Forensic Evidence

Erin Murphy[*]

The field of forensic science has come under increasing scrutiny in the past decades. DNA-exoneration cases revealed the pervasive problem of misuse of forensic evidence, blue-ribbon panels of experts have critiqued common methods of forensic science as fundamentally unsound, and a series of laboratory scandals have called into question the integrity of the institutions and actors who deliver forensic findings. Although this attention reveals a system of scientific evidence that is badly broken, the body of scholarly and governmental criticism of the field, along with innovations and expertise at the state and national level, offer clear pathways to reform. This chapter aims to distill that wide body of work into a broad diagnosis of the problems presented in the current state of forensic science, and synthesize some of the best and most promising proposals for reform.

INTRODUCTION

The task of appraising the treatment of forensic evidence in the criminal justice system and setting out recommendations for reform requires first defining the term "forensic evidence." On its face, forensic evidence means evidence derived from the use of a field of science or the scientific method in order to investigate and prove crimes. Accordingly, the phrase encompasses a broad range of disciplines—ranging from "softer" fields of study like psychology or social science to "harder" methods such as biology or chemistry. "Forensic evidence" thus includes everything from a DNA match to a mental-illness diagnosis to the results of a study that reveals cognitive biases in eyewitness identification.

* Professor of Law, New York University.

Because other chapters in this volume address some of the advances in social science most pertinent to criminal cases,[1] this chapter will focus instead on the methods that make up the heartland of forensic evidence. Specifically, in offering a critique and making recommendations, this chapter reviews fields such as DNA typing, fingerprint, fire science, and firearm, toolmark, fiber, hair, and bite-mark analysis. That said, many of the observations shared in this chapter apply across a wide array of methods and techniques, including those drawn from disciplines as diverse as medicine (e.g., "shaken-baby syndrome") or social science (e.g., forensic psychology).

Fortunately, forensic science has received increased attention in the past decade. Early scholars offered trenchant critiques that highlighted the lack of a scientific foundation to support most familiar forensic methods, the lack of standardized qualifications and skills testing for forensic analysts, and the culture of law enforcement that pervades the field.[2] Those views, once considered outliers, have since been augmented and amplified by a new generation of scholars as well as scientific authorities.[3]

The signature assessment of forensic evidence remains the 2009 report by a blue-ribbon panel convened by the National Research Council of the National Academy of Sciences (NAS), *Strengthening Forensic Science in the United States*,[4] which surveyed a wide array of disciplines and offered a critical assessment of their status, while also proposing concrete suggestions for restructuring the delivery of forensic science to the criminal justice system. One of the clearest dictates of that report was the call to remove forensic science from the management and control of law enforcement, where it has historically settled. In 2016, the President's Council of Advisors on Science and Technology (PCAST) issued a report titled *Forensic Science in the Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* that endorsed and amplified

---

1. *See, e.g.,* Stephen J. Morse, "Mental Disorder and Criminal Justice," in Volume 1 of the present Report; Richard A. Leo, "Police Interrogation and Suspect Confessions," in Volume 2 of the present Report; Gary L. Wells, "Eyewitness Identification," in Volume 2 of the present Report; John Monahan, "Risk Assessment in Sentencing," in Volume 4 of the present Report.

2. Although this work is too voluminous to cite in full, and at the risk of leaving out others, some of the most prominent early scholars include Paul Giannelli, Michael Saks, Michael Risinger, Jane Campbell Moriarty, David Kaye, Jennifer Mnookin, Jonathan Koehler, and David Faigman.

3. *See, e.g.*, Jennifer Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. Rev. 725 (2011) (interdisciplinary document urging the "development of a research culture" in forensic science).

4. Nat'l Research Council, Strengthening Forensic Science in the United States: A Path Forward (2009) [hereinafter 2009 NAS Report].

the critical findings of the 2009 NAS Report.[5] Both reports contain a wealth of resources, including detailed assessments of specific disciplines and overarching recommendations for improving the use of forensic science in criminal courts.

As a result of the 2009 NAS Report, in 2014 the United States Department of Justice and the National Institute of Standards and Technology (NIST) jointly convened the National Commission on Forensic Science, a multidisciplinary body that issues its own recommendations.[6] At the same time, the agencies also jointly launched the Organization of Scientific Area Committees for Forensic Science (OSACs), a series of expert groups nested within NIST and tasked with the crafting of technical standards and guidelines for the practice of forensic science.[7]

The Commission served a critical role in marshaling expertise to provide balanced and reasonable recommendations on a wide array of topics.[8] Although not immune to criticism, it was widely considered a success. Yet sadly, after political changes in the executive branch, the Commission was disbanded by Attorney General Jeff Sessions.[9] It now appears that many of the tasks of the Commission will return to the exclusive control of prosecutors within the Department of Justice rather than a neutral entity that is primarily scientific in character and represents the full array of constituencies. If so, then important

5. President's Council of Advisors on Sci. and Tech., Forensic Science in the Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods (2016) [hereinafter PCAST Report].

6. National Commission on Forensic Science, U.S. Dep't of Justice Archives, https://www.justice.gov/ncfs.

7. Organization of Scientific Area Committees (OSAC) for Forensic Science, National Institute of Standards and Technology, https://www.nist.gov/forensics/organization-scientific-area-committees-forensic-science. The stated goal was to "create a sustainable organizational infrastructure dedicated to identifying and fostering the development of technically sound, consensus-based documentary standards and guidelines for widespread adoption throughout the forensic science community." John M. Butler, The National Commission on Forensic Science and the Organization of Scientific Area Committees: Proceedings of the International Symposium on Human Identification (2014), https://www.nist.gov/sites/default/files/documents/forensics/Butler-ISHI-Proceedings2014.pdf. The OSACs are in some respects an outgrowth of the Federal Bureau of Investigation's Scientific Working Groups, one of which (pertaining to DNA) remains and which similarly were technical, field-specific bodies that set standards. Scientific Working Groups, Federal Bureau of Investigation, https://www2.fbi.gov/hq/lab/html/swg.htm.

8. Recommendations have covered issues such as accreditation and proficiency testing, error correction and reporting, reporting and terminology issues, and pretrial discovery, among other things. *See generally* National Commission on Forensic Science, *Work Products*, U.S. Dep't of Justice Archives, https://www.justice.gov/ncfs/work-products-adopted-commission. The Commission has weathered criticism, much of it grounded in the selection of the initial appointees.

9. *See* Erin E. Murphy, *Sessions is Wrong to Take Science Out of Forensic Science*, N.Y. Times (Apr. 11, 2017), https://www.nytimes.com/2017/04/11/opinion/sessions-is-wrong-to-take-science-out-of-forensic-science.html?_r=0.

progress is at risk of stalling, leaving the criminal justice system vulnerable to the very dynamics that gave rise to the crisis in forensic science in the first place.[10]

In light of these changes, it may be more important than ever to shore up existing improvements in the use of forensic evidence in criminal courts, and fortify against calls to dismiss or disregard the need for continued momentum. Against this background, this chapter addresses each of the major components essential to ensuring the integrity of forensic evidence—the scientific basis of evidence, its execution in a particular case, and the checks in place in the criminal justice system—and makes recommendations as to each.

I. SCIENTIFIC BASIS

The scientific integrity of forensic evidence depends on its successful navigation of three safeguards: its methodological validity, its statistical validity, and its execution in a particular case. Unfortunately, as Table 1 illustrates, many familiar forensic techniques have long been admitted as evidence in criminal cases without meeting all—and often times *any*—of these foundational requirements.

A. METHODOLOGICAL VALIDITY

Methodological validity refers to the method's scientific foundation. It measures whether the discipline is valid and reliable; that is, whether the method accurately measures what it purports to measure, and does so in a manner that generates consistent, reproducible results. Palm-reading may be reproducible in the sense that different readers, examining the same palm, would make the same judgments about the length of different lines and their significance, but it is not valid because there is no evidence showing that those readings in fact measure what they purport to measure (for instance, that the length or quality of your life turns on the lines on your palm). Conversely, a valid measurement (say, determining the height of a person) can be undertaken in an unreliable fashion (say, by checking shoe size). It is the two together—an accurate measure taken in a reliable way—that determines scientific legitimacy.

10. Indeed, the Department of Justice and some prosecutorial professional associations have already occasionally proven a reluctant partner in much-needed reforms. *See, e.g.*, Erin Murphy, *What 'Strengthening Forensic Science' Today Means for Tomorrow: DNA Exceptionalism and the 2009 NAS Report,* 9 J. L. Probability & Risk 7 (2010), https://doi.org/10.1093/lpr/mgp030 (detailing history leading up to creation of NAS Report, as well as reactions to its issuance); *see also infra* note 12.

Nearly all of the forensic techniques familiar to laypeople from television dramas or media reports in fact lack this kind of rigorous foundation.[11] The range of support varies dramatically. Some methods have no scientific basis whatsoever, and future research is unlikely ever to establish such foundations. Other methods have surprisingly thin histories of empirical testing, but may become rigorous upon greater study. Some disciplines—such as fire investigation—have undergone recent revamping, having conducted research that invalidated familiar methods as nonscientific while also refining and improving legitimate aspects of investigation (such as chemical analysis). The 2009 NAS Report and the 2016 PCAST Report, although contested to some extent,[12] provide excellent templates for assessing the current state of the science across a range of disciplines. The 2009 NAS report also describes the history that led courts to regularly admit and rely upon nonscientific methods, placing blame largely at the feet of the historical entwining of law enforcement and forensic science.[13] Table 1 summarizes the state of the science, as recounted in these reports, for the most common disciplines.

Importantly, assessment of methodological validity is a dynamic, not static, process. As scientific knowledge advances, it may overturn long-held beliefs or improve upon prior practice. When such changes occur, the legal system must be able to adapt, rather than cling stubbornly to the old ways. History suggests that tension can arise between scientific culture, with its emphasis on challenge and refinement, and legal culture, with its emphasis on precedent, consistency, and finality. But if the criminal justice system is to properly accommodate scientific knowledge, it must devise structures that can accommodate evolution in scientific knowledge when it occurs.

11. *See generally* 2009 NAS Report, *supra* note 4, at 127-82; PCAST Report, *supra* note 5, at 67-123.

12. Both studies were met with resistance of varying degree, typically from professional organizations in criticized disciplines, or from prosecutorial bodies. For example, the United States DOJ responded to the PCAST Report with a letter that critiqued aspects of the study, although it did not include citations to work it claimed the Commission had overlooked. *Comments on President's Council of Advisors on Science and Technology Report to the President*, Federal Bureau of Investigation (Sept. 20, 2016), https://www.fbi.gov/file-repository/fbi-pcast-response.pdf/view. The National District Attorneys Association released a similar response, adding the tautological argument that the use of such evidence in court offered support for its scientific basis. Press Release, National District Attorneys Association, National District Attorneys Association slams President's Council of Advisors on Science and Technology report (Sept. 2, 2016), http://www.ndaa.org/pdf/NDAA%20Press%20Release%20on%20PCAST%20Report.pdf.

13. 2009 NAS Report, *supra* note 4.

TABLE 1: SUMMARY OF PCAST AND NAS ASSESSMENTS OF SELECT FORENSIC TECHNIQUES

METHOD	METHODOLOGICAL VALIDITY	STATISTICS	COMMENTS
Drug testing	Gas chromatography-mass spectrometry (GC-MS) for most controlled substances other than marijuana; well-established and generally reliable when performed according to established standards.	Not commonly included in reports, although such data should be included.	Field tests may be unreliable. Precise weights require regular calibration of instruments.
DNA typing (STR)	Single-source and simple mixture DNA analysis using STR typing is well-founded and rigorously established. Complex mixtures and mixtures involving low amounts of template require special care, and reliable methodologies for traditional analysis may not work properly on samples of low quantity or quality.	The statistics underpinning single-source and simple mixture DNA matches allow for quantitative probability statements, typically as a random match probability. The Combined Probability of Inclusion (CPI) should not be used in mixture cases, as it may be unfairly biased against the defendant.	Probabilistic genotyping software may provide a promising method for analyzing complex mixtures, but at present such software is validated only for mixtures of no more than 3 persons, where the minor contributor offers at least 20% of the DNA in the mixture.
Fingerprint	ACE-V method most commonly used, and judged foundationally valid. But it does not specify specific measurement criteria or match standards. Instead, it depends on subjective judgments of analyst. Evidence suggests the method is susceptible to analyst bias, and has a high false-positive rate.	Examiners have reported an identification (i.e., an individualized match), exclusion, or inconclusive findings. But there is not yet a statistical basis to assert individualization, although such research is underway.	As studies continue, latent print identification may move from a subjective to objective field. As a subjective discipline, analyst training, proficiency, and protection from bias is critical.
Firearm/ toolmark identification	Methodology identifies using "class characteristics" and "individualizing marks." PCAST declares that "firearms analysis currently falls short of the criteria for foundational validity," and NAS cautions that "not enough is known about the variabilities among individual tools and guns," and there is a lack of established match protocol.	Analysts may testify that an item was the "source" of the mark, but this conclusion is based on subjective intuition rather than rigorous science.	As with other pattern-matching disciplines, research is ongoing to establish foundational validity. As a subjective discipline, analyst training, proficiency, and protection from bias is critical.
Hair analysis	Both PCAST and NAS judge hair analysis as lacking scientific foundation. NAS notes there are "no uniform standards on the number of features on which hairs must agree before an examiner may declare a 'match'."	The DOJ underscores that hair analysis "cannot lead to personal identification," and NAS concludes that "no scientifically accepted statistics exist about the frequency with which particular characteristics of hair are distributed in the population," thus no individualizing statements should be made.	A recent review by the FBI found that 90% of cases involving hair analysis contained erroneous statements, suggesting a need for great caution with respect to hair testimony. In place of hair analysis, nuclear or mitochondrial DNA testing offer more-reliable findings.

Bite-mark identification	PCAST concludes that "bitemark analysis does not meet the scientific standards for foundational validity, and is far from meeting such standards." NAS reports that "there is still no general agreement among practicing forensic odontologists about national or international standards for comparison."	NAS states that "[a]lthough the majority of forensic odontologists are satisfied that bite marks can demonstrate sufficient detail for positive identification, no scientific studies support this assessment, and no large population studies have been conducted."	Bite-mark evidence not only lacks any scientific foundation, it also is unlikely to establish such foundation due to variations in how human skin registers bite marks, the uncertain uniqueness of human dentition, and the variability in how bites transfer patterns. As PCAST reports, "available scientific evidence strongly suggests that examiners cannot consistently agree on whether an injury is a human bitemark and cannot identify the source of bitemark with reasonable accuracy."
Arson/fire investigation	Fire investigation involves a wide variety of methods, including chemical detection of accelerants and observation of fire debris to deduce burn patterns, etc. The use of established principles of chemistry is well-founded; in contrast, many canards of fire investigation (e.g., "alligatoring," stippling, crazed glass) have been wholly discredited.	Fire investigators may testify that certain burn characteristics indicate (or even firmly establish) that a fire was intentionally set; many of those rules of thumb lack scientific basis. Even chemical tests that indicate the presence of an accelerant may be confounded by residue from burned household products with similar signatures.	Experiments are underway to improve the knowledge base and training of fire investigators, but such research already indicates that "flashover"—whole room involvement in a fire—may occur early and complicate analysis.
Impressions/ pattern analysis (shoe, tire, etc.)	Class characteristics (e.g., determining shoe size or make from an impression) and individualizing marks (e.g., "randomly acquired characteristics"). Both lack established methodology and a rigorous empirical basis; PCAST judges individualizing efforts "not scientifically valid."	Typically reported qualitatively not quantitatively, such as "positive identification" or "nonidentification," but such statements unsupported by statistical surveys, and may mislead jurors if not properly contextualized.	Largely a subjective discipline, with a high degree of variation among analysts and little testing of many assumptions underlying the field. Current studies by NIST and FBI underway. As a subjective discipline, analyst training, proficiency, and protection from bias is critical.
Fiber analysis	NAS reports that "[N]o set standards, for the number and quality of characteristics that must correspond in order to conclude that two fibers came from the same manufacturing batch," and "no studies of fibers … on which to base such a threshold" or "whether environmentally related changes … are distinctive enough to reliably individualize."	Matches can provide only "class" based evidence, that fibers could have come from the same type of item, not individualization evidence.	More rigorous studies of the chemistry underpinning fiber analysis would enhance the discipline. As a subjective discipline, analyst training, proficiency, and protection from bias is critical.

B. STATISTICAL VALIDITY

Statistical validity is a shorthand for the value that can properly be ascribed to the results of a forensic matching test. It is an umbrella term meant to capture of range of ideas that underpin how an analyst reports the value of a forensic match. Typically, forensic evidence is introduced to prove the ultimate issue in a criminal case: the guilt of the defendant. This purpose often invites analysts and prosecutors to engage in "source attribution"—the claim that because a piece of evidence matches a characteristic of the defendant, it must have come from the defendant alone, or even more pointedly that it alone provides proof of the defendant's guilt.

But nearly all forensic methods lack the scientific foundation necessary to establish that a piece of evidence in fact came from any one individual person to the exclusion of all others. And our legal culture dictates that the question of guilt is for the factfinder to answer, not any individual witness. Even if evidence may be said to match a person (typically, the defendant), it is necessary to properly contextualize that match within scientific parameters that show how common or uncommon such matches might be. For most nonscientific evidence, this is intuitive: If a witness says the robber had blonde hair, and the factfinder observes the defendant also has naturally blonde hair, the factfinder uses lived experience to gauge the value of that match, which might change if one lives in Minnetonka versus El Paso. Depending on the community, blonde hair may be more or less common and thus render the match more or less powerful evidence. But of course, even where blonde hair is rare, the match alone cannot provide proof that the defendant is the source; there remains the possibility that the defendant matches the evidence by coincidence.

When an expert witness reports on a match between a piece of forensic evidence and the accused, however, these background frequencies—the rate at which such matches occur by coincidence—are all too often not known, much less shared to the jury. Juries do not have intuitive sensibilities about the frequency of the loops and whorls of a fingertip match or the microscopic characteristics of a hair or fiber, and so they are wholly dependent on the analyst's testimony not just to accurately report observed matches, but to contextualize the import of those matches. Yet all too often, analysts write reports or testify in court in a manner that inflates the value of the match beyond its scientific basis. In other words, analysts engage in source attribution, notwithstanding the lack of a statistical basis upon which to make such conclusions. In the scholarly literature, this is known as the "individualization fallacy"—the idea that simply because certain characteristics of a piece of evidence match the defendant, it is possible to conclude that the defendant is the source of that evidence.[14] Forensic analysts may report simply that evidence "matches the defendant" without clarifying that a match alone does not mean the evidence must have come from the defendant, or even make much more powerful declarations—at times even inventing statistics or making unfounded, quantitative claims ("In my 10-year career, I've never seen a match this good.")—notwithstanding the lack of a scientific basis upon which to make such claims.[15]

Both the 2009 NAS Report and the 2016 PCAST Report sharply criticize the practice of reporting the significance of a match in language that either misleads as to the strength of that match or outright fabricates a statistical basis. The National Commission on Forensic Science issued formal recommendations encouraging greater transparency in report-writing,[16] so that, at minimum, analysts make clear whether their judgments rest on identifiable and reliable criteria, or simply on their experience or personal intuitions (which, unless scientifically documented, are of course opaque and thus problematic). The Commission has also expressly discouraged the use of specific common but misleading language. For instance, the Commission urged rejection of the phrase "reasonable degree of scientific certainty," which had been widely adopted by forensic analysts as an invented measure of the significance of their findings, without any scientific support.[17] All consumers of forensic evidence should pay close attention to the language used to express the significance of test

14. The literature is full of scathing critiques and examples of the individualization fallacy. *See, e.g.,* Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 Vand. L. Rev. 199 (2008). The mistaken equation of the probability of a random match with the probability that the defendant is the source even has acquired its own catchphrase: the "prosecutor's fallacy." William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 Law & Hum. Behav. 167 (1987).

15. *See, e.g.*, Brandon L. Garrett & Peter J. Neufeld, *Invalid Forensic Science Testimony and Wrongful Convictions*, 95 Va. L. Rev. 1 (2009) (surveying invalid testimony in wrongful conviction cases).

16. Nat'l Comm'n on Forensic Sci., Recommendation to the Attorney General Documentation, Case Record, and Report Contents (2016), https://www.justice.gov/ncfs/page/file/905536/download ("Reports should clearly state: the purpose of the examination or testing; the method and materials used; a description or summary of the data or results; any conclusions derived from those data or results; any discordant results or conclusions; the estimated uncertainty and variability; and possible sources of error and limitations in the method, data, and conclusions.").

17. Nat'l Comm'n on Forensic Sci., Recommendations to the Attorney General Regarding Use of the Term "Reasonable Scientific Certainty" (2016), https://www.justice.gov/ncfs/file/839726/download ("Forensic discipline conclusions are often testified to as being held 'to a reasonable degree of scientific certainty' or 'to a reasonable degree of [discipline] certainty.' These terms have no scientific meaning and may mislead factfinders about the level of objectivity involved in the analysis, its scientific reliability and limitations, and the ability of the analysis to reach a conclusion.").

results and whether such expressions in fact have sound scientific footing. Not every jurisdiction will be willing or able to conduct original research in order to determine the scientific validity of forensic methods. But every jurisdiction is capable of adopting and endorsing a rigorous approach to forensic evidence, using resources made available at the national level.

II. EXECUTION

Even a methodologically sound, statistically supported technique must also be properly executed in order to produce trustworthy results. Proper execution is particularly important for many of the pattern-matching disciplines, because what little methodological validity exists depends almost entirely on quality of the execution. That is, the integrity of the findings in disciplines grounded primarily on an analyst's experience hinges directly on the integrity of the examination process. But even methods with greater scientific rigor, such as DNA testing, are still only as reliable as the laboratory and analyst that conduct that testing.

Indeed, the soundness of the testing process has as great a bearing on evidentiary reliability as anything that happens within the judicial process itself. Perhaps the single most culpable contributor to our current broken system of forensic science has been the misplaced belief that courts and lawyers would serve as a bulwark against faulty evidence. In fact, the first and most important line of defense against faulty forensics is a well-run laboratory staffed with trained and competent personnel, who engage in founded scientific inquiry using testing protocols designed to minimize and immediately remedy error. Unfortunately, the key components to scientific integrity remain lacking in all too many jurisdictions. Although some states, such as New York, Virginia, and Texas, have attempted to impose greater oversight on crime laboratories, those efforts have met with various degrees of success; more significantly, even these modest efforts are altogether absent from too many jurisdictions.

Generally speaking, there are three key components to effective oversight: meaningful accreditation, imposition of certification or qualification standards for analysts, and regular and effective proficiency testing. As described in greater detail below, accreditation determines the quality of the laboratory itself and its standard operating procedures; certification verifies that analysts have acquired certain skills or technical abilities necessary to perform their job; and proficiency testing ensures that the analyst's actual practice meets the prescribed standards of accuracy and excellence.

A. ACCREDITATION

Accreditation standards for crime laboratories, like those for other institutions, require that the laboratory meet certain pre-established criteria. Clinical testing laboratories that perform medical analysis are by and large reliable in the United States, no doubt in part because they must adhere to demanding standards of oversight.[18] In contrast, crime laboratories historically have not been required to meet any accreditation standards at all. Nevertheless, in the past decades, tremendous progress has occurred in accrediting labs, and as of now, roughly 88% of 409 publicly funded crime laboratories in the nation hold accreditation from a professional forensic-science organization.[19] The problem is that 73% of labs hold accreditation from ASCLSD/LAB,[20] which has proven too lax an accreditor.

ASCLD/LAB is a spinoff from ASCLD, which is a professional organization of crime-laboratory directors.[21] These original accreditation programs were largely decorative, with little by way of arduous review. Eventually, the accreditation arm of ASCLD broke off, forming an independent entity known as ASCLD/LAB. That group aimed to impose more-demanding standards, and eventually implemented an accreditation process that required laboratories to meet the international ISO/IEC 17025 standard.[22] ASCLD/LAB also gained prominence as a result of a rule imposed upon DNA testing laboratories by the FBI, which required that laboratories earn accreditation in order to access the national DNA database.

Unfortunately, the ASCLD/LAB accreditation process has largely proven a failure. Nearly every major lab has weathered a major scandal of incompetence or malfeasance,[23] and almost all were accredited by ASCLD/LAB. Critics charge that the accreditation process lacks the necessary seriousness to truly correct or deter bad practices. Accreditation reviews are primarily reviews of documents submitted by the lab, with advance notice, done by peers who are unlikely to penalize a fellow lab.

18. *See generally* Clinical Laboratory Improvement Amendments of 1988, 42 U.S.C. § 263a; *see also* 42 C.F.R. § 493.

19. Andrea M. Burch et al., Bureau of Justice Statistics, U.S. Dep't of Justice, Publicly Funded Crime Laboratories: Quality Assurance Practices, 2014 (Nov. 2016), https://www.bjs.gov/content/pub/pdf/pffclqap14.pdf.

20. *Id.* at 2.

21. *See* Am. Soc'y of Crime Lab. Dirs., http://www.anab.org/ (last visited Apr. 1, 2017); *see also* Erin Murphy, Inside the Cell: The Dark Side of DNA 59-65 (2015).

22. *See* Murphy, *supra* note 21, at 59-65.

23. 4 David Faigman et al., Modern Scientific Evidence: The Law and Science of Expert Testimony § 30:15 (2016) (listing laboratories involved in just DNA-related scandals).

However, ASCLD/LAB was recently acquired by ANAB,[24] and there are indications that the accreditation process may become more rigorous. Truly meaningful oversight would require that every lab that conducts forensic testing—no matter the kind—be accredited, and by an accrediting agency that conducts intensive reviews. An excellent process of accreditation would involve random, unannounced inspections; regular review (clinical labs generally follow a two-year schedule, for instance, in contrast to five years for crime labs); and standards for alerting the accreditor when major errors are discovered and for determining and correcting the root cause of such errors, among other things.

B. CERTIFICATION

Certification is a process by which an individual examiner demonstrates that she or he has acquired the specialized knowledge and expertise to carry out specific testing duties. Certification is akin to licensing; whereas a license is issued by the state to authorize an individual to engage in a particular practice restricted to licensees, certifications are typically voluntarily undertaken and simply signify that the person has met particular standards of achievement. Certifications or licenses may be awarded on the basis of a variety of assessment methods, including through "exams, proficiency testing, evaluation of education, training, and practical experience, adherence to codes of ethics, and other standards."[25] As expected, any particular certification program can impose more or less demanding standards.

As with accreditation, the fraction of labs with at least one externally certified analyst, as well as the overall percentage of certified analysts, has steadily climbed over the past two decades. As of now, 72% of public crime labs have at least one externally certified analyst,[26] but far too few analysts overall have such certifications. Moreover, no jurisdiction requires that all forensic analysts be licensed, even as most jurisdictions require licensing for occupations as diverse as manicurists or private investigators. Without such requirements, an analyst may be entrusted to perform evaluations, supervise others, or train subordinates without having met any external requirements to ensure that the person is capable of performing the assigned tasks.

24. ANAB stands for the ANSI-ASQ National Accreditation Board, an entity formed in the 1990s in response to the need for an American certification body that would ensure private-sector compliance with international standards developed to facilitate commerce in the nascent European Union. In 2011, it expanded into forensic science with the acquisition of Forensic Quality Services, an established accreditor of forensic laboratories; and in 2015 it acquired ASCLD/LAB. *See generally* ANSI-ASQ National Accreditaiton Board, *About ANAB*, http://www.anab.org/about-anab.

25. Burch et al., *supra* note 19, at 3.

26. *Id.* at 6.

C. PROFICIENCY

Proficiency is a measure of a particular lab employee's actual performance. With regard to a forensic analyst, proficiency typically refers to measures designed to determine whether the analyst is executing forensic tests properly. The term can also refer to supervisors or others in the chain of evidence processing, who may likewise be tested to ensure that they are accurately performing their duties. Proficiency is distinguished from certification in that the latter measures whether the person knows the rules and standards of the job, whereas proficiency measures whether the person actually can fulfill their duties.

Proficiency tests may be conducted many different ways, and those variables affect the degree to which the test actually captures an individual's likely performance in normal working conditions. For instance, tests may be given internally by personnel within a laboratory, or administered by external authorities. They may be declared, so that the analyst is aware that the test is taking place, or blind, so that the analyst is unaware that his or her work will be scrutinized. Tests may mirror casework conditions, which entails the kinds of difficult judgments that an analyst makes in the field, or be conducted using artificially pristine or clear-cut samples designed to determine basic competence. And reviews may involve fabricated samples inserted into the regular routine, or they may involve post-hoc case reanalysis of actual work conducted by the analyst in a real case (called "case re-analysis"). For obvious reasons, each of these variables has significant effect on the degree to which a proficiency test actually measures typical field performance.

Proficiency testing has been a continued source of debate and controversy within the forensic-science community.[27] Without question, the most demanding measure of proficiency would involve blind testing, in fieldwork conditions, by an external tester. Perhaps second would be random case reanalysis, wherein an external reviewer randomly pulls a sampling of analysts' completed files and conducts a full re-analysis. Because these two approaches are more likely to uncover malfeasant or incompetent actors, numerous expert bodies have recommended that either blind testing or random re-analysis become a regular feature of laboratory oversight.[28] But some laboratories have

27. *See* 2009 NAS Report, *supra* note 4, at 207-08 (recounting debate over feasibility of blind proficiency testing).

28. PCAST Report, *supra* note 5, at 58 ("[P]roficiency testing should *ideally* be conducted in a 'test-blind' manner—that is, with samples inserted into the flow of casework such that examiners do not know that they are being tested.").

insisted that such testing is too costly and not feasible, even as others have willingly implemented blind exams.[29]

For reasons that are not clear, the use of random case re-analysis and blind proficiency testing have declined in recent years, even as greater numbers of labs gain accreditation and greater attention has focused on the quality of forensic evidence. In 2014, only 35% of labs conducted random case analysis (down from 54% in 2002), and only 10% conducted blind exams (down from 27% in 2002).[30] These declines are worrying; although 98% of labs conduct some kind of proficiency testing, the vast majority of labs rely on declared tests to gauge proficiency.[31] Yet a declared test—which oftentimes does not even include samples that truly replicate the ambiguity or difficulty inherent in real-world conditions—is a poor means by which to judge an analyst's typical work performance. The reluctance of accreditors and other oversight entities to require blind proficiency testing or regular random case re-analysis may be the single greatest factor contributing to continued laboratory failures.

RECOMMENDATIONS

In nearly every jurisdiction, the governance structures of existing forensic science are inadequate to safeguard the integrity of forensic evidence. That is why so many jurisdictions have endured a major laboratory or analyst scandal of some kind. All too commonly, prosecutors and legislators cite the adversarial system itself as the best safeguard against admission of faulty forensic evidence, expecting that the courtroom or the legal process can somehow substitute for rigor and precision in the testing process itself. But the task of ensuring the integrity of forensic evidence begins at the crime scene, extends through the testing stage, and ends in the courtroom. The judicial branch ought to serve as the last, not first, line of defense against bad science.

Consistent with this vision, a wealth of expertise has emerged that provides guidance and instruction to jurisdictions seeking to overhaul their systems of forensic evidence. The list below draws upon that rich literature.

1. **Statewide oversight commission.** A handful of states, including New York, Virginia, and Texas, have created statewide commissions that oversee forensic science.[32] The duties, composition, and actual function of these commissions vary significantly, and not all are equally successful.

---

29. *Id.* at 59 & nn.139-40.

30. Burch et al., *supra* note 19, at 4.

31. *Id.*

32. *See, e.g.*, Tex. Code Crim. Pro. Ann. art. 38.01 *et seq.*; 37 Tex. Admin. Code § 651.1 *et seq.*; N.Y. Exec. Law § 995 *et seq.*; Va. Code Ann. § 9.1-1109 *et seq.*

One determinant of success seems to be the balance of power on such commissions, which can easily tilt heavily pro-government, given that seats apportioned by constituency may naturally align prosecutors, police, and lab personnel together against a sole criminal defense attorney representative. Another imperative of success seems to be that the commission be housed independent of law enforcement or prosecutor offices, and staffed in an even-handed and independent fashion, including with persons formally trained in the scientific method. Commissions have also faltered as a result of inadequate resources or structures incompatible with the oversight expectations. A commission that meets a handful of times a year, composed of busy professionals supported only by a thin permanent staff, may simply not have the bandwidth to carry out a lengthy roster of duties or conduct searching inquiries. Success requires a substantial commitment—in time, personnel, and resources.

The Texas Commission, established in 2005 and then given expanded powers in 2015, has emerged as a model in many respects. Commission members—nine individuals appointed by the governor, half of whom by designation are academic faculty with scientific expertise[33]—have shown repeated willingness to directly confront shortcomings in forensic science. The Commission has issued guidance documents and ordered the reopening of compromised cases. The Commission manages an impressively transparent website, where it posts its official positions on scientific topics, offers an avenue to lodge complaints against specific labs or analysts, and issues comprehensive annual reports. In addition, as of 2015, the state Legislature has required that testing labs be accredited, and charged the Commission with overseeing that accreditation process.[34] The state likewise requires the Commission to create and execute a licensing program for analysts within certain forensic disciplines. In short, the Commission has proven itself indispensable to the project of ensuring

33. Member Appointments, Texas Forensic Science Commission, http://www.fsc.texas.gov/member-appointments. Interestingly, none of the seats are specifically reserved for law enforcement; and there is only one prosecutor and one defense attorney. By comparison, the New York Commission on Forensic Science has twelve members appointed by the government, and a third are reserved for persons connected with law enforcement or state crime labs. *See* N.Y. Exec. Law § 995-a (requiring two seats for persons connected with crime labs in the state, one for the director of the office of forensic services, and one for law enforcement, along with two criminal defense representatives and one prosecutor).

34. Unfortunately, the legislation exempts several disciplines, such as latent print examination and breath testing, from that requirement. 37 Tex. Admin. Code § 651.6. For allegations involving unaccredited labs or disciplines, the Commission maintains disciplinary authority but cannot make findings of negligence or misconduct; it can only report observations and make recommendations.

methodological soundness, statistical competence, and integrity in the execution of forensic tests.

Texas provides one successful model, but each jurisdiction must respond to local needs in crafting its own commission. In all cases, however, properly staffed and resourced oversight bodies of this kind serve an important institutional role. They function as both repositories of information and watchdogs. In their best iteration, a commission of this kind might set out and enforce standards (or translate nationally set standards) for testing, report-writing and disclosure, and consistent terminology. Commissions can superintend the accreditation process, and flex their muscle in order to encourage outside accrediting agencies to adopt more-demanding review processes. In well-resourced states, such commissions might also identify areas of necessary research, and even foster or apportion funds in support of research activity.

Perhaps most importantly, statewide commissions can serve as coordinators and regulators in a field rife with error and misunderstanding. The Texas Commission, for instance, has undertaken discipline-specific investigative reviews in an array of fields, including microscopic hair analysis, fire investigation, bite-mark analysis, and DNA mixture interpretation. Those reviews may involve official statements about faulty forensics as well as reopening of closed cases. The goal is to generate, or transmit, authoritative findings that reflect a research-oriented approach to forensic disciplines, rather than simply perpetuate baseless methods because they are familial to legal actors.[35] The Commission has also conducted reviews of specific forensic scientists and laboratories,[36] and produced guidance and training documents for an array of actors.

However, it is important to note that a commission is not a stand-in for the difficult kinds of judgments that legislative, executive, and judicial actors must make regarding forensic policy. It exceeds the proper scope of

35. *See, e.g.*, Mnookin et al., *supra* note 3.

36. The Paul Coverdell National Forensic Science Improvement Act was intended to operate as a check on laboratory quality, because it conditioned receipt of grant funds on a laboratory's identification of an outside auditor to investigate allegations of serious negligence or misconduct. But, in keeping with concerns about prosecutorial or law enforcement oversight of lab quality, the Department of Justice—entrusted to enforce this provision—largely ignored its oversight responsibilities until the Inspector General criticized this neglect in an investigation. *See* Murphy, *supra* note 21, at 286; *see also id.* at 288 (detailing criticism of DOJ's seeming bias in awarding research grants).

a commission to set policy about compulsory DNA collection or privacy-impinging searches, for instance. Nor should executive or judicial actors consider the judgments of the commission unimpeachable.

A commission's findings may enhance and embolden a court's understanding of the reliability of a forensic method under the evidentiary rules, but it should not supersede the court's own responsibility to apply evidentiary standards of admissibility. And while a police department might be guided by a commission's dictates as to collection or handling procedures, absent express legislative authority, determinations of that kind ultimately fall to the executive branch.

Notwithstanding clear limits on its authority, an active, empowered commission with adequate resources and proper personnel can achieve great strides in safeguarding forensic integrity. A good commission would engage in methodological and as-applied preventative maintenance and error correction, setting best practices for the future while also demonstrating a willingness to identify past errors and engage in a transparent assessment and correction of their systemic, root causes. It would also serve as a repository and record-keeper, generating important data (about forensic practices in the state) that could be shared widely and made available for analysis and inspection.

2. **Meaningful accreditation requirements for all laboratories.** As the prior section noted, accreditation ought to be compulsory for all testing laboratories. At the same time, current data suggest that most labs are already accredited; the problem is no longer lack of accreditation, but lack of *meaningful* accreditation. At this time, there is an opportunity to influence the manner in which crime-laboratory accreditation processes unfold, because the two major accrediting entities—known as ASCLD/LAB and FQS—have been subsumed into a single new organization, ANAB.[37]

   A full accounting of the components of effective accreditation are too lengthy to recount here, but the critical components include requiring: that laboratories undertake certain bias-reducing strategies, given extensive

37. ANAB even recently sought comments on the accreditation requirements they ought to impose on forensic laboratories. *Comments Sought on ANAB-ASCLD/LAB Requirements*, http://anab.org/news/latest-news/comments-sought-on-anab-ascldlab-requirements/ (Nov. 28, 2016) (last visited Apr. 1, 2017). ANAB also administers the ABFT accreditation program, which accredits toxicology labs. In addition, a longstanding accreditor of non-forensic labs, known as A2LA, has recently sought to expand its forensic laboratory accreditation. *See* https://www.a2la.org/.

scholarly evidence of the problems of unconscious bias in testing;[38] that laboratories adopt a framework for root-cause analysis of significant errors that arise, and report transparently on those findings, rather than dismiss such problems as idiosyncratic or atypical;[39] that analysts undergo certification and regular, effective proficiency testing; and that meaningful inspections include surprise visits, random case re-analysis, or other means to ensure that the laboratory's true face, and not just its best face, is presented for evaluation.

3. **Certification/licensing and meaningful proficiency testing of analysts.** Few states require that any forensic analysts be licensed, even though such standards of demonstrated competence are required for employees tasked with arguably less consequential work, such as manicurists or athletic trainers. Meaningful forensic reform would impose a licensing or certification standard on all analysts. Such requirements would require an initial demonstration of capability and be followed by regular proficiency testing and continuing-education requirements. Proficiency tests and certification requirements not only ensure an analyst's basic competence, they also can provide useful data as to the frequency and probability of error in the testing process.

   In some disciplines, such as drug analysis, blind proficiency testing may be more feasible and common (in part due to the need to ensure that analysts do not substitute or remove controlled substances). In other disciplines, simulating casework may prove more challenging. But even where costs or logistics prohibit regular blind testing, no such limits preclude random spot checks of an analyst's prior work. Random case re-analysis, while not ideal, is preferable to a program wholly dependent on the analyst's performance on declared tests that often little replicate actual casework conditions.

   Finally, it should be acknowledged that, for some disciplines, a certification and proficiency test process is likely not possible, because the underlying methodology is fundamentally unsound. But that is a reason to impose such requirements, not to avoid them. A discipline that cannot be objectively measured and assessed is one that has no place in evidence in the criminal courts.

38. *See, e.g.*, Itiel E. Dror et al., *Cognitive Bias and Its Impact on Expert Witnesses and the Court*, 54 Judges J. 4, 8 (2015). Dr. Dror is likely the most recognized empirical expert in cognitive bias in forensic testing, with a large body of relevant work. *See generally* http://www.ucl.ac.uk/~ucjtidr/.

39. *See, e.g.*, N.Y.C. Admin. Code § 17-207 (requiring that municipal DNA testing laboratory engage in root cause analysis following any "significant event").

4. **Training for legal actors.** Forensic disciplines are subject to change and evolution, as researchers make new findings or unseat received wisdom. Without clear conduits of this information, however, legal actors are unlikely to possess the wisdom necessary to perform their systemic roles. Apart from large jurisdictions with the capacity to develop dedicated dockets, the typical judge, defense attorney, or prosecutor may encounter any single forensic method only sporadically and infrequently.

   Some critics charge that legal actors are simply impervious to education, and thus alternative solutions (such as independent, court-appointed experts or specialist teams of attorneys) pose the only way of ensuring forensic integrity. But such an approach is both unrealistic in the vast majority of jurisdictions, and perhaps too fatalistic. At the federal level, resource manuals and judicial trainings offer an opportunity for judges to learn about specific methods and their shortcomings. Prosecutors often have direct access to laboratory personnel, who typically align themselves with the government's interests. And defense attorneys have increasingly taken up the mantle of peer education through professional programs and continuing legal education trainings.

   A state-level commission might be explicitly tasked with ensuring adequate opportunity for legal actors to acquire scientific expertise, including programs that cover fundamental principles of the scientific method and statistical competence. But even in the absence of a commission, resources must be made available on the state and local level for such education and access to expertise.

5. **Defense access to experts.** The problem of inadequate access to expertise particularly plagues defense counsel. All too often, counsel's ability to secure expert advice turns on the leniency of a court or judicial officer who must grant a request for added expenditures. But that process raises the specter of impropriety, both in that it requires the defense attorney to petition the judge, thereby compromising client confidentiality, as well as by placing an attorney at the mercy of the court in assessing the validity of the claim. What is more, the thin market for defense experts may mean that the kind of resources most valuable to a defense attorney—such as transcripts from prior testimony of the government's expert or consultations regarding the kind of available challenges to the evidence—are not available.

   Ideally, a jurisdiction would make generous awards for defense access to experts, or endow institutional defenders with sufficient funds to provide such services. But in reality, solutions that depend on large capital expenditures by the government, in favor of the defense, often face political obstacles.[40]

In 2010, North Carolina's Office of Indigent Defense Services, which is charged with overseeing indigent defense in the state, created the position of Forensic Resource Counsel (FRC).[41] That office, staffed by a single attorney, provides an array of services to indigent defenders in the state. The FRC oversees a collection of databases that provide information about state experts for both the prosecution and defense, free online training courses in an array of topics, state laboratory protocols and procedure documents, the latest scientific research, and pertinent news stories. The blog, which is regularly updated, serves as a conduit for important information, and the webpage offers a bank of motions and briefs that may serve as templates for everything from discovery to orders to suppress evidence or appoint experts. The FRC also provides limited consultation and referral services.

Of course, perhaps the most critical determinant of the success of a position of this kind is personnel; the attorney who originated the role and still serves in that position has shown remarkable industry and innovation in establishing the office. But apart from conducting a thorough job search, the precise job description could include specific mandates to generate and maintain materials like those outlined above.

Another benefit of a position of this kind is to give a statewide voice to the needs of criminal defendants as regards forensic science. A statewide resource counsel is well positioned to identify structural infirmities in the delivery of forensic evidence in the state, and to provide feedback to regulators and oversight entities—including a statewide commission, which might even appoint the counsel ex officio. A statewide resource counsel's office might also propose or comment on existing legislation, from the express standpoint of a representative of the defense community. This kind of insight may become even more critical as sophisticated forensic evidence continues to feature in ordinary criminal cases. For instance, DNA testing increasingly involves results from private, for-profit software companies that use secret algorithms to determine the significance of a DNA sample, and defense access to DNA databases is

40. *See, e.g.*, Eve Brensike Primus, "Defense Counsel and Public Defense," in the present Volume.

41. *See* North Carolina Office of Indigent Defense Services, http://www.ncids.com/forensic/index.shtml (last visited Apr. 1, 2017).

increasingly a point of contention. Digital forensics likewise relies upon technically sophisticated information that companies are often reluctant to reveal. As such evidence continues to surface in courts, it is imperative that institutional actors hear from those close to the ground—not just on the prosecution but also defense side—as to how best to ensure the integrity of such evidence.

6. **Legal reform to accommodate changes in science.** A final critical area of reform that the criminal justice system must undertake to ensure the reliability of forensic evidence involves the systemic response to changes in scientific knowledge. The problem is twofold: existing, even well-researched scientific methods may become eclipsed by newer, improved methods, and current techniques may be revealed problematic or infallible by future work, thereby calling the integrity of prior convictions into question. Both situations present challenges for a criminal justice system that heavily relies on finality and certainty in deciding cases.

   Again, here, legal reforms in Texas provide a valuable template. In 2013, the state Legislature enacted a series of reforms aimed at providing convicted persons with an avenue of relief under habeas corpus in the event that science upon which the conviction was based was later discredited, or new scientific methods emerged that might establish innocence.[42] But too many jurisdictions still impose procedural hurdles to overturning convictions, even when the science upon which those convictions are based has proven demonstrably false.[43] Statutory reforms are needed to ensure that defendants have adequate access to evidence pretrial, as well as avenues for accessing and testing evidence,[44] or seeking redress for wrongful convictions, when the state of the science changes.

CONCLUSION

This chapter offered several concrete suggestions for reforming forensic science, drawn from the author's own work as well as a rich scholarly literature. However, this list is by no means exhaustive. Some common recommendations were deliberately left off; for instance, the 2009 NAS report and many scholars have called for independent forensic crime laboratories. That recommendation, while fundamentally sound,

42. Tex. Code Crim. Pro. Ann. art. 11.073 (allowing reopening of cases if scientific evidence "was not available to be offered by a convicted person" or that "contradicts scientific evidence relied on by the state at trial"). Around the same time, the state also passed the Michael Morton Act, which enhanced discovery requirements for criminal defendants.

43. *See generally* Brandon L. Garrett, *Claiming Innocence*, 92 Minn. L. Rev. 1629 (2008).

44. With respect to DNA evidence, reforms must include avenues to allow defendants access to DNA database for exculpatory searches. *See* Murphy, *supra* note 21, at 149-50 & nn.39-41.

has met with strong resistance at the state and national level, and thus it appears politically challenging at this time. Similarly, calls to increase funding and support for crime laboratories and their personnel, while also important, rest more upon the fiscal (and political) inclinations of legislators and less upon reasoned argument. In laying out the suggestions above, this chapter aimed to provide feasible, consensus-oriented targets for immediate and meaningful reform.