NISTIR 8225 DRAFT

NIST Scientific Foundation Reviews

John M. Butler
Hari Iyer
Rich Press
Melissa K. Taylor
Peter M. Vallone
Sheila Willis*
*International Associate under contract; former director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8225-draft

NISTIR 8225 DRAFT

NIST Scientific Foundation Reviews

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Associate Director of Laboratory Programs

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs
Director's Office

*International Associate under contract; former director of Forensic Science Ireland

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8225-draft

September 2018

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

National Institute of Standards and Technology Internal Report 8225 DRAFT
(September 2018)

Acknowledgments: Input and suggestions on this document and project were made by Richard Cavanagh, Barbara Guttman, Bill MacCrehan, Kathryn Miller, Kathy Sharpless, Jack Ballantyne, Todd Bille, Jennifer Breaux, Robin Cotton, Roger Frappier, Bruce Heidebrecht, Keith Inman, Eugene Lien, Tamyra Moretti, Lisa Schiermeier-Wood, Joel Sutton, Ray Wickenheiser, and Charlotte Word.

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

Public comment period: September 24, 2018 through November 19, 2018

National Institute of Standards and Technology
Attn: Special Programs Office – Scientific Foundation Review
100 Bureau Drive, MS 4701
Gaithersburg, MD 20899-4701

Email: scientificfoundationreviews@nist.gov

All comments, including commenter name and affiliation, will be published at https://www.nist.gov/topics/forensic-science/draft-nistir-8225-nist-scientific-foundation-reviews.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

Abstract

The National Institute of Standards and Technology (NIST) is a scientific research agency that works to advance measurement science, standards, and technology and that has been working to strengthen forensic science methods for almost a century. In recent years, several scientific advisory bodies [2-4] have expressed the need for scientific foundation reviews of forensic disciplines and identified NIST as an appropriate agency for conducting them. The purpose of a scientific foundation review is to document and consolidate information supporting the methods used in forensic analysis and identify knowledge gaps where they exist. In fiscal year 2018, Congress appropriated funds for NIST to conduct scientific foundation reviews [5], p. 22. NIST has begun reviews of DNA mixture interpretation and bitemark analysis. In addition to providing insights into these specific disciplines, the initial reviews serve as pilot studies which will guide future efforts of this type. This document outlines NIST's approach to conducting scientific foundation reviews, including data sources, evaluation criteria, and expected outputs.

Keywords

forensic science, scientific foundation review, technical merit evaluation

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

Table of Contents

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

1. What is a Scientific Foundation Review?

A scientific foundation review is a study that seeks to document and evaluate the foundations of a scientific discipline, that is, the trusted and established knowledge that supports and underpins the discipline's methods. NIST is conducting scientific foundation reviews in forensic science. These reviews seek to answer the question: "What empirical data exist to support the methods that forensic science practitioners use to analyze evidence?"

The central activity of forensic science is to make associations between pieces of evidence or between evidence and known items in order to shed light on past events and actions. Forensic practitioners do this by comparing and classifying items based on selected features such as the minutiae of a fingerprint, the alleles in a DNA sample, or the toolmarks on a fired bullet.

For each forensic method studied, we will evaluate whether the selected features are characterized and measurable; to what extent the discriminating power of those features is known; and whether the factors that affect the transferability and persistence of those features are understood.

Each foundation review will be different depending on the specifics of the discipline, but all will be based on the following generalized approach.

1.1. What Data Sources Will We Use?

Because peer-reviewed publications are essential building blocks of a respected edifice of scientific knowledge, studies that address the reliability of forensic methods would ideally be present in a discipline's published, peer-reviewed, and well-cited scientific literature. However, a focus on peer-reviewed literature alone may not provide a complete picture of a discipline's available body of knowledge. For instance, data from laboratory validation studies may not be publicly available or published. Therefore, NIST scientific foundation reviews are designed to seek input by:

- collecting and evaluating the peer-reviewed literature
- assessing available data from interlaboratory studies, proficiency tests, and laboratory validation studies
- exploring other available information including position statements and non-peer reviewed literature
- obtaining input from members of the relevant community through interviews, workshops, working groups and other formats for the open exchange of ideas and information.

Obtaining input from experts outside of NIST is an integral component of a NIST scientific foundation review. This will help ensure that these reviews capture the full breadth of knowledge that forensic practitioners and researchers consider foundational to their discipline.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

1.2. How Will We Evaluate the Data?

After gathering information, we will evaluate it against the following criteria:

1) *Retrievable:* Does the information appear in a peer-reviewed journal or book that is indexed? Is the reference citation accessible by search engines? If not, is it published online or otherwise reasonably available for review by others?
2) *Reliable:* Can the information be verified against other sources? Can the reported methods do what they claim to do? Are capabilities and limitations of the methods understood? Are the methods clearly explained so that they can be reproduced? Are experimental materials with known values used to show that accurate conclusions can be made? Are statistically significant sample sizes used?
3) *Respected*: Has the information been cited as being useful by other researchers or practitioners in the scientific literature? Has it been scrutinized or reviewed by others?

Retrievability is among the criteria because transparency and openness are hallmarks of good science [6]. Therefore, we believe that for something to be considered foundational, it must be reasonably accessible to anyone who wishes to review it.

Where peer-reviewed publications are not available, transparency and accessibility can help fill the gap. For instance, publishing validation data from forensic laboratories online would allow for "open peer review" [7].

1.3. What Information Will We Report?

The outcome of each NIST scientific foundation review will be a publicly-available report that may be accompanied by additional online resources. We expect that these reports will include:

1) an introduction to the issues involved
2) historical perspectives of the field and current methods in use
3) a discussion of the NIST review team's efforts to collect and evaluate data sources, literature, and input received from experts in the field
4) a complete list of literature and other sources used
5) a discussion of our findings with regard to scientific foundations
6) key takeaways and considerations for the field

We anticipate that scientific foundations reviews will be useful in a number of ways. First, identifying those methods that are built on a solid scientific foundation will increase trust in those methods. Second, by identifying those parts of the foundation that would benefit from strengthening, a foundation review can provide strategic direction for future research efforts. Third, in an interdisciplinary environment in which legal, academic, and forensic professionals need to understand one another's perspectives, consolidating key points and principles can promote a shared understanding of critical concepts and lead to more effective communication. Fourth, in many disciplines, hundreds of forensic science research articles are published every year, yet time to absorb and discuss those articles is limited. Identifying a

discipline's foundational literature can help a community develop a shared understanding of core principles. In addition, establishing a comprehensive and curated canon can promote a better appreciation for the capabilities and limitations of methods, increase competency, and reduce variability across the field.

What is Science?

When conducting a scientific foundation review, it is important to define the word "science" and what attributes we consider to be "scientific." A succinct statement usually attributed to German philosopher Immanuel Kant is that "science is organized knowledge."[1]

The UK Science Council defines science as "the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence."[2] Thus, science involves data collected, evaluated, and understood in a systematic and logical fashion.[3]

The UK Science Council notes some key attributes of scientific study that include:

- Repetition (a phenomenon can be demonstrated repeatedly)
- Measurements and data
- Experiments
- (Falsifiable) hypotheses
- Critical analyses that consider more than one possibility
- Verification and testing
- Exposure to scrutiny, peer-review, and assessment

Several decades ago the National Academy of Sciences (NAS) prepared a report entitled *On Being a Scientist* that describes "the ethical foundations of scientific practices" and how these foundations "safe-guard the integrity of the scientific enterprise" [1]. This NAS report emphasizes that "science is not done in isolation" and "takes place within a broad social and historical context, which gives substance, direction, and, ultimately, meaning to the work of individual scientists" [1].

On Being a Scientist notes that

> "[scientists] submit their work to be examined by others with the hope that it will be accepted. This process of public, systematic skepticism is critical in science. It minimizes the influence of individual subjectivity by requiring that research results be accepted by other scientists. It also is a powerful inducement for researchers to be critical of their own conclusions, because they know that their objective must be to convince their ablest colleagues, including those with contrasting views… Publication in a scientific journal includes important aspects of quality control – particularly, critical review by peers who can detect mistakes, omissions, and alternative explanations" [1].

[1] https://quoteinvestigator.com/2015/05/18/science/
[2] https://sciencecouncil.org/about-science/our-definition-of-science/
[3] We recognize that methods used in various forensic disciplines may have differing levels of supporting background information available. As noted in the NRC 2009 report (p. 39): "…the term 'forensic science' is used with regard to a broad array of activities, with the recognition that some of these activities might not have a well-developed research base, are not informed by scientific knowledge, or are not developed within the culture

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

2. Why NIST?

NIST was founded in 1901 and is one of the nation's oldest physical science laboratories. Over its long history, NIST has cultivated deep scientific expertise that cuts across a wide range of disciplines. Having that expertise without being a regulatory agency allows NIST to work closely with a broad spectrum of partners.

NIST provides industry, academia, and other government agencies with:

- Expertise in measurement science and best practices in many disciplines, including physics, chemistry, materials science, information technology, and engineering
- World class, unique, cutting-edge research facilities
- Leadership in the development of consensus-based standards, test methods, and specifications that define technical and performance requirements

Drawing on these capabilities, its national networks, international partnerships, and relationship with industry, NIST works to address complex measurement challenges, ranging from the physical (renewable energy sources) to the virtual (cybersecurity and cloud computing), and from fundamental (quantum measurements) to the applied (fire spread rates).

NIST has been involved in forensic science since the 1920s, when physicist Wilmer Souder conducted precision measurements to assist hundreds of investigations involving handwriting, typewriting, and ballistic examinations [8]. NIST's direct involvement in criminal investigations ended in the 1950s, but NIST has been working since then to strengthen the measurements and technologies underpinning methods for analyzing DNA, fingerprints, firearms and toolmarks, and digital evidence, among others. In addition, NIST provides standard reference materials including human DNA, standard bullets, and mass spectral data to U.S. forensic laboratories to help ensure accurate and reliable measurements.

Because NIST is not directly involved in the criminal justice system, its scientists are able to offer an independent perspective on scientific matters bearing on forensic science. NIST furthered its involvement in forensic science following a 2013 Memorandum of Understanding (MOU) between NIST and the Department of Justice (DOJ). This MOU stated that, "Scientifically valid and accurate forensic science strengthens all aspects of our justice system"[10]. Under this MOU, which established the National Commission on Forensic Science (NCFS) and the Organization of Scientific Area Committees for Forensic Science (OSAC), NIST had the following four responsibilities:

1) appoint a senior NIST official to serve as the Co-Chair of the Commission
2) administer and coordinate all necessary support for OSAC
3) conduct research supporting the development and dissemination of methods, standards, and technical guidance for forensic science measurements
4) test and validate select forensic science practices and standards as appropriate

NIST's scientific foundation reviews fulfill the responsibilities outlined in the fourth element of that MOU.

In February 2014, NIST launched OSAC to support the development of documentary standards. Through OSAC, NIST convenes stakeholders and provides technical and scientific guidance and expertise to help stakeholder groups reach a consensus.

2.1. Calls for NIST to Conduct Scientific Foundation Reviews

Several entities have specified the need for scientific foundation reviews. In 2009, the National Research Council published a report entitled *Strengthening Forensic Science in the United States: A Path Forward*, which requested "studies establishing the scientific bases demonstrating the validity of forensic methods" [3], p. 22. More recently, the President's Council of Advisors on Science and Technology (PCAST) [4], the National Commission on Forensic Science (NCFS) [11], and the American Association for the Advancement of Science (AAAS) [12, 13] have published recommendations encouraging further research and studies assessing the scientific foundations of forensic disciplines.

In September 2016, both PCAST and NCFS requested that NIST examine the scientific literature and conduct technical merit evaluations and validation studies of forensic science methods and practices. The NCFS recommended that the results of these technical merit evaluations "be issued by NIST as publicly available resource documents" and that "NIST's evaluation may include but is not limited to: a) research performed by other agencies and laboratories, b) its own intramural research program, or c) research studies documented in already published scientific literature" [14]. NCFS also requested that these evaluation documents "be broadly disseminated in the scientific and criminal justice communities and accompanied by judicial trainings" [14].

During the September 12, 2016 NCFS meeting, NIST leadership announced that the agency would respond to the NCFS requests by conducting a "pilot" scientific foundation review of DNA mixture interpretation, to be followed by reviews of bitemarks and firearms identification [15]. At the final NCFS meeting held on April 10, 2017, then Acting NIST Director Kent Rochford reiterated these plans [16].

In fiscal year 2018, Congress appropriated funding for NIST to conduct "technical merit evaluations." NIST scientific foundation reviews are intended to fulfill this mandate. The first NIST scientific foundation review, a study on DNA mixture interpretation, began in September 2017. A review of bitemark analysis began later that year.

3. Previous Efforts: A Historical Overview

Our approach to the review of forensic science literature builds upon previous efforts and experiences, which are summarized below. These activities, which have often been conducted independent of other on-going or previous efforts, include literature reviews, input from advisory groups, and workshops. Many of these previous efforts have been prospective (i.e., looking to where the field needs to go) rather than introspective (i.e., reflecting on the

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

foundations and support that exist for specific methods). An important goal of our NIST scientific foundation reviews is to consider, compile, and integrate information from previous efforts.

3.1. *Analytical Chemistry* Bi-annual Application Reviews (1983 to 2011)

A number of literature summaries have been gathered over the years to reflect various topics that were published in forensic science. For three decades, the journal *Analytical Chemistry* published a brief review of activities focused primarily in three areas: drugs and poisons; forensic biochemistry; and trace evidence. In the 15 review articles published on alternate years between 1983 and 2011, there were a total of 9263 publications reviewed, of which 1565 were articles related to DNA methods [17].

These *Analytical Chemistry* application reviews surveyed articles published in the *Journal of Forensic Sciences*, *Science and Justice*, *Forensic Science International*, *Forensic Science International: Genetics*, *Journal of the Canadian Society of Forensic Science*, *Journal of Forensic Identification*, *Forensic Science Review*, *Analytical Toxicology*, *The Microscope*, and *Chemical Abstracts*. While each of these reviews provided a nice summary of the breadth of information published in the previous two years, there was no attempt to assess the quality or prioritize the publications in any way. Moreover, as noted previously [17], these reviews were methods-focused to enable readers to find information that might aid forensic laboratory work.

3.2. INTERPOL Literature Review

The International Forensic Science Managers Symposium provides another approach to gathering and discussing forensic science literature. Experts from around the world speak at this symposium, which is held every three years at INTERPOL headquarters in Lyon, France. As part of this gathering, a summary of the published literature from the previous three years is organized into a review article. The approach taken for each discipline varies and the number of publications examined, summarized, and reported on can range from a few dozen to over a thousand.

The 2010-2013 literature summary contains 4832 references from the following disciplines (with number of listed references in parentheses): firearms (159), gunshot residue (49), toolmarks (189), paint (201), fibers and textiles (68), forensic geology (102), arson and fire debris analysis (140), explosives and explosive residues (1341), drug evidence (668), toxicology (324), forensic audio analysis (133), forensic video analysis (31), imaging (256), digital evidence (190), fingermarks and other impressions (472), body fluid identification and DNA typing in forensic biology (114), questioned documents (275), and forensic science management (120). These compiled literature summaries can be accessed on the INTERPOL website at https://www.interpol.int/content/download/21910/206602/version/1/file/IFSMSReviewPapers2013.pdf. Authors of these forensic discipline summaries come from Australia, Belgium, Canada, Finland, France, Hong Kong, Israel, Japan, The Netherlands, Switzerland, the United Kingdom, and the United States (see [17]).

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

The 2013-2016 literature summary contains 4891 references from the following disciplines: firearms (179), gunshot residue (77), toolmarks (104), paint and glass (102), fibers and textiles (92), forensic geosciences (245), fire investigation and debris analysis (194), explosives (646), drugs (1434), toxicology (600), audio analysis (88), video and imaging (108), digital evidence (100), fingermarks and other impressions (536), DNA and biological evidence (75), questioned documents (255), and forensic science management (56). These literature summaries are at https://www.interpol.int/content/download/33314/426506/version/1/file/INTERPOL%2018th%20IFSMS%20Review%20Papers.pdf.

The reviews on DNA involve only 114 articles from 2010 to 2013 and 75 articles from 2013 to 2016. These reviews are minimal in nature, typically involve just a summary listing of the material, and are focused on topics of interest to the authors rather than attempting to be comprehensive. For example, in the 75 articles discussed in the 2013 to 2016 review, selected topics include rapid DNA analysis (11 references), analysis of complex DNA profiles including mixtures and low-template DNA (4 references), and the development of next-generation sequencing and its application to DNA phenotyping (60 references). There are many, many more references on DNA mixture interpretation during this time period that were not covered, which points to the fact that it can be a challenge with any review of the literature to be both effective and thorough.

3.3. Measurement Science Workshops

Since 2012, NIST has conducted or sponsored a number of measurement science workshops to assist the transition of research in specific forensic fields into more effective practice. These workshops are typically webcast from the NIST campus in Gaithersburg, Maryland. Topics have included firearms analysis, DNA mixture interpretation, emerging trends in synthetic drugs, handwriting analysis, cloud computing, mobile forensics, probabilistic genotyping, validation, and forensic science error management (e.g., see Table 7 in [17])[1]. Forensics@NIST conferences have been held bi-annually since 2012 to share research conducted at NIST with the forensic community.

3.4. National Research Council 2009 Report

In November 2005, the United States Congress authorized the National Academy of Sciences (NAS) to conduct a study on forensic science [3]. From January 2007 to November 2008 a 17-member committee met eight times, heard from 70 presenters, and discussed the information received. In February 2009 the National Research Council (NRC) arm of the NAS issued a 352-page report entitled *Strengthening Forensic Science in the United States: A Path Forward*.

This 2009 NRC report, which is often referred to in forensic circles as “the NAS report,” proposed 13 recommendations to improve forensic science in the United States. Recommendation #3 emphasized the need for research “to address issues of accuracy,

[1] An updated list of past NIST events is available at https://www.nist.gov/topics/forensic-science/conferences-and-events.

reliability, and validity in the forensic science disciplines" and encourages funding of peer-reviewed research involving "(a) studies establishing the scientific bases demonstrating the validity of forensic methods, (b) the development and establishment of quantifiable measure of the reliability and accuracy of forensic analyses [that can be expected as forensic evidence conditions vary]…, (c) the development of quantifiable measures of uncertainty in the conclusions of forensic analyses, and (d) automated techniques capable of enhancing forensic technologies" ([3], pp. 22-23). This recommendation also emphasized that research results should be published in respected scientific journals.

In this report, nuclear DNA testing from single-source, high-quality samples are given high marks with statements like: "Among existing forensic methods, only nuclear DNA analysis has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between an evidentiary sample and a specific individual or source" ([3], p. 100; see also p. 7). The 2009 NRC report does not discuss DNA mixture interpretation beyond a brief mention on page 100: "There may be problems in a particular case with how the DNA was collected, examined in the laboratory, or interpreted, such as when there are mixed samples, limited amounts of DNA, or biases due to the statistical interpretation of data from partial profiles."

The 2009 NRC assessment of bitemark analysis notes on page 176: "Despite the inherent weaknesses involved in bitemark comparison, it is reasonable to assume that the process can sometimes reliably exclude suspects." However, "[t]he committee received no evidence of an existing scientific basis for identifying an individual to the exclusion of all others." They emphasize "[s]ome research is warranted in order to identify the circumstances within which the methods of forensic odontology can provide probative value."

The report opines that "the interpretation of forensic evidence is not always based on scientific studies to determine its validity" and that "a body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias" (p. 8).

3.5. White House Subcommittee on Forensic Science (SoFS)

From July 2009 to December 2012, the White House Office of Science and Technology Policy (OSTP) established a federal government effort – a Subcommittee on Forensic Science (SoFS) under the National Science and Technology Council (NSTC) – to work towards potential solutions that could help address the 2009 NRC report recommendations. For a brief timeline of recent U.S. efforts to strengthen forensic science, see Ref. [17].

Five interagency working groups (IWGs) met on almost a monthly basis during the time period in which the SoFS was in existence. The IWG activities involved nearly 200 subject matter experts from 23 Federal departments and agencies as well as 49 participants representing state and local forensic laboratories. One of these IWGs covered Research, Development, Testing, and Evaluation (RDT&E) and as part of their work wrote to the then-existing Scientific Working Groups (SWGs) to request information on literature supporting the scientific foundations of their disciplines.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

In 2011 and 2012, annotated bibliographies were provided to the SoFS RDT&E IWG for 10 forensic disciplines in responses to questions raised (see Ref. [17]). The forensic disciplines represented include: (1) firearms and toolmarks, (2) bloodstain pattern analysis, (3) bitemark (odontology) analysis, (4) fiber analysis, (5) shoeprint and tire tread, (6) latent print analysis, (7) arson investigation and burn pattern analysis, (8) digital evidence, (9) hair analysis, and (10) paints and other coatings. Links to these bibliographies can be found at https://www.nist.gov/topics/forensic-science/working-groups/legacy-scientific-working-groups.

For example, with firearms and toolmarks, SWGGUN and AFTE (Association of Firearms and Toolmark Examiners) prepared a 94-page response to 25 questions on the foundations of their field [18]. These questions ranged from "What literature documents the scientific domains used to inform the foundations of firearm/toolmark analysis?" to "What statistical research has been conducted and applied to firearm and toolmark examinations? What statistical models for firearms and toolmarks have been published?"

Some of the provided bibliographies contain only meeting presentation abstracts to address questions raised on foundational issues. The limited responses obtained by the SoFS RDT&E IWG on some of the foundational questions ultimately led to the National Commission on Forensic Science (NCFS) position statements on scientific literature, the request for NIST to perform technical merit reviews, and the American Association for the Advancement of Science (AAAS) studies described below.

3.6. NSF/NIJ-Funded Workshop

In May 2015, a workshop was held at the American Association for the Advancement of Science (AAAS) in Washington, D.C. that was co-funded by the National Science Foundation (NSF) and the National Institute of Justice (NIJ) [19]. This workshop, entitled "Forensic Science Research and Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature," brought together 17 experts to cover topics of experimental design and statistics, interpretation and assessment, and policy implications regarding scientific foundations of forensic disciplines. Each participant submitted a short essay on their presented topic at the workshop. An important output from this workshop was a 122-page report, which is available from NIJ [20].

A purpose of this workshop was to inform AAAS regarding approaches to examining the literature for foundational studies on selected forensic disciplines (see below). Concurrent with the AAAS forensic science assessments that began in 2015, the Department of Justice and NIST had begun discussing many of these issues and had begun working on other efforts to strengthen forensic science via the National Commission on Forensic Science.

3.7. Insights from the National Commission on Forensic Science

The National Commission on Forensic Science (NCFS), which served as a Federal Advisory Committee to the U.S. Department of Justice from 2013 to 2017, held 13 meetings and

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

approved 43 work products. The adopted work products were either Recommendations to the Attorney General [14] or Views of the Commission [21]. There were seven subcommittees that prepared and presented the work products to the full NCFS.

The Scientific Inquiry and Research Subcommittee drafted and championed five documents that were approved by NCFS:

1) Views on Scientific Literature in Support of Forensic Science and Practice (January 2015) [22]
2) Recommendation to Fund Post-Doctoral Projects to Facilitate Translation of Research into Forensic Science Practice (March 2016) [23]
3) Views on Identifying and Evaluating Literature that Supports the Basic Principles of a Forensic Science Method or Forensic Science Discipline (March 2016) [24]
4) Views on Technical Merit Evaluation of Forensic Science Methods and Practices (June 2016) [25]
5) Recommendation on Technical Merit Evaluation of Forensic Science Methods and Practice (September 2016) [26]

Four of these documents (the only exception being the one on funding post-doctoral projects) apply directly to scientific foundation reviews. Each of these four will be discussed further.

3.7.1. Appropriate Scientific Literature

Some members of the NCFS Scientific Inquiry and Research Subcommittee had been part of the SoFS RDT&E IWG and were familiar with the submissions made a few years before in response to inquiries about foundational literature. As stated in the January 2015 Views document: “A cursory review of the literature citations raised concerns within the NCFS that extend beyond these specific bibliographies: (1) In some cases, it was unclear which literature citations are crucial to support the foundation of a particular forensic science discipline. (2) Some of the cited literature had not undergone a rigorous peer-review process.”

These observations fueled a desire to describe what is appropriate scientific literature to provide support for methods used in forensic practice. The NCFS states: “The goal of this [January 2015] Views document is to provide the framework necessary to address these and broader concerns regarding the status of the scientific foundation of forensic science across its many disciplines and practices.”

In January 2015, the NCFS unanimously approved *Views of the Commission on Scientific Literature in Support of Forensic Science and Practice*: “The NCFS believes that a comprehensive evaluation of the scientific literature is critical for the advancement of forensic science policy and practice in the United States. While other forms of dissemination of research and practice (e.g., oral and poster presentations at meetings, workshops, personal communications, editorials, dissertations, theses, and letters to editors) play an important role in science, the open, peer-reviewed literature is what endures and forms a foundation for further advancements” [27].

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

The NCFS Views document states that "foundational, scientific literature supportive of forensic practice should meet criteria such as the following:

- Peer-reviewed in the form of original research, substantive reviews of the original research, clinical trial reports, or reports of consensus development conferences.
- Published in a journal or book that has an International Standard Number…and recognized expert(s) as authors (for books) or on its Editorial Board (for journals).
- Published in a journal that maintains a clear and publicly available statement of purpose that encourages ethical conduct such as disclosure of potential conflicts of interest integral to the peer review process.
- Published in a journal that utilizes rigorous peer review with independent external reviewers to validate the accuracy in its publications and their overall consistency with scientific norms of practice.
- Published in a journal that is searchable using free, publicly available search engines (e.g., PubMed, Google Scholar, National Criminal Justice Reference Service) that search major databases of scientific literature (e.g., Medline, …).
- Published in a journal that is indexed in databases that are available through academic libraries and other services (e.g., JSTOR, Web of Science, …)."

This Views of the Commission document points out that "the term 'foundation' was used no less than thirty times [in the 2009 NRC report [3]] to emphasize that each forensic discipline must have a scientifically robust and validated basis to its methods, its technologies, and its process of interpreting data." It also notes: "…each forensic discipline must have an underlying foundation that is the result of a rigorous vetting process and that is ultimately captured in the peer-reviewed scientific literature."

It continues: "Scientific literature comprises manuscripts that report empirical data and have been independently peer-reviewed for quality, originality, and relevance to the discipline. To strengthen confidence in results obtained in forensic examinations, each forensic discipline must identify resources that are scientifically credible, valid and with a clear scientific foundation. Such foundational literature in forensic practice should conform to norms across all scientific disciplines. Accordingly, the [NCFS] proposes criteria [those listed above] by which scientific literature can be assessed for its consistency with principles of scientific validity."

3.7.2. Identifying and Evaluating Literature

In March 2016, the NCFS approved "Views of the Commission Regarding Identifying and Evaluating Literature that Supports the Basic Principles of a Forensic Science Method or Forensic Science Discipline" [2].

This Views document states: "In any scientific discipline, an on-going process to evaluate the weight and merit of published materials must be established. The NCFS is aware of past and on-going efforts to establish the scientific foundation of forensic discipline[s] through literature reviews and generation of bibliographies. As part of these efforts, it is the view of

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

the NCFS that scientific literature must be evaluated and be vetted through an objective and critical review process using tenets based on general scientific principles and practices. These tenets must be satisfied before any form of scientific literature is included in, and considered part of, a forensic discipline's scientific foundation. Herein, foundational literature is intended to refer to that upon which a discipline has derived, developed, or defined practices and procedures examined and validated by a given discipline and applied within a legal, medicolegal, or judicial setting."

The Commission provides some specific guidance by asking 15 questions to provide a framework for an objective and critical review. The document states: "The following tenets of literature review should be considered in a critical review process that evaluates the merit of an individual article:

- Does the publication adhere to the guidelines stated in the Views Document "Scientific Literature in Support of Forensic Science and Practice"?
- Is the problem or hypothesis clearly stated?
- Is the scope of the article clearly stated as appropriate (article, case study, review, technical note, etc.)?
- Is the literature review current, thorough, and relevant to the problem being studied?
- Does this work fill a clear gap in the literature or is it confirmatory and/or incremental?
- Are the experimental procedures clear and complete such that the work could be easily reproduced?
- Are the experimental methods appropriate to the problem?
- Are the methods fully validated to the necessary level of rigor (fit for purpose)?
- Are the data analysis and statistical methodology appropriate for the problem, and explained clearly so it can be reproduced?
- Are the experimental results clearly and completely presented and discussed?
- Are omissions and limitations to the study discussed and explained?
- Are the results and conclusions reasonable and defensible based on the work and the supporting literature?
- Are the citations and references complete and accurate?
- Are the references original (primary) and not secondary?
- Are funding sources and other potential sources of conflict of interest clearly stated?"

The document also points out: "Evaluations of the literature using a universal systematic process will provide a means to determine which studies are truly foundational. As an on-going effort, these reviews will document the evolution of a given discipline with respect to the expectations outlined in the National Research Council Report on Forensic Science in 2009. Such an approach could allow for strengths and weaknesses of a given discipline to be discovered which could result in systematic exploration of these weaknesses through future research."

The Commission document further notes: "Compilations of accepted foundational literature serves additional purposes. First, compilations generated under stringent review criteria define general scientific acceptance and should be used to assist in admissibility decisions

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

and gatekeeping functions. Second, priorities can be established for translational studies designed to bring the most promising developments into mainstream forensic practice. Third, research needs can be identified and used to develop initiatives and calls for proposals to fill these needs and to spur investigator-initiated research. Success in these endeavors depends on current and complete understanding of the foundational literature."

In this same document, the NCFS opined: "Documentation of the literature that supports the underlying scientific foundation for each forensic discipline is a critical component in determining if methods, technologies, interpretation guidelines and conclusions are supported by science."

3.7.3. A Proposal for NIST to Perform Scientific Foundation Reviews

Following the NSF/NIJ-funded workshop described earlier, the NCFS Scientific Inquiry and Research Subcommittee reached out to NIST leadership with a request for NIST to perform what was referred to as "technical merit" reviews of forensic disciplines. As described previously, the MOU between DOJ and NIST that established NCFS and OSAC had agreed that NIST would "test and validate select forensic science practices and standards as appropriate." NCFS felt that their request fell within NIST's agreed upon responsibilities under the MOU.

During the September 2016 NCFS meeting, Dr. Richard Cavanagh, as Director of the NIST Special Programs Office, responded to the NCFS request by reviewing how NIST might approach the issue of examining the scientific foundations of forensic disciplines.

3.7.4. NIST Announcement at the September 2016 NCFS Meeting

A "Technical Merit" panel was held on September 12, 2016 as part of the 11th meeting of the NCFS. The proceedings can be viewed at https://www.nist.gov/topics/forensic-science/ncfs-meeting-11-webcast (Meeting 11, Part 2 [1:20:37]; the NIST plan is described from 4:40 to 17:45 and the Q&A portion begins at 1:08:30). Slides for the NIST plan are available on the archived NCFS website: https://www.justice.gov/archives/ncfs/page/file/893966/download.

The proposed NIST plan presented at that time called for performing three pilot studies (dependent on available funding) involving DNA, bitemarks, and firearms and toolmark identification. These three diverse examples were selected in order to learn if the approach(es) taken could be effective. The stated goals involved examining the scientific maturity and technical merit of selected methods and practices through considering research performed by other agencies and laboratories, NIST research, and studies documented in the literature.

For each area studied, the NIST proposal involved (1) assembling a NIST review team with a range of expertise in order to view issues from multiple perspectives, (2) seeking input on issues to consider from a variety of outside experts, (3) examining the scientific literature to evaluate available support for claims made, (4) conducting interlaboratory studies where appropriate and possible, (5) publishing a written report of findings and recommendations,

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

and (6) sharing findings with the scientific and criminal justice communities to convey the capabilities and limitations of studied forensic disciplines to practitioners, judges, lawyers, jurors, and other stakeholders.

The NIST scientific foundation reviews are the outcome of what began with the NIST plan presented at the September 2016 NCFS meeting.

During the question and answer portion of this September 2016 panel discussion, several members of the Commission discussed issues involved in pursuing technical merit (scientific foundation) reviews. There was a desire stated to avoid duplication of effort by being aware of and learning from other ongoing efforts, such as the AAAS forensic science assessments, which will be discussed below. Concern was expressed regarding the amount of time required to perform studies as well as the idea articulated by some that a field should not and could not move forward until a foundation review was completed.

The challenge of being "independent" in assessments performed was shared given that a certain level of expertise and connection to the community is needed to evaluate scientific details of any method. One Commissioner stated that there were different perceptions regarding what "methodology" can mean and the extent to which a forensic method or entire discipline might be reviewed. Finally, there was a desire expressed for open access to published reports of findings so that the information could be freely and widely available.

3.7.5. NCFS Technical Merit Review Documents

The NCFS approved two documents expressing its desires regarding technical merit reviews: (1) "Views of the Commission: Technical Merit Evaluation of Forensic Science Methods and Practices," which was published in June 2016 and (2) "Recommendation to the Attorney General: Technical Merit Evaluation of Forensic Science Methods and Practices," which was approved in September 2016 following the technical merit panel discussion mentioned above.

The Views document begins: "Forensic data, results, interpretations, and conclusions have life-changing consequences for individuals and society. It is vital that the analytical data be generated through reliable methods and practices build upon valid core scientific principles and methodology." Three views of the Commission are stated in the document:

> "(1) All forensic science methodologies should be evaluated by an independent scientific body to characterize their capabilities and limitations in order to accurately and reliably answer a specific and clearly-defined forensic question. The independent scientific body should evaluate how forensic science test methods and practices meet the standards of technical merit as defined in the OSAC Technical Merit Worksheet[2].
> (2) The National Institute of Standards and Technology (NIST) should assume the role of independent scientific evaluator within the justice system for this purpose.

[2] The OSAC Technical Merit Worksheet has evolved over time. Version 4 was the one available at the time the NCFS voted: https://www.nist.gov/sites/default/files/documents/forensics/osac/4-OSAC-QIC-Form-01-Technical-Merit-Worksheet-Form-V4.pdf. For a more recent version, see https://www.nist.gov/sites/default/files/documents/2018/01/05/technical_merit_guide_and_worksheet_january_3_2018.pdf.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

(3) Additional resources should be made available to support this new capacity."

The word "independent" as defined in this document "refers to a body that is fair, impartial, and without conflict of interest in the results of the evaluation." It is also noted that "an entity's independence does not imply that this work will be conducted without the contribution of individuals who are knowledgeable of a specific discipline. It is expected that an independent scientific body will be able to retain the relevant experts to advise the independent body as to the real life forensic application of the science." A DNA Mixture Resource Group, which provides expert input to the current DNA mixture interpretation review at regular intervals during the study, helps fulfill this vital role. Other NIST scientific foundation reviews may seek input from relevant experts using this approach or perhaps by gathering a larger group of perspectives in a single workshop near the start of the study, such as is anticipated for the bitemark effort.

The Views document defines "technical merit" as "the process that ensures the accuracy, capabilities, and limitations of forensic science tests" and states "the data and research that need to be gathered to support technical merit include, but are not limited to, clearly defined terminology, quality control, uncertainty, limitations, validation, fitness-for-purpose, and general acceptance in both the forensic and the general scientific communities." It continues: "While NIST may have a centralized evaluative role, the Commission envisions that the data and research NIST will evaluate will be generated by the robust and diverse scientific research community as well as by NIST. The resulting resource documents will be continually updated as the state of the science develops. Centralizing the evaluative role will facilitate the development of a knowledge base at NIST that will build over time."

The Views document concludes: "It is the view of the NCFS that an institutional entity assigned a permanent independent scientific evaluation function would facilitate the gathering of scientific research, knowledge, and expertise over time, creating a service resource for forensic science, technology research, and user communities. Development of a trusted and impartial process of evaluating technical merit of forensic practices and the presentation of data will ensure that all decisions rendered by the justice system are based on sound and current science."

The second document approved by the NCFS on technical merit evaluation proposed "that the Attorney General endorse and refer to the Director of NIST the following [three] recommendations:"

> "Recommendation #1: NIST should establish an in-house entity with the capacity to conduct independent scientific evaluations of the technical merit of test methods and practices used in forensic science disciplines.
>
> "Recommendation #2: The results of the evaluations will be issued by NIST as publicly available resource documents. NIST's evaluation may include but is not limited to: a) research performed by other agencies and laboratories, b) its own intramural research program, or c) research studies documented in already published scientific literature. NIST should begin its work by piloting three resource documents to establish their design

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

and requirements. The release of these documents should be broadly disseminated in the scientific and criminal justice communities and accompanied by judicial trainings.

"Recommendation #3: The Organization of Scientific Area Committees for Forensic Science (OSAC) leadership, the Forensic Science Standards Board (FSSB), should commit to placing consensus documentary standards on the OSAC Registry of Approved Standards for only those forensic science test methods and practices where technical merit has been established by NIST, or in the interim, established by an independent scientific body. An example of an interim independent scientific body could be an OSAC created Technical Merit Resource Committee composed of measurement scientists and statisticians appointed by NIST and tasked with the evaluation of technical merit."

In providing these recommendations, NCFS recognized "that NIST is a non-regulatory agency and is not recommending that NIST's function here will be regulatory in nature." The document concludes with "the vision and hope of the NCFS is that NIST will develop resource documents for all forensic science disciplines, but [recognizes] that process will take time."

The NCFS concluded its appointed role as a Federal Advisory Committee to DOJ in April 2017. However, the deliberations held, insights provided, and documents approved serve as important background material and as a roadmap for NIST scientific foundation reviews.

3.8. PCAST 2016 Report

In September 2016 a report entitled "Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods" [4] was provided to President Barack Obama from the President's Council of Advisors on Science and Technology (PCAST). This PCAST group was led by co-chairs John P. Holdren (Assistant to the President for Science and Technology and Director of the White House Office of Science and Technology) and Eric S. Lander (President of the Broad Institute of Harvard and Massachusetts Institute of Technology and one of the leaders of the Human Genome Project).

The Executive Summary notes that "in the course of its study, PCAST compiled and reviewed a set of more than 2,000 papers from various sources – including bibliographies prepared by the Subcommittee on Forensic Science of the National Science and Technology Council and the relevant Working Groups organized by the National Institute of Standards and Technology (NIST); submissions in response to PCAST's request for information from the forensic science stakeholder community; and PCAST's own literature searches" (p. 2). See Ref. [28] for full reference list.

During their study, "PCAST concluded that there are two important gaps: (1) the need for clarity about the scientific standards for the validity and reliability of forensic methods and (2) the need to evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable" (p. 1). The PCAST report examines and comments on "foundational validity" and "validity as applied" for six forensic feature-comparison methods: (1) DNA analysis of single-source and simple-mixture samples, (2)

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

DNA analysis of complex-mixture samples, (3) bitemarks, (4) latent fingerprints, (5) firearms identification, and (6) footwear analysis. Expressing the desire for peer-reviewed research publications with data to support claims, PCAST notes "the publication and critical review of methods and data is an essential component in establishing scientific validity" (p. 68).

Commenting on bitemark analysis: "In its own review of the literature [involving 407 entries] PCAST found few empirical studies that attempted to study the validity and reliability of the methods to identify the source of a bitemark" (p. 85). They conclude: "Among those studies that have been undertaken, the observed false positive rates were so high that the method is clearly scientifically unreliable at present… [A]vailable scientific evidence strongly suggests that examiners cannot consistently agree on whether an injury is a human bitemark and cannot identify the source of bitemark with reasonable accuracy." (p. 87).

PCAST found "that DNA analysis of single-source samples or simple mixtures of two individuals, such as from many rape kits, is an objective method that has been established to be foundationally valid," (p. 75) but expressed some concerns with complex mixtures (pp. 75-83). PCAST concludes that, "NIST should play a leadership role in this process [of conducting scientific studies], by ensuring the creation and dissemination of materials and stimulating studies by independent groups through grants, contracts, and prizes; and by evaluating the results of these studies" (p. 83).

Regarding the need for assessments of foundational validity, PCAST recommended:

> "It is important that scientific evaluations of the foundational validity be conducted, on an ongoing basis, to assess the foundational validity of current and newly developed forensic feature-comparison technologies. To ensure the scientific judgments are unbiased and independent, such evaluations must be conducted by a science agency which has no stake in the outcome. (A) The National Institute of Standards and Technology (NIST) should perform such evaluations and should issue an annual public report evaluating the foundational validity of key forensic feature-comparison methods. (B) The President should request and Congress should provide increased appropriations to NIST of (a) $4 million to support the evaluation activities described above and (b) $10 million to support increased research activities in forensic science, including on complex DNA mixtures, latent fingerprints, voice/speaker recognition, and face/iris biometrics" (pp. 128-129).

It is important to keep in mind that funding levels are determined by Congress regardless of recommendations made by PCAST or any other group. In fiscal year 2018, Congress provided funding to NIST to perform "technical merit evaluations," which we have termed "scientific foundation reviews."

There were numerous reactions to the PCAST report, with some applauding its findings, some ignoring its findings, and some criticizing them. Critics raised at least six distinct points as noted by one legal scholar [29]: (1) The PCAST committee was biased against forensic science, (2) PCAST offered an overly narrow and idiosyncratic definition of scientific validity, (3) PCAST ignored strong evidence that proves the scientific validity of various

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

forensic sciences, (4) PCAST usurped the role of judges and juries by inserting its own opinions about forensic science, (5) forensic science evidence should not be held to scientific standards of validity because the evidence includes technical or specialized knowledge, and (6) practitioners' personal experiences and observations should be given weight when assessing the scientific validity of forensic science.

An Addendum to the PCAST Report on Forensic Science in Criminal Courts released on January 6, 2017 [30] emphasized that "an empirical claim cannot be considered scientifically valid until it has been empirically tested" (p. 1) and continues that "while scientists may debate the precise design of a study, there is no room for debate about the absolute requirement for empirical testing" (p. 2). This addendum further notes that "the test problems used in the empirical study define the specific bounds within which the validity and reliability of the method has been established (e.g., is a DNA analysis method reliable for identifying a sample that comprises only 1% of a complex mixture?)" (p. 2).

3.9. AAAS Studies

The American Association for the Advancement of Science (AAAS) announced a partnership with the Laura and John Arnold Foundation in 2015, with plans to explore the "underlying scientific bases for the forensic tools and methods currently used in the criminal justice system." AAAS planned to begin with ten forensic disciplines: (1) bloodstain pattern analysis, (2) digital evidence, (3) fire investigations, (4) firearms and toolmarks/ballistics, (5) footwear and tire tracks, (6) forensic odontology – bitemark analysis, (7) latent fingerprints, (8) trace evidence – fibers, (9) trace evidence – hair, and (10) trace evidence – paint and other coatings. Their website notes that the project goals were to evaluate "the scientific underpinnings the forensic community relies on to support their practices and, where these fall short, recommend areas requiring further study" [31].

Reports were released for fire investigations (in July 2017, [12]) and latent fingerprint examination (in September 2017, [13]). The fire investigation report offers 25 recommendations that provide a roadmap for future research efforts [12] while the latent fingerprint examination report provides 14 recommendations to assist future research [13]. These reports were authored by a small group (e.g., four authors – William Thompson, John Black, Anil Jain, and Joseph Kadane – for the latent fingerprint examination report; five authors – Jose Almirall, Hal Arkes, John Lentini, Frederick Mowrer, and Janusz Pauliszyn – for the fire investigation report) with three contributing AAAS staff and a seven-member advisory committee.

Apparently, any future work by AAAS with their forensic science assessments is subject to availability of funding (personal communication from Deborah Runkle, AAAS).

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

4. Other Similar International Activities

4.1. Australian NIFS Forensic Fundamentals

In 2016, the Australia New Zealand Policing Advisory Agency (ANZPAA) National Institute of Forensic Science (NIFS) [32] released "A Guideline to Forensic Fundamentals" that describes their plan for evaluating the scientific foundations for human-based forensic disciplines where comparisons of features by an expert inform the final result obtained. This guideline document is available at http://www.anzpaa.org.au/ArticleDocuments/220/A%20Guideline%20to%20Forensic%20Fundamentals.pdf.aspx. The goal of this effort is to help forensic managers, researchers, and practitioners "to assess the validity of current methods and opinions and to consider the suitability of new techniques being considered for implementation in forensic casework" (p. 3).

The ANZPAA NIFS effort notes that "the application of human-based forensic disciplines is based on underlying feature set assumptions which should be quantified and assessed as they form the basis of all methods and opinions that are derived. These assumptions relate not only to the nature and frequency of the feature set, but also to whether they can be used as a means to distinguish between groups or individuals" (p. 5). Therefore, their effort is focusing on eight areas: "(1) how the features originate and whether they are random or ordered, (2) the persistence of the features, (3) the transference of the features, (4) the potential for something foreign/unrelated to be mistaken as a feature, (5) the dependence or independence of the subcomponents of the feature set, (6) whether unrelated items have the potential to resemble one another, (7) population studies to determine the level of variation and frequency of variants, and (8) whether there are established databases to determine the frequency of concurring features" (p. 5).

ANZPAA NIFS encourages a review of published empirical studies available in the literature and states: "a good published scientific validation study would include the following: [(1)] explanation of the methodology and the opinions that can be derived, [(2)] publication in a recognized, peer reviewed scientific journal, [(3)] use of ground truth known experimental materials, and [(4)] use of a statistically significant sample size" (p. 6).

The Forensic Fundamentals guidelines point out that "acceptance in court does not provide confirmation that a method is scientifically valid" (p. 7). This document emphasizes that "appropriate experimental design is important to ensure that the correct processes are validated" and provides examples of types of factors that need to be tested including: "accuracy, precision, specificity, sensitivity, reliability, and reproducibility" (p. 7). These guidelines stress: "The test materials should be prepared based on studies of how closely unrelated items may resemble one another. Experimental design should include an equal mixture of randomly presented test materials that include: items that are related [and] items that are unrelated with the highest degree of similarity" (p. 7). This section of their guidelines concludes: "The ground truth of test items should be known" (p. 7).

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

A section on limitations encourages acknowledgment of the limitations of a method "to ensure that the evidence provided can be appropriately assessed by the Court" (p. 7). Some examples are provided including: "element or general discipline-specific limitations, case-specific limitations (where appropriate), and applicable error rates that may exist" (p. 7). A section on assumptions stresses the importance of "acknowledge[ing] any assumptions that have been made" and the need "to disclose [any] assumptions in any scientific report that is prepared" (p. 8). Examples provided include: "underlying principles of the feature set on which the basis for the analysis is being performed and case-specific assumptions required to perform the analysis, where appropriate" (p. 8).

These guidelines also cover implementation considerations for proficiency testing, accreditation, presenting opinions, reporting scales, propositions, peer-review, and human bias. The document concludes: "Forensic science evidence has served the Courts well for many years and its continued success will be dependent on ensuring that there is empirical support for the validity and reliability of the underlying science. It is anticipated that if each of the considerations presented in this document can be satisfied, for each of the elements identified within a given forensic science discipline, a sound scientific basis will be available for the Court to assess the strengthen of the forensic evidence appropriately" (p. 10).

4.2. UK Forensic Science Regulator

Since 2008 in the United Kingdom, a Forensic Science Regulator has been appointed to oversee and coordinate quality efforts in forensic science in serving the entire criminal justice system. Codes of Practice and Conduct have been developed over the years with the fourth version issued in October 2017 [33]. This document notes: "This Code of Conduct provides a clear statement to customers and the public of what they have a right to expect" (p. 12). For example, the tenth requirement for a practitioner is to "conduct casework using methods of demonstrable validity and comply with the quality standards set by the Regulator relevant to the area in which you work" (p. 12).

The UK Forensic Science Regulator has published over 140 documents with guidance on a variety of topics including interpreting DNA evidence (December 2012), DNA contamination detection (September 2014), validation (November 2014), cognitive bias effects relevant to forensic science examinations (October 2015), laboratory DNA: anti-contamination (December 2015), crime scene DNA: anti-contamination (July 2016), expert report content (October 2017), DNA mixture interpretation (July 2018), and software validation for DNA mixture interpretation (July 2018). These publications are available at https://www.gov.uk/government/publications?departments%5B%5D=forensic-science-regulator.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

5. Terminology and Concepts

Information is conveyed in how terms are defined. Having a common vocabulary is important for communicating ideas and developing a shared understanding of concepts. Therefore, the terms and concepts below are defined as we are using them in NIST scientific foundation reviews.

Repeatability: a measure of the variability that exists even when measurement conditions are kept as constant as possible—same laboratory, same protocol, same technician, same method, same batch of materials, same temperature, and so on. Repeatability represents the smallest amount of variability that can be achieved when all influential factors are kept as constant as possible.

Reproducibility: a measure of the variability that can exist under field application conditions—different operators, different labs, different equipment, different software programs, etc. Reproducibility represents a more relevant quantity than repeatability when assessing variability that may be present in practice.

Measurement error: the difference between a reported value and the true value when the true value is known and can be calculated. Otherwise this is a conceptual quantity).

Degree of reliability: a quantity that summarizes the average magnitude of the measurement errors. The degree of reliability is often reported as a root mean square error or mean absolute error for continuous measurements using known test cases of known value. For binary decisions (present/absent, positive/negative, etc.), reliability may be judged using error rates. The test cases with known values used to assess the degree of reliability must be representative of cases that may be encountered in practice.

False positive error rate: the proportion of times a known negative sample is classified as positive by a binary decision rule over a large number of independent tests that are representative of casework.
.

False negative error rate: the proportion of times a known positive sample is classified as negative by a binary decision rule over a large number of independent tests that are representative of casework.

Error rates: false positive and false negative rates are often reported as global averages, i.e., average error rates across all labs, all examiners, samples of various complexities, etc. For such error rates to be useful in casework, it is important to assess error rates in cases similar to the current case samples being considered, which may be called case-specific error rates. Attempts to use case-specific error rates still involves some subjectivity in the sense that someone has to make the decision of what it means to be similar to the current case, but it is important to note that global error rates may not be relevant in a particular application.

Validation: the process of empirically demonstrating the suitability or fitness for purpose of a method of analysis. A validation exercise should explicitly state the criteria that are

required to be met to demonstrate fitness for purpose. In the absence of clearly stated criteria that need to be met for a method to be regarded as validated, it is possible to calculate and share clearly defined metrics such as repeatability, reproducibility, degree of reliability, error rates, etc. When such metrics are available, the user can determine whether the method is fit for purpose. This is the preferred output from a validation exercise. Since it is impractical to carry out test runs to exhaustively cover all use cases, sufficient information needs to be made available for the user to determine the reasonable limits of extrapolation from actually conducted test scenarios to new scenarios not explicitly considered as part of a validation study.

Note that validation is neither a universal nor binary concept (i.e., validated versus not validated) because the same method may be considered to be valid for one application and not for another application. Even for two similar applications, a method may be considered valid or not valid depending on the seriousness of the consequences of errors. Therefore, it is good practice to report metrics that allow assessment of fitness for purpose rather than to report suitability for each specific application.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

6. References

[1] National Academy of Sciences (NAS) , Committee on the Conduct of Science (1989) On being a scientist. *Proceedings of the National Academy of Sciences of the United States of America* 86(23):9053–9074.

[2] National Commission on Forensic Science (2016) *Views of the Commission Regarding Identifying and Evaluating Literature that Supports the Basic Principles of a Forensic Science Method or Forensic Science Discipline*. Available at https://www.justice.gov/archives/ncfs/file/839716/download. Accessed September 18, 2018.

[3] National Research Council (2009) Strengthening Forensic Science in the United States: A Path Forward. National Academies Press, Washington, D.C.

[4] President's Council of Advisors on Science and Technology (PCAST) (September 20, 2016) *Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf Accessed June 28, 2018.

[5] Committee on Appropriations (2017) *Departments of Commerce and Justice, Science, and Related Agencies Appropriations Bill, 2018. Report 115-139*. Available at https://www.congress.gov/115/crpt/srpt139/CRPT-115srpt139.pdf. Accessed September 18, 2018.

[6] Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, Dafoe A, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hesse B, Humphreys M, Ishiyama J, Karlan D, Kraut A, Lupia A, Mabry P, Madon T, Malhotra N, Mayo-Wilson E, McNutt M, Miguel E, Paluck EL, Simonsohn U, Soderberg C, Spellman BA, Turitto J, VandenBos G, Vazire S, Wagenmakers EJ, Wilson R, Yarkoni T (2015) Promoting an open research culture. *Science* 348(6242):1422-1425. https://doi.org/10.1126/science.aab2374

[7] Ross-Hellauer T (2017) What is open peer review? A systematic review. *F1000Research* 6:588. https://doi.org/10.12688/f1000research.11369.2

[8] Press R (2017) *Who was Detective X?* Available at https://www.nist.gov/featured-stories/who-was-detective-x. Accessed August 30, 2018.

[9] Federal Bureau of Investigation (2011) *Quality Assurance Standards for Forensic DNA Testing Laboratories*. Available at https://www.fbi.gov/file-repository/quality-

assurance-standards-for-forensic-dna-testing-laboratories.pdf/view. Accessed August 30, 2018.

[10] Memorandum of Understanding (MOU) between NIST and the Department of Justice (DOJ) (2013) Available at https://www.justice.gov/archives/ncfs/file/761051/download. Accessed September 19, 2018.

[11] *National Commission on Forensic Science (NCFS).* Available at https://www.justice.gov/archives/ncfs. Accessed June 28, 2018.

[12] American Association for the Advancement of Science (AAAS) (July 11, 2017) *Forensic Science Assessments: A Quality and Gap Analysis – Fire Investigation.* Available at https://www.aaas.org/page/forensic-science-assessments-quality-and-gap-analysis and https://www.aaas.org/report/fire-investigation. Accessed June 28, 2018.

[13] American Association for the Advancement of Science (AAAS) (September 15, 2017) *Forensic Science Assessments: A Quality and Gap Analysis – Latent Fingerprint Examination.* Available at https://www.aaas.org/page/forensic-science-assessments-quality-and-gap-analysis and https://www.aaas.org/report/latent-fingerprint-examination. Accessed June 28, 2018.

[14] National Commission on Forensic Science , Recommendation to the Attorney General (September 12, 2016) *Technical Merit Evaluation of Forensic Science Method and Practices.* Available at https://www.justice.gov/archives/ncfs/page/file/905541/download. Accessed June 28, 2018.

[15] 2016) *NCFS Meeting #11 webcast.* Available at https://www.nist.gov/topics/forensic-science/ncfs-meeting-11-webcast (see part 2). Accessed July 9, 2018.

[16] 2017) *NCFS Meeting #13 webcast.* Available at https://www.nist.gov/topics/forensic-science/ncfs-meeting-13-webcast; transcript available at https://www.nist.gov/sites/default/files/documents/2017/04/27/1_april_10_doj_part_1.pdf (see page 8). Accessed July 9, 2018.

[17] Butler JM (2015) U.S. initiatives to strengthen forensic science & international standards in forensic DNA. *Forensic Science International: Genetics* 18:4-20. https://doi.org/10.1016/j.fsigen.2015.06.008

[18] Scientific Working Group for Firearms and Toolmarks , Association of Firearms and Toolmark Examiners (June 14, 2011) *Response to 25 foundational firearm and toolmark examination questions received from the Subcommittee on Forensic Science*

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

(SoFS), Research, Development, Testing, & Evaluation Interagency Working Group (RDT&E IWG). Available at https://www.nist.gov/sites/default/files/documents/forensics/Annotated-Bibliography-Firearms-Toolmarks.pdf. Accessed September 18, 2018.

[19] Bartick EG & Floyd MA eds (2016) Forensic Science Research and Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature. May 26-27, 2015. Washington, D.C. Available at https://www.ncjrs.gov/pdffiles1/nij/250088.pdf. Accessed June 28, 2018.

[20] National Institute of Justice (May 26–27, 2015) *Forensic Science Research and Evaluation Workshop: A Discussion on the Fundamentals of Research Design and an Evaluation of Available Literature*. Available at https://www.ncjrs.gov/pdffiles1/nij/250088.pdf. Accessed September 18, 2018.

[21] National Commission on Forensic Science , Views of the Commission (May 1, 2015) *Defining Forensic Science and Related Terms*. Available at https://www.justice.gov/archives/ncfs/file/786571/download. Accessed July 11, 2018.

[22] National Commission on Forensic Science (NCFS) , Scientific Inquiry and Research Subcommittee (January 2015) *Views on Scientific Literature in Support of Forensic Science and Practice*. Available at https://www.justice.gov/archives/ncfs/file/786591/download. Accessed September 18, 2018.

[23] National Commission on Forensic Science (NCFS) , Scientific Inquiry and Research Subcommittee (March 2016) *Recommendation to Fund Post-Doctoral Projects to Facilitate Translation of Research into Forensic Science Practice*. Available at https://www.justice.gov/archives/ncfs/page/file/839721/download. Accessed September 18, 2018.

[24] National Commission on Forensic Science (NCFS) , Scientific Inquiry and Research Subcommittee (March 2016) *Views on Identifying and Evaluating Literature that Supports the Basic Principles of a Forensic Science Method or Forensic Science Discipline* Available at https://www.justice.gov/archives/ncfs/file/839716/download. Accessed September 18, 2018.

[25] National Commission on Forensic Science (NCFS) , Scientific Inquiry and Research Subcommittee (June 2016) *Views on Technical Merit Evaluation of Forensic Science Methods and Practices*. Available at https://www.justice.gov/archives/ncfs/file/881796/download. Accessed September 18, 2018.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft

[26] National Commission on Forensic Science (NCFS) , Scientific Inquiry and Research Subcommittee (September 2016) *Recommendation on Technical Merit Evaluation of Forensic Science Methods and Practice* Available at https://www.justice.gov/archives/ncfs/page/file/905541/download. Accessed September 18, 2018.

[27] National Commission on Forensic Science (2015) *Views of the Commission Scientific Literature in Support of Forensic Science and Practice*. Available at https://www.justice.gov/archives/ncfs/file/786591/download. Accessed September 18, 2018.

[28] PCAST (2016) *References in Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf. Accessed June 28, 2018.

[29] Koehler JJ (2018) How Trial Judges Should Think About Forensic Science Evidence. *Judicature.* 102(1):28-38

[30] PCAST (January 2017) *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf. Accessed June 28, 2018.

[31] American Association for the Advancement of Science (AAAS) (2015) *Forensic Science Assessments: A Quality and Gap Analysis*. Available at https://www.aaas.org/page/forensic-science-assessments-quality-and-gap-analysis. Accessed September 18, 2018.

[32] *ANZPAA NIFS publications*. Available at http://www.anzpaa.org.au/forensic-science/our-work/products/publications. Accessed June 28, 2018.

[33] *UK Forensic Science Regulator publications*. Available at https://www.gov.uk/government/publications?departments%5B%5D=forensic-science-regulator. Accessed June 28, 2018.

This publication is available free of charge from: https://doi.org/10.6028/NIST.IR.8225-draft