Check for updates

NIST Special Publication 2100
NIST SP 2100-06

Long-Term Vision and Strategic Priorities for Forensic Science in the United States: Summary Report of a Roundtable Discussion with Thought Leaders

September 6-7, 2023

Henry Swofford, Ph.D.
Special Programs Office
Laboratory Programs
National Institute of Standards and Technology

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.2100-06

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

NIST Special Publication 2100
NIST SP 2100-06

Long-Term Vision and Strategic Priorities for Forensic Science in the United States: Summary Report of a Roundtable Discussion with Thought Leaders

September 6-7, 2023

Henry Swofford, Ph.D.
Special Programs Office
Laboratory Programs
National Institute of Standards and Technology

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.2100-06

March 2024

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST SP 2100-06
March 2024

Disclaimers

Certain entities, equipment, instruments, software, or materials, commercial or non-commercial, are identified in this report to specify an experimental procedure or concept 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.

Publications in the SP 2100 subseries are proceedings from conferences organized predominately by NIST scientific and technical staff. This publication may include external perspectives from industry, academia, government, and others. The opinions, recommendations, findings, and conclusions in this publication do not necessarily reflect the views or policies of NIST or the United States Government.

NIST Technical Series Policies
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

Publication History
Approved by the NIST Editorial Review Board on 2024-03-11

How to Cite this NIST Technical Series Publication
Swofford, HJ (2024) Long-Term Vision and Strategic Priorities for Forensic Science in the United States: Summary Report of a Roundtable Discussion with Thought Leaders. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) NIST SP 2100-06. https://doi.org/10.6028/NIST.SP.2100-06

NIST Author ORCID iDs
Henry Swofford: 0009-0003-3114-5958

Abstract

On September 6 and 7, 2023, the National Institute of Standards and Technology (NIST) convened a roundtable discussion and workshop with a diverse group of forensic science thought leaders representing various local, state, federal, and private forensic laboratories, academia, and the legal system throughout the United States. This meeting gave participants the opportunity to provide their perspectives on major challenges to forensic science practices in the United States and how research and standards can address those challenges. This report summarizes the views and opinions expressed by participants during the plenary and breakout sessions.[1]

Keywords

Forensic Science; Roundtable; Strategic Priorities.

[1] The opinions, recommendations, findings, and conclusions in this publication do not necessarily reflect the views or policies of NIST or the United States Government.

Table of Contents

Acknowledgements

This report was made possible through the contributions and support of many individuals. Special appreciation is extended to the presenters and everyone who participated in the roundtable discussion and workshop. Additionally, special recognition is given to NIST staff who helped with the event and this report: John Butler, Timothy Czakoczi, Crystal DeGrange, Vincent Desiderio, Greg Fiumara, Allison Getz, Barbara Guttman, Joseph Hynes, Christian Incognito, John Paul Jones, II, Sandra Koch, Corrine Lloyd, Steven Lund, Sabina Mohan, Marcela Najarro, Melissa Phillips, Donna Ramkissoon, Robert Ramotowski, Karen Reczek, Christina Reed, Erica Romsos, Kelly Sauerwein, Katherine Sharpless, S. Shyam-Sunder, Donna Sirk, Ed Sisco, Austin Snelick, Johannes Soons, Melissa Taylor, Robert Thompson, Dalia Travis, Pauline Truong, Peter Vallone, and Alan Zheng.

1. Executive Summary

Background

On September 6 and 7, 2023, NIST convened a roundtable discussion and workshop with a diverse group of about 50 thought leaders in forensic science, representing local, state, federal, and private forensic laboratories, academia, and the legal system throughout the United States.[2] The two-day meeting allowed participants to describe, discuss, and prioritize critical long-term and near-term challenges to forensic science practices in the United States that they believe should be addressed through research and standards.[3] The insights gained from the roundtable discussion and workshop are intended to inform a long-term vision and strategic priorities for the NIST Forensic Science Program over the next several years.

On the first day, plenary sessions focused on the challenges faced by practitioners, leadership, legal teams, researchers, and quality system managers. The second day featured breakout group discussions in five areas: standards and practices; validity and reliability; forensic algorithms; research, development, testing, and evaluation; and non-technical challenges. Brief summaries from each and key takeaways are provided below.

Practitioner Perspectives

In this session participants focused on issues affecting the day-to-day work of forensic practitioners. Major points expressed during the discussion were:

(a) Changes are being proposed at rates that outpace the capacity for forensic service providers to incorporate them.

(b) Requests for forensic services are outpacing the capacity of many forensic service providers. This has caused backlogs to grow, cases being delayed, evidence or cases being prioritized over others, and other duties (such as research, innovation, and standardization) being neglected.

(c) Certification programs need to be more robust and better aligned to the knowledge, skills, and abilities that forensic practitioners are expected to possess so that those

[2] The opinions, recommendations, findings, and conclusions in this publication do not necessarily reflect the views or policies of NIST or the United States Government.

[3] In this report, the term "standards" refers to "documentary standards and guidelines." The term includes standard test methods, standard practices, standard guides, and best practice recommendations published by a standards development organization (SDO) or proposed by the Organization of Scientific Area Committees for Forensic Science (OSAC) or Scientific Working Groups (SWGs). Documentary standards and guidelines are distinct from physical standards, such as standard reference materials (SRMs) and standard reference data (SRD).

programs can provide a more meaningful and universally accepted way for forensic service providers to assess and demonstrate competency and proficiency.

(d) Private-sector forensic service providers compete with free services from public-sector providers. Innovation is key to competitiveness as it can lead to new analytical methods and capabilities, faster turnaround times, and lower costs for services. Public-sector forensic service providers often lack a comparable incentive to continuous innovation.

Leadership Perspectives

In this session participants focused on issues affecting the overall management of forensic laboratories and the broader forensic science community. Major points expressed during the discussion were:

(a) Forensic service providers often face an uphill battle for resources, which has the potential to adversely affect the quality and timeliness of forensic testing services that are critical for public safety.

(b) Forensic service providers often face a deficit in the number of technically trained personnel able to address case throughput demands, largely due to a lack of standardized training programs and universally accepted criteria to test and demonstrate competency and proficiency.

(c) The cumulative stresses of too few people, large caseloads, high mental strains, tough testimony requirements, and an adversarial culture within the forensic science and legal communities may adversely impact the quality of forensic testing and have led to record numbers of personnel seeking wellness assistance.

(d) Traditional analytical capabilities have limitations, and research is needed on the pharmacology of new drugs, statistics in pattern disciplines, multi-biometric identification systems, and performance benchmarking through black-box and white-box testing.[4]

(e) Consistency and standardization are important but progress toward standards implementation has been slow due to a proliferation of standards, wide-ranging implications of many requirements, and confusing and multiple sources of guidance.

[4] Black-box testing refers to studies designed to evaluate the output of a method or process (i.e., analyst results) without regard to how those outputs are produced. White-box testing refers to studies designed to evaluate the basis for the output (i.e., analyst results) of a method or process.

(f) Translation and implementation[5] of new methods, technologies, and practices require significant time and resources, and personnel often struggle to understand how to properly design and execute validation studies and often lack knowledge on the concepts and principles underlying the methods or technologies to use them effectively.

(g) Forensic service providers often lack a fully integrated laboratory information management system (LIMS) and connectivity with other entities thereby creating challenges with data accessibility, disclosure notifications, and strategic and investigative insights.

Legal Perspectives

In this session participants focused on issues affecting the admissibility and presentation of forensic evidence in litigation. Major points expressed during the discussion were:

(a) The validity and reliability of forensic science methods and practices are often disputed in court. Research efforts should be prioritized to assess and improve accuracy, reproducibility, and repeatability of forensic results.

(b) Standards need to be more rigorous, robust, consistent, and specific. They should establish requirements for all forensic service providers and not be retrofitted or reduced to the lowest level that accommodates the status quo.

(c) A criminal justice system without transparency lacks legitimacy. The forensic science community needs to create a culture of transparency, that acknowledges error and uncertainty, so that when issues arise, meaningful (and safe) conversations can be had about addressing the issues in practice to promote systemic improvements.

Researcher Perspectives

In this session participants focused on issues affecting the execution and implications of forensic science research. Major points expressed during the discussion were:

(a) Variabilities in policies and method protocols in some disciplines have contributed to disagreements and different results among analyses conducted by different forensic service providers. These differences can be significant and consequential, even when the same types of instruments, technologies, or products are used.

(b) Unless outputs from research can be effectively translated and implemented into operational practice, there is little utility or impact. Researchers and practitioners need

[5] Translation and implementation refer to the uptake and use in an operational setting.

to strengthen collaborations to ensure the research is relevant to real-world problems and that barriers to translation and implementation are accounted for in the design and execution of the research.

(c) The prevalence of digital technologies and emergence of artificial intelligence (AI) have created opportunities for applications in forensic science. There is also a growing need for forensic analyses of AI systems. Understanding how AI systems work and when failures occur are critical to mitigating bias and error and ensuring fair and appropriate applications.

Quality Management Systems Perspectives

In this session participants focused on issues affecting quality assurance in forensic science practice. Major points expressed during the discussion were:

(a) Forensic service providers must employ practices that have the capacity not only to address traditional quality assurance measures, but also to adapt, evolve, and embrace current and emerging challenges, such as AI-based systems, blind proficiency testing, personnel wellness, human error, and institutional or systematic biases affecting evidence analysis and interpretation.

(b) Performance monitoring through blind testing enables forensic service providers to monitor the entire process—from evidence submission to reporting results—and provide a real-time evaluation of analytical procedures, identify areas for improvement, and provide a direct assessment of performance and reliability.

(c) Achieving consistency and standardization through the implementation of standards has been challenging. The standards available have been considered "too much" by some members of the forensic science community and "not doing enough" by others.

Breakout I: Standards and Practices

Breakout group I focused on the development and implementation of standards in forensic science practice. Major points expressed during the discussion were:

(a) Priorities for enhancing the quality, consistency, and efficiency of forensic science practice include establishing standards for reporting and testimony across all disciplines; standardizing training programs to enable better alignment of certification programs; and breaking down discipline-specific silos to promote more interdisciplinary collaboration and consistency.

(b) Means for assessing and assuring the quality of standards include developing model standards and frameworks; strengthening training relating to standards development; providing access to technical editors to improve document clarity and consistency; and establishing a feedback loop for ongoing standards development and review.

(c) Mechanisms for promoting and facilitating the adoption and implementation of standards include identifying major barriers; establishing financial incentives; strengthening education and outreach; and creating means for auditing and assessing conformance.

Breakout II: Validity and Reliability

Breakout group II focused on strengthening public trust and confidence in forensic science practice. Major points expressed during the discussion were:

(a) Priorities for strengthening the validity, reliability, and public trust and confidence of forensic science methods, practices, and disciplines include promoting widespread adoption of standards; demonstrating conformance to standards through third-party auditing schemes; enabling public access and sharing of standard operating procedures and validation information; and providing model procedures and validation plans for forensic service providers.

(b) Means for determining fitness for purpose and assuring the validity and reliability of forensic science methods, practices, and disciplines include establishing clear criteria and requirements for validation studies and enabling public access and sharing of validation information and quality incidents.[6]

(c) Mechanisms for assessing the foundational validity and reliability of forensic science methods, practices, and disciplines include establishing clear definitions and criteria; centralizing and providing open-access to published research and validation data; and conducting regular black-box and interlaboratory studies to monitor performance.

Breakout III: Forensic Algorithms

Breakout group III focused on the responsible use of computational algorithms in forensic science practice. Major points expressed during the discussion were:

(a) Priorities for enabling the use of algorithms in forensic science include developing guidelines for software developers and end users to enhance transparency; establishing frameworks for determining sample appropriateness and testing various algorithm

[6] E.g., incidents relating to errors or deviations from established standard operating procedures.

types; creating standard application programming interfaces (APIs); and identifying areas for automation (and augmentation) to improve efficiency and productivity.

(b) Factors relating to the design and use of algorithms to promote responsible applications in forensic science include improving transparency and explainability; promoting and supporting multi-tiered testing schemes to establish conditions for appropriate use; and consideration of sociotechnical implications and impacts caused by the use of algorithms.

(c) Mechanisms for assessing and assuring the validity, reliability, and fairness of algorithms in forensic science include developing standards addressing algorithmic data needs; establishing requirements for algorithmic transparency; and promoting and supporting multi-tiered testing schemes with diverse datasets.

Breakout IV: Research, Development, Testing, and Evaluation

Breakout group IV focused on advancing critical research, development, testing, and evaluation relating to forensic science practice. Major points expressed during the discussion were:

(a) Priorities for research, development, testing, and evaluation include developing an interdisciplinary and cross-cutting strategic research plan; advancing the role of technology in addressing capacity challenges; and facilitating research to bridge technical gaps between standards and scientific foundation reviews.

(b) Means for strengthening researcher and practitioner partnerships (including data sharing) to advance research include increasing outreach and implementing convergence models to encourage interdisciplinary collaboration; establishing and promoting the use of centralized databases for connecting practitioners and researchers; and creating and curating datasets, databases, and other resources to enable and promote data sharing to support research, development, testing, and evaluation.

(c) Research products needed to improve the validity, reliability, and standardization of forensic science practice include the development of datasets to establish method limitations; the development of scalable databases to archive temporally relevant datasets and promote data sharing; and the development and promotion of statistical tools and educational resources addressing statistical methods and metrological principles.

Breakout V: Non-Technical

Breakout group V focused on identifying and overcoming non-technical challenges affecting the capabilities, quality, and reliability of forensic science practice. Major points expressed during the discussion were:

(a) Priorities for non-technical challenges include issues relating to translation and implementation of new methods; standards; LIMS; quality assurance (quality incident reporting and blind proficiency testing); and training and staffing.

(b) Mechanisms for addressing non-technical challenges include providing greater support for the translation and implementation of new methods; developing interdisciplinary standards; promoting standardization and interoperability of LIMS; creating a centralized database for quality incident reports; supporting the development of blind proficiency testing programs; and expanding curricula for training and continuing education programs with a focus on online modules for greater accessibility.

(c) Metrics for assessing the effectiveness and impact of research and standards on addressing non-technical challenges include measures relating to the scope, scale, and context for which tools, resources, and materials have been accessed, used, or referenced by forensic service providers, researchers, and other members of the forensic science community.

Key Takeaways

The following key takeaways represent common challenges and persistent themes expressed among participants during presentations and discussions.

1. Critical advancements in forensic science are often stifled by resource and capacity limitations faced by forensic service providers. Stronger partnerships are needed to help lower barriers to the translation and implementation of new methods, technologies, and practices and to ensure that outputs from research and standards programs are impactful to strengthening forensic science practice.

2. Consistency and standardization of forensic science practices are priorities shared across the forensic science community. However, progress toward achieving these goals has been slow and challenging. Standards play an important role, but the number of requirements and recommendations proposed have been overwhelming for many forensic service providers to implement. Furthermore, shortfalls in the quality, clarity, and rigor of those requirements and recommendations allow for flexibility in application and conformance. Achieving the goals of consistency and standardization requires consolidation of requirements and recommendations into interdisciplinary standards and improvements to the quality, clarity, and rigor of the documents.

3. The validity and reliability of forensic science methods and practices are often disputed. A greater emphasis on blind testing, black-box testing, and interlaboratory studies is needed to strengthen performance monitoring programs for existing practices as well as to identify priorities for research and standards programs to address critical challenges or limitations.

4. Interdisciplinary collaboration and standardization are key priorities for improving forensic science practice. These efforts can enable harmonization across disciplines, lower barriers to adoption and implementation of standards, provide a rubric for evaluating acceptable practices, and promote consistency in analytical methods and results among forensic service providers.

5. Transparency and accountability are essential for ensuring validity and reliability of forensic science methods. Public access to standard operating procedures, validation methods and data, and quality incidents promote public trust and confidence in forensic methods, accelerate research to advance forensic practices, and enable greater coordination, collaboration, and resource sharing among forensic service providers to alleviate translation and implementation challenges.

6. Ongoing research, development, testing, and evaluation are critical for advancing forensic science practices. These efforts provide the means for monitoring performance, identifying systemic challenges, and addressing evolving needs to improve the validity, reliability, and consistency of forensic science methods and practices. Priorities include stronger researcher-practitioner partnerships, creation of centralized data and databases, and use of computational technologies and methods.

2. Overview

2.1. Background

On September 6 and 7, 2023, NIST convened an in-person roundtable discussion and workshop with a diverse group of about 50 thought leaders in forensic science, representing local, state, federal, and private forensic laboratories, academia, and the legal system throughout the United States.

Attendees described major challenges facing the forensic science community from their various perspectives and how the forensic science community can strengthen domestic forensic science practice, with particular emphasis on the following topics:

- Developing and implementing standards
- Strengthening public trust and confidence
- Responsible use of computational algorithms
- Advancing critical research, development, testing, and evaluation
- Identifying and overcoming non-technical challenges affecting the quality and reliability of forensic science practice

The insights gained from the roundtable discussion and workshop are intended to inform the long-term vision and strategic priorities for the NIST Forensic Science Program over the coming years.

2.2. Meeting Structure

The meeting agenda is provided in Appendix A. Day one of the two-day meeting consisted of five plenary sessions involving presentations and panel discussions. The different perspectives represented across the five plenary sessions included those of practitioners, leaders, litigators, researchers, and quality managers. Panelists were each given 15 minutes to share their views and were asked to frame their presentation around two questions:

(a) What are the major cross-cutting (not primarily discipline-specific) forensic science research challenges that must be addressed to significantly strengthen forensic science practice in the United States today?

(b) What are the major forensic science standards challenges that must be addressed to significantly strengthen forensic science practice in the United States today?

On day two, attendees were divided into five breakout groups to discuss the following topics: (i) standards and practices, (ii) validity and reliability, (iii) forensic algorithms, (iv) research, development, testing, and evaluation, and (v) non-technical issues. Each breakout

group consisted of 10 to 15 participants, including a group chair (a non-NIST participant) and a recording secretary (a member of NIST's forensic science program team). The group chair was responsible for (a) facilitating the discussion, ensuring multiple viewpoints were captured, the objectives were met, and the timeframes were adhered to, and (b) presenting a summary of the group findings and discussions in a plenary session following the breakout discussions. The recording secretary was responsible for (a) keeping track of the proceedings for their assigned breakout group, including major discussion points, counter viewpoints, and common themes identified by participants, and (b) in collaboration with the group chair, consolidating the substance and outcome of the discussions into a short summary presentation that was presented by the group chair following the breakout discussion.

Participants were assigned to specific breakout groups by the NIST organizers based on participants' expertise, experiences, and preferences, taking care to provide diversity and balance in terms of participants' professional backgrounds, roles, and responsibilities. During registration, participants were asked to rank their preferred assignments based on their expertise and experiences for which they believe they could offer the most constructive contribution. Most participants were assigned to breakout groups that aligned with their first or second choices.

Breakout groups were each provided with one overarching question and three specific objectives to address related to identifying and prioritizing near- and long-term strategic priorities in each topic area. Breakout groups had up to three hours to convene during the morning of the second day. The summary of discussions and outcomes from each breakout group were presented by the group chairs that afternoon. The overarching question and subtopics for each breakout group that participants were asked to address are provided below.

<u>Breakout I: Standards and Practices</u>

How should the forensic science community address the standards and practices challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term priority areas for the development of documentary standards and guidelines that will improve quality, consistency, and efficiency of forensic science practice.*

b. *Identify ways to assess and assure the quality of documentary standards and guidelines used in forensic science practice.*

c. *Identify mechanisms to promote and facilitate the adoption and implementation of documentary standards and guidelines in forensic science practice, including mechanisms for demonstrating conformance.*

Breakout II: Validity and Reliability

How should the forensic science community significantly strengthen public trust and confidence in forensic science practice in the United States today?

a. *Identify critical near-term and long-term priorities for forensic science methods, practices, and disciplines relating to validity, reliability, and public trust and confidence that will benefit from greater emphasis on research or standards.*

b. *Identify relevant criteria for determining fitness for purpose and assuring the validity and reliability of forensic science methods, practices, and disciplines.*

c. *Identify methods for assessing the foundational validity and reliability of forensic science methods, practices, and disciplines.*

Breakout III: Forensic Algorithms

How should the forensic science community significantly strengthen the responsible use of computational algorithms in forensic science practice in the United States today?

a. *Identify critical near-term and long-term priority areas for the use of computational (both procedural and artificial intelligence / machine learning [AI/ML]-based) algorithms in forensic science practice.*

b. *Identify factors relating to the design and use of computational algorithms to promote responsible applications in forensic science practice.*

c. *Identify mechanisms for assessing and assuring the validity, reliability, and fairness of computational algorithms for use in forensic science practice.*

Breakout IV: Research, Development, Testing, and Evaluation

How should the forensic science community address the research challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term research priorities, and methods for establishing such priorities, to strengthen forensic science practice.*

b. *Identify partnering and collaboration models to strengthen researcher and practitioner partnerships (including data sharing) to advance critical research.*

c. *Identify critical classes of research products (standard reference materials or data, research reports, tools, and technologies) to improve the validity, reliability, and standardization of forensic science practice.*

Breakout V: Non-Technical Topics

How should the forensic science community address non-technical challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term non-technical priorities to strengthen forensic science practice (e.g., information/data sharing, improving accreditation, quality management, proficiency testing, training, throughput efficiency, translation and implementation of research and standards).*

b. *Identify mechanisms to address critical non-technical priorities to strengthen forensic science practice.*

c. *Identify metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice.*

2.3. Participation

The list of participants is provided in Appendix B. Participants included individuals invited by meeting organizers as well as members of the public who registered to attend prior to maximum capacity being reached. Information about the event and a registration link were posted publicly on the NIST website.[7] Participants were invited based on their experience most directly related to the practice or litigation of forensic science in federal, state, and local jurisdictions across the United States. Presenters and panelists varied in terms of their roles, responsibilities, backgrounds, and experiences.

2.4. Organization of the Report

This report provides a narrative summary and description of the views and opinions expressed by participants during the two-day roundtable and workshop. Summaries of plenary presentations and panel discussions (day 1) and the outcomes of breakout group discussions

[7] "NIST Forensic Science Program Long-Term Vision and Strategic Priorities Roundtable Discussion" [Online]. Available: https://www.nist.gov/news-events/events/2023/09/nist-forensic-science-program-long-term-vision-and-strategic-priorities.

(on day 2) are provided in sections 3 and 4, respectively. Appendices to this report include the agenda (Appendix A) and the list of participants (Appendix B).

3. Panel Presentations and Discussion

3.1. Practitioner Perspectives

In this session participants focused on issues affecting the day-to-day work of forensic practitioners. The panel members were:

- Mr. David Kanaris, Forensic Laboratory Chief for the Alaska Scientific Crime Detection Laboratory
- Mr. Jeremy Triplett, Director of the Kentucky State Police Central Forensic Laboratory
- Dr. Michael Garvey, Director of the Philadelphia (PA) Police Department Forensic Science Laboratory
- Dr. Barry Logan, Chief Scientist and Senior Vice President for Forensic Sciences at NMS Labs

Presentations and discussions primarily focused on the following topics: the rapid pace at which changes are being introduced relating to forensic science methods, technologies, and practices; capacity demands; certification programs; and private versus public sector priorities.

3.1.1. Rapid pace of change

New methods, technologies, and practices have been called for by legal and scientific commentators, such as forensic investigative genetic genealogy, statistical applications and reporting frameworks, and blind testing (including blind verification and blind proficiency testing). However, participants highlighted that changes are being introduced at rates that exceed the ability of forensic service providers to keep up. The proposed new methods, technologies, and practices require additional funding, training, and personnel as well as cultural buy-in and prioritization. Specific concerns highlighted during the discussion included:

- Lack of funding and resources (e.g., time and personnel)
- Personnel feeling overwhelmed (e.g., number of standards and case backlogs)
- Lack of awareness and understanding of underlying research or principles
- Lack of availability or accessibility of training on validation and implementation of new methods, technologies, and practices
- Recruitment practices that focus on hiring people to do casework, not on people with the requisite knowledge or skills in research or validations

- Aversion to being on the "cutting edge" with the adoption of new methods or technologies due to the associated burden of *Daubert* hearings and admissibility challenges
- Perceptions that change means previous methods or practices were "wrong" or "bad" and will face scrutiny
- Defensiveness attributable to the adversarial nature of the criminal justice system
- Differing perceptions of what various groups within the criminal justice system want from forensic service providers

Access to resources and personnel dedicated to research and development, validation, and training may alleviate many of these concerns. Currently, these roles are often additional duties assigned to personnel and secondary to their primary job function of casework. Suggested ways in which these issues could be addressed included:

- Development of roles for support personnel not involved in day-to-day casework whose duties focus on research and development, integration of research findings into practice, validation studies, and training
- Creation of training programs and training resources that can be used by the support personnel
- Increased education opportunities for the criminal justice community so that they have a better grasp of the issues and methods for improvement

3.1.2. Capacity demands

Forensic service providers face a common set of demands driven by the expectations of the communities they serve. Several of the demands identified during the discussion included:

- Accreditation of all forensic service providers
- Relevant and validated standards in all forensic disciplines
- Organizational structures that allow forensic service providers to work with scientific independence *and* public safety priorities (noting that this does not necessarily mean separation from a law enforcement organization)
- Ability to provide timely analyses
- Research and innovation

While the value of these demands and associated expectations was widely recognized, participants emphasized that many forensic service providers have limited capacity to address them.

With increases in casework volume and demands for shorter turnaround times, capacity challenges quickly emerged as the most significant issue facing forensic service providers. Participants noted that requests for forensic services are outpacing the capacity of many

forensic service providers—both when fully staffed but especially when there are absences, vacancies, and people in training. This causes backlogs to grow, cases to be delayed, evidence or cases to be prioritized over others, and other important duties (such as research, innovation, and standardization) to be neglected. Casework, research, validation, implementation, and training all take time and forensic practitioners do not have control over their own schedules, priorities, timelines, or workloads. Instead, cases and courts dominate schedules. Consequently, practitioners often do not have the time to pause and consider innovative or alternative methods and therefore maintain status quo.

Although forensic practitioners are expected to stay abreast of the latest research and trends, forensic service providers are driven by productivity metrics. Consequently, dedicated personnel who have the requisite knowledge, skills, and time to commit to research, validation, implementation, and training on use of new methods, technologies, or practices are rare. Casework personnel and leaders often lack robust knowledge and skills in project management since it is not a core competency required for casework activities and is often not prioritized in recruitment efforts or training. Furthermore, many forensic service providers do not have access to scientific literature. The lack of exposure to current research can impact the ability of forensic service providers to identify, validate, and implement new technologies, methods, or practices. Even if forensic service providers have access to information about new methods, managers may be reluctant to make changes to existing systems because that would require validation, implementation, training, and changes to quality management systems—for which resources are often not available.

Because capacity is limited, participants noted that many laboratories prioritize tasks to maximize efficiency. Managers try to strike a balance between cost and impact. When the concept of "impact" is raised, it is necessary to ask "…on what?" Some might measure impact related to throughput, backlogs, and productivity—but not necessarily to the bigger picture: public safety and justice. With the primary mission of forensic service providers being public safety and justice, participants emphasized that until these resource limitations and capacity challenges are addressed, forensic service providers will continue to lack the resources and critical infrastructure to accomplish their primary mission, putting public safety and justice at risk. The resources that participants felt were most needed relate to people (analytical, technical, administrative, and other support staff), facilities (adequately sized and designed to account for personnel, equipment, and practices), training (new staff and existing staff), and validations (new methods, practices, and technologies).

Many issues could be addressed through cooperation, collaboration, and partnerships to provide support for research, validation, and implementation. Suggestions included:

- Research and validation partnerships, such as through visiting scientists in forensic laboratories and practitioner "residencies" in research laboratories (e.g., for six to twelve months)

- Access to scientific literature to keep casework practitioners informed of advances and to lower the barriers to the translation of research findings into practice
- Development and sharing of sound and robust model validation plans from a trusted entity with relevant scientific expertise
- Development and distribution of validation sample sets to facilitate validation and standardization
- Creation of teams with the expertise and resources that could be deployed to focus specifically on the validation of new or emerging technologies or methods onsite with forensic service providers
- Access to specialized expertise and support to help address admissibility challenges relating to the implementation of new or emerging technologies or methods

Another challenge noted by participants is the growing number of graduates with forensic science degrees who lack the foundational knowledge, skills, and experience necessary to perform casework without extensive training. There is a need to strengthen the education and training forensic practitioners receive prior to being hired. Suggestions included:

- Improvements to undergraduate and graduate programs so that staff can be qualified to conduct independent casework much sooner after hiring; currently qualification can take up to two years
- Additions to curricula of discipline- and test-specific (rather than generalized) coursework and "hands-on" work with analytical equipment and instrumentation
- Additions to curricula of real-world context and experiences relating to casework to reduce on-the-job training needs (e.g., varying types, conditions, and qualities of analytical materials)

3.1.3. Certification programs

Although there is increasing demand for practitioner certification, certification programs are not mandated or incentivized and lack consistency in terms of their requirements, test designs, quality, availability, and administration among disciplines. Thus, in their current form, some participants questioned their overall value and purpose. If certification programs were more robust and better aligned to the knowledge, skills, and abilities that forensic practitioners are expected to possess, those programs could provide a meaningful and universally accepted way for forensic service providers to assess and demonstrate competency and proficiency. Suggested ways the issues could be addressed included:

- Development of a clear and defined purpose of certification and a better alignment to the knowledge, skills, and abilities expected of forensic practitioners
- Establishment of universal and discipline-specific standardized requirements for certification programs
- Establishment of incentives and/or mandates for certification

3.1.4. Private sector priorities

Private sector forensic service providers[8] often compete with "free" services that are available from public sector providers. Competing with "free" requires a compelling value proposition, which includes innovation, quality, value, and service. Among those, innovation is foundational as it can lead to new analytical methods and capabilities, faster turnaround times, and lower costs for services. Consequently, participants felt that operational research tends to be more emphasized and prevalent in private sector forensic service providers. Current focus areas for operational research highlighted by some participants to help provide a competitive edge and stronger value proposition include automation, multiplexing technologies, and AI/ML for data screening and data mining.

Looking specifically at research, some participants indicated that private sector forensic service providers can have several strengths compared to their public sector counterparts: dedicated resources for research and method improvement/validation activities; greater insights into changing market trends and customers' needs; prioritized engagements with peers and customers; and nimbleness in technology adoption and acquisition. However, these strengths are not without their own challenges, which are also experienced by public forensic service providers, such as the need for early awareness of demands or opportunities for new tests; rapid development and validation of new methods or technologies; overcoming natural friction points in translating outputs from research and development into operational practice; sustaining the resources required for research and development; and continuous vetting of new technology solutions.

The value that private sector forensic service providers place on standards was also discussed. Not only do standards promote consistency in terms of quality expectations across jurisdictions, but they also provide safeguards for quality degradation, objective benchmarks for comparisons, and competitive differentiation. However, implementation can be challenging: consensus-based standards may not reflect laboratory priorities or unique capabilities, may be costly to implement (particularly if research and development resources are required or updates to existing methods and practices are needed), and justification for compliance can be difficult when standards are voluntary.

Suggested ways to address these issues included focusing greater attention and support toward the following topics:

- Development and validation of standards for automation
- Development of robust portable technologies for field-deployable applications

[8] Private sector forensic service providers refer to entities that are not under direct government control and include both for-profit and non-profit providers.

- Development of methods for leveraging data analytics on instrument data (e.g., chromatographic or mass spectrum data)
- Support for validation and implementation of novel methods and technologies
- Standards relating to AI/ML for assisting with common tasks faced by forensic service providers (e.g., data review)

3.2. Leadership Perspectives

In this session participants focused on issues affecting the overall management of the laboratory and broader forensic science community. The panel members were:

- Ms. Linda Jackson, Director of the Virginia Department of Forensic Science
- Mr. Brady Mills, Chief of the Texas Department of Public Safety Crime Laboratory Division
- Mr. Matthew Gamette, Director of the Idaho State Police Forensic Services Laboratory System
- Mr. Jason Bundy, Director of the Florida Department of Law Enforcement Forensic Services

Presentations and discussions primarily focused on the following topics: resource limitations; training; personnel wellness; analytical capabilities; consistency and standardization; translation and implementation; and information management and infrastructure.

3.2.1. Resource limitations

Participants noted that most forensic service providers are faced with a common situation:

- Their mission is to provide accurate and timely forensic testing services
- Their resources are generally predictable, with the majority coming from within their jurisdiction and a smaller portion from grant funding, but those resources are often not flexible when needs change or do not scale with cost increases, workload increases, or increased staffing requirements
- Their workload is increasing, but varies among disciplines
- Their staffing is largely based on the needs at a particular time (often at least one to two years in the past) and not necessarily flexible when needs change
- Their training is typically done in-house by personnel also responsible for casework
- Their development, validation, and implementation of new methods and practices are done in-house by personnel also responsible for casework

Consequently, forensic service providers face an uphill battle in the acquisition and allocation of resources to support increasing workloads, training needs, and validation and implementation of new methods and practices. The respective justification data is often outdated. Several participants felt that the impacts that resource limitations have on the effectiveness of the overall system has the potential to adversely affect the quality and timeliness of forensic testing services that are critical for public safety and justice.

3.2.2. Training

Training newly hired forensic practitioners typically requires two years, which places a strain on forensic service providers and creates a deficit in the number of technically trained personnel able to address incoming cases. Many participants indicated that even analysts who have prior training and experience from other service providers need additional training. This is largely due to variations in workflows, methods, technologies, and practices among forensic service providers as well as a lack of universally accepted criteria and test methods competency and proficiency. These issues create two challenges for many forensic service providers—not only is the trainee unable to immediately perform casework, but the personnel providing the training are also pulled away from casework. Consequently, some participants felt that the typical expectation that practitioners receive two years of training before they can be authorized for independent casework is not practical in the long-term given the resource limitations that many forensic service providers continue to face today. Suggestions for alternative solutions raised during the discussion included:

- Development and validation of tools that could be used during recruitment and interview processes to predict practitioners' ability to perform the job functions, particularly in the pattern comparison disciplines (e.g., visual acuity)
- Development of educational curricula within colleges and universities that provide greater depth and focus on more real-world and hands-on experiences with specialized instrumentation and technologies
- Creation of joint partnerships and regional deployments of practitioner training programs among forensic service providers, academia, or other entities providing shared training facilities, samples, and materials that are representative of casework
- Development of standards that clarify the breadth, depth of knowledge, and skillsets necessary to perform basic job functions

In addition to training newly hired personnel, training existing personnel can also be a challenge. Training topics of greatest interest included:

- Measurement uncertainty evaluation
- Designing validation studies
- Technical and scientific writing and publishing
- Data analysis and statistics

- Method performance measures and monitoring (e.g., sensitivity, specificity)
- Process mapping and gap assessments
- Instrument selection and evaluation
- Leadership development concepts and strategies

The importance of training relating to leadership and validation and implementation of new methods, technologies, and practices was emphasized. Training deficits in these areas often lead to inefficient or ineffective outcomes and hesitancy among forensic practitioners and leaders to embrace new methods and technologies—particularly because of the perceived difficulties in properly designing and executing defensible validation studies and handling of the legal challenges that might result. While several different entities offer continuing education opportunities within the forensic science community, those curricula need to be expanded to include these additional topics and provide better coordination, collaboration, and consolidation to eliminate duplicative efforts and promote more consistency in the information being taught.

3.2.3. Personnel wellness

Challenges relating to staff wellbeing have grown in recent years. Participants noted that the combined effects of staff shortages, case backlogs, tough testimony requirements, and an adversarial culture within the forensic science and legal communities have led to record numbers of personnel seeking assistance in dealing with stress, with some even considering demotions to avoid roles that require testimony. Many forensic service providers are struggling to understand the full effects of cumulative stressors and ways those stressors can be reduced. Participants discussed the urgent need for attention to these issues as they can have implications on the quality of forensic testing. Participants highlighted the need for better training, tools, and resources to identify and mitigate extreme stress and help build a strong, resilient workforce.

3.2.4. Analytical capabilities

Overarching and discipline-specific research needs were highlighted during the discussion focused on strengthening current analytical capabilities, including:

- Uses, effects, and modes of action of new and emerging drugs, including the selection of best instrumentation and analysis methods, identification of metabolites, reference sample development, and testing of impairing effects
- Statistics and applications in pattern evidence disciplines (e.g., friction ridge, footwear, tire track, and firearms and toolmarks), including method and technology development as well as validation and implementation (i.e., technologies must be accessible,

operationally relevant, and accompanied by sample protocols, training guides, and validation materials)
- Applications and field-use of multi-biometric identification systems (MBIS) to enable real-time identification capabilities
- Black-box and white-box studies to enable benchmarking and monitoring performance for key forensic disciplines
- Evaluations of new and emerging instruments and technologies—not only in terms of characterizing performance but also to understand instrument lifecycles and to inform the return on investment for costly service contracts—so that forensic service providers can make data-driven procurement decisions

Greater coordination, collaboration, and integration between researchers and forensic practitioners are also important for advancing analytical capabilities. Suggestions included embedding researchers (e.g., graduate-level or post-doctoral researchers) in operational environments alongside forensic service providers to jointly address specific scientific or technical challenges that have immediate impacts on day-to-day operations, policies, or practices (e.g., method development and validation) as well as engaging more forensic practitioners in discipline-specific scientific foundation reviews.

Finally, while discipline-specific scientific foundation reviews help identify important limitations to traditional analytical capabilities, some participants suggested that the reviews should also outline practical solutions to promote stronger technical foundations that are practical to implement given the scientific evidence available. Further, participants suggested greater input be solicited from the forensic science community regarding the topics and prioritization of future reviews.

3.2.5. Consistency and standardization

Participants emphasized the need to promote greater consistency and standardization across the forensic science community, but noted that progress has been slow. The number of standards that have been produced in recent years has made it difficult for forensic service providers to keep up. This has been compounded by the wide-ranging implications of many requirements. Standards relating to one discipline may include requirements that impact other disciplines, which often necessitates an interdisciplinary approach for implementation, while confusing language or too many requirements clumped together create additional challenges. Further, with multiple groups and organizations developing standards, participants noted that there can be differing requirements or guidance leading to confusion among forensic service providers on which standards to implement and when.

To alleviate some of these issues, participants suggested to consider consolidating, reorganizing, and developing a multi-organizational strategy within and between the various entities working on standards development. Questions were raised as to whether everything

within forensic workflows require standards. Efforts were suggested to map and monitor workflows and discipline-specific processes across forensic service providers to evaluate inconsistencies and identify optimal processes and priority areas for research and standardization. A "core" set of topics should be identified and prioritized that have interdisciplinary applicability and that are critical for the different scientific areas and disciplines to standardize (e.g., initial training, continual education training, certification, methods and instrumentation, reporting results). While interdisciplinary standards might help reduce the implementation burden, participants cautioned that those documents would need to be written in a way that ensures the requirements don't become overly broad such that the content becomes diluted and vague.

The implementation of standards available requires resources, which forensic service providers often do not have readily available. Participants suggested developing fiscal implementation estimates for each standard that the forensic science community is encouraged to implement, such as the standards listed on the Organization of Scientific Area Committees for Forensic Science (OSAC) Registry, to provide greater clarity and transparency into the fiscal implications of implementation so that forensic service providers can plan and budget more effectively. Furthermore, given the resource limitations faced by many forensic service providers, the importance for standards to remain free and publicly accessible was emphasized.

3.2.6. Translation and implementation

The implementation of new methods, technologies, and practices often requires significant time and resources for procurement of the technology; design and execution of validation studies; training of personnel; and development of standard operating procedures, competency testing, and on-going proficiency testing. Additionally, participants noted that implementation can have a heavy toll on personnel. Implementation activities stress the system by diverting resources away from casework, causing caseloads and backlogs to grow. Furthermore, personnel (including the discipline-specific technical leaders) often struggle to understand how to properly design and execute validation studies or are unfamiliar with concepts and principles underlying the methods or technologies to use them effectively. Even more, forensic service providers often vary in their approach to validation, instrument parameter settings, or operating procedures, resulting in differences among forensic service providers which can lead to legal implications when challenged during testimony. Suggested ways issues could be addressed to improve translation and implementation included:

- Dedicated training on validation (designing and executing validation studies, technical and scientific writing and publishing, data analysis and statistics, performance measures and performance monitoring, etc.)
- Better coordination, support, and development of test samples or data for validation efforts, specifically those that include multi-laboratory collaborations

- Development of explicit criteria and guidelines for conducting validations, such as guidance on how to approach validations (and verifications), model validation protocols (including guidelines for minimum breadth and depth of testing), model standard operating procedures, and recommended training materials
- Development of a centralized repository for sharing validation related materials and resources among service providers
- Means for providing technical review and guidance for validation plans to ensure technical rigor and soundness before validation studies are carried out

Finally, participants emphasized the importance of implementing blind proficiency testing programs as a means of monitoring and evaluating the performance of methods, technologies, and practices to ensure that their usage remains appropriate.

3.2.7. Information management and infrastructure

Many forensic service providers lack a fully integrated LIMS in their operational workflows. Participants indicated that forensic service providers are oftentimes forced to tailor their workflows to the capabilities of their LIMS and design ad hoc means for managing critical information and data. Further, the lack of integration and connectivity with other entities create challenges with data accessibility, disclosure notifications, and obtaining strategic and investigative insights. Suggested ways to address these issues included:

- Designing next-generation LIMS and information technology infrastructures that have better application programming interfaces (APIs) to permit information exchanges and data connectivity across multiple systems and the ability to support a distributed workforce (e.g., telework, remote work)
- Augmenting LIMS capabilities using embedded AI to promote data insights and efficiencies
- Enabling “forensic connectivity” across multiple databases, entities, and people (e.g., first responders, public heath, forensic labs, investigators, policymakers, litigators) to permit real-time access to critical information and data relevant to public safety and litigation
- Developing best practices related to case management procedures, workflows, and decisions
- Ensuring LIMS designs enable flexibility to adapt to operational workflows that align with best practices
- Enabling interoperability to allow for interchange of data between different information technology systems (e.g., LIMS and automated biometric identification systems)

3.3. Legal Perspectives

In this session participants focused on issues affecting the admissibility and presentation of forensic evidence in litigation. The panel members were:

- Mr. Raymond Valerio, Assistant District Attorney and Director of Forensic Sciences for the Queens County (NY) District Attorney's Office
- Ms. Jennifer Friedman, Deputy Federal Public Defender for the Central District of California
- Dr. Sarah Chu, Director of Policy and Reform, Perlmutter Center for Legal Justice, Cardozo Law
- Hon. Kent Cattani, Chief Judge for the Arizona Court of Appeals
- Hon. Ronald Reinstein, Judicial Consultant to the Arizona Supreme Court and Judge of the Superior Court of Arizona (ret)

Presentations and discussions primarily focused on the following topics: validity and reliability, consistency and standardization, and transparency and accessibility.

3.3.1. Validity and reliability

The importance of ensuring forensic science methods and practices are valid, reliable, and grounded by a rigorous evidence-base was universally agreed upon. However, participants noted that there can be disagreement in terms of the extent to which existing methods and practices are valid and reliable.

Some participants felt that many prosecutors are generally satisfied with current forensic science disciplines, and hold the view that the disciplines are well-supported by research and that examiners are accurate in their expert opinions. However, they felt that there will always be criticism given the adversarial nature of the judicial system. An example that was raised is the recent discourse on firearms and toolmarks. In 2016, the President's Council of Advisors on Science and Technology (PCAST)[9] raised concerns that there were not enough well-designed black-box studies to assess performance. However, once additional studies were completed, critics shifted focus toward the ecological validity[10] of those studies, then to the prevalence and treatment of inconclusive decisions, and, more recently, to sample "missingness" and the implications on generalizable performance data. This is not to say that prosecutors don't recognize that there are opportunities for improvement and that a discourse on challenges can be healthy and help identify blind spots to improve the discipline.

[9] President's Council of Advisors on Science and Technology (2016) *Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods* (U.S. Executive Office of the President, Washington, D.C., USA).
[10] Ecological validity refers to the ability to generalize research results to real-world settings.

When considering the significance of concerns relating to validity and reliability from prosecutors' perspectives and associated research needs, some participants suggested the issue can be framed in terms of questions that have remained unanswered. For example, in pattern evidence disciplines, questions that were highlighted included:

- What are the next-generation technologies that have been discussed over the last several years?
- Why has there not been wide adoption of recently introduced technologies (such as three-dimensional (3D) technologies for firearms and toolmark examinations)—is it resource limitations, lack of confidence that the analysts understand the technologies, or something else?
- Where are the statistical and algorithmic methods that have been purported to enhance objectivity? Why is it taking so long?
- Would white-box studies be helpful to identify how forensic science works and to see why examiners draw their conclusions?

When reviewing defense attorneys' viewpoints, a different perspective emerged. Defense attorneys recognized that calls for greater attention to forensic science issues have been heard, and they are generally appreciative of the work that has been done in recent years. However, participants identified several issues that still need to be addressed related to the evidence-base supporting claims about the validity and reliability of forensic science evidence introduced in criminal courts, including:

- More consistency in how discipline-specific scientific foundation reviews are conducted
- Research focused on establishing the validity of analytical methods in many forensic science disciplines to be conducted by independent entities and outside of the context of existing litigation or for purposes of furthering specific litigation
- Stronger experimental designs to avoid research that might produce biased or misleading information
- Increased investments relating to the impacts and mitigation strategies of cognitive bias
- Assistance with the development and implementation of proficiency testing schemes that have greater ecological validity, such as blind proficiency testing programs

From a judicial perspective, participants indicated that valid and reliable forensic results can only be achieved if the forensic science methods and practices are traceable to valid and reliable evidence-based research. Although courts are often end-users of results produced from forensic science methods, courts are not the appropriate place for resolving scientific disputes. Judges need valid and reliable results presented in an understandable manner. If any of those conditions are not met, then erroneous decisions can result.

A significant challenge to litigating scientific information raised during the discussion is that judges and attorneys are not scientists. They are unlikely to have backgrounds in science or the expertise to interpret complex scientific principles or independently assess questions of

reliability. Even more, some participants pointed out that most judges and attorneys are still not aware of the 2009 report on forensic science by the National Research Council[11] that identified scientific limitations of forensic disciplines. This challenge is exacerbated by the fact that judges must rely on the attorneys for each side to argue the issues; however, those arguments are often made by parties who are advocating for a specific result and cases are often handled by overworked and underfunded counsel. Further, some participants noted that, by nature of the adversarial system, experts, and at times, research is cited that is not "independent." Thus, the difficultly for judges to distinguish between methods that have valid and reliable foundations versus those that do not was emphasized. To that point, participants noted that although *Daubert* imposes a "gatekeeping function" and is intended to provide a framework for ensuring reliable (but perhaps not yet "generally accepted") methods are admitted while ensuring previously accepted (but not reliable) methods are excluded, in practice, it is not always effective. Participants noted that some judges tend to be overly lenient and admit testimony from experts who have what appear to be "good" credentials and point to cross-examination as the means by which concerns about the reliability of the testimony can be established, suggesting that reliability issues go to the weight of the testimony versus admissibility.

Within this context, participants highlighted the need for research that focuses on the validity and reliability of methods and practices (e.g., estimating error rates and determining method limitations). Such research should prioritize efforts designed to assess and improve accuracy, reproducibility, and repeatability of forensic disciplines. Importantly, participants stressed that the research should be conducted by entities independent of judicial issues and without a vested interest in the outcomes of litigation.

Suggested ways to address concerns regarding the validity and reliability of forensic science evidence included:

- Developing a single federal research agenda that:
    - consists of an interagency group of federal scientific agencies focused on basic, applied, clinical, and social science research (including sociotechnical analyses of methods and technologies)
    - addresses both extant and emerging research challenges
    - prioritizes solidifying the scientific fundamentals of existing methods and practices before addressing scaling issues
- Establishing guidelines and criteria for what constitutes a "well-designed" validation and reliability study
- Identifying and promoting means for strengthening the rigor of peer review for forensic science research

[11] National Research Council Committee on Identifying the Needs of the Forensic Sciences Community (2009) Strengthening Forensic Science in the United States: A Path Forward (The National Academies Press, Washington, D.C. USA). https://doi.org/10.17226/12589.

- Ensuring forensic science methods and practices are traceable to rigorous evidence-based research
- Establishing and applying an objective criterion-based rubric to evaluate the "rigor" of the scientific evidence base underlying forensic methods and practices (e.g., such as in healthcare)[12]
- Increasing investments in research relating to large-scale and well-designed interlaboratory studies, black-box studies, and white-box studies by neutral, impartial scientific entities
- Elevating and emphasizing the importance for research relating to the effects and mitigation of cognitive biases effects in forensic science
- Providing support for the development and implementation of blind proficiency testing programs in forensic science operations
- Creating a national program to provide independent third-party responses that states and localities could leverage when adverse events occur relating to forensic science and that could help identify systemic vulnerabilities and recommendations to strengthen practice moving forward (e.g., similar to the NIST-administered National Construction Safety Team)[13]

3.3.2. Consistency and standardization

The importance of consistency and standardization was universally recognized among participants. However, inconsistencies between discipline-specific standards have led to calls for overarching requirements for the contents of these standards. Additionally, participants raised the need for greater transparency of the process for approving or endorsing standards by the OSAC.

Participants suggested that standards need to be more rigorous, robust, and consistent. They should contain specific requirements, rather than recommendations, and represent the benchmark that all forensic service providers ought to adhere to without diluting requirements to accommodate the practices of the lowest-performing forensic service providers. The need for a better system of vetting standards and comments for technical rigor was also highlighted.

[12] Specific references cited included: Saran, A. (2023). What is a systematic review? Campbell Collaboration. Retrieved September 3, 2023, from https://www.campbellcollaboration.org/what-is-a-systematic-review.html; Berkman ND, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290- 2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm; Clair, J.S. (2005). A New Model of Tracheostomy Care: Closing the Research–Practice Gap. In K. Henriksen, J. B. Battles, E. S. Marks, & D. I. Lewin (Eds.), Advances in Patient Safety: From Research to Implementation (Volume 3: Implementation Issues). Agency for Healthcare Research and Quality (US). http://www.ncbi.nlm.nih.gov/books/NBK20542/.

[13] "National Construction Safety Team" [Online]. Available: https://www.nist.gov/disaster-failure-studies/national-construction-safety-team-ncst.

The current practice of relying solely on volunteers to commit the time necessary to fully evaluate a standard is not believed to be effective. Instead, participants suggested that there should be a committee of experts (e.g., technical writers, statisticians, and experimental scientists) who are compensated to ensure their attention is focused on the quality of the standards. Additionally, the standards should not only address analytical methods but also the reporting of results and testimony. Further, participants indicated that standards should include explicit requirements for blind testing (including blind proficiency testing) by forensic service providers as well as requirements to ensure the ecological validity of proficiency testing schemes, representativeness of samples, and access to the samples (or data underlying the samples) for independent review and scrutiny.

Also discussed were challenges in accelerating standards development and adoption. The consensus process used in standards development is believed to contribute to these challenges. Some participants expressed the view that consensus can lead to compromise, and compromise has the potential to dilute the power of science. Given that use of forensics standards is voluntary, participants recognized that not all will be adopted. However, in those situations, experts need to be prepared to handle challenges during cross examination relating to their lack of conformance.

3.3.3. Transparency and accessibility

Participants discussed the need for the forensic science community to formally adopt a set of core values, because values drive behaviors and behaviors reinforce values. Core values for forensic science should reflect the guiding principles of both science and law: evidence-based, transparent, just, and equitable. While evidence-based decisions are integral to ensuring validity and reliability of forensic science, the values of transparency, justice, and equity align to a broader theme of ensuring access. Participants noted that forensic practitioners are often socialized in an adversarial system where transparency and accessibility to information is equated with conflict. Thus, to move forward, participants suggested that the forensic science community will need to create a culture of transparency that acknowledges errors and uncertainty so that when issues arise, meaningful (and safe) conversations can be had about how those issues can be properly addressed to promote systemic improvements. Further, the need was discussed to adopt an industry model that embraces the duty and a systematic approach for reporting and responding to adverse events that impact the integrity of the forensic science product. When considering current and future challenges, participants also stressed that the forensic science community must address sociotechnical implications, such as systematic and structural disparities of research, technologies, methods, and practices proposed for investigative and criminal justice purposes.

Limitations in the transparency and accessibility of information can undercut the fairness and effectiveness of the criminal justice system. Participants discussed the need for forensic service providers and practitioners to have a better understanding and appreciation of

Brady[14] rules and other legal requirements affecting disclosure of information relating to forensic practices. Suggested ways to improve transparency and accessibility of information included:

- Ensuring that research activities provide:
  - more information about the research and access to the underlying data
  - better insights into the demographics of study participants (solicitation methods, backgrounds, experiences, etc.)
  - easier access to validation data and protocols underlying the performance of methods used by forensic service providers in operational contexts
- Ensuring that forensic science testimony and litigation processes:
  - include recognitions of the limitations of the methods
  - avoid overstatements of what is scientifically supported
  - provide method performance data or acknowledge method uncertainties when performance data are not available
- Establishing publicly accessible centralized repositories so that data and information are available that are relevant and foundational to ensuring sound forensic science practices (e.g., national forensic science repositories or libraries for validation data and studies, method protocols, and quality incident reports)
- Providing explicit detail relating to method protocols in standards so that courts and attorneys can evaluate whether methods were reliability applied in the case at hand

3.4. Researcher Perspectives

In this session participants focused on issues affecting the execution and implications of forensic science research. The panel members were:

- Dr. Austin Hicklin, Director of the Forensic Science Group at Noblis
- Dr. Keith Morris, Professor of Forensic and Investigative Science at West Virginia University
- Dr. Simson Garfinkel, Chief Scientist at BasisTech

Presentations and discussions primarily focused on the following topics: consistency and standardization, translation and implementation, and artificial intelligence and digital technologies.

[14] *Brady v. Maryland*, 373 U.S. 83 (1963).

3.4.1. Consistency and standardization

Research relating to black-box studies, white-box studies, and interlaboratory studies over the last fifteen years has been important for providing estimates of error rates and other measures of performance for several forensic science disciplines and has yielded insights with broad implications. The most significant of those insights is the extent to which there is consistency and standardization among forensic service providers and forensic practitioners. Key findings from those research activities highlighted during the discussion include the prevalence of outlier examiners, limited understanding of the overall population of participants, and variations in methods and practices, each of which are discussed in turn below.

First, a finding that is common across several black-box, white-box, and interlaboratory studies is that a small number of examiners made a disproportionate number of errors. Consequently, the performance of one examiner is not necessarily predictive of the performance of another examiner and, therefore, some participants suggested that the results of such studies should only be taken as imperfect estimates of the discipline. Participants also noted that, with few exceptions, errors were generally not associated with a specific forensic service provider, experience, or training.

Second, there is a lack of meaningful data relating to the size and demographics of the overall population of practitioners across various disciplines. Further, due to study protocols and related requirements for conducting the research, the studies must rely on volunteers (i.e., "opportunity sampling schemes"). Thus, participants noted that it is challenging to know the extent to which study participants are representative of the overall population, which can limit how broadly the results are generalizable.

Third, wide variabilities in policies and method protocols exist in some disciplines, which have contributed to disagreements and different results from forensic analyses conducted by different forensic service providers. Participants noted that these differences can be significant and consequential, despite the same types of instruments, technologies, or products being used.

Suggested ways to improve consistency and standardization included:

- Strengthening verification procedures and proficiency testing schemes so that performance issues relating to "outlier examiners" can be detected and addressed
- Gaining a better understanding of the overall examiner population (e.g., size, demographics) for each forensic science discipline in the United States to allow for more detailed and generalizable inferences to be made about the performance of a discipline
- Prioritizing the development of standards that relate to analytical methods and interpretation of results while ensuring that those documents are specific enough to enable greater consistency in practices among forensic service providers and reduce the variability in results

3.4.2. Translation and implementation

Research is motivated by several factors, including demands for new forensic analyses, opportunities provided by new technologies, improvements and characterization of method performance, improvements in efficiency, establishing theoretical foundations of disciplines or practices, and addressing legal challenges and requirements. Opportunities for research must be prioritized with consideration of factors such as case load demands, relative cost, or potential impact. Participants emphasized that impact requires the outputs from the research to be effectively translated and implemented into operational practice. Thus, once a research need has been identified, it is critically important that the research be designed with translation and implementation in mind. While challenging, various suggestions were discussed to help improve translation and implementation.

First, researchers must collaborate with forensic practitioners to ensure their research addresses a relevant issue and is designed with the context of operational use or applications. However, most practitioners do not have the time or resources to engage with researchers in a meaningful way. Consequently, the forensic science community needs to consider ways of strengthening collaboration between researchers and practitioners despite resource limitations.

Second, many people are resistant to change, and many forensic service providers can be risk averse—both of which are conditions that create nontrivial barriers to the voluntary adoption of new methods, technologies, or practices. Participants noted that this reticence is further compounded when the research involves complex principles or technologies (e.g., algorithmic methods) or when it may result in new legal challenges. Consequently, researchers need to be sensitive to these aversions and focus on strategies that reduce the risk (or perception of risk) and help the forensic service providers and practitioners feel more comfortable and confident with the proposed changes. Participants stressed that researchers need to work alongside practitioners, be trusted by practitioners, and be sensitive to the challenges and concerns raised by forensic practitioners.

Third, validations are necessary prior to adoption of new methods and technologies but have been consistently shown to be challenging for forensic service providers to undertake, particularly for first adopters, because of resources as well as a limited understanding of what validation entails. Thus, participants emphasized the importance for researchers to ensure that validation and implementation factors are accounted for in their design and execution of the research to ease the translation and implementation into practice.

Research is often conducted through short-term projects. While those can be effective under certain circumstances, there is also a need to prioritize longer-term research and provide the continuity and long-term support necessary to meaningfully address some of the more complex challenges facing the community.

3.4.3. Artificial intelligence and digital technologies

Artificial intelligence (AI) and digital technologies are being introduced at accelerating rates. This produces two different issues: the need for AI for (digital and conventional) forensics and the need for digital forensics for AI systems. Several different ways for AI to impact forensic science were discussed, particularly in areas where forensic practitioners lack the human capacity to perform the functions, such as by helping forensic practitioners become more efficient and effective in their tasks, by providing a means for evidence interpretation, or by leveraging data and information to perform investigations. Pattern evidence evaluation and digital forensics were given as examples of areas where the use of AI could be impactful. Examples in digital forensics include language translation of human text and machine text, data interpretation, cyber investigation tools and recommendations.

While AI systems can be helpful for forensics, participants recognized that there also is an emerging need for forensic analyses of AI systems. AI systems have remarkable potential; however, there will be failures from AI systems. Questions will be raised as to why the AI failed and whether the failure was idiosyncratic or systematic. As AI systems become more prevalent, researchers and practitioners will need to develop new techniques designed to address these issues. Specifically, participants discussed the need to:

- Develop and validate approaches to acquire and stabilize forensic evidence from AI-enabled cyber-physical systems[15]
- Develop the equivalent of "file hashing" for AI models to detect identical or equivalent models and to quantify model divergence and impact of model changes
- Distinguish model changes and associated behaviors that are benign versus malicious and emergent versus directed[16]
- Develop standards and policies around the use and scope of AI systems as well as standards for validation and evaluations of bias and error
- Identify strategies for mitigating biases and errors through pre-deployment certification or other accountability mechanisms

Some of these requirements apply to AI systems designed for conventional forensics. Participants discussed the importance of ensuring that these AI systems, including training and validation datasets, can be audited for fairness and that results can be explained, which is critical for their use for criminal justice purposes.

[15] Cyber-physical systems refer to engineered systems that integrate sensing, computation, control and networking into physical objects and infrastructure, connecting them to the internet and each other (e.g., https://www.nsf.gov/news/special_reports/cyber-physical/).

[16] Emergent changes refer to modifications that occur without a priori intentions by automatically adapting and learning from a changing environment. Directed changes refer to modifications that occur due to intentional interventions.

3.5. Quality Management Systems Perspectives

In this session participants focused on issues affecting quality assurance in forensic science practice. The panel members were:

- Ms. Lynn Garcia, General Counsel for the Texas Forensic Science Commission
- Dr. Peter Stout, President/CEO for the Houston (TX) Forensic Science Center
- Ms. Eva M.L. King, Quality Assurance Director for the Wisconsin State Crime Laboratories

Presentations and discussions primarily focused on the following topics: robustness and resilience, performance monitoring, and consistency and standardization.

3.5.1. Robustness and resilience

Traditional factors that must be accounted for in a robust quality management system include facilities; proficiency testing; internal and external audits; management reviews; records management; and standard operating procedures. The latter includes both technical and non-technical manuals (e.g., policy and procedures, validations, non-conformances and corrective action, training), technical (peer) review, and court testimony. Within the context of traditional quality assurance measures, topics of particular interest included:

- Improving training and education, such as through partnerships with local academia
- Strengthening evidence handling, testing, and interpretation practices by identifying and pursuing opportunities to review and implement standards recommended by the OSAC
- Implementing new instrumentation and equipment with appropriate technical training

In addition to traditional quality assurance measures, participants noted that an effective quality management system must also be resilient to changes and embody the principle of continuous improvement to mitigate risks to quality and impartiality and leverage opportunities for a better work product and more effective and efficient services. Areas of emerging interests that were highlighted included:

- Development of new software performance measures (including for AI-based systems)
- Implementation of blind proficiency testing
- Greater emphasis on trainings related to bias, risks to impartiality, and ethics
- Development of quality controls for transitioning hardcopy documentation to digital records

On the issue of record digitization, participants noted that data is a valuable commodity if it can be leveraged for actionable insights. Thus, as forensic service providers continue to digitize their methods and processes, there is also a need for the development of practical methods to

more efficiently and effectively analyze large amounts of often unstructured, incomplete, and unconnected data and reports for forensic intelligence purposes or to respond to specific data interrogation requests. Handling this via manual methods was considered by some participants to be impractical.

With these interests in mind, several other challenges were raised that have highlighted the need for:

- Additional training support due to personnel turnover
- More opportunities to support personnel wellness
- Better ways to mitigate the risks associated with human error
- Stronger collaborative engagements among forensic service providers
- Greater support and proactive efforts from academic institutions to address emerging challenges in examination and interpretation
- More frequent and effective training relating to bias and improvement of practices to reduce the possibility of biases and disparities, e.g.:
    - removal of demographic information from LIMS, such as gender, race, or age when it is irrelevant to evidence examination and interpretation
    - considerations for institutional and systematic biases that might be embedded in laws forensic service providers must follow, policies instituted by forensic service providers, or other norms practiced by forensic service providers

3.5.2. Performance monitoring

Participants noted that there is not a single resource for providing information to attorneys and judges in response to questions about the reliability of specific forensic science disciplines. Instead, the only option is to send various articles and references. However, this approach is not efficient in providing the requested answers and may cause recipients to be overwhelmed. Judges and attorneys want (and need) a "bench book" to assist them in understanding the reliability or limitations of forensic methods.

Recognizing that attorneys and judges are primarily interested in validity and reliability, participants underscored the importance of establishing performance metrics through expanded quality control programs that incorporate blind testing. Such programs would allow forensic service providers to assess and monitor the performance and reliability of the entire process—from evidence submission to reporting results—and provide a real-time evaluation of analytical procedures and identify areas for improvement. Some participants stressed that quality control programs should also include routine blind testing of systems for searching large national databases (e.g., CODIS, AFIS, and NIBIN). The extent to which these databases, search engines, and related processes (e.g., reporting and follow-up of investigative leads) are routinely tested is unclear, as are the results from such testing. Significant resources have been directed toward the development and implementation of these systems. However, their

reliability remains unknown until these systems undergo large-scale and systematic testing using blind samples as part of a continuous performance monitoring program.

Participants noted that establishing blind quality control programs would also help shine a light on the issue of inconclusive results—what they mean, when they are appropriate, and how they should be reported—particularly in pattern evidence disciplines where comparison methods are often subjective and often lack clear criteria for when to report inconclusive results. While inconclusive results might be appropriate for a particular case, inappropriate decisions of inconclusive could have inculpatory or exculpatory implications. Factfinders need to understand and account for measures of method performance, including the occurrence of inconclusive results, when weighing the information they are given. Although blind intramural quality control programs have been piloted by larger forensic service providers, participants recognized that smaller providers need help, particularly with program design, sample creation, and inter-laboratory coordination.

Suggested ways to strengthen performance monitoring included:

- Allocating resources toward assisting forensic service providers with the design, development, and implementation of quality control programs that include blind testing, for example:
    - coordinating collaborative expertise
    - providing centralized sample creation (though sample preparation and packaging need to be localized)
    - providing third-party testing to establish the "assigned value" or "expected result" for the sample
- Providing better clarity and consistency around metrics for assessing reliability of forensic practices, particularly when "inconclusive" results are possible

3.5.3. Consistency and standardization

Consistency and standardization are foundations for a strong quality management system. An overarching concern that was raised is the need to have a better understanding of the challenges with implementing standards. Participants noted that there is a growing discourse in the forensic science community as to whether emerging standards address too many or too few issues and the appropriate level of detail. There needs to be a better understanding of why these concerns are being raised so that practical solutions can be developed. Suggested ways the issues could be addressed included:

- Providing greater transparency and clarity around the standards development process (e.g., through public meetings of the Forensic Science Standards Board [FSSB])

- Strengthening the rigor and accountability of the standards development process (e.g., to ensure that standards address perspectives from different groups within the forensic science community, such as legal, quality, human factors, and statistics)
- Establishing a core menu of requirements and consolidating the number of documents—both intra- and inter-disciplinary (e.g., similar to NFPA 921 - Guide for Fire and Explosion Investigations by the National Fire Protection Association [NFPA])
- Expanding efforts to provide resources and support to assist in the implementation of standards (e.g., fostering regional/localized implementation cohorts)

Another issue that was highlighted is the need to develop standards that establish minimum requirements for evidence submission, including collection, packaging, and preservation (i.e., storage and stability). Some participants noted that the quality of evidence collected and submitted to forensic service providers has been problematic with many of the issues preventable if better practices were applied. For example, some participants observed that blind samples were easily identified as quality control samples because they happened to be properly packaged, stored, and labeled. Without standards to establish minimum requirements or clear direction for best practices, forensic service providers often do not have leverage to enforce better practices. Consequently, participants felt that forensic service providers often have little choice except to spend their limited resources on "doing the best they can" with the quality of the evidence they receive.

4. Breakout Groups

4.1. Standards and Practices

Standards are critical for facilitating consistency within and among forensic service providers and to ensure that forensic science evidence is collected, analyzed, and communicated according to accepted practices. To achieve these goals, there must be a clear process to identify priority topic areas for the development of standards, methods to assess and assure the quality of those standards, and mechanisms to support their implementation. The overarching question and subtopics that this breakout group was asked to address were:

How should the forensic science community address the standards and practices challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term priority areas for the development of documentary standards and guidelines that will improve quality, consistency, and efficiency of forensic science practice.*

b. *Identify ways to assess and assure the quality of documentary standards and guidelines used in forensic science practice.*

c. Identify mechanisms to promote and facilitate the adoption and implementation of documentary standards and guidelines in forensic science practice, including mechanisms for demonstrating conformance.

4.1.1. Subtopic A: Near-term and long-term priority areas

The objective for this discussion was to identify priority areas for developing standards in forensic science to enhance the quality, consistency, and efficiency of practice. Discussions primarily focused on topics most in need of standards (e.g., reporting, testimony, training, methods, technology, terms and definitions); the importance of greater interdisciplinary coordination, collaboration, and harmonization; and challenges relating to the processes used to develop standards. Several priorities were identified by the breakout group (Box 4.1.1).

One priority identified by participants is to develop standards for the reporting of results of forensic examinations. Another priority is standards for training and certification programs. Participants highlighted that many forensic service providers feel as they have no choice but to re-train forensic practitioners, even those with prior casework experience, because of inconsistencies in training programs and the lack of a standardized way of verifying competency. Participants also highlighted the need to establish clear terms and definitions within and across disciplines.

Another major issue that was discussed is the lack of apparent coordination and collaboration among committees developing standards. This affects interdisciplinary consistency and harmonization among documents and contributes to a high number of standards. There is a need to break down discipline-specific silos and consider ways to strengthen coordination and consolidate documents. Participants noted that not everything a forensic service provider does needs standards. However, for those activities where standardization is important, there needs to be greater consistency. Suggestions included:

- A single set of interdisciplinary requirements that could be augmented by discipline-specific requirements in lower-tier documents
- A systematic approach to the identification of priority topic areas for standards development, such as through interdisciplinary and discipline-specific process mapping exercises that highlight specific practices that need standardization the most
- Liaisons within committees who can strengthen connections and provide relevant context between committees and organizations (i.e., OSAC and applicable private sector standards development organizations, SDOs)

Another challenge raised is the time currently needed to develop standards, which is of particular concern for emerging technologies. This led to discussions over the effectiveness of the voluntary consensus process, particularly in the absence of mechanisms to prioritize and streamline development activities.

Box 4.1.1: Priorities relating to the development of standards to improve quality, consistency, and efficiency of forensic science practice:

(a) Develop reporting and testimony standards for all forensic disciplines

(b) Standardize training programs among forensic service providers to enable better alignment between discipline-specific training programs and certification programs

(c) Break down discipline-specific silos to increase consistency and cooperation among disciplines and identify core (common) topics for standards development

4.1.2. Subtopic B: Ways to assess and assure the quality of standards

The objective of this discussion was to explore ways to assess and ensure the quality of standards used in forensic science practice. Discussions primarily focused on the quality of existing standards that have been produced; the extent that standards developing entities ought to provide more stringent top-down directives and frameworks for document development; and how well the standards achieve their desired purpose. Several priorities were identified by the breakout group (Box 4.1.2).

The quality of standards that have been produced to date was a major point of discussion—not only in terms of technical rigor, but also in terms of the clarity and coherence of the writing. To the first point, participants indicated that there are no clearly defined criteria by which the technical rigor of documents can be assessed when considering inclusion on the OSAC Registry. While everyone has the opportunity to express their concerns and viewpoints during the approval process, participants felt that the effectiveness of these reviews are negatively impacted by the overwhelming volume of documents being produced, the lack of resources available to ensure meaningful reviews, and the lack of enforcement to ensure concerns that have been raised are appropriately addressed. To the second point, participants noted that some documents could benefit from technical editing to strengthen clarity and coherence of the writing to ensure proper interpretation of the meaning and intent of the requirements and recommendations. Currently, most standards are written by subject matter experts. However, participants felt that few have the background knowledge, skills, training, or experiences necessary to write clear, coherent, and effective documents.

Consideration was given to the type of documents being developed, distinguishing between requirements and recommendations. There was discussion on the language used, “shalls” and “shoulds” in particular, and whether the community would benefit more from focusing resources on writing requirements alone or also on writing recommendations. Concerns were raised that some (smaller) forensic service providers will only focus on requirements (i.e., “shalls”) and might not give the same consideration to recommendations

(i.e., “shoulds”). Given the nature of the consensus process, however, some participants questioned whether top-down directives of this nature would be permissible (e.g., directives from the FSSB to the OSAC committees and subcommittees).

Discussions also addressed the extent to which current standards have achieved their desired purpose. Some participants suggested that this could be assessed by measuring the impact standards have on a practice and by soliciting feedback from the forensic science community about the standards available for implementation. Both mechanisms could provide general context about the quality and utility of documents that have been produced and inform specific areas for improvement during periodic reviews.

Box 4.1.2: Priorities relating to assessing and assuring the quality of standards used in forensic science practice:

(a) Develop model standards and frameworks (i.e., “a standard for standards”) to guide the development of standards that include minimum requirements and recommendations relating to topics and content

(b) Develop a more robust training program on how to develop standards (i.e., drafting, commenting, and adjudication)

(c) Provide access to technical editors tasked with improving the clarity and consistency of documents as they are being developed

(d) Survey members of the forensic science community about the standards available to implement and create a feedback loop for future standards development and review

4.1.3. Subtopic C: Mechanisms to promote and facilitate adoption and implementation

The objective of this discussion was to explore mechanisms to promote and facilitate the adoption and implementation of standards in forensic science practice, including mechanisms for demonstrating conformance. Discussions primarily focused on challenges relating to adoption and implementation; strengthening outreach and training relating to the use of standards; and developing auditing mechanisms that enable forensic service providers and others to assess the extent forensic service providers conform to applicable standards. Several priorities were identified by the breakout group (Box 4.1.3).

Several participants noted that the adoption and implementation of standards has been a significant challenge for many forensic service providers, including those that employ personnel who actively participate in standards developing activities. Not only do forensic service providers often lack the personnel with the project management skillsets needed to effectively execute the implementation, but they also lack the capacity and resources to do so.

Further, the number of standards that forensic service providers are expected to implement has become overwhelming, even for better resourced entities.

Participants discussed the need to better understand barriers to implementation and possible incentives to accelerate implementation. Suggestions included:

- Financial incentives (i.e., grants)
- Better communication of the value of implementation and its impact on laboratory leadership, practitioners, and others within the criminal justice system
- More effective outreach and training relating to the use of the standards, such as by integrating them into:
    - training and educational programs for forensic practitioners and litigators—both as part of continuing education programs as well as college curricula
    - certification programs for forensic practitioners

Discussions also addressed mechanisms for forensic service providers to demonstrate conformance to standards that have been recommended for forensic science practice (e.g., such as those on the OSAC Registry). While voluntary disclosure of implementation is an important step forward, some participants indicated that adoption and implementation efforts will likely be accelerated if internal and external auditing mechanisms are developed to provide greater transparency around the issue of conformance. Such auditing mechanisms could be used as a tool to help with implementation and assess the extent of conformance. They can also serve as a precursor to future accreditation and as a means for other entities (e.g., forensic science commissions, regulatory agencies, or legislative bodies) to promote adoption and implementation through influence mechanisms or mandates.

Box 4.1.3: Priorities relating to mechanisms to promote and facilitate the adoption and implementation of standards used in forensic science practice:

(a) Evaluate the extent to which standards have been adopted by forensic service providers, particularly by those that employ personnel who actively participate in standards developing activities, and determine the reasons for lack of adoption and implementation among those entities to date

(b) Establish financial incentives (e.g., grant funding that does not necessarily have accreditation requirements) to help offset resource limitations and incentivize the adoption and implementation of standards

(c) Better communicate the value of standardization and provide continuing education programs relating to the use of standards to forensic service providers and others within the forensic science community (e.g., litigators, judges, law enforcement organizations), so that they understand the role of standards and can create the necessary conditions and pressures to foster adoption and implementation

(d) Strengthen educational and training opportunities and integrate the use of standards in continuing education programs for forensic practitioners and leadership and in college curricula

(e) Create auditing mechanisms (internal, external, third-party) so that conformance to standards can be assessed and demonstrated

(f) Provide the means for advisory, oversight, or regulatory bodies (e.g., forensic science commissions, regulatory agencies, or legislative bodies) to promote adoption and implementation of standards through influence mechanisms or mandates

4.2. Validity and Reliability

The validity and reliability of forensic science practices are foundational elements for the admissibility of forensic science evidence in criminal and civil litigation and serve as the backbone for public trust and confidence in forensic science. To ensure the criminal justice system is fair and effective when it relies on forensic science results, there must be clarity around the priorities relating to research or standards to strengthen the validity, reliability, and public trust and confidence of forensic science methods, practices, and disciplines; a common understanding of relevant criteria for determining fitness for purpose; and methods for assessing the foundational validity and reliability of forensic science methods, practices, and disciplines. The overarching question and subtopics that this breakout group was asked to address were:

How should the forensic science community significantly strengthen public trust and confidence in forensic science practice in the United States today?

a. *Identify critical near-term and long-term priorities for forensic science methods, practices, and disciplines relating to validity, reliability, and public trust and confidence that will benefit from greater emphasis on research or standards.*

b. *Identify relevant criteria for determining fitness for purpose and assuring the validity and reliability of forensic science methods, practices, and disciplines.*

c. *Identify methods for assessing the foundational validity and reliability of forensic science methods, practices, and disciplines.*

4.2.1. Subtopic A: Near-term and long-term priority areas

The objective for this discussion was to identify priority areas for methods, practices, and disciplines relating to validity, reliability, and public trust and confidence that will benefit from greater emphasis on research or standards. Discussions primarily focused on approaches that would promote and facilitate greater transparency and accountability of forensic science practices, including adoption and implementation of standards; demonstration of conformity to standards; and enabling greater public accessibility of methods and practices, including studies and data supporting their validity and reliability (to the extent allowable by existing privacy laws or regulations). Several priorities were identified by the breakout group (Box 4.2.1).

The adoption and implementation of standards into forensic science practice was raised as an immediate priority. Standards not only promote consistency within and among forensic service providers, but also provide a means of demonstrating conformance to requirements and recommendations that have been formulated through an open, consensus-based process by different groups within the criminal justice system. Further, participants highlighted that as standards are adopted and implemented by an increasing number of forensic service providers, a significant amount of experience and data can be accumulated, which strengthens the empirical foundation underpinning the validity and reliability of methods, practices, and disciplines. This not only bolsters legitimacy but also strengthens public trust and confidence in forensic science as an institution and the credibility of information relied upon by the criminal justice system.

While adoption and implementation of standards was widely recognized as a significant and near-term priority, several challenges to achieving this across the community were identified, including:

- Conformance to discipline-specific standards remains voluntary in the United States

- Adoption and implementation of discipline-specific standards are not uniform among forensic service providers (e.g., no implementation of some standards and partial implementation of others)
- Many forensic service providers lack the resources and capacity to make the necessary changes for adoption and implementation (e.g., overhaul of quality manuals and procedures, equipment acquisitions, method validations, personnel training, competency testing, and proficiency testing)

Suggested ways to mitigate some of these challenges included:

- Financial incentives to support adoption and implementation
- Creation of influence mechanisms or mandates, particularly for some set of "core" standards that are most impactful to the validity and reliability of forensic results
- Evaluation and grading of the obstacles to adoption and implementation of each standard so that forensic service providers can better prioritize and plan their efforts

The need to establish mechanisms for forensic service providers to demonstrate conformance to discipline-specific standards was another key issue raised. While self-declaration of conformity is important, public trust and confidence require transparent demonstration of conformance. Some suggestions encouraged incorporating discipline-specific standards into supplemental requirements for accreditation so that conformity assessment bodies can provide third-party assurances that requirements are met. Other suggestions encouraged forensic service providers to publish their methods (e.g., standard operating procedures) online so that interested parties can directly evaluate the extent to which forensic service providers are conforming to standards. Participants noted that greater accessibility of standard operating procedures would not only provide the transparency necessary for promoting public trust and confidence, but it would also contribute to improving the efficiency and effectiveness of pre-trial discovery processes.

In addition to providing public access to standard operating procedures, participants emphasized the need to publish studies and data supporting the validity and reliability of methods. Suggested ways to address these needs included:

- Development of a centralized, open access repository for studies and data
- Creation of model standard operating procedures, validation plans, and reports for forensic service providers to consider
- Recommendations for handling of personally identifiable information such that data relevant to the validation studies can be made public while adhering to applicable privacy laws and regulations

Box 4.2.1: Priorities relating to strengthening the validity, reliability, and public trust and confidence in forensic science methods, practices, and disciplines:

(a) Wide-scale adoption and implementation of standards by forensic service providers

(b) Demonstrate conformity to standards (i.e., conformity assessment through third-party auditing schemes and integrating into requirements for accreditation)

(c) Provide public access to standard operating procedures and validation information (data and studies)

(d) Develop a centralized repository with a sustainable and scalable infrastructure to provide public access to validation information (data and studies)

(e) Produce model standard operating procedures, validation plans, and reports that are publicly available for forensic service providers to consider

4.2.2. Subtopic B: Criteria for determining fitness for purpose

The objective for this discussion was to identify relevant criteria for determining fitness for purpose and assuring the validity and reliability of forensic science methods, practices, and disciplines. Discussions primarily focused on strengthening guidance and infrastructure for validations; improving transparency and accessibility to validation information, including the occurrence and mitigation of quality-related incidents; and instituting continuous quality assurance principles. Several priorities were identified by the breakout group (Box 4.2.2).

While the concept of fitness for purpose is often used to describe whether a method actually works for its intended application, participants noted that there is no universal definition for acceptable performance or minimum number of samples that must be tested to determine whether this has been achieved. Instead, the level of confidence in a method is often characterized based on the testing that has been completed, which may vary among disciplines and forensic service providers. Concerns were raised that validation studies are generally not done well across the community and the confidence that one can have in the result of a method given the validation that has been performed is not well understood or communicated. Participants indicated that many forensic service providers do not have sufficient expertise relevant to the principles of validation, including on how to plan and execute well-designed validation studies. Suggested ways to help address these concerns included:

- Development of standards, including minimum requirements, on validation for each discipline
- Creation of model validation plans and summary reports illustrating how to design and document validation studies

- Development of more robust training on validation in all disciplines
- Establishment of mechanisms for validation plans to be reviewed by external entities with the relevant expertise to provide feedback or advice prior to execution

The lack of transparency and access to validation information was also raised as a concern. Participants discussed that the quality of validation studies varies, and it often takes significant effort to determine whether there are issues with validation of a particular method without access to the information underpinning such claims. Consequently, participants suggested that forensic service providers make their validation information publicly accessible in a centralized repository, particularly for methods that are currently being used. Then, as resources permit, participants suggested forensic service providers also post archived information for methods used in older cases. Participants recognized that many forensic service providers might find this challenging to accomplish given the limited knowledge among providers on how to effectively write validation summaries and continued cultural resistance to sharing information. For example, participants noted that the American Society of Crime Laboratory Directors (ASCLD) Forensic Research Committee (FRC) has created a repository to help promote this objective; however, only a small number of forensic service providers have used it. Nevertheless, participants felt the culture is starting to change, and more forensic service providers are changing their mindset toward openness and transparency. Suggested ways these efforts might be accelerated include providing incentives or recognitions to forensic service providers who make their validation information publicly accessible.

Another factor relevant to the concept of fitness for purpose that was discussed is the need to acknowledge errors and "quality incidents" so that forensic service providers are able to learn from them. Participants called for quality incidents to be more transparent and publicly accessible in a centralized location so that they can be monitored and considered in aggregate when evaluating trends or systemic vulnerabilities in methods and practices. This enables forensic service providers to be proactive in addressing issues when they arise in the community rather than reactive after they occur in their own operations. Participants felt that posting this information publicly would allow the community to learn from each other and grow stronger together. How to accomplish this in a way that protects the privacy of the individuals involved while still complying with legal requirements became a point of discussion. Some participants suggested this topic be addressed in standards to ensure the relevant information is available in a consistent format for aggregate monitoring and trend analyses.

Box 4.2.2: Priorities relating to determining fitness for purpose and assuring the validity and reliability of forensic science methods, practices, and disciplines:

(a) Clarity around the criteria and requirements for planning, executing, and reporting validation studies, including the development of standards (what is required, how to accomplish it, what it means in terms of confidence), model validation plans and summaries, and training relating to the principles of validation and execution of validation studies for each discipline

(b) Publicly accessible validation information (data and studies) and reporting of quality incidents

(c) A centralized repository with a sustainable and scalable infrastructure to provide public access to validation information (data and studies) and quality incidents

4.2.3. Subtopic C: Methods for assessing foundational validity and reliability

The objective for this discussion was to identify methods for assessing the foundational validity and reliability of forensic science methods, practices, and disciplines. Discussions primarily focused on establishing a clear definition for "foundational validity" and the requirements that must be met to achieve such a designation as well as the challenges relating to different approaches for assessing foundational validity and reliability. Several priorities were identified by the breakout group (Box 4.2.3).

While the concept of "foundational validity" requires empirical evidence of accuracy, repeatability, and reproducibility, questions were raised about whether there will ever be a point where a forensic discipline is considered to have a solid foundation, and whether such foundations would endure over time. This is particularly important because scientific methods and practices evolve, as do the circumstances to which those methods and practices are applied. Where relevant data exists, it often has limitations. Participants emphasized, however, that there is a distinction between the lack of available data to demonstrate validity versus the lack of validity demonstrated by available data. Some participants pointed to bitemark analysis as an example where data exists that does not support foundational validity.

Further complicating the issue of determining "foundational validity" is that while black-box studies and interlaboratory studies are often considered critical to evaluating the reliability of a discipline, those studies can vary in terms of quality and ecological validity (i.e., representativeness of samples, participants, test conditions), which can limit the extent to which those studies are rigorous enough to validate specific methods on their own. Participants highlighted that this is especially true when variability exists among forensic service providers. In such situations, black-box studies and interlaboratory studies often provide a general estimate of the performance of the discipline in aggregate, combining different

methods (and derivations) together. While informative, it is often difficult to extrapolate the estimates of performance from black-box studies or interlaboratory studies to the performance of a specific method applied by a specific forensic service provider. Although this is an important limitation, participants emphasized that published data should be accounted for, even if it is not comprehensive, and that data produced by black-box studies and interlaboratory studies provide an effective and practical means of assessing the performance of a discipline overall. Limitations on the quality of such data and the inferences that the data support can be considered separately rather than discounting the data altogether. Participants suggested that committees involved in standards development activities could catalogue the published data relevant to supporting the validity of methods and practices referenced in standards since this information should already have been considered when drafting such documents. If relevant data does not exist to support a method or practice referenced in a standard, then this should likewise be noted.

While black-box and interlaboratory studies are an effective and practical means for evaluating foundational validity and reliability for a discipline, those studies require resources. Participants noted that there should be a balance between the burden imposed on study organizers and participants and the need for the data (i.e., its potential impact on the discipline, forensic service providers, and others within the criminal justice system). Potential incentives were suggested, such as providing financial stipends to support overtime pay or to help offset other resource challenges faced by forensic service providers.

Box 4.2.3: Priorities relating to assessments of foundational validity and reliability of forensic science methods, practices, and disciplines:

(a) Establish a clear definition for what constitutes foundational validity and reliability

(b) Catalogue published data relating to foundational validity and reliability of forensic science methods, practices, or disciplines to facilitate assessments of data quality and support for foundational validity and reliability

(c) Conduct more black-box and interlaboratory studies relating to different forensic science methods, practices, and disciplines, and do so in a recurring fashion to ensure data is both relevant and applicable to current circumstances

(d) Incentivize forensic service providers to participate in black-box and interlaboratory studies, such as by providing support to offset resources required for participation, so that the data gathered is representative and useful to assessments of foundational validity and reliability

4.3. Forensic Algorithms

Practitioners of many forensic science disciplines are facing increased calls to provide more objective interpretations of forensic evidence using empirical measurements, statistical data, and probabilistic models. Forensic algorithms are often a means for providing these capabilities. They may furthermore enable automation to improve quality and efficiency as well as new analytical approaches. However, algorithms and their implementation in software can be complex, and come with their own sets of challenges when used for criminal justice purposes. To ensure that forensic algorithms are used appropriately within the criminal justice system, there must be clarity around their purpose; design considerations that promote responsible use; and methods for assessing and assuring the validity, reliability, and fairness of the algorithms. The overarching question and subtopics that this breakout group was asked to address were:

How should the forensic science community significantly strengthen the responsible use of computational algorithms in forensic science practice in the United States today?

a. *Identify critical near-term and long-term priority areas for the use of computational algorithms (both procedural and AI/ML-based) in forensic science practice.*

b. *Identify factors relating to the design and use of computational algorithms to promote responsible applications in forensic science practice.*

c. *Identify mechanisms for assessing and assuring the validity, reliability, and fairness of computational algorithms for use in forensic science practice.*

4.3.1. Subtopic A: Near-term and long-term priority areas

The objective for this discussion was to identify priority areas for the use of computational algorithms (both procedural and AI/ML-based) in forensic science practice. Discussions primarily focused on the strengths of algorithms and the potential value they can provide to forensic applications; challenges associated with the use of algorithms in forensic science; and the creation of guidance and testing frameworks to promote greater transparency in the design, development, and use of algorithms in forensic science. Several priorities were identified by the breakout group (Box 4.3.1).

Topic areas identified by participants that need to be addressed include data needs, development, implementation, testing, output presentation (in the legal system), legal challenges, and transparency. Transparency was an overarching issue that transcended each of the other topic areas (i.e., the need for transparency in data, transparency in development, transparency in implementation, etc.). The way those issues are approached may vary depending on the specific type of algorithm (i.e., procedural versus AI/ML), but all remain applicable.

The current generation of AI/ML algorithms are data intensive, often requiring large amounts of training data. Despite these requirements, algorithms are at a transitional point and, in some circumstances, have still been shown to perform similarly to that of human experts (e.g., face identification). Participants recognized that the performance of algorithms will only continue to improve and offer incredible potential but have new challenges.

Applications where algorithms could provide immediate value to forensic service providers include procedural data analysis, automation, and augmentation to improve efficiency and productivity. Examples discussed include the use of algorithms to:

- Provide recommendations relating to evidence collection, sampling, and preservation
- Automate image searching
- Augment image comparisons
- Enable complex data computation, analysis, and interpretation

While algorithms have immense potential, participants felt that there needs to be careful consideration for how they are used in forensic science since there can be implications to civil liberties and challenges from the criminal justice system. Importantly, algorithms cannot replace human testimony—the sixth amendment to the Constitution of the United States provides the right for criminal defendants to confront the witnesses against them. Participants noted that there should always be a human to serve as that witness. Whether that person is the user of the algorithm, developer of the algorithm, or tester of the algorithm, someone will need to be able to address questions and concerns relating to the use of the algorithm and associated validity and reliability if the outcome of the algorithm produces information that is used as evidence in court. Discussions relating to possible risks of constitutional infringements and questions pertaining to civil liberties extended beyond just how the algorithms are used, and included issues concerning the design and development of the algorithms as well as the data used for training and testing the algorithms.

Participants noted that the extent to which an algorithm will be relied upon will often depend on the extent to which it is understood and trusted. Algorithms are inherently complex systems. Further, there are fundamental differences among different types of algorithms, such as procedural algorithms and AI/ML-based algorithms. Thus, the importance of transparency relating to how algorithms are designed, developed, tested, implemented, and used was emphasized. A common issue highlighted by some participants is the lack of transparency around the internal operation of an algorithm and the data used in its development. Intellectual property protections and licenses that prohibit reverse engineering often contribute to limited transparency. Some participants noted that algorithms developed by commercial vendors can be challenging from a litigation standpoint, particularly if the vendor does not understand the needs of the legal system and the importance of due process and disclosure.

The Secure Software Development Framework[17] published by NIST was highlighted as a possible resource for commercial vendors to help mitigate some of these concerns.

Like transparency, the importance of empirical testing also emerged as a central theme in the discussion. Empirical testing in this context means applying various inputs to the algorithm and evaluating the outputs. Such testing should be done by experts with knowledge of computer software and the context of the intended use case. The more testing that is done under different conditions, the better the capabilities and limitations of the algorithm can be understood, particularly when the testing reveals situations for which the algorithm fails so that the application boundaries of the algorithm can be established. To that point, participants emphasized that multi-tiered testing schemes are critical (i.e., testing in development and deployment, testing by the end user, and testing by independent parties) and should go hand in hand with code reviews, where practical. However, code reviews might not be practical or useful with AI/ML-based algorithms, requiring more emphasis on the review of the data sets used in their training and validation.

Another important point that was raised concerns population sampling represented by the data. Training data must be collected in equitable ways and accurately represent the population of interest for the intended purpose to avoid the potential for algorithmic biases to emerge. A major challenge, particularly with AI/ML-based algorithms, is that the training data needs to over-represent smaller segments of the population to ensure the algorithms perform equally well for those segments. However, when algorithms are used to express the weight of evidence, for example, the data must provide a statistically adequate sampling of the population, such that all segments are proportionately represented. Further complicating this is that populations can change over time. Thus, data needs will also likely evolve, requiring the algorithms to be able to routinely adapt to those changes. Participants noted that these data challenges underscore the complexities involved with the use of algorithms in sensitive domains such as criminal justice.

The need for better application programming interfaces (APIs) to permit multi-system integrations and data connectivity was also highlighted. Oftentimes, the same high-level algorithms are used across multiple software applications. APIs would help standardize data representation and permit better data exchange, thereby helping provide the infrastructure for algorithms to build on one another. Standardized data representation not only helps facilitate transparency, but having the ability to connect multiple systems through APIs also allows leveraging well-established and well-tested algorithms to accelerate innovation rather than require developers to start from scratch each time. APIs also promote accessibility to different algorithms and software applications, as well as data exchange between different systems. For example, standardized APIs could allow for different LIMS software applications to exchange data across different platforms and entities.

[17] "Secure Software Development Framework" [Online]. Available: https://csrc.nist.gov/projects/ssdf.

Box 4.3.1: Priorities relating to the use of algorithms in forensic science practice:

(a) Identify where automation (and augmentation) can improve efficiency and support the development and implementation of the respective algorithms

(b) Develop guidelines for software developers in how basic information can be provided to improve transparency, including:
   i. Developer identity
   ii. Software categorizations
   iii. Certification for conforming to the NIST Secure Software Development Framework (SSDF)
   iv. First-party training dataset coverage
   v. First-party testing structure and dataset coverage
   vi. Data provenance/collection and retention policies

(c) Develop guidelines for end users of algorithms on how basic information can be provided to improve transparency, including second-party testing structure, dataset coverage, and deployment plan

(d) Develop frameworks for determining when a sample is not appropriate for submission to an algorithmic method

(e) Develop frameworks for testing different types of algorithms

(f) Develop standardized data representations and application programming interfaces (APIs)

4.3.2. Subtopic B: Factors relating to design and use for responsible applications

The objective for this discussion was to identify factors relating to the design and use of algorithms to promote responsible applications in forensic science practice. Discussions primarily focused on the importance of transparency, explainability, and understandability; ensuring the use of algorithms are limited to the scope of their validation; and consideration of sociotechnical implications and bias impact assessments relating to their use. Several priorities were identified by the breakout group (Box 4.3.2).

Participants noted that it is important to design and develop algorithms in a way that ensures they are secure, transparent (both in terms of documentation and function), explainable, and understandable by forensic practitioners and factfinders. Not only are these factors important from due process and disclosure standpoints, but they also could increase the comfort and willingness of forensic practitioners and others within the criminal justice system to apply algorithms and leverage the benefits that they can provide. Further, participants

suggested that these principles can help reduce the potential for adverse outcomes to occur due to, for example, incorrect uses of an algorithm, inappropriate applications of an algorithm, or incorrect or misleading interpretation or communication of the output from an algorithm.

Ensuring that the use of algorithms is limited to the scope of their validation was also highlighted. Issues that were raised included:

- Consideration of factors relating to human-computer interactions in the design and development of the algorithm
- Training on the appropriate use of the algorithm, including conditions and scope of acceptable use
- Thorough and multi-tiered testing and validation of the algorithm to establish the conditions and scope of acceptable use and the conditions for which it should not be used (i.e., conditions for which the outputs are not reliable)
- The development of standard operating procedures and protocols governing algorithm use, including disclosure of datasets used during development and testing
- The creation of appropriately representative datasets to support development and enable testing and validation
- The use of audit mechanisms, version controls, digital signatures, and security frameworks to monitor or mitigate issues that could impact the integrity of the algorithm or responsible use of the algorithm

Finally, participants emphasized the need to consider sociotechnical implications and conduct bias impact assessments as part of an implementation strategy and before deploying the algorithm operationally. Some participants felt that algorithms have been deployed in the criminal justice system without consideration of the impacts that their uses can have on communities and, in some situations, algorithms have served as a means of systematically perpetuating certain social inequalities. The risks of these issues manifesting will vary depending on the type of algorithm, its intended use, and factors relating to its development (e.g., source, type, and representativeness of training data).

Box 4.3.2: Priorities relating to the design and use of algorithms to promote responsible applications in forensic science practice:

(a) Design and develop algorithms in ways that improve explainability and understandability

(b) Ensure a transparent record of documentation is maintained relating to the development and use of algorithms

(c) Implement proper software development methods during the design and development of algorithms (e.g., as recommended by the Secure Software Development Framework published by NIST)

(d) Engage human-computer interaction specialists and incorporate the respective best practices in the design and development of algorithms

(e) Conduct multi-tiered testing of algorithms to establish the conditions for acceptable use and ensure testing data is different from data used during development

(f) Promote transparency and disclosure relating to datasets used during development, testing, and validation of algorithms

(g) Develop appropriately representative datasets that can be used for development and testing of algorithms

(h) Promote broad-based adoption of digital signatures in algorithms to enable authenticity and integrity verification

(i) Conduct bias impact studies prior to deployment of algorithms

4.3.3. Subtopic C: Mechanisms for assuring validity, reliability, and fairness

The objective for this discussion was to identify mechanisms for assessing and assuring the validity, reliability, and fairness of algorithms for use in forensic science practice. Discussions primarily focused on data needs and the importance of considering transparency, testing, explainability, and understandability in the design, development, implementation, and deployment of algorithms. Several priorities were identified by the breakout group (Box 4.3.3).

Participants emphasized the importance of ensuring that data is appropriately collected, representative, and unbiased. Some participants expressed concern that algorithms are trained using data that are not appropriately representative of the population of interest and for the

intended application. In some cases, the algorithms are then tested using that same data. Participants emphasized that such practices are not appropriate and could lead to misleading information about the validity, reliability, and fairness of the algorithms. Participants noted that data practices raise important questions about civil liberties that developers and users of algorithms cannot ignore. The lack of oversight that currently exists around the development, validation, and use of algorithms in sensitive domains such as criminal justice was highlighted, leading to calls by some participants for better regulation, standards, and best practices relating to data collection, data retention, and strategies to mitigate bias.

Additionally, the importance of transparency, testing, explainability, and understandability became recurring themes in this discussion. First, participants emphasized that transparency is a factor that spans across the entire pipeline of algorithm design, development, testing, implementation, and deployment. Transparency is fundamental for any algorithm that is intended for use in the criminal justice system given the concerns surrounding civil liberties and constitutional guarantees. Second, participants pointed to testing as a pillar to evaluating and demonstrating validity, reliability, and fairness. They recommended that algorithms be subject to multi-tiered testing schemes, including testing by second and third parties (end users and independent parties) to establish the conditions and scope of acceptable use and conditions for which the algorithms should not be used (i.e., conditions for which the outputs are not reliable). These testing schemes should be thorough and robust, specifically focused on identifying the conditions for which the algorithm fails. Finally, participants stressed that algorithms should be designed and developed in a way that promotes explainability and understandability. These principles are important to ensure that forensic practitioners can accurately convey the information produced by the algorithm and to ensure that factfinders can properly interpret the information. Understandability is furthermore important to ensure that forensic practitioners know when use of the algorithm is appropriate. For example, if a forensic practitioner is not able to explain how an algorithm works, then participants questioned whether they understand the circumstances for which the algorithm should or should not be applied and whether end-users of the algorithm output (e.g., investigators, litigators, factfinders) can make appropriate decisions based on that output.

Box 4.3.3: Priorities relating to mechanisms for assessing and assuring the validity, reliability, and fairness of algorithms for use in forensic science practice:

(a) Create standards relating to algorithmic data needs, including currently unregulated practices of broad-based collection and retention as well as bias and lack of representativeness

(b) Create standards relating to the issue of transparency in algorithm design, development, implementation, testing, use, and output presentation

(c) Perform multi-tiered testing of algorithms to establish the conditions for acceptable use, and ensure data used for validation is different from data used during development

(d) Create standards that ensure algorithms are explainable and understandable

4.4. Research, Development, Testing and Evaluation

Research, development, testing, and evaluation (RDTE) is the vehicle by which forensic science advances and is the foundation on which public trust and confidence in forensic science rests. RDTE activities:

- contribute to improvements in accuracy, reliability, and efficiency;
- promote innovation through interdisciplinary solutions, produce new capabilities, and enable adaptations to emerging challenges;
- help strengthen quality assurance practices;
- explore social, legal, and ethical issues relating to technologies and techniques; and
- provide the scientific rigor underpinning forensic science methods, practices, and disciplines.

To ensure forensic science maintains a rigorous scientific footing and can adapt to evolving demands, there must be clarity around the priorities relating to research and development (including methods for establishing those priorities); consideration of optimal models to strengthen researcher and practitioner partnerships; and determination of the research products that are most needed to promote greater validity, reliability, and consistency of practices. The overarching question and subtopics that this breakout group was asked to address were:

How should the forensic science community address the research challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term research priorities, and methods for establishing such priorities, to strengthen forensic science practice.*

b. *Identify partnering and collaboration models to strengthen researcher and practitioner partnerships (including data sharing) to advance critical research.*

c. *Identify critical classes of research products (standard reference materials or data, research reports, tools, and technologies) to improve the validity, reliability, and standardization of forensic science practice.*

4.4.1. Subtopic A: Near-term and long-term priority areas

The objective for this discussion was to identify priority areas for research to strengthen forensic science practice and methods for establishing such priorities. Discussions primarily focused on the need for a coordinated and unified strategic research plan and the importance of addressing critical research challenges relating to capacity, efficiency, validity, and reliability of current practices. Several priorities were identified by the breakout group (Box 4.4.1).

The need for a strategic research plan emerged as a key priority. Although several entities have identified research needs over the years, some participants felt that those needs are not well organized across disciplines, lack prioritization, and do not necessarily align to a broader strategy. Rather, research needs are often perceived as being identified ad hoc based on current operational challenges, capacity limitations, and present demands. The research strategy should identify the overarching goals to be achieved over the next 5 to 10 years and outline the steps that are necessary to achieve them. Participants suggested that the focus be on longer-term strategic goals instead of short-term research needs that can be accomplished in 1 to 2 years. Short-term research needs were described as the incremental steps that contribute to achieving the larger strategic goals. Finally, participants recommended that the research strategy represent needs of the entire forensic science community, encompassing multiple disciplines and perspectives.

Major topics that the strategic research plan should address were threefold. First, the research should focus on leveraging technologies to improve efficiencies and address challenges due to capacity and resource limitations. Second, the research should focus on improving the scientific rigor and evidence-base of current practices to ensure their validity and reliability. Third, the research should focus on identifying and addressing opportunities and challenges provided by emerging technologies. While emerging technologies have potential to help address current limitations (e.g., automation and augmentation technologies), they bring their own challenges that need to be addressed. Thus, forensic service providers and researchers cannot be narrowly focused on the development of the technologies without consideration and research relating to downstream implications caused by the use of those technologies.

Several different ideas were suggested about how research needs could be prioritized, including identifying common and cross-cutting themes and convening focus groups comprised of subject matter experts and other members of the forensic science community. Factors that could be taken into consideration included:

- Throughput demands
- Prevalence of errors
- Gaps in efficiencies
- Technical gaps
- Reliability
- Implementation costs
- Financial considerations (e.g., return on investment)
- Future impacts of research or technology
- Court/legal challenges
- Crime trends
- Capacity limitations

Ultimately, participants recognized that there are several viable approaches that could be taken and rubrics that could be used, with varied benefits and limitations. What is most important is that a strategic research plan be created that includes a prioritization and alignment with longer-term goals.

Box 4.4.1: Priorities relating to research, development, testing, and evaluation to strengthen forensic science practice:

(a) Develop a strategic research plan for forensic science that accounts for research needs across disciplines and includes input from members of a task group or focus group with diverse backgrounds and perspectives responsible for:

  i. Reviewing the list of research priorities that have been proposed to date and consider whether those priorities are still current

  ii. Providing recommendations on methods to prioritize the research needs

(b) Assess how new technology (e.g., AI/ML algorithms) can address staffing issues and efficiency challenges, including:

  i. Needs for additional or improved infrastructure to support the use of the technology

  ii. Needs to address future impacts and challenges associated with the new technology (e.g., staffing, testing, bias)

(c) Facilitate research to address gaps between practices referenced in standards and scientific foundation reviews.

4.4.2. Subtopic B: Researcher and practitioner partnerships

The objective for this discussion was to identify partnering and collaboration models to strengthen researcher and practitioner partnerships (including data sharing) to advance critical research. Discussions primarily focused on expanding outreach efforts and developing and promoting resources for data sharing. Several priorities were identified by the breakout group (Box 4.4.2).

Participants observed that when researchers and practitioners interact, it is often among a similar group of people. There need to be dedicated efforts to expand outreach efforts to reach all practitioners, including those who may not traditionally interact with researchers. Otherwise, there is risk that some research and development challenges will be overlooked. Likewise, participants noted that there is value in expanding outreach to other types of researchers as well to promote greater diversity and collaboration in research perspectives, such as through a convergence research model used by the National Science Foundation (NSF). The convergence research model promotes multi-disciplinary collaboration and provides a means for researchers to share their knowledge, theories, methods, data, and research toward a common purpose.

Discussions relating to the expansion of outreach efforts led to suggestions for the creation of a centralized database of practitioners and researchers who are willing to collaborate on research and development challenges of mutual interest. Examples that were highlighted include the Laboratories and Educators Alliance Program (LEAP) administered as a joint effort between the ASCLD FRC and the Council of Forensic Science Educators (COFSE) as well as the Collaboration Hub established by the ASCLD FRC. Other suggestions included opportunities for researchers to work alongside practitioners in an operational setting (e.g., within the forensic service provider facility) and for practitioners to work alongside researchers in a research setting (e.g., within the research laboratory). These engagements could be done for a specific project or for general purposes (i.e., residencies). To be most effective, participants recommended that residency-based models last at least 6 to 12 months to ensure the engagement is meaningful.

Another topic raised was the need for collaborative validation models. Not only do these models help bring together practitioners and researchers to solve real challenges relating to translation and implementation of research information and technologies, but they have several other benefits, including:

- Reducing duplication of effort by forensic service providers validating the same or similar methods or technologies
- Reducing the resource burden on any single forensic service provider
- Promoting greater scientific rigor in the validation studies
- Enabling sharing of data, protocols, and training among forensic service providers

Data sharing was an important issue raised. Research can be stifled when access to data is limited. This is particularly true for issues relating to forensic science and criminal justice, which often require access to sensitive information, such as controlled or private chemical, biologic, or biometric data. The development of forensic algorithms and tools leveraging AI/ML-based technologies were highlighted as some examples where access to large and operationally relevant datasets is critical. Data sharing limitations also impact collaborative quality assurance programs, such as blind proficiency testing. There has been an increased emphasis for more data sharing in recent years, particularly relating to research supported by federal funding (e.g., mandates from the White House Office of Science and Technology Policy [OSTP] that research supported by federal funding be made publicly available). However, participants noted several challenges to sharing forensic data due to privacy considerations, Institutional Review Board (IRB) requirements, and informed consent limitations. These challenges led to calls for resources to promote and support data sharing, such as:

- Centralized repositories and databases to facilitate data sharing
- Templated language that could be included in interagency agreements or other Memoranda of Understandings (MOUs) addressing practitioner-researcher or practitioner-practitioner collaborations

- Training and educational resources relating to data sharing

Box 4.4.2: Priorities relating to partnering and collaboration models to strengthen researcher and practitioner partnerships (including data sharing) to advance critical research:

(a) Increase outreach to populations of practitioners not currently engaged in research or interlaboratory collaboration efforts

(b) Implement convergence models that bring together different types of researchers from different disciplines to address research challenges and strengthen implementation and technology transfer

(c) Evaluate current practitioner-researcher collaboration models and curate a centralized database of practitioners and organizations willing to participate in research

(d) Establish a centralized, curated, reliable, sustained, and shareable database for algorithm development and testing (e.g., training datasets and validation datasets) across disciplines that are both human-readable and machine-readable

(e) Facilitate and promote resources to strengthen education and training for forensic practitioners relating to data sharing, including the development of example templates and language that could be included in MOUs for data sharing and collaboration

4.4.3. Subtopic C: Classes of research products

The objective for this discussion was to identify classes of research products to improve the validity, reliability, and standardization of forensic science practice. Discussions primarily focused on the development of different datasets and databases to support research and validation activities; tools and technologies to promote greater rigor and operational efficiencies; and educational and training resources to help translate complex scientific and statistical concepts into easily understandable terms. Several priorities were identified by the breakout group (Box 4.4.3).

The importance of data became a central theme of this discussion, with an emphasis on the need for both real and synthetic datasets to be collected and curated to help establish the application boundaries of various methods, such as:

- Datasets that approach the limit of detection (LOD) of current technologies

- Datasets containing error-prone samples (e.g., close non-matches for pattern evidence disciplines)
- Datasets containing known mixtures of substances (chemical, biologic) to assess vulnerabilities of methods and practitioners

Also discussed was the need for datasets to enable future testing of current or legacy tools and technologies. Digital evidence is one major discipline that commonly faces these challenges. An example provided was location services in cellular phones. In 2015, the technologies that provided location services were different than those used today, and testing was limited in 2015. Thus, without legacy data, some participants indicated that there is no way to assess validity and reliability retrospectively, which can have implications to current or future litigation. These challenges are not limited to digital evidence. Participants also pointed to the need to capture data where populations might change over time (e.g., firearms, footwear, tires, and manufactured items). These needs for continuous collection and storage of data led to calls for creating and maintaining databases with scalable architectures that could archive these datasets, maintain version controls, and ensure their authenticity. Suggestions also included expanding those databases to include software tools and algorithms.

In addition to data, participants emphasized the need to focus on the development of tools and technologies that can help strengthen scientific and technical rigor and improve operational efficiencies (e.g., provide stronger statistical foundations and promote automation and augmentation). To this end, participants recommended that any strategic research plan that is developed (e.g., as proposed in Subtopic A of this breakout group discussion) also include tangible outputs (i.e., tools and technologies) that can be leveraged by the forensic science community and translated and implemented into operational practice.

Finally, participants discussed the importance of developing and promoting educational and training resources for forensic practitioners, litigators, and others who might be responsible for analyses and decision-making based on forensic evidence. By helping translate complex scientific, technological, and statistical concepts into plain and clear language, such resources could help ensure that the capabilities and limitations of the results from forensic examinations are properly understood and that decisions are appropriately informed. Suggestions for educational and training resources included:

- Statistical tools that help practitioners better understand and perform data analyses (e.g., statistical sampling, probabilistic modeling, calibration, and uncertainty estimation)
- Short educational reports and primers aimed at practitioners and other members of the forensic science community (e.g., investigators, litigators, and judges)

Box 4.4.3: Priorities relating to classes of research products to improve the validity, reliability, and standardization of forensic science practice:

(a) Develop datasets that can establish method limitations and help distinguish between human and technology performance challenges (e.g., known mixtures of analytes, synthetic data sets, and complex matrices)

(b) Evaluate challenges that the forensic science community will face in the future relating to technologies and algorithms currently in use and be proactive in addressing those issues, including:

i. The creation of temporally relevant datasets relating to rapidly changing technologies or evolving populations (e.g., digital evidence tools and technologies, software applications, AI/ML-based algorithms, manufactured items)

ii. The development of databases with scalable architectures to archive datasets, software, and algorithms; employ version control techniques; and apply methods to ensure data authenticity.

(c) Develop and promote statistical tools that help practitioners better understand and perform data analyses

(d) Develop and promote educational and training resources for forensic practitioners, litigators, and other members of the forensic science community addressing:

i. Statistical methods and concepts relating to forensic practices, tools, and technologies

ii. Metrological principles and capabilities that could be leveraged to address systematic or idiosyncratic measurement challenges

4.5. Non-Technical Topics

Forensic service providers face a wide range of scientific and technological challenges relating to research and standards development and implementation. Oftentimes, discussions relating to the needs of the forensic science community focus narrowly on technical challenges. However, non-technical challenges are a common impediment to forensic science practice. To ensure the forensic science community can chart a path forward, there must be clarity around the non-technical challenges and priorities to strengthen forensic science practice; mechanisms to address those non-technical challenges and priorities; and methods to assess their

effectiveness and impact. The overarching question and subtopics that this breakout group was asked to address were:

How should the forensic science community address non-technical challenges to significantly strengthen forensic science practice in the United States today?

a. *Identify critical near-term and long-term non-technical to strengthen forensic science practice (e.g., information/data sharing, improving accreditation, quality management, proficiency testing, training, throughput efficiency, translation and implementation of research and standards).*

b. *Identify mechanisms to address critical non-technical priorities to strengthen forensic science practice.*

c. *Identify metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice.*

4.5.1. Subtopic A: Near-term and long-term priority areas

The objective for this discussion was to identify critical non-technical challenges and priority areas to strengthen forensic science practice. Discussions touched on issues encompassing a wide range of areas, including translation and implementation, interdisciplinary standards, LIMS, quality assurance, and training and staffing. Several priorities were identified by the breakout group (Box 4.5.1).

Significant resources are required to translate and implement information, methods, and technologies produced from research and development activities. This led to calls for providing more support for validation planning, validation studies (including verification and testing), and training. Model validation plans, Standard Reference Materials, best practice guides, and availability of expert advice were highlighted as examples of ways those challenges could be mitigated.

Participants raised the need for high-level interdisciplinary standards that address quality, reporting, testimony, proficiency testing, and packaging. ASTM E2917 (Standard Practice for Forensic Science Practitioner Training, Continuing Education, and Professional Development Programs) was identified as an example interdisciplinary standard. High-level standards would not only promote greater harmonization and consistency among disciplines, but also help reduce burdens for forensic service providers attempting to adopt and implement them in their standard operating procedures.

Challenges relating to LIMS and data sharing were other issues raised during the discussion. These challenges include enabling controlled access for customers to streamline

discovery processes and for the criminal justice community to review and analyze data for more actionable insights and trends. Accomplishing this requires coordination across LIMS vendors. Such coordination would aim to support the development of interoperability standards for data exchange of LIMS information, integration of LIMS modules through standardized APIs and tracking mechanisms, and enable greater connectivity among different groups, including forensic service providers, law enforcement, litigators, courts, public health, and other applicable government and non-government entities.

Quality assurance emerged as a major discussion topic, resulting in the recommendation to create a national database for quality incidents as well as blind proficiency testing results. The creation of a national database would permit greater insights into systemic vulnerabilities and challenges faced by forensic service providers across the country and bring greater awareness to the effectiveness of different methods and practices. The database would also promote standardized tracking of non-conformities, better assessments and understanding of errors, and more consistency in conducting risk assessments and severity ratings among forensic service providers. Participants recommended that such databases include anonymized information to promote self-disclosure and protect personally identifiable information.

Blind proficiency testing was highlighted as a priority for the community, but participants recognized that progress has been slow due to implementation challenges. Participants proposed that data and samples be created to support blind proficiency testing programs, which could then be packaged locally for submission to participating forensic service providers in accordance with their typical packaging and submission procedures. Additionally, the development of standards relating to packaging specifications was also highlighted. Standardizing such practices would not only help lower the barriers for forensic service providers to participate in systematically administered blind proficiency testing schemes and interlaboratory comparisons, but also provide a means for forensic service providers to enforce conformance to a common set of best practices.

Training and staffing issues were also discussed. On the topic of training, participants emphasized the importance of promoting better integration and two-way communication between legal and scientific communities and cross-disciplinary knowledge, along with the need to address issues relating to human factors, biases, and error mitigation. Additionally, continuing education curricula need to be expanded to include topics such as evidence management, testimony, report writing, and measurement uncertainty. On the topic of staffing, participants highlighted challenges relating to recruitment, retention, and training of personnel. The importance of initiatives to promote employee wellness was also underscored as many leaders within the forensic science community have observed increases in employee stress levels that can adversely affect quality, efficiency, and retention. Moving forward, participants recommended that the forensic science community embrace a stronger culture of quality assurance and transparency where errors are acknowledged and accepted so that occurrences can be discussed more openly and lead to more constructive and effective solutions.

Box 4.5.1: Priorities relating to non-technical challenges that need to be addressed to strengthen forensic science practice:

(a) Provide greater support for translation and implementation of new methods and technologies (e.g., guidance and model plans relating to validation, verification, and testing)

(b) Promote and develop interdisciplinary standards

(c) Promote and facilitate standardization and interoperability of LIMS

(d) Develop a centralized database for quality incident reports and tools to promote reporting of quality incidents

(e) Promote and support the development of blind proficiency testing programs within forensic service providers (e.g., guidance and test samples)

(f) Promote and support expanded curricula for training and continuing education programs, including the development of online modules for greater accessibility

4.5.2. Subtopic B: Mechanisms for addressing non-technical priorities

The objective for this discussion was to identify mechanisms to address critical non-technical challenges and priorities to strengthen forensic science practice. Discussions primarily focused on mechanisms relating to translation and implementation, interdisciplinary standards, LIMS, quality assurance, and training and staffing. Several priorities were identified by the breakout group (Boxes 4.5.2a, 4.5.2b, and 4.5.2c).

Suggested mechanisms for addressing translation and implementation challenges relating to new methods and practices included:

- Development of standards to provide direction on how to properly construct and conduct validation studies
- Development of model specification documents, validation plans, and reference materials
- Coordination of collaborative testing models to promote information sharing among forensic service providers
- Comparative evaluations of emerging technology options to promote data-driven procurement decisions
- More interlaboratory studies and support for developing blind proficiency testing programs to monitor the effectiveness of new methods and technologies

- Establishment of a risk management framework prior to adoption of new methods and technologies, which includes consideration of sociotechnical impacts

Also discussed was the need for greater top-down direction to standards development committees. For example, pointing specifically to the OSAC, suggestions were made for the FSSB to define a core set of standards that committees are directed to prioritize, including those that should be developed in an interdisciplinary capacity. Several standards have been produced to date that have a discipline-specific focus but include interdisciplinary content. Participants recommended that such documents be evaluated and consolidated where possible.

Mechanisms for addressing challenges relating to LIMS focused on the development of interoperability guides and toolkits, as well as standards to ensure compatibility, data sharing, and connectivity. Standards relating to interoperability of Automated Fingerprint Identification Systems (AFIS) and the *Roadmap for Using LIMS* published by the NIST Material Measurement Laboratory[18] were highlighted as examples. Participants stressed that collaboration with commercial vendors and the broader criminal justice community will be essential, and efforts should be directed toward facilitating engagements that convene the different groups.

For quality incidents, suggestions were made to establish a centralized national database for anonymous reporting, data analysis, and trendspotting. Such data would inform the development of standards, accreditation requirements, and annual reports identifying challenges and emerging trends relating to quality assurance. As part of the creation of a centralized database, participants emphasized the need to standardize error categories, develop error taxonomies, and establish risk and severity ratings. Also highlighted was the importance of conducting error symposia to foster transparency and collaboration among forensic service providers, quality assurance experts, and accreditation bodies.

With respect to blind proficiency testing, participants noted the importance of convening participant groups, creating representative sample sets, and expanding training initiatives (including online modules) to inform the design and execution of blind proficiency testing programs. Participants recommended that guidance materials and training initiatives be developed to inform forensic service providers and others within the criminal justice community on the importance of blind proficiency testing programs, provide recommendations for how to conduct blind proficiency testing schemes (including "lessons learned" from past experiences), and summarize outcomes and impacts relating to the implementation of blind proficiency testing programs. Participants also suggested that annual reports be produced to consolidate and summarize key challenges and emerging trends relating to and resulting from blind proficiency testing programs to provide a macro view of issues that could inform future strategic priorities focused on improving quality assurance.

[18] "A Roadmap for LIMS at NIST Material Measurement Laboratory" [Online]. Available: https://www.nist.gov/publications/roadmap-lims-nist-material-measurement-laboratory.

Means of addressing training and staffing challenges included calls for greater coordination with other entities (e.g., ASCLD and the Forensic Technology Center of Excellence) and increased emphasis on leveraging online modules and platforms to improve accessibility and engagement opportunities. Participants recommended training efforts be geared toward forensic service provider leaders, discipline-specific technical leaders, practitioners, and officers of the court (e.g., litigators and judges). Participants also highlighted a need to establish core competencies for different disciplines and roles. Specific training topics identified as priorities include measurement traceability, measurement uncertainty, reporting, testimony, and employee wellness.

Finally, participants stressed the need to continually measure the impact of the suggested changes.

Box 4.5.2a: Priorities relating to mechanisms to address non-technical challenges to strengthen forensic science practice:

(a) Translation and implementation:

i. Conduct comparative studies of available technology options

ii. Develop model specifications documents, validation plans, and reference materials (i.e., validation toolkits) supporting validation and implementation of new methods and technologies

iii. Conduct interlaboratory studies

iv. Develop a model Risk Management Framework that includes sociotechnical implications prior to the adoption of new methods and technologies

v. Provide support for collaborative validation initiatives, and development of information sharing clearinghouses

(b) Interdisciplinary Standards:

i. Define a core set of standards that should be prioritized

ii. Define a core set of standards that should be developed in an interdisciplinary capacity

iii. Evaluate and consolidate extant standards that are currently discipline-specific but contain information that is interdisciplinary

Continued in Box 4.5.2b . . .

Box 4.5.2b: Priorities relating to mechanisms to address non-technical challenges to strengthen forensic science practice *(continued from Box 4.5.2a)***:**

(c) LIMS:

i. Develop standards, specification documents, and toolkits relating to LIMS interoperability

ii. Convene forensic service providers, commercial LIMS vendors, and other entities to promote and facilitate pathways toward better data sharing and connectivity

iii. Develop a roadmap for using LIMS in forensic science

(d) Quality Assurance (Quality Incident Reporting)

i. Create a centralized national database for anonymous reporting, data analysis, and trendspotting of quality incidents

ii. Standardize error categories and error taxonomy, and establish risk and severity ratings

iii. Convene members of the forensic science community and host national symposia focused on error and error management

(e) Quality Assurance (Blind Proficiency Testing)

i. Facilitate the creation and coordination of participant groups comprised of forensic service providers, law enforcement, and other interested parties to inform sample types for blind proficiency testing programs

ii. Create representative sample sets to support blind proficiency testing programs

iii. Promote and develop training modules relating to the design, development, and implementation of blind proficiency testing programs

iv. Develop mechanisms to facilitate collection, analysis, and consolidation of information relating to the outcomes and impacts of blind proficiency testing programs through annual reports

Continued in Box 4.5.2c . . .

Box 4.5.2c: Priorities relating to mechanisms to address non-technical challenges to strengthen forensic science practice *(continued from Box 4.5.2b)***:**

(f) Training and Staffing:

i. Improve coordination with various entities providing training relating to forensic science to promote greater consistency, collaboration and synergy

ii. Increase training and continuing education opportunities using online modules and platforms for forensic service provider leaders, discipline-specific technical leaders, practitioners, and officers of the court (e.g., litigators and judges)

iii. Establish core competencies for personnel in different disciplines and roles

iv. Expand the curricula for training and continuing education to include traceability, measurement uncertainty, reporting, testimony, and employee wellness

(g) Impact:

i. Evaluate the extent to which the enactment of different mechanisms for addressing non-technical priorities affected policy and practice

ii. Conduct economic impact studies to measure the return on investment relating to the enactment of different mechanisms for addressing non-technical priorities

4.5.3. Subtopic C: Metrics for assessing effectiveness and impact

The objective for this discussion was to identify metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice. Discussions primarily focused on metrics relating to translation and implementation, interdisciplinary standards, LIMS, quality assurance, and training and staffing. Several priorities were identified by the breakout group (Boxes 4.5.3a, 4.5.3b, and 4.5.3c).

Participants noted that several metrics could be used to measure the effectiveness and impact of translation and implementation of research and standards. The most direct measure of impact is the number of forensic service providers that are using validation toolkits (e.g., model specification documents, validation plans, and reference materials supporting validation

and implementation of new methods and technologies). Other indirect metrics proposed included:

- The number of instances that published information and materials relating to translation and implementation have been accessed or downloaded (e.g., web traffic statistics and download counts)
- The number of reference materials and other test samples that have been prepared and distributed to support validation and implementation of new methods and technologies
- Comparative analyses of the time required to implement new methods and technologies before and after having access to the validation toolkits and related resources

On the topic of interdisciplinary standards, discussions were focused on encouraging greater alignment of priorities, coordination among disciplines, and reducing the focus on the number of standards developed. Metrics proposed included:

- The number of comments received when documents are available for public comment
- The number of core topics defined for interdisciplinary standards
- The number of disciplines that have adopted interdisciplinary standards versus those that have not

Discussions relating to LIMS included consideration of the extent to which commercial vendors are willing to collaborate to achieve interoperability through standardization, as well as the extent to which members within the forensic science community leverage materials relating to LIMS interoperability (e.g., standards, specification documents, and toolkits) in their procurement processes. Suggested metrics included:

- The number of commercial LIMS vendors that engage in collaborative activities to achieve interoperability (e.g., participate in development of standards relating to interoperability and design and develop their systems to be interoperable)
- The number of commercial LIMS vendors that include interoperability in their advertising materials, quotations, and proposals
- The number of forensic service providers that reference standards and technical specifications relating to interoperability in their requests for proposals and other processes to secure funding
- The number of forensic service providers that receive funding geared toward the procurement of interoperable LIMS

Next, participants discussed the importance of evaluating the extent to which the forensic science community has embraced a culture of transparency by engaging in open conversations about quality incidents and error and by becoming more consistent in their tracking and categorization of quality incidents and error. Such a culture was considered critical to enable future research and standards aimed at identifying sources of error and eliminating

reoccurrence. The most direct measures of impact proposed by participants include the number of organizations that voluntarily self-report quality incidents and errors as well as the number of quality incidents reported overall. Other suggested metrics for measuring impact included:

- The number of organizations and individuals that attended symposia focused on error and error management distinguished by type of organization or attendee (e.g., forensic service provider, prosecutor, defense attorney, researcher)
- The number of organizations that work on developing or adopting standardized error categories, error taxonomies, and risk and severity ratings

Impact measures relating to blind proficiency testing were similar to those proposed for translation and implementation. Importantly, blind proficiency testing provides a framework for detecting systemic vulnerabilities and identifying areas where research and standardization are most important. The number of forensic service providers actively participating in blind proficiency testing programs was identified as the most direct measure of impact, particularly when also considering the rate at which participation grows over time. Other suggested indirect metrics included:

- The number and types of samples distributed for blind proficiency testing schemes
- The number of instances that published information and materials relating to blind proficiency testing have been accessed or downloaded (e.g., web traffic statistics and download counts)
- The number of participants who have attended training relating to the design, development, and implementation of blind proficiency testing programs
- The number and type of quality incidents identified because of blind proficiency testing
- Measures relating to changes in the perceived rigor and robustness of quality assurance programs that include blind proficiency testing schemes

The discussion on metrics relating to training and staffing focused on actionable insights that could be gleaned from those metrics that might inform future priorities and ways to be more effective through research and standards. For example, understanding the number and types of training programs offered and analyzing them based on participant demographics could help tailor educational efforts to improve competencies. Suggested metrics included:

- Types of training opportunities available and demographic information of participants
- The types and number of individuals attending training opportunities in various disciplines
- The types and number of training opportunities that have been developed in various disciplines

Box 4.5.3a: Priorities relating to metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice:

(a) Translation and implementation:

i. Measure the number of forensic service providers using toolkits designed to support validation and implementation of new methods and technologies

ii. Measure the number of instances where published information and materials relating to translation and implementation have been accessed or downloaded (e.g., web traffic statistics and download counts)

iii. Measure the number of reference materials and other test samples that have been prepared and distributed to support validation and implementation of new methods and technologies

iv. Measure the time required to implement new methods and technologies before and after having access to toolkits and other related resources

(b) Interdisciplinary Standards:

i. Measure the number of comments received when interdisciplinary standards are available for public comment

ii. Measure the number of core topics defined for interdisciplinary standards

iii. Measure the number of disciplines that have adopted core interdisciplinary standards versus those that have not

Continued in Box 4.5.3b . . .

Box 4.5.3b: Priorities relating to metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice *(continued from Box 4.5.3a)***:**

(c) LIMS:

i. Measure the number of commercial LIMS vendors that engage in collaborative activities to achieve interoperability (e.g., participate in development of standards relating to interoperability and design and develop their systems to be interoperable)

ii. Measure the number of commercial LIMS vendors that include interoperability in their advertising materials, quotations, and proposals

iii. Measure the number of forensic service providers that include interoperability in their requests for proposals and other processes to secure funding

iv. Measure the number of forensic service providers receiving funding or awards to support the procurement of interoperable LIMS

(d) Quality Assurance (Quality Incident Reporting)

i. Measure the number of forensic service providers that participate in voluntary self-reporting of quality incidents

ii. Measure the number of quality incidents reported by forensic service providers

iii. Measure attendance at symposiums on error and error management

iv. Measure the number of forensic service providers and other organizations working on developing or adopting standardized error categories, error taxonomies, and risk and severity ratings

Continued in Box 4.5.3c . . .

Box 4.5.3c: Priorities relating to metrics to assess the effectiveness and impact of research and standards on strengthening forensic science practice *(continued from Box 4.5.3b)***:**

(e) Quality Assurance (Blind Proficiency Testing)

i. Measure the number of forensic service providers actively participating in blind proficiency testing programs

ii. Measure the number and types of samples distributed for blind proficiency testing schemes

iii. Measure the number of instances that published information and materials relating to blind proficiency testing have been accessed or downloaded (e.g., web traffic statistics and download counts)

iv. Measure the number of participants that have attended training relating to the design, development, and implementation of blind proficiency testing programs

(f) Training and Staffing:

i. Measure the types of training offered based on demographics (e.g., participant background and employment)

ii. Measure the types and number of individuals attending training opportunities in various disciplines

iii. Measure the types and number of training opportunities that have been developed in various disciplines

5. Key Takeaways

The following key takeaways represent common challenges and persistent themes expressed among participants during presentations and discussions.

1. Critical advancements in forensic science are often stifled by resource and capacity limitations faced by forensic service providers. Stronger partnerships are needed to help lower barriers to the translation and implementation of new methods, technologies, and practices and to ensure that outputs from research and standards programs are impactful to strengthening forensic science practice.

2. Consistency and standardization of forensic science practices are priorities shared across the forensic science community. However, progress toward achieving these goals has been slow and challenging. Standards play an important role, but the number of requirements and recommendations proposed have been overwhelming for many forensic service providers to implement. Furthermore, shortfalls in the quality, clarity, and rigor of those requirements and recommendations allow for flexibility in application and conformance. Achieving the goals of consistency and standardization requires consolidation of requirements and recommendations into interdisciplinary standards and improvements to the quality, clarity, and rigor of the documents.

3. The validity and reliability of forensic science methods and practices are often disputed. A greater emphasis on blind testing, black-box testing, and interlaboratory studies is needed to strengthen performance monitoring programs for existing practices as well as to identify priorities for research and standards programs to address critical challenges or limitations.

4. Interdisciplinary collaboration and standardization are key priorities for improving forensic science practice. These efforts can enable harmonization across disciplines, lower barriers to adoption and implementation of standards, provide a rubric for evaluating acceptable practices, and promote consistency in analytical methods and results among forensic service providers.

5. Transparency and accountability are essential for ensuring validity and reliability of forensic science methods. Public access to standard operating procedures, validation methods and data, and quality incidents promote public trust and confidence in forensic methods, accelerate research to advance forensic practices, and enable greater coordination, collaboration, and resource sharing among forensic service providers to alleviate translation and implementation challenges.

6. Ongoing research, development, testing, and evaluation are critical for advancing forensic science practices. These efforts provide the means for monitoring performance, identifying systemic challenges, and addressing evolving needs to improve the validity, reliability, and consistency of forensic science methods and practices.

Priorities include stronger researcher-practitioner partnerships, creation of centralized data and databases, and use of computational technologies and methods.

Appendix A. Agenda

WEDNESDAY, SEPTEMBER 6, 2023		
7:30am – 8:30am	**Registration**	
8:30am – 8:45am	**Welcome and Introduction**	S. Shyam Sunder
8:45am – 9:00am	**Opening Remarks** ➢ Roundtable Overview ➢ NIST FSP Overview ➢ Goals for this Roundtable	Henry Swofford
9:00am – 10:15am	**Plenary I** Practitioner Perspectives	David Kanaris Jeremy Triplett Mike Garvey Barry Logan ➢ Discussion
10:15am – 10:30am	**Break**	
10:30am – 11:45am	**Plenary II** Leadership Perspectives	Linda Jackson Brady Mills Matthew Gamette Jason Bundy ➢ Discussion
11:45am – 1:00pm	**Lunch** (on your own)	
1:00pm – 2:15pm	**Plenary III** Legal Perspectives	Raymond Valerio Jennifer Friedman Sarah Chu Kent Cattani & Ronald Reinstein ➢ Discussion
2:15pm – 2:30pm	**Break**	
2:30pm – 3:30pm	**Plenary IV** Researcher Perspectives	Austin Hicklin Keith Morris Simson Garfinkel ➢ Discussion
3:30pm – 3:45pm	**Break**	
3:45pm – 4:45pm	**Plenary V** Quality Management System Perspectives	Lynn Garcia Peter Stout Eva King ➢ Discussion
4:45pm – 5:00pm	**Day 1 Wrap-up**	Henry Swofford

THURSDAY, SEPTEMBER 7, 2023		
8:30am – 8:45am	**Opening Remarks** ➢ Day 1 Recap ➢ Breakout Guidance	Henry Swofford
8:45am – 11:45am	**Breakout Discussions** I. Standards and Practices II. Validity and Reliability III. Forensic Algorithms IV. RDTE V. Non-Technical	
11:45am – 1:00pm	**Lunch** (on your own)	
1:00pm – 2:30pm	**Plenary VI**	Breakout I Rep. ➢ Discussion Breakout II Rep. ➢ Discussion Breakout III Rep. ➢ Discussion
2:30pm – 2:45pm	**Break**	
2:45pm – 3:45pm	**Plenary VII**	Breakout IV Rep. ➢ Discussion Breakout V Rep. ➢ Discussion
3:45pm – 4:00pm	**Closing Remarks** ➢ Roundtable Review ➢ Next Steps	Henry Swofford
4:00pm	**Meeting Ends**	

Appendix B. Participant List

1. **Mr. Brian Bataille**
   Director, National Media Exploitation Center, U.S. Department of Defense

2. **Ms. Tebah Browne**
   Forensic Science Policy Specialist, Innocence Project

3. **Mr. Jason Bundy**
   Director, Florida Department of Law Enforcement Forensic Services

4. **Dr. JoAnn Buscaglia**
   Research Chemist, FBI Laboratory

5. **Dr. Eoghan Casey**
   Vice President, Cybersecurity Strategy and Product Development, OwnBackup

6. **Hon. Kent Cattani**
   Chief Judge, Arizona Court of Appeals

7. **Dr. Elizabeth Chin**
   Assistant Professor, Department of Biostatistics, Johns Hopkins University

8. **Dr. Sarah Chu**
   Director, Policy and Reform, Perlmutter Center for Legal Justice, Cardozo Law

9. **Ms. Elizabeth Daniel Vasquez**
   Director, Science and Surveillance Project, Brooklyn Defender Services

10. **Mr. James Darnell**
    Chief Operations Officer, VTO Labs

11. **Mr. William Demuth, II**
    Director, Training and Applications Laboratory, Illinois State Police Forensic Sciences Command

12. **Mr. Brian Dew**
    Forensic Laboratory Manager, New Hanover County (NC) Sheriff's Office

13. **Dr. Rebecca Ferrell**
    Program Director, Biological Anthropology Program, National Science Foundation

14. **Ms. Jennifer Friedman**

Deputy Federal Public Defender for the Central District of California

15. **Mr. Matthew Gamette**
    Director, Idaho State Police Forensic Services Laboratory System

16. **Ms. Lynn Garcia**
    General Counsel, Texas Forensic Science Commission

17. **Dr. Simson Garfinkel**
    Chief Scientist, BasisTech, LLC

18. **Dr. Michael Garvey**
    Director, Philadelphia (PA) Police Department Forensic Science Laboratory

19. **Mr. John Grassel**
    Senior Manager, Investigative Sciences Program, Center for Forensic Science Advancement and Applications, RTI International

20. **Dr. Austin Hicklin**
    Director, Forensic Science Group, NOBLIS

21. **Ms. Linda Jackson**
    Director, Virginia Department of Forensic Science

22. **Mr. David Kanaris**
    Chief, Alaska Scientific Crime Detection Laboratory

23. **Ms. Eva King**
    Quality Assurance Director, Wisconsin State Crime Laboratories

24. **Mr. Christopher Krug**
    Vice Chair, Forensic Science Standards Board, OSAC

25. **Mr. Timothy Kupferschmid**
    Chief of Laboratories, New York City Office of Chief Medical Examiner

26. **Ms. Beth Lavach**
    Director of Government Relations, Consortium of Forensic Science Organizations

27. **Ms. Jennifer Leach**
    Chief, Media-Data Exploitation Branch, National Media Exploitation Center, U.S. Department of Defense

28. **Dr. Barry Logan**
    Chief Scientist and Senior Vice President for Forensic Sciences, NMS Labs

29. **Dr. Jennifer Love**
    Physical Scientist, Office of Investigative and Forensic Sciences, National Institute of Justice

30. **Mr. Brendan Max**
    Chief, Forensic Science Division, Cook County (IL) Public Defender Office

31. **Dr. Jonathan McGrath**
    Principal Technical Advisor, U.S. Customs and Border Protection (CBP) Laboratories and Scientific Services (LSS) Southwest Regional Science Center

32. **Ms. Danielle McLeod-Henning**
    Program Manager/Physical Scientist, Office of Investigative and Forensic Sciences, National Institute of Justice

33. **Mr. Brady Mills**
    Chief, Texas Department of Public Safety Crime Laboratory Division

34. **Dr. Keith Morris**
    Professor of Forensic and Investigative Science, West Virginia University

35. **Mr. Jeff Nye**
    Director, Forensic Science Division, Michigan State Police

36. **Dr. Christopher Palenik**
    Senior Research Microscopist and Vice President, Microtrace LLC

37. **Ms. Elizabeth Pascual**
    Quality Assurance Manager, Office of Forensic Sciences, Drug Enforcement Administration

38. **Mr. Eric Pokorak**
    Assistant Director, FBI Laboratory Division

39. **Ms. Mindi Ramage**
    Forensic Science Program Manager, U.S. Department of Homeland Security (DHS), Customs and Border Protection (CBP), Laboratories and Scientific Services (LSS)

40. **Hon. Ron Reinstein**
    Judicial consultant, Arizona Supreme Court

41. **Dr. Jeri Ropero-Miller**
    Principal Scientist, Justice Practice Area, RTI International

42. **Dr. Jeff Salyards**
    Forensic Research Scientist IV CSAFE, Iowa State University; Principal Analyst, Compass Scientific Consulting

43. **Dr. Peter Stout**
    President/CEO, Houston (TX) Forensic Science Center

44. **Ms. Roxanne Taylor**
    Section Chief, Laboratory Management and Operations, Office of Forensic Sciences, Drug Enforcement Administration

45. **Mr. Jeremy Triplett**
    Director, Kentucky State Police Central Crime Laboratory

46. **Mr. Raymond Valerio**
    Assistant District Attorney and Director of Forensic Sciences, Queens County (NY) District Attorney's Office

47. **Ms. Amy Watroba**
    Assistant State's Attorney, DuPage County (IL) State's Attorney's Office

48. **Mr. C. Ken Williams**
    Admin Manager, New Jersey State Police Office of Forensic Services

49. **Dr. Matthew Wood**
    Laboratory Director, Ocean County (NJ) Sheriff's Office Forensic Science Laboratory

50. **Mr. Tate Yeatman**
    Director, Palm Beach County (FL) Sheriff's Office Crime Laboratory

51. **Ms. Erika Ziemak**
    Director, Special Projects, University of North Texas Health Science Center, Center for Human Identification