

Another Harsh Spotlight on Forensic Sciences

By Judge Herbert B. Dixon Jr.

In a 2009 issue of this magazine,¹ I reported the issuance of a press release by the National Academy of Sciences entitled “‘Badly Fragmented’ Forensic Science System Needs Overhaul; Evidence to Support Reliability of Many Techniques Is Lacking.” That press release concerned a congressionally mandated report by the National Research Council (NRC) that called for major reforms to correct serious deficiencies in the nation’s forensic science system.

Following the NRC report, the U.S. Department of Justice (DOJ) and the National Institute of Standards and Technology (NIST) conceived and established specialty organizations to coordinate development of forensic science standards and guidelines and provide policy recommendations to improve the quality and consistency of work in the forensic science community; specifically, the DOJ in collaboration with NIST established the National Commission on Forensic Science (NCFS). After that, NIST separately established the Organization for Scientific Area Committees for Forensic Science (OSAC). In September 2015, after the formation of NCFS and OSAC, President Obama asked the President’s Council of Advisors on Science and Technology (PCAST) whether there are additional steps on the scientific side that would strengthen the forensic science disciplines and ensure the validity of forensic evidence used in the nation’s legal system.

In response to the president’s inquiry to PCAST, that council of advisors undertook evaluations of various forensic science methodologies and thereafter issued a report in September 2016 that, as occurred with the NRC report, critically analyzed the state of forensic science evidence presentations in the nation’s courtrooms. The report, “Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods,”² was damning and dismal. One particular description of PCAST findings



in the opening paragraph of the report’s executive summary is illustrative.

Developments over the past two decades—including the exoneration of defendants who had been wrongfully convicted based in part on forensic-science evidence, a variety of studies of the scientific underpinnings of the forensic disciplines, reviews of expert testimony based on forensic findings, and scandals in state crime laboratories—have called increasing attention to the question of the validity and reliability of some important forms of forensic evidence and of testimony based upon them.

The report is replete with examples of faulty forensic science presentations and conclusions, each of which might aptly be described as a miscarriage of justice. The report focuses on objective and subjective

“feature-comparison” forensic science methods, procedures undertaken to subjectively or objectively compare measurements of two or more samples; e.g., to determine whether an evidentiary sample from a crime scene is or is not associated with a sample from a known individual. Such samples can include DNA, hair, fingerprints, bitemarks, toolmarks, bullets, tire tracks, voiceprints, and other samples or attributes that may be candidates for subjective or objective comparison. Next, the report adopts an evaluation standard to evaluate the validity and reliability of these forensic science methodologies.

Although this column attempts to summarize salient features of the PCAST report and evaluation of several forensic sciences, the full 160-page report is “must-read” preparation material prior to confronting any issue related to the scientific validity of any feature-comparison forensic science methodology.

Standards Used for Evaluating the Validity and Reliability of Forensic Science Methods

In discussing the validity and reliability of feature-comparison methodologies, the report distinguishes between two types of scientific validity, namely, foundational validity and validity as applied. Both are important concepts utilized to evaluate feature-comparison forensic methods.

Foundational Validity

Foundational validity for forensic science requires reliability. It must be based on empirical studies, repeatable, reproducible, and accurate at appropriate levels—a scientific concept intended to correspond to the legal requirement in Rule 702(c) of “reliable principles and methods.” The report concludes that evaluations of validity and reliability must be based on “black-box studies” in which many examiners render decisions about many independent tests and the error rates are determined. Significantly, the report states that an expert’s confidence based on personal, professional experience or expressions of consensus among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies and that nothing can substitute for the requirement to establish foundational validity based on empirical evidence. According to the report, statements claiming or implying greater certainty than demonstrated by empirical evidence are scientifically invalid.

Validity as Applied

Validity as applied means that the method has been reliably applied in practice—a scientific concept intended to correspond to the legal requirement in Rule 702(d) that an expert “has reliably applied the principles and methods to the facts of the case.” The report states that the expert should report the overall false-positive rate and sensitivity for the method established in the studies of foundational validity and should not make claims or implications that go beyond the empirical evidence and the applications of valid statistical principles to that evidence.

Error Rates and Good Professional Practices

The report emphasizes that neither experience nor judgment nor good professional practices can substitute for actual evidence of foundational validity and reliability. In addition, the report condemns testimony by any expert that the expert’s conclusion is “100 percent certain”; has a “zero,” “essentially zero,” “vanishingly small,” “negligible,” “minimal,” or “microscopic” error rate; or has a chance of error so remote as to be a “practical impossibility” as not scientifically defensible. The PCAST report emphasizes that all laboratory test and feature-comparison analyses have nonzero error rates, including highly automated tests. This list of not scientifically defensible claims also includes the often-used phrase “to a reasonable degree of scientific certainty,” which has no generally accepted meaning in science and is open to widely differing interpretations by different scientists and may be understood by the factfinder as implying certainty.

Evaluations of Forensic Science Methods

After establishing the criteria for scientific validity, foundational validity, and validity as applied, the report provides an extensive description of the scientific criteria for establishing the foundational validity and reliability of six well-known forensic science feature-comparison methods:

1. DNA analysis of single-source and simple-mixture samples,
2. DNA analysis of complex-mixture samples,
3. bitemarks,
4. latent fingerprints,
5. firearms identification, and
6. footwear analysis.

In addition, the report provides a brief discussion of a seventh forensic science feature-comparison method, hair analysis, by reviewing recent DOJ guidelines and supporting documents concerning testimony on hair examination.

By far, the most provocative portions of the report are the sections discussing the scientific validity of feature-comparison

forensic science methodologies. What may have been in the past a routine issue of evidence admissibility may in the future be met with frequent challenges based on the conclusions expressed in the PCAST report, including the extent a witness may be permitted to express confidence in the expert opinion being offered. In summary, here are the report’s evaluations of specific feature-comparison methods of forensic analysis.

DNA Analysis of Single-Source and Simple-Mixture Samples

As did the NRC study and report, this study commended DNA analysis of single-source and simple-mixture samples as an excellent example of objective methods whose foundational validity has been properly established. The report notes that DNA analysis, like all forensic analyses, is not infallible and that errors can and do occur, such as errors from sample mix-ups, contamination, incorrect interpretation, and errors in reporting.

DNA Analysis of Complex-Mixture Samples

The PCAST advisors were critical of DNA analyses of complex-mixture samples, noting that the fundamental difference between the analysis of a complex-mixture sample and the analysis of single-source and simple-mixture samples lies not in the laboratory processing, but in the interpretation of the resulting DNA profile. For example, it is often impossible to tell with certainty



Judge Herbert B. Dixon Jr. is the technology columnist for *The Judges’ Journal*, a member of the ABA Board of Governors, and a former chair of the

National Conference of State Trial Judges. He sits on the Superior Court of the District of Columbia. He can be reached at Herbert.Dixon@dcsc.gov. Follow Judge Dixon on Twitter @Jhbdixon.

which genetic variants are present in the mixture or how many separate individuals contributed to the mixture or accurately determine the DNA profile of each contributor. The report concludes that, except in limited circumstances,³ subjective analysis of complex DNA mixtures has not been established to be foundationally valid and is not a reliable technology. The report concludes that substantially more evidence is needed across broader settings to establish foundational validity for DNA analyses of mixture samples.

Bitemark Analysis

The report notes that bitemark analysis is a subjective forensic feature-comparison method based on the premises that dental characteristics differ substantially among people and skin or other marked surfaces can reliably capture these distinctive features. The report notes that current protocols do not provide well-defined standards to support a reliable comparison analysis. Further, no appropriate black-box studies have been undertaken to study the ability of examiners to accurately identify the source of a bitemark. In what may be a death knell for bitemark analysis,

the report concludes that the prospect of developing bitemark analysis into a scientifically valid method is low and recommends against devoting significant resources to such efforts.

Latent-Fingerprint Analysis

Latent-fingerprint analysis, a feature-comparison method that has been in use for over a century, was the subject of significant criticism in the 2009 NRC report. However, due to the implementation of subsequent black-box studies, the PCAST report notes emerging efforts to move the field from a purely subjective method toward an objective method. The report finds that fingerprint analysis is a foundationally valid subjective methodology, but with a false positive rate higher than expected due to longstanding claims about the infallibility of fingerprint analysis. In applying the validity as applied evaluation, the report notes previously identified factors that have affected the accuracy of an examiner's comparison, namely (1) confirmation bias where examiners often alter the features they initially mark in an unknown print based on comparison with an apparently matching print; (2) contextual bias where an examiner's judgment is influenced by other information, irrelevant to the print comparison, about the facts of a case; and (3) proficiency testing for assessing an examiner's capability and performance in making accurate judgments. In addition, the report notes the important move to convert latent print analysis from a subjective method to an objective method due to advances in automated image analysis that holds the promise of making fully automated latent-fingerprint analysis possible in the near future.

Firearms Analysis

Firearms analysis concerns an examiner's attempt to determine whether ammunition is or is not associated with a specific firearm based on toolmarks produced by guns on the ammunition. Although firearms analysts have often stated their discipline has near-perfect accuracy, the 2009 NRC study concluded that sufficient studies had not been done to understand the reliability and reproducibility of the methods; i.e.,

foundational validity of the field had not been established. The PCAST report notes the existence of one appropriately designed black-box study since 2009 that estimates a false-positive rate, but affirmatively states that the scientific criteria for foundational validity require more than one such study. In addition, the validity as applied criteria require that an expert in this field (1) undergo rigorous proficiency testing and disclose the results and (2) disclose awareness of any other facts of the case that might influence the conclusion. Lastly, the report positively notes progress in image analysis over the last several years that gives cause for optimism that, similar to latent-print analysis, fully automated firearms analysis may be possible in the near future. In other areas, e.g., medicine, it is expected that automated image analysis will become the gold standard for many applications involving interpretation of X-rays, MRIs, funduscopy, and dermatological images.

Footwear Analysis

The process of comparing a known object, such as a shoe, to a complete or partial impression found at a crime scene, to assess whether the object is likely to be the source of the impression, is the working definition of footwear analysis in the report. No doubt, the advisors note the absence of appropriate black-box studies. The report states that comparison efforts to undertake footwear analysis are unsupported by any meaning evidence or estimates of their accuracy and are not scientifically valid based on the criteria established for evaluation of feature-comparison methodologies.

Hair Analysis

Hair analysis is a process by which examiners compare microscopic features of hair to determine whether a particular person may be the source of a questioned hair. The PCAST advisors did not undertake a full evaluation of hair analysis scientific validity, but reviewed DOJ guidelines and supporting documents concerning testimony on hair examination.⁴ The DOJ guidelines were released for comment as the advisors were completing their PCAST report to the president.

In the supporting materials, the PCAST

advisors noted a 2002 FBI study comparing microscopic hair examination and DNA analysis where Federal Bureau of Investigation (FBI) analysts used mitochondrial DNA analysis to reexamine 170 samples from previous cases in which the FBI Laboratory had performed microscopic hair examination. The authors found that, in 9 of 80 cases (11 percent) in which the FBI Laboratory had found the hairs to be microscopically indistinguishable, the DNA analysis showed that the hairs actually came from different individuals.

In sum, the PCAST advisors concluded that based on the methodology and results, the DOJ supporting documents provide no scientific basis for concluding that microscopic hair examination is a valid and reliable process.

Conclusion

The report's conclusion is clear that the accuracy of many forensic

feature-comparison methods has been assumed rather than scientifically established on empirical evidence. Over the past decade, the report notes, the forensic science community has recognized the need to empirically test whether specific methods meet the scientific criteria for scientific validity. For instance, there have been appropriate studies that establish the foundational validity and measure the reliability of latent-fingerprint analysis. For most subjective methods, according to the report, there are no appropriate black-box studies. Absent valid black-box studies, there is little or no appropriate evidence of the method's foundational validity or estimates of its reliability. PCAST expects, partly based on the strength of its evaluations of scientific validity in this report, that some forensic feature-comparison methods may be determined inadmissible because they lack adequate evidence of scientific validity. ■

Endnotes

1. Herbert B. Dixon Jr., *Forensic Science Under the Spotlight*, 48 JUDGES' JOURNAL, no. 4, Fall 2009.
2. https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.
3. The limited circumstances exception referenced in the report involves a three-person mixture in which the minor contributor constitutes at least 20 percent of the intact DNA in the mixture.
4. DEP'T OF JUSTICE, PROPOSED UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR THE FORENSIC HAIR EXAMINATION DISCIPLINE, *available at* <http://www.justice.gov/dag/file/877736/download>; DEP'T OF JUSTICE, SUPPORTING DOCUMENTATION FOR DEPARTMENT OF JUSTICE PROPOSED UNIFORM LANGUAGE FOR TESTIMONY AND REPORTS FOR THE FORENSIC HAIR EXAMINATION DISCIPLINE, *available at* <http://www.justice.gov/dag/file/877741/download>.