, the sources of uncertainty are either addressed within a technical report or through the quality assurance system. For uncertainties beyond laboratory control (e.g. case-specific characteristics), the uncertainty is expressed within an evaluative report. Examples for evaluative reporting of uncertainty would be for questions of rarity or

commonness of a material when the question is source attribution and mechanisms of transfer, persistence, and background levels of the material of interest.

- Likelihoods are defined for a particular value of a parameter or a simple hypothesis. A challenge occurs when hypotheses are composite, which would require several sub-hypotheses to be put together, as the data does not indicate how much weight to give to each sub-hypothesis. The Bayes factor is then a weighted average and requires more information about priors.
- The NFC uses a symmetric ordinal scale from -4 to +4 representing nine different ranges of likelihood ratios, with likelihood ratios above 1 for levels +1 to +4 and below 1 for levels -1 to -4. The ranges are based on interpretations of posterior probabilities when even prior odds are assumed. Level +2 then gives a posterior probability of at least 0.99, which is generally assumed to be sufficient for probable cause (but not beyond reasonable doubt).
- The NFC has training programs for forensic experts, investigators, prosecutors, judges, and defense attorneys on evaluative reporting.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-when-how-and-whom

7.3. Communicating LR Conclusions in Forensic Reports – Marjan Sjerps and Rolf Ypma, Netherlands Forensic Institute (Netherlands)

At the Netherlands Forensic Institute (NFI), statisticians are embedded in the laboratories to advise colleagues and provide training to judges, prosecutors, police, etc. LRs are used in reports where inferences are made. Numerical LRs are used in all DNA reports where numbers can be calculated and occasionally for other disciplines (e.g., glass, fingerprints, automatic speaker recognition). Verbal LRs are based on knowledge not (only) expressed in datasets and rest on personal probabilities. The order of magnitude of the 'personal' LR is verbalized using a scale relating words to numbers (used NFI-wide). When a verbal LR is reported in a discipline that does not have numerical data, there is a footnote in the reports to explain how the LR is used, define the verbal terms numerically, and warn against the prosecutor fallacy. The footnote also provides a link to a professional annex with more information about the LR framework.

Some specific challenges of note include:

- Potentially misleading database search results – When using database searches, there is a potential that the searches could generate false leads that must be clearly communicated. The laboratory should be contacted if more leads need to be found. The NFI adds a text box warning in these reports.
- The potential to encounter fallacies and other forms of misleading reasoning with any form of conclusion (not just LR) including: the "prosecutor's fallacy" (most prevalent), the "comparison fallacy", the "relevancy fallacy", "rule-based reasoning", and potential issues when LRs need to be combined with prior odds.

Examples provided with respect to what works well included:

- Explaining the LR framework in different ways (e.g., numbers, pictures, words, formulas) because the audience is diverse and may comprehend information differently.
- Using everyday comparisons (e.g., relating to facial features – do they have a nose or an uncommon feature such as a scar)

- Being transparent about errors and uncertainty (expected from scientists)

Main message from European perspective (Biedermann, Nordgaard, Sjerps and Ypma):

- Communicating LRs is difficult but not impossible with the logic of LRs gaining traction and becoming very persuasive; numbers are unequivocal, words are not; training is important; and statisticians should be embedded in laboratories.

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-communicating-lr-conclusions-forensic-reports

7.4. Session 6 Discussion

There was a question on how to get end users to understand what the language in a forensic scientist's report means, as interpretations are personal. There were concerns regarding how numbers are perceived by others. The speakers acknowledged that lay people may experience difficulty with understanding what the numbers mean, making it necessary to explain what LRs mean in lay terms. Suggestions for how to do this included using plain language (less technical jargon) and everyday examples to help everyone understand. The personal nature of interpretation was discussed as humans are all different and their experiences can impact interpretations. Whether interpretations are sound is an ongoing area for research and improvement, as is figuring out ways to improve peoples' ability to assign probabilities in a robust and defendable manner. Although getting a precise likelihood ratio value is not the main goal, it is getting the right order of magnitude that is most important. For pattern comparisons, it comes down to an expert's understanding of the LR and getting to a point where you can assign probabilities in a robust and defensible manner. The assignment of an LR allows one to proceed in a logical, balanced, and transparent way. The availability of task-relevant data is also important. Disclosures about the lack of case-specific studies to support interpretations and the impact that missing data can have on probability assignment will help the forensic community to improve and be transparent.

How to get everyone to use the same scale was another topic of interest. The speakers shared that they found it was often easier to train new examiners, rather than experienced laboratory personnel. Prosecutors were not always interested in going through training because they thought it was the job of the forensic scientist to explain their interpretation of the findings, while defense lawyers were very interested in the training because they wanted to find where any issues were and identify potential gaps in knowledge. Short workshops for 1 to 2 days were found to be useful for practitioners to gain an interest in the topic of interpretation but were not enough to have them fully understand and be able to implement probabilistic reasoning and evaluative reporting. Longer-term learning is necessary for acquiring proficiency in evaluative reporting, with personalized feedback necessary to understand in which areas to focus on for improvement. The workshop attendees found that they obtained the most feedback during informal moments in classes or workshops (e.g., on how scales are understood), otherwise it was hard to tell how well the training was understood. They also indicated that they had limited information on how a universal scale was to be applied in court.

8. Session 7: Breakout Group Discussions

Workshop attendees were assigned to smaller breakout groups based on professions and similar questions were asked of the different groups. The group assignments were designed to provide information on whether there are perspectives more prevalent in one community or another and to provide an opportunity for people with similar backgrounds to find common ground and share their experiences.

8.1. Group 1: Legal Community

The representatives of the legal community present at the workshop were asked about communication of forensic findings, specifically how they are understood and used in court. This group shared that pre-trial dialogue was an illustration of good communication between analysts and attorneys that helped the attorneys to understand what is important and what is not, as well as to work through the questions that should be asked to help clarify the information for the court. They felt that having attorneys who are familiar with forensic issues was beneficial to cases but that there is a high turnover of lawyers so training on forensic disciplines and issues is a continual need. They thought that there is a need for a non-adversarial/neutral resource to help vet evidence and train lawyers about evidence that will be presented in court; while there are some examples of this regionally, more is needed.

When asked what happens if someone uses the information relayed by an analyst incorrectly, this breakout group noted it was a blind spot between the legal community and the laboratory as the laboratory will rarely hear about cases where the information was misrepresented in closing. Testimony monitoring of forensic scientists by supervisory analysts or technical leaders can address technical mistakes made before a witness is excused but there is little accountability for a lawyer's miscommunication of information during closing arguments, especially as transcript reviews by laboratories focus mostly on their analyst's testimony and not the lawyers' closing arguments. Improved communication between parties ahead of time and more resources (e.g., training) were suggested as ways to mitigate/prevent this from happening.

When asked whether jurors are made aware that there is a range of opinions that could be reached and that there can be differences in opinion between experts, this group varied on how such information might come out at trial. While opinion scales and any explanations for the range of conclusions that could be reached are typically included in reports and those reports are put into evidence, not all of the potential conclusions are typically brought out during direct or cross examination. It was noted that there is a lot of information in reports that recipients do not necessarily understand so the limitations to forensic findings may not be fully understood by all parties.

Other topics briefly touched upon included:

- Use of demonstrative aids to improve communication
  - These were generally deemed useful, but such tools are not often used. Seeing and hearing improve understandability, but it is up to judges to allow these aids to be used in court, the defense should have the opportunity to see them prior to the trial, and demonstrative aids need to be universal.
- Performance data
  - Performance data does not come up routinely in trials. Jurors are not expected to look at that information, and it is mostly used in admissibility hearings. It is hard to

interrogate a witness about published studies they might not be aware of, and it is a hard topic to introduce in court.

- How jurors use information presented at trial
  - Studies show jurors rely on non-data related characteristics of the witness: likeability, perceived confidence, and competency. This group felt jurors evaluate witnesses on whether the expert is trustworthy and likeable, and confidence in testimony makes them think the person is competent. Jurors may use David Kaye's "likelihood-ism" that something is more or less likely but not in a true Bayes format. Instead, jurors often try to map testimony about evidence on to the stories in their heads and rely on who told the better story that best fits the facts.

8.2. Group 2: Statisticians and Academic Researchers

Statisticians and researchers focused on data supporting communication efforts. They noted the critical importance of both relevant supporting data and the amount of different performance data that are available, but stressed such data needs to be linked to the case at hand and that there needs to be a greater willingness to talk about uncertainty and error rates. Multiple types of data, not just proficiency test data, are needed.

This group thought that a narrative description of results can be an effective communication strategy but that there is also a need to consider how to communicate with the different generations that make up juries now and in the future. There are tradeoffs between complete and comprehensible reports and limits to what jurors understand. It was also noted that there are peripheral effects (e.g., appearance, clothing, demeanor, juror prior knowledge) that impact what is understood. Communication efforts should not just focus on jurors since most cases do not go to trial, but also on law enforcement, lawyers, and judges so opinion results are not taken as set facts but understood within the framework of limitations for each discipline, the case, and the findings being reported.

When discussing the reporting of results, the statistician and researcher breakout group felt that the forensic community should aim toward standardization of objective opinions. When there is not data to support an opinion then an expert can still provide valuable insights that are informative to a case, but it should be qualified, and that experts should more often say when they *do not know* something. Regarding opinion results, this group had a discussion on the differences in professional opinion versus personal opinion, the subjectivity of these, and how validation data provides more communal information on how results should be interpreted.

- Experts should be transparent about the basis of findings: they should be justifiable, accurate, and complete.
- Communication efforts should focus on reports for law enforcement, lawyers, and judges before focusing on jurors so opinion findings are not taken as set facts.

8.3. Group 3: Forensic Practitioners and Discipline-Focused Researchers

The forensic practitioners and discipline-focused researchers shared their sense that there was growing self-awareness on issues related to interpretation and reporting among their colleagues. They thought forensic scientists need to move out of discipline-focused silos to speak more with each other within and across laboratory systems to help solve these issues as a community. OSAC was pointed to as a way several felt connected with the broader forensic community, but they believed there is a greater need

for the broader practitioner community to have this dialog too so more people can see how common some of these issues are and to learn possible solutions from others. While change is generally perceived as scary, this group found it heartening to see how far they have come in interpretation and report writing.

LRs were acknowledged as hard to understand so an emphasis on numbers to the exclusion of anything else is the wrong way to proceed. Instead, the focus should be on the evidence and interpretation in relation to competing propositions rather than categorical results. The LR framework can include verbal statements indicating a finding is explained better by same source than different sources. This group recognized there is currently a lack of standardized scales; even within a discipline there can be multiple scales and different ways of describing levels depending on the laboratory system. This is affecting recipients of information and practitioners. When it is known that multiple recipients will receive a report, the amount of information and how it is described can increase if the report writer does not know which audience is being addressed. Guidance documents being developed in OSAC subcommittees are taking a transition approach to using an LR framework for interpreting results, with several disciplines not focusing so much on numbers but more on evaluating results in relation to competing propositions.

Specific issues discussed related to likelihood ratios focused on understanding and relaying LRs.

- Discussion focused on the way LRs are currently used in the U.S. by the DNA discipline. Attendees felt the current use of LR was incorrect, because instead of interpreting the results within a framework themselves, DNA examiners are using numbers provided by software. Instead, they should evaluate the evidence and report their own likelihood ratio.
- Engaging in these types of discussions was seen as an opportunity for other disciplines to get the correct fundamentals in place and to better understand the principles before trying to assign LR numbers themselves. If this is difficult for DNA with all the population data that they have available, then other disciplines will face additional challenges.
- There needs to be a greater focus on the evidence, not just numbers. LRs and the principles of interpretation are not understood well among practitioners and jurors, so if practitioners manage to teach themselves then they will gain the tools to teach others.

9. Session 8: Why This Matters?

9.1. Why It Matters? – Lynn Garcia, Texas Forensic Science Commission

There are gaps in the conversations being held here and to what laboratory personnel (analysts, technical leads, directors, quality managers) are exposed. There needs to be extensive training to better understand statistics and how to apply LRs correctly. Testimony transcripts show the importance of communication and how data can be misrepresented. Transitioning to evaluative reporting using LRs will take time, and there are aspects that need to be addressed as a community before they are more widely adopted in the United States. Right now, "is it possible?" is the question, but this can be transitioned to talking about the probability of the evidence given two mutually exclusive activity propositions.

Issues to work on include*:*

- Evaluative reporting in all the other [non-DNA] disciplines, but the foundation for how the strength of evidence is assessed and where the support for the findings comes from is key.
- Activity-level reporting: everyone will need training before using as there is a risk of ad hoc pseudo-evaluations on the stand.
- Traceability! Transparency! Communication!

The presentation can be accessed through this link: https://www.nist.gov/document/communicating-forensic-findings-why-it-matters

10. Wrap Up and Next Steps

10.1. Next Steps: Focus of Foundation Review and Team to be Developed – Sandy Koch, NIST

The NIST team will take some time to digest all the information that was gathered and develop a focus for the next Foundation Review. Anyone interested in serving on the foundation review team was asked to get in touch and to communicate their level of interest and availability. Suggestions were requested for references that should either be included as part of the workshop summary or that should be reviewed during the foundation study. Suggested references submitted by the attendees at the workshop are posted on the NIST website along with the presentation slides at https://www.nist.gov/news-events/events/2024/06/communicating-forensic-findings-current-practices-and-future-directions

11. Works Cited

[1] Taylor MK, Romsos E, Ballantyne K, Dawn Moore Boswell D, Busey T (2024) Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach. https://doi.org/10.6028/NIST.IR.8503

[2] Van Der Bles AM, Van Der Linden S, Freeman ALJ, Mitchell J, Galvao AB, Zaval L, Spiegelhalter DJ (2019) Communicating uncertainty about facts, numbers and science. *Royal Society Open Science* 6(5). https://doi.org/10.1098/rsos.181870

[3] Willis S, McKenna L, McDermott S, O'Donnel G, Barrett A, Rasmusson B, Nordgaard A, Berger C, Sjerps M, Molina JJL, Zadora G, Aitken C, Lunt L, Champod C, Biedermann A, Hicks T, Taroni F (2015) ENFSI guideline evaluative reporting.

[4] Carter KE, Vogelsang MD, Vanderkolk J, Busey T (2020) The Utility of Expanded Conclusion Scales During Latent Print Examinations. *Journal of Forensic Sciences* 65(4):1141–1154. https://doi.org/10.1111/1556-4029.14298

[5] Bali AS, Edmond G, Ballantyne KN, Kemp RI, Martire KA (2020) Communicating forensic science opinion: An examination of expert reporting practices. *Science and Justice* 60(3):216–224. https://doi.org/10.1016/j.scijus.2019.12.005

[6] Trejos T, Koch S, Mehltretter A (2020) Scientific foundations and current state of trace evidence—A review. *Forensic Chemistry* 18. https://doi.org/10.1016/j.forc.2020.100223

[7] Mehltretter A, Prusinowski M, Arkes H, Flohr D, Neumann C, Ryland S, Sirk D, Trejos T (2023) Interpretation and report writing in forensic comparisons of paint evidence: An interlaboratory exercise. *Forensic Chemistry* 35. https://doi.org/10.1016/j.forc.2023.100513

[8] National Research Council (2009) Strengthening forensic science in the United States: a path forward. (Washington, D.C).

[9] Thompson WC (2023) Shifting decision thresholds can undermine the probative value and legal utility of forensic pattern-matching evidence. *Proceedings of the National Academy of Sciences of the United States of America* 120(41). https://doi.org/10.1073/pnas.2301844120

[10] Biedermann A, Taroni F, Thompson WC (2011) Using graphical probability analysis (Bayes Nets) to evaluate a conditional DNA inclusion. *Law, Probability and Risk* 10(2):89–121. https://doi.org/10.1093/lpr/mgr007

[11] Kalafut T, Curran JM, Coble MD, Buckleton J (2024) *Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: a case study comparing STRmix™ and TrueAllele™.* (John Wiley and Sons Inc), Vol. 69. https://doi.org/10.1111/1556-4029.15405

[12] Aitken C, Roberts P, Jackson G (2010) *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings Fundamentals of Probability and Statistical Evidence in Criminal Proceedings Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses* (RSS).

[13] Royal Statistical Society (2019) Statistics and probability for advocates.

[14] Royal Statistical Society (2020) The use of statistics in legal proceedings: a primer for courts. (The Royal Society: The Royal Society of Edinburgh).

Suggested References (from Speakers and Attendees):

[1] Almazrouei MA, Kukucka J, Morgan RM, Levy I (2024). Unpacking workplace stress and forensic expert decision-making: From theory to practice. *Forensic Science International: Synergy*, 8, 100473. https://doi.org/10.1016/j.fsisyn.2024.100473

[2] Almazrouei MA, Dror IE, Morgan RM (2019). The forensic disclosure model: What should be disclosed to, and by, forensic experts? *International Journal of Law, Crime and Justice*, *59*, 100330. https://doi.org/10.1016/j.ijlcj.2019.05.003

[3] Bali AS, Edmond G, Ballantyne KN, Kemp RI, Martire KA (2021). Corrigendum to Communicating forensic science opinion: An examination of expert reporting practices [Sci. Justice 60 (3) (2020) 216-224]. *Science & Justice, 61*(4), 449-450. https://doi.org/10.1016/j.scijus.2021.04.001

[4] Bali AS, Edmond G, Ballantyne KN, Kemp RI, Martire KA (2020). Communicating forensic science opinion: An examination of expert reporting practices. *Science & Justice*, *60*(3), 216-224. https://doi.org/10.1016/j.scijus.2019.12.005

[5] Bali AS, Martire KA, Edmond G (2021). Lay comprehension of statistical evidence: A novel measurement approach. *Law and Human Behavior, 45*(4), 370–390. https://doi.org/10.1037/lhb0000457

[6] Biedermann A, Taroni F, Thompson WC (2011). Using graphical probability analysis (Bayes Nets) to evaluate a conditional DNA inclusion. *Law, Probability & Risk*, *10*(2), 89-121. https://doi.org/10.1093/lpr/mgr007

[7] Blastland M, Freeman AL, van der Linden S, Marteau TM, Spiegelhalter D (2020). Five rules for evidence communication. *Nature*, 587(7834), 362-364. https://doi.org/10.1038/d41586-020-03189-1

[8] Buckleton J, Bright JA, Taylor D, Evett I, Hicks T, Jackson G, Curran JM (2014). Helping formulate propositions in forensic DNA analysis. *Science & Justice*, *54*(4), 258-261. http://dx.doi.org/10.1016/j.scijus.2014.02.007

[9] Carlson L, Kennedy J, Zeller KA, Busey T (2022). Describing communication during a forensic investigation using the Pebbles on a Scale metaphor. *Forensic Science International: Synergy*, 4, 100199. https://doi.org/10.1016/j.fsisyn.2021.100199.

[10] Champod C, Biedermann A, Vuille J, Willis S, De Kinder J (2016). ENFSI guideline for evaluative reporting in forensic science: A primer for legal practitioners. *Criminal Law and Justice Weekly*, 180(10), 189-193. http://www.criminallawandjustice.co.uk/features/ENFSI-Guideline-Evaluative-Reporting-Forensic-Science

[11] Champod C, Eldridge H, Lambert S (2020). A primer on error rates in fingerprint examination. *Zenodo*. https://doi.org/10.5281/zenodo.3734560

[12] Chin JM, Ibaviosa MC (2022). Beyond CSI: Calibrating public beliefs about the reliability of forensic science through openness and transparency. *Science & Justice, 62*(3), 272-283. https://doi.org/10.1016/j.scijus.2022.02.006.

[13] Chin JM, Cullen JH, Clarke B (2023). The prejudices of expert evidence. *Monash Law Review*, 48(2), 1-39. https://doi.org/10.31222/osf.io/nxcvy.

[14] Chin JM, Growns B, Sebastian J, Page MJ, Nakagawa S (2022). The transparency and reproducibility of systematic reviews in forensic science. *Forensic Science International,* 340*,* 111472. https://doi.org/10.1016/j.forsciint.2022.111472.

[15] Chin JM, McFadden R (2020). Expert witness codes of conduct for forensic practitioners: a review and proposal for reform" *Canadian Journal of Law and Justice*. 2, 23. https://doi.org/10.31228/osf.io/54jau

[16] Chin JM, Ribeiro G, Reardon A (2019). Open Forensic Science. *The Journal of Law and the Biosciences*. 6(1), 255. https://doi.org/10.1093/jlb/lsz009
[17] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998). A hierarchy of propositions: deciding which level to address in casework. *Science & Justice*, 38(4), 231-239. https://doi.org/10.1016/S1355-0306(98)72117-3
[18] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998). A model for case assessment and interpretation. *Science & Justice*, 38(3), 151-156. https://doi.org/10.1016/S1355-0306(98)72099-4
[19] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1999). Case pre-assessment and review in a two-way transfer case. *Science & Justice*, 39(2), 103-111. https://doi.org/10.1016/S1355-0306(99)72028-9
[20] De Keijser J, Elffers H (2012). Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. *Psychology, Crime & Law*, *18*(2), 191-207. https://doi.org/10.1080/10683161003736744
[21] Edmond G, Martire KA, San Roque M (2017). Expert reports and the forensic sciences. *University of New South Wales Law Journal*, *40*(2), 590–637. 637. https://search.informit.org/doi/10.3316/agis_archive.20172705
[22] Edmond G, Thompson MB, Tangen JM (2014). A guide to interpreting forensic testimony: Scientific approaches to fingerprint evidence. *Law, Probability & Risk, 13*(1), 1-25. https://doi.org/10.1093/lpr/mgt011
[23] Eldridge H (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*, 1, 24-34. https://doi.org/10.1016/j.fsisyn.2019.03.001
[24] Evett IW (1995). Avoiding the transposed conditional. *Science & Justice*, 35(2), 127-132. https://doi.org/10.1016/S1355-0306(95)72645-4
[25] Evett IW (1998). Towards a uniform framework for reporting opinions in forensic science casework. *Science & Justice*, 3(38), 198-202. https://doi.org/10.1016/S1355-0306(98)72105-7
[26] Evett IW, & Buckleton, J. (1990). The interpretation of glass evidence. A practical approach. *Journal of the Forensic Science Society*, 30(4), 215-223. https://doi.org/10.1016/S0015-7368(90)73342-7
[27] Evett IW, Gill, P. D., Jackson, G., Whitaker, J., & Champod, C. (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*, 47(3), 520-530. https://doi.org/10.1520/JFS15291J
[28] Evett IW, Jackson G, Lambert JA (2000). More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science & Justice*, 40(1), 3-10. https://doi.org/10.1016/s1355-0306(00)71926-5
[29] Evett IW, Jackson G, Lambert JA, & McCrossan, S. (2000). The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice* 40(4), 233-239. https://doi.org/10.1016/s1355-0306(00)71993-9
[30] Evett I (2015). The logical foundations of forensic science: towards reliable knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1674), 20140263. https://doi.org/10.1098/rstb.2014.0263
[31] Expert Working Group on Human Factors in Forensic DNA Interpretation (2024). Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach. National Institute of Standards and Technology, Gaithersburg, MD, NIST IR 8503. https://doi.org/10.6028/NIST.IR.8503

[32] Georgiou N, Morgan RM, French JC (2020). Conceptualising, evaluating and communicating uncertainty in forensic science: Identifying commonly used tools through an interdisciplinary configurative review. *Science & Justice*, 60(4), 313-326. https://doi.org/10.1016/j.scijus.2020.04.002

[33] Hackman L (2021). Communication, forensic science, and the law. Wiley Interdisciplinary Reviews: Forensic Science, 3(2), e1396. https://doi.org/10.1002/wfs2.1396

[34] Hackman, L. (2021). Communication, forensic science, and the law. *Wiley Interdisciplinary Reviews: Forensic Science*, 3(2), e1396. https://doi.org/10.1002/wfs2.1396

[35] Heavey AL, Turbett G R, Houck MM, Lewis S W (2023). Management and disclosure of quality issues in forensic science: A survey of current practice in Australia and New Zealand. *Forensic Science International: Synergy*, *7*, 100339. https://doi.org/10.1016/j.fsisyn.2023.100339

[36] Hicklin RA, Ulery BT, Ausdemore M, Buscaglia J (2020). Why do latent fingerprint examiners differ in their conclusions? *Forensic Science International*, 316, 110542. https://doi.org/10.1016/j.forsciint.2020.110542

[37] Hicks T, Biedermann A, De Koeijer JA, Taroni F, Champod C, Evett IW (2015). The importance of distinguishing information from evidence/observations when formulating propositions. *Science & Justice*, 55, 520-525. https://doi.org/10.1016/j.scijus.2015.06.008

[38] Hicks T, Buckleton J, Castella V, Evett I, Jackson G (2022). A logical framework for forensic DNA interpretation. *Genes*, 13(6), 957. https://doi.org/10.3390/genes13060957

[39] Houck MM (1999). Statistics and Trace Evidence: The Tyranny of Numbers. *Forensic Science Communications*, 1(3). https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct1999/houck.htm

[40] Houck MM, Chin JM, Swofford H, Gibb C (2022). Registered reports in forensic science. *Royal Society Open Science, 9*:(11), 221076. https://doi.org/10.1098/rsos.221076

[41] Howes LM (2015). The communication of forensic science in the criminal justice system: A review of theory and proposed directions for research. *Science & Justice*, *55*(2), 145–154. https://doi.org/10.1016/j.scijus.2014.11.002

[42] Howes LM, Martire KA, Kelty SF (2014). Response to recommendation 2 of the 2009 NAS report – standards for formatting and reporting expert evaluative opinions: Where do we stand? *Forensic Science Policy & Management: An International Journal, 5*(1-2), 1-14. https://doi.org/10.1080/19409044.2014.880973

[43] Jackson G (2013). Understanding forensic science opinions. In *Handbook of Forensic Science* (pp. 419-445). J. Fraser, R. Williams (Eds.), Handbook of Forensic Science, Willan, Devon, UK (2009), pp. 419-445 Willan. https://doi.org/10.4324/9781843927327.ch16

[44] Jackson G, Biedermann A (2019). “Source” or “activity”. What is the level of issue in a criminal trial? *Significance* 2019, 16, 36–39. https://doi.org/10.1111/j.1740-9713.2019.01253.x

[45] Kalafut T, Curran JM, Coble MD, Buckleton J. (2024) Commentary on: Thompson WC. Uncertainty in probabilistic genotyping of low template DNA: a case study comparing STRmix™ and TrueAllele™. *Journal of Forensic Science*. 2023; 68 (3): 1049-63. *Journal of forensic sciences*, *69*(1), 371-377. https://doi.org/10.1111/1556-4029.15405

[46] Kaye DH (2016). Hypothesis testing in law and forensic science: a memorandum. *Harvard. Law. Review. Forum*., vol. 130(, no. 5), March 2017, pp. 127-136. https://heinonline.org/HOL/P?h=hein.journals/forharoc130&i=130.

[47] Kaye DH, Koehler JJ (1991) Can Jurors Understand Probabilistic Evidence? *Journal of the Royal Statistical Society. Series A*, Vol. 154, (No. 1), pp. 75-81. https://doi.org/10.2307/2982696

[48] Kind SS 1986. Doctrine, science, belief, evidence. *Journal of Forensic Science Society*, 26, 85-94. https://doi.org/10.1016/S0015-7368(86)72452-3

[49] Kloosterman A, Sjerps M, Quak A (2014). Error rates in forensic DNA analysis: Definition, numbers, impact and communication. *Forensic Science International: Genetics*, *12*, 77-85. https://doi.org/10.1016/j.fsigen.2014.04.014

[50] Koehler JJ (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior* 25(5):493– 513. https://doi.org/10.1023/A:1012892815916

[51] Lavine M, Schervish MJ (1999). Bayes factors: What they are and what they are not. *The American Statistician*, 53(2), 119-122. https://doi.org/10.1080/00031305.1999.10474443

[52] Lindley D (2004a). Bayesian thoughts. *Significance*, 1(2), 73-75. https://doi.org/10.1111/j.1740-9713.2004.027.x

[53] Lindley D (2004b). That wretched prior. *Significance*, 1(2), 85-87. https://doi.org/10.1111/j.1740-9713.2004.026.x

[54] Lindsey S, Hertwig R, Gigerenzer G (2002) Communicating statistical DNA evidence. *Jurimetrics*. 43:147–163. https://doi.org/10.1126/science.290.5500.2261

[55] Lund SP, Iyer H (2017). Likelihood ratio as weight of forensic evidence: a closer look. *Journal of Research of the National Institute of Standards and Technology*, *122*, 1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7339646/

[56] Lund SP, Iyer H (2024). The Influence of Validation Data on Logical and Scientific Interpretations of Forensic Expert Opinions. arXiv preprint arXiv:2403.02663. https://doi.org/10.48550/arXiv.2403.02663

[57] Marquis R, Biedermann A, Cadola L, Champod C, Gueissaz L, Massonnet G, Mazzella WD, Taroni F, Hicks T (2016). Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. *Science & Justice*, 56(5), 364-370. https://doi.org/10.1016/j.scijus.2016.05.009

[58] Martire KA (2018). Clear communication through clear purpose: understanding statistical statements made by forensic scientists. *Australian Journal of Forensic Sciences*, 50(6), 619-627. https://doi.org/10.1080/00450618.2018.1439101

[59] Martire KA, Edmond G (2020). How well do lay people comprehend statistical statements from forensic scientists. *Handbook of Forensic Statistics*, 201-224. Chapman and Hall/CRC. https://doi.org/10.1201/9780367527709-9

[60] Martire KA, Watkins I (2015). Perception problems of the verbal scale: A reanalysis and application of a membership function approach. *Science & Justice, 55*(4)*,* 264-273. https://doi.org/10.1016/j.scijus.2015.01.002

[61] Martire KA, Edmond G, Navarro D (2020). Exploring juror evaluations of expert opinions using the Expert Persuasion Expectancy framework. *Legal and Criminological Psychology*, *25*(2), 90-110. https://doi.org/10.1111/lcrp.12165

[62] Martire KA, Edmond G, Navarro DJ, Newell BR (2017). On the likelihood of "encapsulating all uncertainty. *Science & Justice, 57*(1), 76 -79. https://doi.org/10.1016/j.scijus.2016.10.004

[63] Martire KA, Kemp RI, Newell BR (2013). The psychology of interpreting expert evaluative opinions. *Australian Journal of Forensic Sciences, 45*(3), 305-314. https://doi.org/10.1080/00450618.2013.784361

[64] Martire KA, Kemp RI, Sayle MA, Newell BR (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Sciences International, 240*, 61-68. https://doi.org/10.1016/j.forsciint.2014.04.005

[65] Martire KA, Kemp RI, Watkins I, Sayle MA, Newell BR (2013). The expression and interpretation of uncertain forensic science evidence: Verbal equivalence, evidence strength and the weak evidence effect. *Law & Human Behavior, 37*(3), 197-207. https://doi.org/10.1037/lhb0000027
[66] McQuiston-Surrett D, Saks MJ (2017). Communicating opinion evidence in the forensic identification sciences: Accuracy and impact. In *Expert Evidence and Scientific Proof in Criminal Trials* (pp. 421-451). Routledge. https://doi.org/10.4324/9781315094205
[67] National Institute of Forensic Science Australia New Zealand (NIFS) (2017). An Introductory Guide to Evaluative Reporting. www.nifs.org.au
[68] National Research Council (2009). Strengthening forensic science in the United States: a path forward. National Academies Press.
[69] Neal TM, Martire KA, Johan JL, Mathers EM, Otto RK (2022). The law meets psychological expertise: Eight best practices to improve forensic psychological assessment. *Annual review of law and social science*, *18*(1), 169-192. https://doi.org/10.1146/annurev-lawsocsci-050420-010148
[70] Nordgaard A, Ansell R, Drotz W, Jaeger L (2011). Scale of conclusions for the value of evidence. Law, Probability and Risk, 11, 1-24. https://doi.org/10.1093/lpr/mgr020
[71] Nordgaard A, Ansell R, Drotz W, Jaeger L (2012). Scale of conclusions for the value of evidence. *Law, Probability & Risk*, 11(1), 1-24. https://doi.org/10.1093/lpr/mgr020
[72] Nordgaard A, Hedell R, Ansell R (2012). Assessment of forensic findings when alternative explanations have different likelihoods-"Blame-the-brother"-syndrome. *Science & Justice*, 52 (4) 226-36. https://doi.org/10.1016/j.scijus.2011.12.001
[73] Pfister R (2022). Towards a theory of abduction based on conditionals. *Synthese*, 200(3), 206. https://doi.org/10.1007/s11229-022-03581-6
[74] Rassin E, Arbiyah N, Boskovic I, Otgaar H, Merckelbach H (2022). Likelihood ratios in psychological expert opinion, and their reception by professional judges. *The International Journal of Evidence & Proof*, *26*(4), 325-341. https://doi.org/10.1177/13657127221119545
[75] Reid CA, Howes LM (2020). Communicating forensic scientific expertise: An analysis of expert reports and corresponding testimony in Tasmanian courts. *Science & Justice*, *60*(2), 108-119. https://doi.org/10.1016/j.scijus.2019.09.007
[76] Ribeiro, G., Likwornik, H., & Chin, J. M. (2023). Visual decision aids: Improving laypeople's understanding of forensic science evidence. Journal of Applied Research in Memory and Cognition, 12(2), 230–240. https://doi.org/10.1037/mac0000026
[77] Ribeiro G, Tangen JM, McKimmie BM (2019). Beliefs about error rates and human judgment in forensic science. *Forensic Science International, 297*(1), 138–147. https://doi.org/10.1016/j.forsciint.2019.01.034
[78] Ribeiro G, Tangen J, McKimmie B (2020). Does DNA evidence in the form of a likelihood ratio affect perceivers' sensitivity to the strength of a suspect's alibi? *Psychonomic Bulletin & Review, 27*(6), 1325-1332. https://doi.org/10.3758/s13423-020-01784-x
[79] Ristenbatt III RR, Hietpas J, De Forest PR, Margot PA (2022). Traceology, criminalistics, and forensic science. *Journal of Forensic Sciences*, 67(1), 28-32. https://doi.org/10.1111/1556-4029.14860
[80] Ross R, Martire KA, Kramer K (2017). Consistent with: What doctors say and jurors hear. *Australian Journal of Forensic Sciences, 51*(1),109–116. https://doi.org/10.1080/00450618.2017.1324583
[81] Samie L, Champod C, Delémont S, Basset P, Hicks T, Castella V (2022). Use of Bayesian Networks for the investigation of the nature of biological material in casework. *Forensic Science International*, 331, 111174. https://doi.org/10.1016/j.forsciint.2022.111174

[82] Searston RA, Thompson MB, Robson SG, Corbett BJ, Ribeiro G, Edmond G, Tangen JM (2019). Truth and transparency in expertise research. *Journal of Expertise, 2*(4), 199–209. https://doi.org/10.31234/osf.io/bn85g

[83] Sjerps M (2020). Probabilistic considerations when interpreting database search and selection effects. In *Handbook of Forensic Statistics* (pp. 325-340). Chapman and Hall/CRC.

[84] Sjerps MJ, Berger CE (2012). How clear is transparent? Reporting expert reasoning in legal cases. *Law, Probability and Risk*, *11*(4), 317-329. https://doi.org/10.1093/lpr/mgs017

[85] Sjerps M, Biesheuvel DB (1999). The interpretation of conventional and 'Bayesian' verbal scales for expressing expert opinion: A small experiment among jurists. *International Journal of Speech, Language and the Law*, *6*(2), 214-227. https://doi.org/10.1558/ijsll.v6i2.214

[86] Smith AM, Neal TMS (2021). The distinction between discriminability and reliability in forensic science. *Science & Justice, 61*(4), 319-331. https://doi.org/10.1016/j.scijus.2021.04.002

[87] Stoney DA (1991). What made us ever think we could individualize using statistics?. *Journal of the Forensic Science Society*, 31(2), 197-199. https://doi.org/10.1016/s0015-7368(91)73138-1

[88] Summersby S, Edmond G, Kemp RI, Ballantyne KN, Martire KA (2024). The effect of following best practice reporting recommendations on legal and community evaluations of forensic examiners reports. *Forensic Science International*, 359, 112034. https://doi.org/10.1016/j.forsciint.2024.112034

[89] Thompson WC (2009). Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. *Law, Probability & Risk*, 8(3), 257-276. https://doi.org/10.1093/lpr/mgp013

[90] Thompson WC, Grady RH, Lai E, Stern H (2018). Perceived strength of forensic scientists' reporting statements about source conclusions, *Law, Probability and Risk*, 17(2), 133–155. https://doi.org/10.1093/lpr/mgy012

[91] Thompson W, Black J, Jain A, Kadane J (2017). Forensic science assessments: A quality and gap analysis–Latent fingerprint examination. Washington, DC: *American Association for the Advancement of* Science (AAAS). DOI: 10.1126/srhrl.aag2874. https://www.aaas.org/resources/latent-fingerprint-examination

[92] Thompson WC, Schumann EL (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, *11,* 167-187. https://doi.org/10.1007/BF01044641

[93] Thompson WC, Scurich N (2019). How cross-examination on subjectivity and bias affects jurors' evaluations of forensic science evidence. *Journal of Forensic Science*s, 64(5): 1379-1388. https://doi.org/10.1111/1556-4029.14031.

[94] Thompson WC (2018). How should forensic scientists present source conclusions? *Seton Hall Law Review*, 48(3): 774-813. http://scholarship.shu.edu/shlr/vol48/iss3/9

[95] Thompson WC (2023). Shifting decision thresholds can undermine the probative value and legal utility of forensic pattern-matching evidence. *Proceedings of the National Academy of Sciences, USA*, 120(41): e2301844120e. https://doi.org/10.1073/pnas.2301844120

[96] Thompson WC (2023). Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRMix™ and TrueAllele™ *Journal of Forensic Sciences*, 68(3): 1049-1063. https://doi.org/10.1111/1556-4029.15225

[97] Thompson WC, Grady RH, Lai E, Stern H (2018). Perceived strength of forensic scientists' reporting statements about source conclusions. *Law, Probability & Risk,* 17(2): 133-155. https://doi.org/10.1093/lpr/mgy012

[98] Thompson WC, Kaasa SO, Peterson T (2013). Do jurors give appropriate weight to forensic identification evidence? *Journal of Empirical Legal Studies*, *10(2):*359-97. https://doi.org/10.1111/jels.12013

[99] Thompson WC, Taroni F, Aitken CGG (2003). How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences*, *48(1),* 47-54. https://doi.org/10.1520/JFS2001171

[100] Thompson WC, Vuille J, Biedermann A, Taroni F (2013). The role of prior probability in forensic assessments. *Frontiers in Genetics*, 4: 220-223. https://doi.org/10.3389/fgene.2013.00220

[101] Thompson WC, Vuille J, Taroni F, Biedermann A (2018). After Uniqueness: The Evolution of Forensic Science Opinion. *Judicature,* 102(1): 18-27. ISSN: 00225800

[102] Thompson WC Newman EJ (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law & Human Behavior. 39(4):* 332-349. https://doi.org/10.1037/lhb0000134

[103] Wray-Jones N, Chin J (2021). "Can ADR improve expert evidence?" (2021) 95 *The Australian Law Journal* 467., https://doi.org/10.21428/cb6ab371.fc7e92d5

[104] Younan M, Martire KA (2021). Likeability and expert persuasion: Dislikeability reduces the perceived persuasiveness of expert evidence. *Frontiers in Psychology*, *12*, 785677. https://doi.org/10.3389/fpsyg.2021.785677

Appendix A. List of Symbols, Abbreviations, and Acronyms

CSAFE – Center for Statistics and Applications in Forensic Evidence
CPI – Combined probability of inclusion
DNA – Deoxyribonucleic acid
ENFSI – European Network of Forensic Science Institutes
FRE – Federal Rules of Evidence
LR – Likelihood ratio
MML – Material Measurement Laboratory
NFI – Netherlands Forensic Institute
NIST – National Institute of Standards and Technology
OSAC – Organization of Scientific Area Committees for Forensic Science
RMP – Random match probability
SED – Statistical Engineering Division
SOP – Standard operating procedure
SPO – Special Programs Office
SWGDAM – Scientific Working Group for DNA Analysis Methods
SWGMAT – Scientific Working Group for Materials Analysis

Appendix B. List of Invited Speakers and Registered Attendees

Role	Name (alphabetical by role)	Affiliation
Speaker	Miriam Angel	Los Angeles Police Department Laboratory
Speaker	Alex Biedermann	University of Lausanne
Speaker	Julia Lloyd Burrill	Alan Alda Center for Communicating Science + Leverhulme Research Centre for Forensic Science
Speaker	Heidi Eldridge	George Washington University
Speaker	Lynn Garcia	Texas Forensic Science Commission
Speaker	Clinton Hughes	Brooklyn Defender Services
Speaker	David Kanaris	Alaska State Crime Lab
Speaker	Jarrah Kennedy	Kansas City Police Crime Laboratory
Speaker	Kristy Martire	University of New South Wales, Sydney
Speaker	Anders Nordgaard	Swedish Police Authority - National Forensic Centre
Speaker	Kate Philpott	Virginia Commonwealth University/RFCT
Speaker	Dan Rabinowitz	Columbia University
Speaker	J. Michael Ryan	Superior Court for the District of Columbia
Speaker	John D. Schmid	Sixth District (Minnesota) Public Defender's Office
Speaker	Lora Sims	Ideal Innovations, Inc.
Speaker	Maria Johanna Sjerps	Netherlands Forensic Institute
Speaker	Hal Stern	University of California, Irvine
Speaker	William Thompson	University of California, Irvine
Speaker	Tatiana Trejos	West Virginia University, Department of Forensic and Investigative Science
Moderator	Sanne Aalbers	NIST Forensic Science Research Applied Genetics Focus Area
Moderator	John Butler	NIST Special Programs Office/Forensic Foundations Program
Moderator	Will Guthrie	NIST Statistical Engineering Division
Planning team	Hari Iyer	NIST Statistical Engineering Division
Moderator	Sandra Koch	NIST Special Programs Office/Forensic Foundations Program
Speaker	Steve Lund	NIST Statistical Engineering Division
Moderator	Melissa Taylor	NIST Special Programs Office/Forensic Science Research Program
Moderator	Sheila Willis	Private consultant
Notetaker	Darby Harris	George Washington University
Notetaker	Kelsey Johns	George Washington University
Notetaker	Kelly Sauerwein	NIST Special Programs Office/Forensic Foundations Program
Notetaker	Holly Zhao	George Washington University

Attendee	Mohammed Almazrouei	Yale University
Attendee	Thomas Baird	U.S. Department of Health and Human Services (HHS)
Attendee	Catherine Brown	Collaborative Testing Services
Attendee	JoAnn Buscaglia	FBI Laboratory
Attendee	Crystal DeGrange	NIST Special Programs Office/Forensic Science Standards Program
Attendee	Vinny Desiderio	NIST Special Programs Office/Forensic Science Standards Program
Attendee	Wendy Dinova-Wimmer	Adobe
Attendee	Allison Getz	NIST Special Programs Office/Forensic Science Standards Program
Attendee	Jeff Gilleran	MD Office of the Public Defender
Attendee	Melissa Gische	FBI Laboratory
Notetaker	Darby Harris	George Washington University
Attendee	Jack Hietpas	John Jay College- City University of New York
Attendee	Clinton Hughes	Brooklyn Defender Services
Attendee	Ted Hunt	FBI Laboratory
Attendee	Amy Jenkins	Virginia Department of Forensic Science
Attendee	David Kaye	Pennsylvania State University
Attendee	John Paul Jones	NIST Special Programs Office/Forensic Science Standards Program
Attendee	Susannah Kehl	FBI Laboratory
Attendee	Jeffrey Kukucka	Towson University
Attendee	Julia Leighton	Retired & OSAC
Attendee	Kristen Mertins	Houston Forensic Science Center
Attendee	Dawn Moore Boswell	University of North Texas Health Science Center
Attendee	Raymond Valerio	Queens County District Attorney
Attendee	Johannes Soons	NIST Forensic Science Research Firearms and Toolmarks Focus Area
Attendee	Erica Ramsos	NIST Forensic Science Research Quality Assurance Focus Area
Attendee	Peter Valone	NIST Forensic Science Research Applied Genetics Focus Area
Attendee	Alan Zheng	NIST Forensic Science Research Firearms and Toolmarks Focus Area