The Roles of Science and Statistics in Advancing Forensic Science and Standards

Karen Kafadar
Chair and Commonwealth Professor
Dept of Statistics, University of Virginia
President, American Statistical Association
House Committee on Science, Space, and Technology
September 10, 2019

Abstract

The National Academy of Sciences' report, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009), emphasized the need for more collaborative research between forensic scientists and the physical, chemical, biological, and statistical scientists. The costs of *not* doing so are high (false convictions and false acquittals). The scientific method recognizes the transfer of research across multiple domains and the importance of continuing advancement in knowledge with increasing research. Statisticians have advanced knowledge and practice in nearly all aspects of science for centuries, using their expertise to ensure safe medical treatments, quantify uncertainties in scientific measurements, design experiments to validate the effects of interventions and validate claims of error rates. In this testimony I describe: (1) the role of the sciences, including statistics, in advancing forensic science and standards; (2) progress in forensic science since the 2009 NAS report; (3) challenges that remain; (4) recommendations on roles of Federal Agencies to ensure reliable forensic science that is needed to minimize false convictions and false acquittals, namely: centralized leadership, funding for research, and guidance to the Courts.

Key Words: scientific collaborations, analysis of forensic evidence, experimental design, statistical methodology, reliability, validation, error rates

Introduction

Madame Chair and Members of the Committee, thank you for inviting me to appear today. I am Karen Kafadar, Chair and Commonwealth Professor of Statistics at the University of Virginia. I also am 2019 President of the American Statistical Association (ASA), founded in 1839 and is the world's largest professional society of statisticians with 18,000 members from 93 countries.

I am appearing in my capacity as ASA President and as a statistical scientist who has conducted research and taught statistics at the university level to diverse audiences for 26 years. My collaborations have involved scientists in biology, chemistry, physics, medicine, engineering, psychology, anthropology, and sociology. At the Committee's request, I address four topics:

1. the role of scientists and statisticians in advancing forensic science and standards;

2. progress in forensic science since the 2009 report from the National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward*, of which I was a co-author;

3. challenges that remain;

4. recommendations on roles of Federal Agencies in ensuring continued research needed for the fair administration of justice, namely: centralized leadership, funding for research, and guidance to the Courts.

A brief story: Involvement of Statisticians in Forensic Science

In my written testimony here, I will start by explaining my involvement with forensic science. In 2003 I was invited to serve on the National Academy of Sciences' Committee on Scientific Assessment of Bullet Lead Elemental Composition Comparison, which led to the 2004 report, *Weighing the Evidence: Forensic Analysis of Bullet Lead* (National Academies Press, 2004). The FBI asked the NAS to review its procedures for comparing bullets found at a crime scene with bullets found in the possession of a suspect, based on the concentrations of seven trace elements measured in the lead of the bullets from the two locations. Two of the four aspects of the Committee's charge clearly involved statistical inferences from the measurements, so I was one of two statisticians[1] on the Committee. The shortcomings in the statistical inference procedures were obvious; statisticians advised on statistical analysis of data as well as on data collection and experimental designs, validation, risks of false claims of association and exclusion, and other concepts of statistical thinking. The Committee's report did not recommend the discontinuation of the FBI's Bullet Lead Comparison operation, but it did emphasize that lab reports and courtroom testimony would have to acknowledge that bullets found at a crime scene could not be guaranteed to have come from a specific box (e.g., one found in the possession of a suspect), but rather may have come from one of perhaps thousands of boxes. The FBI recognized that such testimony would greatly limit, or even eliminate, its probative value, and hence discontinued the operation. What surprised me, as a statistician, was the total absence of statisticians in the development inference procedures for data from bullet lead analyses and the value that statisticians could have added. No one would dream of asking a statistician to develop analytical chemical procedures to measure trace element concentrations: if you gave a statistician some manuals, she may have been able to come up with some kind of procedure, but no one would dream of asking a statistician to perform chemistry. But, for some reason, the FBI felt perfectly comfortable asking chemists to develop statistical procedures for drawing inferences from the data they collected on trace element concentrations — and then allowing them to testify in courtrooms on the results of those (improper) statistical procedures. Statistical input could have reduced substantially the number of false accusations based on faulty statistical procedures.

[1]Professor Clifford Spiegelman from Texas A&M University was the other statistician on the Committee.

1 ASA's role in forensic science

The ASA's mission is to promote the profession and practice of statistics to advance science, inform policy, and enhance statistical literacy, to make the world a better place. It is hard to find an area where statistical thinking and methodology does *not* contribute. ASA members participate on committees and sections that interface with many disciplines, including public health, environment, energy, education, law and justice, and forensic science. The ASA Board of Directors issued a statement that endorsed the 2009 NAS report's recommendations for additional research in forensic science, including statistics, and a statement with guidelines for expressing statistical conclusions in forensic science reports and in courtroom testimony.

2 Progress since the 2009 NAS report

The 2009 report (and the Jon Oliver piece on Forensic Science) called the public's attention to the lack of validity and reliability in the analysis and conclusions from forence evidence. Despite well-intentioned employees throughout the forensic science system, the field was relatively isolated from non-forensic scientists in the development of procedures for analyzing, and drawing inferences from, forensic evidence.[2] The acceptance by forensic scientists of their non-forensic counterparts is certainly not complete. But the fact that it has started, notably among the younger generation of forensic scientists, in just 10 years, emphasizes the impact of the 2009 report and the progress it engendered. Three specific projects where funding made possible research collaborations among scientists in different fields that led to progress in the field are: (1) creating a measure of strength in the comparison of two fingerprint images,[3] Use

[2]An exception was DNA evidence, which engaged early with research biologists and geneticists. This collaboration was nicely described in the testimony from Dr. Eric Lander, a world-class molecular biologist, to the Senate Committee on Commerce, Science, and Transportation, 28 March 2012; see `govinfo.gov/content/pkg/CHRG-112shrg77701/html/CHRG-112shrg77701.htm`

[3]H.J. Swofford, A.J. Koertner, F. Zemp, M. Ausdemore, A. Liu, M.J. Salyards, A method for the statistical interpretation of friction ridge skin impression evidence: Method development and validation, *Forensic Science International*, Volume 287, 2018, Pages 113-126, ISSN 0379-0738, `doi.org/10.1016/j.forsciint.2018.03.043`.

of this metric, called *FRstat*, has been limited; (2) a statistical algorithm for matching bullet land impressions,[4] (3) a quantitative approach to blood spatter evidence using fluid dynamics.[5]

This enhanced collaboration, virtually absent ten years ago, was facilitated by several activities; I mention four of them here. First, none of the progress to date would have been possible without the attention to this issue brought by Chair Johnson and other members of Congress. Second, the National Institute of Science and Technology (NIST, where I worked for three years after my doctorate) collaborated with the Department of Justice (DoJ), in forming a National Commission on Forensic Science (NCFS), to "establish and operate a competitively selected Center of Excellence focusing on measurement sciences, technology, and standards in forensic science," and to create a forensic science standards organization. DoJ led the NCFS effort, and the cooperation among forensic scientists, crime lab directors, research scientists, attorneys, and judges resulted in several NCFS statements. That communication barriers were lowered and trust was enhanced among these parties, in just two years, shows the mutual respect and cooperation that can be achieved to advance forensic science.[6] Regrattably, DoJ disbanded this Commission after only 13 meetings.

A third activity was NIST's Organization of Scientific Area Committees (OSAC) to confirm forensic standards.[7] OSAC includes 550+ forensic specialists (about 2/3) and other scientists, judges, and lawyers, on its 25 subcommittees representing forensic disciplines in a triage structure where the final approval of standards occurs at the third level, the Forensic Science Standards Board. Unfortunately, the composition

[4]E. Hare, H. Hofmann, A. Carriquiry (2017), Automatic matching of bullet land impressions, *The Annals of Applied Statistics* 11(4):2332-2358, `projecteuclid.org/euclid.aoas/1514430288`.

[5]P.M. Comiskey, A.L. Yarin, D. Attinger (2019), Hydrodynamics of forward blood spattering caused by a bullet of general shape, *AIP Physics of Fluids* 31, 084103, `doi.org/10.1063/1.5111835`; P.M. Comiskey, A.L. Yarin (2019), Self-similar turbulent vortex rings: interaction of propellant gases with blood backspatter and the transport of gunshot residue, *J. Fluid Mechanics* 876: 859-880, `doi.org/10.1017/jfm.2019.564`. See also `arstechnica.com/science/2019/08/physicists-now-have-even-better-models-for-blood-spatter-from-gunshot-wounds/`

[6]See NCFS products at `www.justice.gov/archives/ncfs`

[7]`https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science`

of the OSAC units remains primarily forensic scientists and practitioners, and the non-regulatory status of NIST requires OSAC to simply approve or reject *existing* standards without allowance for modifications.

A fourth activity is the Congressional allocation of funds to NIST to award a competitively-selected Center of Excellence, housed at Iowa State University and involving Carnegie Mellon, Duke, University of California-Irvine, and University of Virginia.[8] This Center for Statistical Applications in Forensic Evidence (CSAFE) was established in June 2015 and conducts cooperative research among NIST scientists, forensic practitioners in crime labs, and statisticians. Unfortunately, its mandate is limited to research in only *pattern and digital evidence* (which includes fingerprints, shoe and tire track impressions, handwriting, blood spatter patterns, digital signatures) – important areas to be sure, but the mandate does not allow funds to be used for important problems that exist outside those two forensic disciplines. (See Appendix for some of its projects that have yielded important advances.) I currently participate in this Center of Excellence, and formerly was a member of the OSAC's Forensic Science Standards Board and an NCFS subcommittee.

3 Challenges

Shortly after the 2009 NAS report was released, the Honorable Judge Harry T. Edwards, who co-chaired the report's authoring committee, testified before the Senate Judiciary Committee. In that testimony, he said,

> "*There are scores of talented and dedicated people in the forensic science community, and the work that they perform is very important. They are often strapped in their work, however, because of (1) a paucity of strong scientific research, (2) a lack of adequate resources and national support and (3) the absence of unified and meaningful regulation of crime laboratories and practitioners. It is clear that change and advancements, both systemic and scientific, are needed in a number of forensic science*

[8]Center for Statistical Applications in Forensic Evidence; see `forensicstats.org`.

disciplines to ensure the reliability of the disciplines, establish enforceable standards, and promote best practices and their consistent application."

The community has started work in all three areas that Judge Edwards mentions. But challenges remain:

1. Lack of centralized leadership by an agency that has the expertise and commitment to cause forensic disciplines to develop and enforce standards that are grounded in valid and reliable scientific studies and accurate statistical analyses;

2. Lack of studies confirming the validity and reliability of forensic disciplines by domain-specific scientists (versus forensic practitioners) not necessarily tied to a law enforcement agency;

3. Lack of funding to support an independent agency and the studies necessary to confirm the validity and reliability of forensic disciplines;

4. Lack of guidance for the courts on how to handle forensic evidence that is relevant but at best inconclusive. How should a judge instruct a jury in such a case? Until the courts have authoritative guidance, they are likely to continue admitting the evidence because it cannot be declared irrelevant. This would not be a problem if the courts knew how to effectively cabin the testimony of forensic experts: a jury may be inclined to give too much weight to inconclusive evidence merely because an "expert" is testifying to its efficacy.

Non-forensic scientists in chemistry, physics, imaging, statistics, and computer science, can be invaluable in the perspectives that they bring to a given problem (e.g., chemistry and drug toxicology; computer scientists and images of pattern evidence).

4 Recommendations

With my Center of Excellence colleagues and some of my co-authors of the NAS report, I offer the following recommendations. I do not think any of them will be new to NIST.

1. *Charge:* Some agency must take the lead. The 2009 report emphasized that the agency cannot be law enforcement, whose mission has only a small part of what is needed. In fact, since then, the principal law enforcement agency, DoJ, abandoned the National Commission. So another agency must step in. In 2009, NIST was not seen as "a natural leader by the scholars, scientists, and practitioners in the field" (p17) — but that has been true of NIST before, and NIST rose to the challenge to develop that respect and expertise, particularly in measurement science, standards, and technology. NIST can develop mechanisms to support the research in validation and reliability studies needed in forensic science, and remain independent of law enforcement. Its agenda should be informed by the forensic community, but not be beholden to it. Its results need to be communicated to other agencies that have a stake in reliable forensic science, including DoJ (which has NIJ/OIFS), DNI (which has IARPA), DOD (which has DARPA), NIH (which funds biology-related forensics) and presumably others.

2. *OSAC composition:* OSAC, as part of NIST, currently is composed of mostly forensic practitioners with few "arm's length" researchers. It has been directed to approve standards without being allowed into the regulatory space of developing their own. As a result, OSAC primarily approves standards that have been based on past practice. This is not progress. Forensic standards cannot be issued if the research underlying them has not been conducted. Real progress is more likely if OSAC units were closer to 50-50 between forensic community representatives and "arm's length" scientists who together can identify needed studies for validation of existing standards and research to improve them. This balance is especially critical for the top-level OSAC unit, the Forensic Science Standards Board. where presently at most only 3 or 4 members could be considered as "arm's length" from the existing system. "Arm's length" researchers are few, as many are unaware of the scientific challenges involved. Until more of them are made aware, the OSAC units may have to reduce in size, so the balance is 50-50.

3. *Centers of Excellence*: The present NIST-funded Center of Excellence has achieved much in four years, but the work to be done is too extensive to be accomplished by a single Center, which presently is limited to research in only pattern and digital evidence. Research in other disciplines may require more funded Centers whose research agendas are coordinated. Work arising from these Centers, especially regarding validation and reliability, should inform decisions made by the OSAC's Forensic Science Standards Board, not swept aside because it would entail changes to existing practice. (See the OSAC organizational chart for the list of other disciplines outside of pattern and digital evidence.)

   Incidentally, along the lines of promoting research, a policy of transparency — availability of software developed and data collected with federal research grants and agreements — would be sensible. Presently, grantees' data are hard to obtain, even when collected with federal grants. (Research in differential privacy, being conducted for surveys, may apply here, to ensure low risk of identifications from such databases.)

4. *Forensic Science Research and the Courts*: If these recommendations are adopted, then the studies and statistical analysis of validity and reliability of forensic procedures, endorsed by the Forensic Science Standards Board, can be used in judicial challenges to the admission of faulty forensic science evidence and in promoting proper standards on which the Courts can rely with confidence.

The costs of continuing business as usual are too high: Courts remain directed, and the lack of reform leads to both false convictions and false acquittals. But, with proper leadership, the situation can be reversed. Thank you.

Appendix

This Appendix describes four of several projects where forensic scientists collaborated with non-forensic scientists and statisticians to advance research in forensic science.

1. Communication among forensic science stakeholders

   The National Commission for Forensic Science included representatives from areas affected by forensic science: law enforcement, prosecutors, defense attorneys, scientists, judges, and forensic practitioners. In its short 2+ years, commissioners together issued several statements that reflected best practice and recommended new directions. This Commission agreed upon the inappropriateness of definitive conclusions such as "100% match," "match to a reasonable degree of scientific uncertainty," and "zero error rate." Courtroom testimony and lab reports have been successfully discouraged from using such terms.

2. Research in Pattern Evidence

   A cooperative agreement between NIST and Iowa State funds a Center of Excellence in Forensic Science, called CSAFE, for Center for Statistical Applications in Forensic Evidence. NIST limits the Center's funded activities to pattern and digital evidence: latent prints, blood stain pattern analysis, firearms & toolmarks, footwear and tire impressions, handwriting; recovery of material in digital media. Successful projects have included: an improved objective method of comparing ballistic markings, methods for assessing similarity in handwriting samples, and objective metrics of quality in images of pattern evidence (fingerprints, shoe prints, etc.). The Center also studies methods of communicating the strength of forensic evidence to laypersons such as those who sit on juries, to minimize distortion of the information being presented. Researchers affiliated with CSAFE spend far more time on their projects, and in interactions with other CSAFE researchers, than they receive in compensation, so, for the federal government, this investment has had substantial pay-offs.

3. Organization of Scientific Area Committees (OSAC)

   Figure 1 shows a diagram that describes the Organization of Scientific Area Committees (OSAC). Standards are proposed in the subcommittees, and approved by the respective Scientific Area Committees, and finally by the 18-member Forensic Science Standards Board, before being posted on the OSAC registry of standards. The organization has been useful in introducing researchers to forensic scientists and practitioners. Full trust and collaboration remains elusive for many members, which likely will erode over time and as both sides see the value of such collaboration. Because OSAC lies within the NIST structure, and NIST is not a standards developing organization (SDO), OSAC is limited to posting standards that already exist, not in developing or modifying existing standards. This limitation has significant consequences (e.g., for standards that may be satisfactory in all but a few paragraphs).

4. Fingerprint Community: Efforts to estimate error rates

   The fingerprint community studied the risk of errors in one part of the latent print identification process: volunteer latent print examiners compared two fingerprint images and reported whether they matched or not (or "inconclusive"). While the reported error rate was low, a more thorough study is needed, one that evaluates the risk of errors resulting from the entire process, with latent print examiners who do not know that the case is a "test."[9]

   In addition, researchers at the Defense Forensic Science Center, then under the direction of Dr. Salyards two years ago, published an article on an approach to developing a semi-objective metric for the strength of the association between a latent print found at a crime scene and a suspect's print. Researchers at the Center of Excellence (CSAFE) were invited to comment on this work. That research, like all scientific research, is in the process of being made fully public, so that it can be improved with further study.[10] Use of this metric, called *FRstat*, has been limited; with increased use, we can compare the risks of false positives and false negatives using *FRStat* versus current practice.

[9]Reporting an error rate in latent fingerprint identification based on results obtained by latent print examiners who volunteer for the study and conduct the side-by-side image comparisons in a lab (and hence know they are being tested), without consideration of the full examination process, including image processing of the print, marked features, number of other candidates, etc., is rather like reporting an automobile accident rate based on drivers who come to a simulator and are tested by a video of highway driving.

[10]See footnote 3.

Note: In September 2016, the President's Council of Advisors on Science and Technology (PCAST) issued a Report to the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*[11]. Apart from some studies of error rates in fingerprint identification, it largely endorsed the findings of the 2009 NAS report that was issued seven years earlier.

[11] obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf

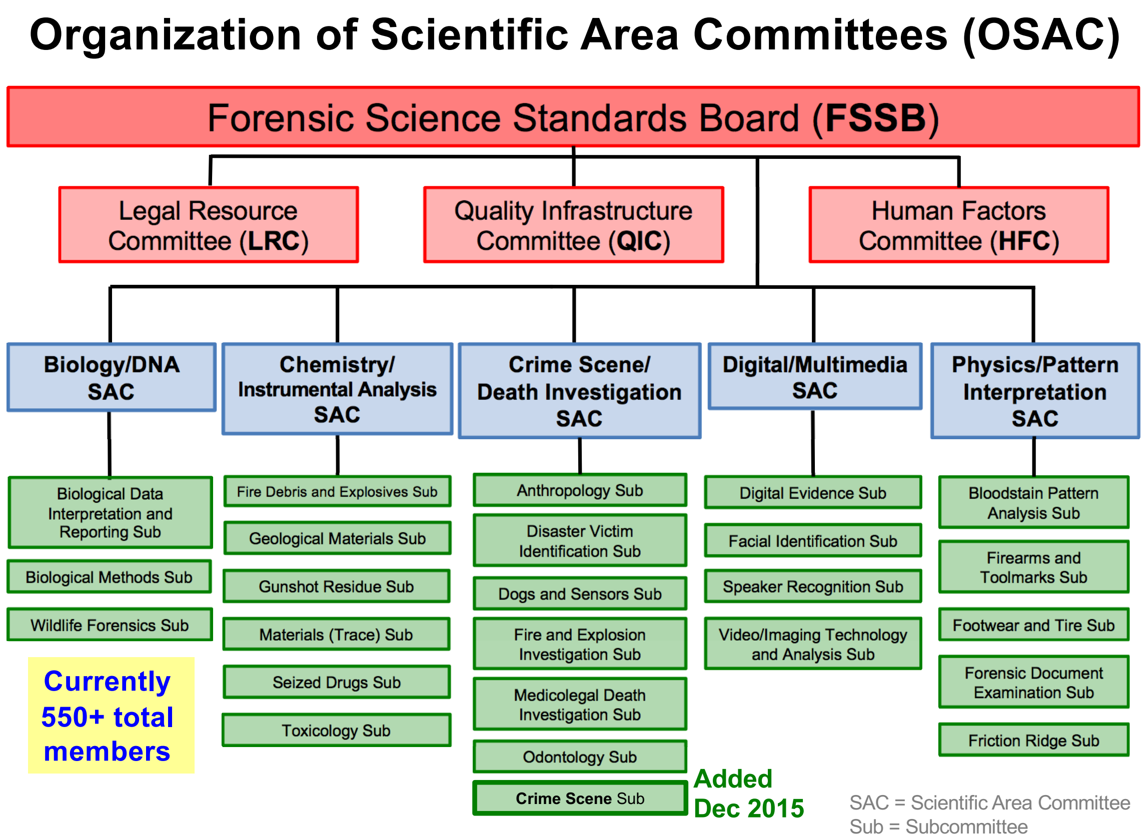

Figure 1: OSAC organizational chart as of October 2018

KAREN KAFADAR

Karen Kafadar is Commonwealth Professor & Chair of Statistics at University of Virginia. She received her B.S. (Mathematics) and M.S. (Statistics) from Stanford University, and her Ph.D. (Statistics) from Princeton University, She previously held positions at National Bureau of Standards (now National Institute of Standards and Technology), Hewlett Packard's RF/Microwave R&D Division, National Cancer Institute, and University of Colorado-Denver. Her research focuses on robust methods, exploratory data analysis, characterization of uncertainty in the physical, chemical, biological, and engineering sciences, and methodology for the analysis of screening trials. She served on the National Academy of Sciences' Committees that led to *Weighing Bullet Lead Evidence* (2004), *Strenthening the Forensic Science System in the United States: A Path Forward* (2009), *Review of the Scientific Approaches Used During the FBI's Investigation of the Anthrax Letters* (2011), *Evaluating Testing, Costs, and Benefits of Advanced Spectroscopic Portals* (2011), and *Identifying the Culprit: Assessing Eyewitness Reliability* (2014). She also served on the governing boards for the American Statistical Association (ASA), Institute of Mathematical Statistics, International Statistical Institute (ISI), and National Institute of Statistical Science. She was a member of NIST's inaugural Forensic Science Standards Board (part of Organization of Scientific Area Committees, or OSAC), and chaired OSAC's Statistical Task Group and ASA's Advisory Committee on Statistics in Forensic Science. She is past Editor of *JASA Reviews* (1996-98) and *Technometrics* (1999-2001), and is current Editor-in-Chief for *The Annals of Applied Statistics*. She is an Elected Fellow of the ASA, ISI, and AAAS, has authored over 100 journal articles and book chapters, and has advised numerous M.S. and Ph.D. students. She is 2019 President of the American Statistical Association.