Fingerprint Error Rates and Proficiency Tests: What They are and Why They Matter

Jonathan J. Koehler
Fingerprint Error Rates and Proficiency Tests: What They Are and Why They Matter

JONATHAN J. KOEHLER*

INTRODUCTION

When a fingerprint examiner declares a match between a print from a known source and a latent print recovered from a crime scene,1 his word may seal a defendant's fate like no other form of evidence save, perhaps, DNA. At trial the fingerprint examiner will offer little in the way of data, statistical tests, or uncertainty. Instead, he will say that latent print could only have been made by the source of the known print, that he is 100% certain, that he has never erred, and that the method he used to make this and other identifications has an error rate of zero.2 In recent years, the broader scientific community has objected to this form of testimony. Critics charge that fingerprint analysis lacks an empirical foundation and that examiners make exaggerated claims that are likely to mislead jurors.3

Regardless of one's views about the scientific underpinnings of fingerprint examination, all agree that jurors need to understand the

---

1. A latent is a print of unknown source that is recovered from a crime scene.
2. See, e.g., State v. Rose, Case No. K06-0545, at 25 (Md. Cir. Ct. Oct. 19, 2007) (order granting motion to exclude testimony) ("Mr. Meagher has stated that the FBI testifies to 'a 100 percent certainty that we have an identification.' . . . Mr. Meagher claimed that there is no error rate for ACE-V [the fingerprint technique]."); Office of the Inspector Gen., U.S. Dep't of Justice, A Review of the FBI's Handling of the Brandon Mayfield Case 8 (2006) ("Latent fingerprint identifications are subject to a standard of 100 percent certainty.").
probative value of fingerprint evidence at trial. Drawing on research by decision theorist David Schum, researchers at U.C. Irvine explain that probative value is a combination of the diagnosticity and reliability of a reported match. The diagnosticity of a reported fingerprint match is the value of shared characteristics for establishing that two prints share a common source. If prints from many different fingers share the observed characteristics, then the match report will be correspondingly less diagnostic of the claim that the two prints share a common source. If the observed characteristics are rare, then the match report will be more diagnostic of the common source claim. The reliability of a reported fingerprint match pertains to whether two reportedly matching prints actually do share a common set of characteristics. That is, is the reported match a true match, or has the examiner made an error? Taken together, the diagnosticity and reliability of a reported match provide rational jurors with the information they need to assess the probative value of a fingerprint examiner’s opinion that two prints share a common source.

The sad truth is that there is a dearth of good data that directly pertains to either diagnosticity or reliability. On the diagnosticity side, the fingerprinting profession has made remarkably few attempts to test its uniqueness assumptions or to develop an empirical foundation from which to offer probabilistic claims. We simply do not know the frequency with which various characteristics relied on by fingerprint examiners exist in various populations.

More important, however, is the risk that an examiner has erred, thereby implicating reliability. An examiner is unreliable if he frequently or perhaps even occasionally concludes that: (a) prints made from different sources were made by a common source, or (b) prints made from common sources were made by different sources. In its landmark

4. The evidence is best characterized as a “reported match” rather than a “match” because the former description captures the uncertainty associated with the possibility that an examiner mistakenly failed to notice differences between two prints that should have produced a conclusion that the two prints do not match. See Jonathan J. Koehler, On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates, 67 U. COLO. L. REV. 859, 870 (1996) (“The major difference between treating DNA evidence as a certain match and treating DNA evidence as a reported match is that only the latter requires consideration and explicit incorporation of the possibility of error.”).
6. Id.
7. In DNA testing, the random match probability (RMP) captures the diagnosticity of a match report. The RMP is the frequency with which a genetic profile (i.e., the relevant set of alleles) exists in a reference population. William C. Thompson & Simon A. Cole, Psychological Aspects of Forensic Identification Evidence, in EXPERT PSYCHOLOGICAL TESTIMONY FOR THE COURTS 31, 34 (Mark Costanzo, Daniel Krauss & Kathy Pezdek eds., 2007).
8. Kaasa et al., supra note 5.
Daubert” and Kumho” decisions, the Supreme Court stated that trial courts “ordinarily should consider the known or potential rate of error” when evaluating the reliability of scientific techniques and other forms of expert testimony. Though the Supreme Court did not require trial courts to seek error rates, fingerprint examination is surely the poster child for the centrality of this Daubert factor. Without information about error rates, fact finders have an insufficient basis for assessing the examiner’s reliability and assigning weight to his opinion.

Though sufficient data on the diagnosticity and reliability of fingerprint match reports do not exist, there is reason to expect that their diagnosticity will generally be high and that their reliability (as given by rates of error) will be substantially lower. If so, then the issue of error rate is even more important because the probative value of the reported fingerprint match is restricted by the chance that a false positive error has occurred. In other words, if experts make false positive errors, say, one time in 200, then it does not matter whether the chance that two randomly selected prints match is one in a million, billion, or trillion. It does not even matter whether the chance of a coincidental match is zero (as implausible and unscientific as this value is) because in these situations, the false positive error rate limits and controls the probative value of the match report. The relevance of this observation for our

12. Kumho Tire, 526 U.S. at 147 (interpreting Daubert to apply not only to “scientific” testimony, but to all expert testimony); Daubert, 509 U.S. at 594.
13. On the diagnosticity side, the chance of a coincidental DNA match is often extremely small (e.g., one in many millions, billions, or trillions). On the reliability side, the chance of a false match that arises from, say, a sample handling mistake is much larger. If fingerprint coincidental match probabilities also turn out to be quite remote, then the rate at which false matches due to lab errors will probably be much larger.
14. False positive errors are defined infra.
15. A similar point has been made many times in the context of reported DNA matches. See, e.g., David J. Balding, Weight-of-Evidence for Forensic DNA Profiles 35 (2005) (“If the false-match probability (ii) is judged to be much larger than the chance-match probability (i), then the latter probability is effectively irrelevant to evidential weight [and] it is not the absolute but the relative magnitude of the false-match to the chance-match probabilities that determines whether the former can be safely neglected.”); Jonathan J. Koehler et al., The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?, 35 Jurimetrics J. 201, 201 (1995) (“[Random Match Probabilities (RMPs)] contribute little to an assessment of the diagnostic significance of a reported DNA match beyond that given by the false positive laboratory error rate when RMPs are several orders of magnitude smaller than this error rate.”); Richard Lempert, After the DNA Wars: Skirmishing with NRC II, 37 Jurimetrics J. 439, 447 (1997) (“The probative value of a DNA match is always limited by the chance of false positive error.”); William C. Thompson et al., How the Probability of a False Positive Affects the Value of DNA Evidence, 48 J. Forensic Sci. 47, 47 (2003) (“Having accurate estimates [of] the false positive probabilities can be crucial for assessing the value of DNA evidence.”); see also Steve Gutowski, Error Rates in the Identification Sciences, Forensic Bull., Summer 2005, at 23 (“An estimate of actual or potential error rate is crucial to the probative value of all evidence.”).
purposes here is to note that error rate is so central to an assessment of
the reliability and probative value of fingerprint evidence that it is not
sufficient to know, simply, that errors might occur or that errors have
occurred. Instead, judges who make admissibility decisions and jurors
who assess the probative value of fingerprint evidence need to have
better information on the risk that an error has occurred in the instant
case.

Does scientific data related to the risk of error exist? If not, how
shall we go about getting those data? There is much confusion
surrounding these questions. In this Article, I use a question and answer
style to address key issues related to fingerprint error rates and the
proficiency tests that are sometimes used to estimate those rates. My
focus throughout is on how to assess the various error rates, why they
matter, and how we might go about collecting the requisite data. I do not
examine individual cases of error, incompetence or fraud; this
information can be found elsewhere.16 In Part I, I identify the different
types of errors and error rates and explain why knowledge of error rates
is important. In Part II, I discuss the connection between proficiency tests
and estimated error rates. In Part III, I identify the features proficiency
tests must include to ensure that the resultant data can help estimate
casework error rates.

I. ERRORS AND ERROR RATES

A. WHAT CONSTITUTES AN ERROR IN FINGERPRINT EXAMINATION?

There is no agreed-upon answer to what counts as an error. Obvious
errors occur when an examiner either matches a print to a person other
than the one who made it or affirmatively excludes the true source of a
print. Less obviously, some would argue that an examiner who correctly
identifies the source of a print but incorrectly identifies the finger that
produced it has made an error. Others would argue that such an error is
so inconsequential relative to other types of errors that it does not
deserve the "error" label.

Similarly, when an examiner offers an "inconclusive" opinion about
whether two prints match, there is a sense in which he has erred. After
all, he did not get the answer right, and the consequences of this failure
may be serious (e.g., missed opportunity to exonerate a suspect).
However, in the more usual sense of the meaning of error, an

16. For reviews of fingerprint errors, see Simon A. Cole, Suspect Identities: A History of
Fingerprinting and Criminal Identification (2001), and Simon A. Cole, The Prevalence and
Potential Causes of Wrongful Conviction by Fingerprint Evidence, 37 Golden Gate U. L. Rev. 39, 57–
60 (2006). For a review of sloppy laboratory procedures and fraud see, e.g., Paul Giannelli, The Abuse
Pol'y & Law 439 (1997).
inconclusive is not an error. It is a pass. An inconclusive means that the examiner offers no judgment about whether two prints do or do not share a common source. Therefore, for purpose of computing the errors and error rates (see below), I set inconclusives aside.

B. What Are the Different Types of Errors?

Table I below offers a visual aid that helps clarify the different types of errors. Table I crosses the two states of nature pertaining to a pair of fingerprints (\(S\) = same source, \(-S\) = different source) with the two types of decisions that fingerprint examiners offer for their relation ("\(S\)" = says common source, "\(-S\)" = says different source). The four aggregate states described in cells A to D are mutually exclusive (non-overlapping) and exhaustive (all situations described). Cells A and B represent states where the examiner reports that two prints share a common source. Cells C and D represent states where the examiner reports that two prints were left by different sources.

<table>
<thead>
<tr>
<th>Examiner's Judgment</th>
<th>States of Nature (i.e., Truth)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SAME SOURCE ((S))</td>
</tr>
<tr>
<td>Common Source (&quot;(S)&quot;&quot;)</td>
<td>A True Positives</td>
</tr>
<tr>
<td>Different Source (&quot;(-S)&quot;&quot;)</td>
<td>C False Negatives</td>
</tr>
</tbody>
</table>

Two important types of errors are false positive errors and false negative errors. A false positive error occurs when an examiner concludes that two prints share a common source when, in fact, they do not. This error is captured by judgments in cell B. A false negative error occurs when an examiner concludes that two prints do not share a common source when, in fact, they do. This error is captured by judgments in cell C.

C. What Are the Different Types of Error Rates?

Three error rates are of central importance: the false positive error rate, the false negative error rate and the false discovery rate. None of these error rates lay claim to being "the" error rate because the term
The false positive error rate identifies the rate at which an examiner concludes that two prints share a common source when, in fact, they do not. It is computed by dividing the number of false positive errors by the number of examined pairs of prints that are from different sources. In terms of the cells in Table I, the false positive error rate

\[ P(\text{"S"} | -S) = \frac{B}{B + D}. \]

The false negative error rate is the rate at which an examiner concludes that two prints do not share a common source when, in fact, they do. It is computed by dividing the number of false negative errors by the number of examined pairs of prints that are from common sources. In terms of the cells in Table I, the false negative error rate

\[ P(-\text{"S"} | \text{"S"}) = \frac{C}{A + C}. \]

The false discovery is identifies the rate at which an examiner's claim that two markings share a common source is wrong. It is computed by dividing the number of false positive errors by the number of examined pairs of prints that the examiner judged to share a common source. In terms of the cells in Table I, the false discovery rate

\[ P(-\text{"S"} | \text{"S"}) = \frac{B}{A + B}. \]

The false discovery rate is the inverse of the false positive error rate. Whereas the false positive error rate is \( P(\text{"S"} | -S) \), the false discovery rate is \( P(-S | \text{"S"}) \). Both are error rates and both provide important

17. IRVING B. WEINER ET AL., HANDBOOK OF FORENSIC PSYCHOLOGY 552 (“In scientific writing and research, error rate has no uniform definition.”).

18. There is no widely agreed-upon terminology for describing this particular error rate. In using the term “false discovery rate,” to describe \( P(-S | \text{"S"}) \), I follow the recent statistics literature. See Yoav Benjamin & Yosef Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, 57 J. ROYAL STAT. Soc’y 289, 291 (1995) (the false discovery rate is “the proportion of the rejected null hypotheses which are erroneously rejected). The epidemiological literature, which spawned the terms “false positive” and “false negative,” does not identify this error rate (which is computationally identical to “one minus the positive predictive value”). Others have referred to it as the “percent false positives.” See JOHN MONAHAN, THE CLINICAL PREDICTION OF VIOLENT BEHAVIOR: AN ASSESSMENT OF CLINICAL TECHNIQUES 79 (1981); Stephen D. Hart et al., A Note on Portraying the Accuracy of Violence Predictions, 17 LAW & HUM. BEHAV. 695, 697 (1993). However, “percent false positives” may be confused with its inverse, the “false positive error rate.” Their term “false alarm rate” would seem to describe this error, but this phrase is often used it as a synonym for the false positive error rate. Terminological ambiguities are yet another reason why there is confusion over the meaning of error rate.
information for assessing the accuracy of fingerprint examiners.

It is important to reiterate that there are other ways to report error rates and that there are even different ways to report the false positive error rate. However, in the interests of simplicity and clarity, I limit the discussion of error rates to the central notions mentioned here.

D. WHICH ERROR RATE IS MORE USEFUL: THE FALSE DISCOVERY RATE OR THE FALSE POSITIVE ERROR RATE?

There is no easy answer. All error rates are useful, but each has its own unique features and limitations that legal actors must understand.

The false discovery rate is probably what decision makers most want to know and what they assume that an error rate tells them: how often is the examiner wrong when he calls a match? The central problem with false discovery rate is that it varies as a function of the baserate chance that an unknown latent print and known prints handled by the examiner share a common source. Figure I illustrates the point. Suppose that two equally skilled fingerprint examiners each have a 5% false positive error rate, and a 5% false negative error rate. If the first examiner reviews 1,000 pairs of prints, 900 of which share a common source, his false discovery rate will be (on average) 5 out of 860 or 0.6%. In other words, when this examiner concludes that two prints share a common source, he is wrong less than 1% of the time.

Now suppose that the second examiner reviews 1,000 pairs of prints, in which only 100 pairs share a common source. This examiner’s false discovery rate will be (on average) 45 out of 140 or 32%! Is the second examiner worse than the first? No. They are identical in terms of their skill level as evidenced by their identical false positive and false negative error rates. However, the second examiner makes more mistakes because he is subjected to a tougher task. The point is that when the false

20. Simon A. Cole, More Than Zero: Accounting for Error in Latent Fingerprint Identification, 95 J. Crim. L. & Criminology 985, 1030 (2005) ("There are a number of different ways of reporting false positives. Often the false positive rate has been reported as the number of participants who committed at least one false positive divided by the total number of participants."); William C. Thompson, Subjective Interpretation, Laboratory Error and the Value of Forensic DNA Evidence: Three Case Studies, 96 GENETICA 153, 155 (1995) ("The false positive rate of a test is most usefully stated as the ratio of false positives to the sum of true positives and false positives."") (citation omitted)). Thompson's definition of the most useful form of the false positive error rate is identical to the "false discovery rate." Id.
21. Lempert, supra note 15, at 449 ("[T]he jury wants to know, given that a match has been reported, how likely is it that it is false.").
22. For simplicity, ignore the rate of inconclusive judgments, which we will assume to be similar for the two examiners.
23. The task is tougher for the second examiner because the match baserate is lower (100 out of
discovery rate is available, special attention should be paid to the mapping between the baserate chance that pairs of prints share a common source in the target case versus in the task (usually a test or series of tests) that generated the false discovery rate. If these chances are similar, then the false discovery rate provides a simple, direct measure of a key error rate in the instant case. If not, then the false discovery rate should be adjusted to account for the difference in chances.

1,000 = 10% for the second examiner versus 900 out of 1,000 = 90% for the first examiner).
FIGURE I: EFFECT OF BASERATES ON THE FALSE DISCOVERY RATE

FALSE POS ERROR RATE = 5%
FALSE NEG ERROR RATE = 5%

1,000 Sample Pairs

Common Source

Different Source

900

85
45

5
95

P (Different Source / Examiner Says "Common Source") =

\[
\frac{5}{855 + 5} = \frac{5}{860} = 0.6\%
\]

FALSE POS ERROR RATE = 5%
FALSE NEG ERROR RATE = 5%

1,000 Sample Pairs

Common Source

Different Source

100

95
5

45
855

P (Different Source / Examiner Says "Common Source") =

\[
\frac{45}{95 + 45} = \frac{45}{140} = 32.1\%
\]
The false positive error rate does not rely on an assumption about the rate at which prints handled by an examiner share a common source. This is both a strength and a weakness. It is a strength because the mapping issue described above does not threaten the applicability of the statistic to the target case. It is a weakness because the implications of this statistic depend on the baserate chance that latent and known prints in the target case share a common source. In short, there is no getting around the problem of baserate match chance when it comes to applying error rates to individual cases. Either that baserate is built into the statistic as it is for the false discovery rate, or a case-specific baserate must be combined with the "purer" false positive error rate to determine how likely it is that the match call in a particular case is erroneous.

E. IS NOT ONE HUNDRED YEARS OF ADVERSARIAL CASEWORK TESTING PROOF ENOUGH THAT THE RISK OF ERROR IN FINGERPRINT EXAMINATION IS EXTRAORDINARILY LOW?

No. The fact that fingerprint evidence has long been admitted in courtrooms is insufficient reason to conclude that fingerprint evidence must have an extremely low error rate. Adversarial testing bears no relation to the type of scientific testing that can yield information about error rates. Cross-examination, presentation of opposing experts, and other tools of the adversarial method cannot reveal rates of error because there is no ground truth available against which to compare examiners' conclusions. That is, in order to know if an examiner's conclusion is correct, we would need to know what the truth was (e.g., is that really the defendant's fingerprint?) as given by reliable information other than that provided by the examiner. Because truth in any particular case...
case is rarely known, the best we can do is estimate the rate at which examiners commit errors in similar situations. Such estimates can only be obtained by subjecting examiners to controlled scientific testing.

The lack of ground truth, in combination with a presumptive belief in the accuracy of fingerprint evidence, stacks the deck against finding fingerprint errors when they are made. Sometimes highly credible and probative information comes to light that provides assurance that the fingerprint examiner's judgment was right or wrong. But those cases are the exception. Most often, there is no independent proof of the accuracy of an examiner's call. Consequently, the absence of large numbers of known errors in fingerprint cases says little about the accuracy of fingerprint match declarations.

F. Does the Training and Experience of the Fingerprint Examiner Provide an Appropriate Basis on Which to Assess the Risk of Error in a Given Case?

No. Whereas an examiner's training and experience are relevant for assessing his expert qualifications, they tell us little about the risk of error. Expertise does not guarantee reliability or translate into a low risk of error. Studies in psychology show that people with the extensive experience and training commit enormous judgment errors and, in some instances, make worse decisions than those with less experience and training. In some forensic sciences, experienced analysts do not commit fewer errors than inexperienced analysts or even, on some tasks, laypeople. The point is that training and experience alone are

28. Cole, supra note 27, at 1211 ("[I]t is difficult to imagine what evidence would possibly convince any criminal justice system actor, including the defendant's own counsel, that the fingerprint evidence was erroneous."); Interview by Adrian Cho with Simon Cole, NEW SCIENTIST, June 16, 2001, at 42 ("Remember a trial is not an experiment because you have no way of validating your results. If you falsely match a fingerprint, then the guy is convicted and goes to jail. Who's going to believe he's really innocent unless there's some extraordinary circumstance, like DNA comes along or someone else confesses? There's no way you would ever know.").

29. Michael J. Saks et al., Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science, 49 HASTINGS L.J. 1069, 1102 n.169 (1998) ("The allusion to 'thousands of cases without error' begs the question of validity. In actual disputed cases it rarely, if ever, is possible to tell whether the identification was correct or not; that is why the issue was before the fact finder.").


insufficient criteria on which to base the risk of error.

G. **Does Not Introduction of an Error Rate Place an Unfair Burden on Fingerprint Evidence?**

Some people have argued that requiring error rates for fingerprint examiners and other forensic scientists is unfair because we don't require error rates for other types of evidence such as eyewitness testimony, evidence of motive, or business records. So why can't we trust juries to apply their collective common sense and experience to fingerprint evidence, much as they apply their judgment to other types of evidence? The reason is that common sense and experience do not prepare fact finders to assess the reliability of a fingerprint expert's judgment. Two reportedly matching prints do not necessarily share a common source just because an expert says so. Yet there is a real risk that this is exactly what jurors will assume after they hear an expert who purportedly has specialized knowledge testify to a 100% certain match.

Common sense and experience do prepare jurors to assess the significance of a finding that a defendant's fingerprint is found on a drinking glass at the scene of a crime. But before they reach this inferential step, jurors must understand that there is uncertainty regarding the identification of the defendant's fingerprint in the first place. Unlike empirically derived error rate estimates, common sense does not illuminate this source of uncertainty.

H. **Is It Not Inappropriate to Use General, Industrywide Examiner Error Rates to Estimate the Chance of Errors by a Specific Examiner in a Specific Case?**

This argument has intuitive appeal. After all, no one in his right mind would assume that all examiners, in all laboratories, working on different cases have identical chances of making a false identification when those examiners have different skill levels, use different subjective matching criteria, and are working with different sets of prints. So doesn't it follow that industrywide examiner error rates are inapplicable to case-specific examiner error rates? No.

No one has ever argued that the unique features associated with a particular forensic science laboratory, examiner, or case should be ignored when these features are demonstrably related to a reduced error rate. But the mere existence of such unique features does not deny relevance to reliably produced industrywide error-rate estimates. Instead, the industrywide error-rate estimates provide anchors for

---

32. Margaret A. Berger, *Laboratory Error Seen Through the Lens of Science and Policy*, 30 U.C. DAVIS L. REV. 1081, 1088 (1997) ("We regularly admit other kinds of evidence, such as business records, without any thought of quantifying the error that undoubtedly attends their creation.").
judgments about the risks of error in individual cases. This is an elementary and critical point: if the industrywide false positive error rate is 20% for technique A and 2% for technique B, then the risk of this type of error in any given case is generally higher when technique A is used rather than technique B.

Wait, you say. What if the examiner in the target case is unusually proficient when using technique A and there is solid evidence to support this contention? Why should he be tagged with a 20% error rate when, in all probability, his error rate is lower? The answer is that he should not be tagged with the 20% error rate as a final estimate. Instead, the 20% industrywide estimate should be adjusted downward in light of this, and all other information. But, as a rule, the 20% industrywide estimate associated with technique A should serve as the starting point for any specific adjustments, just as the 2% estimate serves as the starting point for any adjustments associated with technique B.

Having said this, one must be careful not to presume that industrywide error rate estimates should be adjusted downward in most cases simply because some favorable case-specific features exist (e.g., unusually talented examiner). If the industrywide error rate estimate is produced reliably, then the frequency and magnitude of downward and upward error rate adjustments across all cases should be about the same. Thus, just as a decision maker should adjust a general error rate downward when the examiner is unusually talented, he should adjust it upward when the latent print is unusually faint. All factors should be considered and adjustments should be made around the estimated industrywide error rate.

33. Faigman, supra note 27, at 74 (“It should matter to a factfinder whether the error rate of an identification technique is a fraction of a percent or 20 percent, even if the court regards neither level of error sufficient to exclude the expert testimony.”).

34. Those who contend that a 20% industry-wide error rate should be ignored when some examiner-specific information is not accounted for in the industry-wide rate are committing a logical error called the base rate fallacy. The base rate ordinarily remains relevant even in the face of additional individuating information. A more sophisticated challenge to the base rate argument in this context occurs when there is credible information about the specific cause of testing errors (or the absence of testing errors) that either no longer exists or has changed significantly. If, for example, a disproportionate number of proficiency-test errors were committed by those who had less than one year of experience, but new rules forbid testimony from such relatively inexperienced analysts, then the error rate should be recomputed using the narrower examiner reference class (i.e., examiners who have at least one year of experience). That is, rather than rely on the error rate for all examiners, a more appropriate error rate would be the subset of examiners who have at least one year of experience (subject to sample size concerns). Paul E. Meehl & A. Rosen, Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores, 52 Psych. Bull. 194, 194 (1955). This point is less an indictment of the use of error rates than it is a reminder that error rates based on narrower reference classes are preferable to those based on broad reference classes when both are available. Jonathan J. Koehler, The Base Rate Fallacy Reconsidered: Normative, Descriptive and Methodological Challenges, 19 Behav. & Brain Sci. 1, 10 (1996).

35. Koehler, supra note 4, at 874 (“It cannot be that every laboratory in the industry has an error
I. **Is There a Difference Between the Human Error Rate and a Methodological Error Rate?**

No. In a misguided effort to salvage the strange and misleading "zero error rate" claim, some in the forensic science community have suggested that there are two broad classes of errors: those that arise due to method and those that arise due to the humans who employ those methods. According to this view, the method never has and never will err. When mistakes occur, they are due to incompetent and/or poorly trained people.

But as Simon Cole has pointed out: "There is no methodology without a practitioner, any more than there is automobile without a driver, and claiming to have an error rate without the practitioner is akin to calculating the crash rate of an automobile, provided it is not driven." When this issue of automobile crash rates arises, drivers and cars are part and parcel of the automobiles that occasionally crash. Likewise, fingerprint examiners and the observation methods they use are "inextricably linked" to the fingerprint examination process. Indeed, where a method depends as heavily on subjective human judgment as does fingerprint examination, the method literally is the people who employ it.

To summarize, there are many different types of errors and error rates. Knowledge of error rates is necessary to assess the probative value rate that is less than the industry-wide average. Yet this is exactly the claim that is made every time laboratory personnel reject the industry-wide figures, and argue that their own error rate is substantially less or zero.

36. See, e.g., United States v. Havvard, 117 F. Supp. 2d, 848, 854 (S.D. Ind. 2000) ("The government claims the error rate for the method is zero. The claim is breathtaking, but it is qualified by the reasonable concession that an individual examiner can of course make an error in a particular case.").

37. Id.; Joint Appendix at 168, United States v. Rogers, No. CR-90-1-BR (E.D.N.C. Dec. 7, 2000) ("[I]t's the government's position, with respect to fingerprints, there's no rate of error. One, [sic] the fingerprints match. There's no rate of error. It's a matter of separating the methodology from the practitioner. The method is foolproof. If there's a mistake it's with the practitioner and that goes to the qualifications of the witness."); see also Simon A. Cole, More Than Zero: Accounting for Error in Latent Fingerprint Identification, 95 J. CRIM. L. & CRIMINOLOGY 985, 1034–37 (2005) (criticizing the concept of methodological error rate).

38. Cole, supra note 37, at 1039.

39. David A. Stoney, Fingerprint Identification: Scientific Status, in 4 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 82, 103 (Faigman et al. eds., 2002) ("Current fingerprint examination processes are inextricably linked with the human examiner and one cannot separate the human error of the examiner (whatever the cause) from the reliability of fingerprint evidence.").

of a reported fingerprint match, but this inference is not an easy one. Some might be tempted to eschew error rates in favor of such unscientific arguments as “we would know by now if errors were a problem,” “error rates don’t apply to individual cases,” or “the fingerprint method is infallible.” These arguments should be rejected. A better approach is to focus on understanding what error rates are and how to obtain the best possible error rate estimates. The following section on proficiency tests addresses this issue.

II. PROFICIENCY TESTS

A proficiency test is an assessment of the performance of laboratory personnel using samples whose sources are known to the proficiency test administrator but unknown to the examinee. Such tests can serve many purposes including but not limited to training and testing personnel, improving laboratory practices and procedures, and identifying future needs for a laboratory. Properly designed, proficiency tests may also provide a reasonable estimate of the rate at which false discoveries, false positive errors, and false negative errors occur. My comments below focus on the value of proficiency tests as an indicator of error rates.

A. HOW ARE PROFICIENCY TESTS CURRENTLY CONDUCTED?

Collaborative Testing Services (CTS) offers hundreds of laboratories the opportunity to participate in two fingerprint proficiency tests each year. Test participation is voluntary, examinees know that they are participating in a test, and it is not clear whether examinees work by themselves, in groups, or with assistance from supervisors. A 2007 test in which 351 response sheets were returned to CTS is typical. In this test, participants received photographs of eleven latent prints from an alleged crime scene, four sets of known finger and palm imprints, and a short scenario that described a bank robbery. Nine of the eleven latents matched some of the known prints, two did not. Examiners were not told whether any of the knowns produced any of the latents. Once they reached their conclusions, examinees filled out a results form, mailed the

41. Ronald G. Nichols, Defending the Scientific Foundations of the Firearms and Tool Mark Identification Discipline: Responding to Recent Challenges, 52 J. FORENSIC SCI. 586, 592 (2007) (“[P]roficiency tests can offer to the court a reliable practical indicator of how often the profession, using accepted procedures, practices, and controls, makes a false identification.”).
42. See Forensics Program—Collaborative Testing Services, Inc., http://www.collaborativetesting.com/forensics/index.html (last visited Apr. 20, 2008). CTS is the primary provider of external forensic science proficiency tests across a variety of subfields, including fingerprint examination.
43. Id.
45. Id. at 2.
46. Id. at 3.
data to CTS, and CTS published the performance report (laboratories and examinees were anonymous).

B. DO EXISTING PROFICIENCY TESTS TELL US ANYTHING ABOUT THE POSSIBILITY OF ERROR AND THE RATES WITH WHICH ERRORS ARE LIKELY TO OCCUR IN CASEWORK?

Yes and no. Existing proficiency tests provide some information about the risk of error, though they fall far short of offering convincing scientific proof of the rates at which errors occur. Consider, once again, the 2007 CTS fingerprint proficiency test. The test materials included two latent prints that were produced by people whose known prints were not provided. However, four examiners (1.1%) incorrectly matched one of the latent prints to one of the knowns; three examiners matched a latent to an innocent suspect; a fourth examiner matched a latent to an innocent bank employee.

This test demonstrates that some examiners are likely to commit false positive errors on occasion. But this does not mean we can trust CTS tests to identify the rates with which errors occur. First, there is the base rate problem noted in question D above. Second, the CTS tests tend to be conducted under unreasonable test conditions (e.g., non-blind conditions that use relatively easy materials). This means that the error rates on existing proficiency tests are probably lower than those encountered in ordinary casework.

47. See id. at 19-20.
48. Cole, supra note 37, at 1033 ("The existing data are inadequate to calculate a meaningful error rate for forensic fingerprint identification."); Lyn Haber & Ralph Norman Haber, Error Rates for Human Latent Fingerprint Examiners, in Automatic Fingerprint Recognition Systems 339, 339 (Nalini K. Ratha & Ruud Bolle, eds., 2004) ("It is impossible to determine from existing data whether true error rates are miniscule or substantial."); William A. Tobin & William C. Thompson, Evaluating and Challenging Forensic Identification Evidence, CHAMPION, July 2006, at 12, 19-20 ("[P]roficiency testing in forensic science is frequently worthless as a true indicator of examiner proficiency.").
49. CTS Test No. 07-516, supra note 44, at 2.
50. Id.
51. This is a frequent criticism of DNA proficiency tests. See United States v. Llera Plaza, 188 F. Supp. 2d 549, 558 (E.D. Pa. 2002) ("Llera Plaza II") (containing testimony from fingerprint expert Allen Bayle who claimed that the FBI's fingerprint proficiency tests are easier than casework because they contain too many matching prints and because the test prints had greater than usual clarity); Jonathan J. Koehler, Why DNA Likelihood Ratios Should Account for Error (Even When a National Research Council Report Says They Should Not), 57 Jurimetrics 425, 429-30 (1997) ("[B]ecause most proficiency tests are relatively easy in the sense that they are nonblind, internal tests that use large, clean stains, performance on these tests would appear to provide a generously low estimate of (criminal) casework error rates.").
52. Lempert, supra note 15, at 447 (noting that existing proficiency tests "can be expected to underestimate error rates, for laboratory technicians can take special care when they know they are dealing with test samples"); Adina Schwartz, A Systematic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification, 6 COLUM. SCI. & TECH. REV. 2, 27 (2005) (noting that CTS proficiency tests "are likely to have understated day-to-day laboratory rates because the testing was declared, rather than blind"). But see Steve Gutowski, A Response to: A Systematic
as the lower boundary for actual casework error rates.53

**C. CAN PROFICIENCY TESTS BE DESIGNED TO PROVIDE REASONABLE ESTIMATES FOR THE RATES AT WHICH EXAMINERS COMMIT ERRORS IN CASEWORK?**

Yes. The key lies in the planning and execution of the tests. From a planning standpoint, careful thought must be given to issues of print difficulty and the frequency with which available known prints are the source of the sample latents. The participants should be representative of examiners who testify in court.55 Further, the courts should weigh in to ensure participation.55 From an execution standpoint, disinterested administrators should oversee the testing process. The tests should be double blind in the sense that neither the examinees nor examinees' supervisors should be aware that the materials being analyzed are part of

---

53. When fingerprint proficiency tests are more difficult—as they were in a test provided by CTS in 1995 which include a latent print from the identical twin of one of the knowns—the number of false identifications can increase dramatically. In this test, 33 out of 150 reporting laboratories (22%) made at least one misidentification on the seven samples. Twenty-nine labs made exactly one misidentification, two labs made two misidentifications, and two labs made six misidentifications. **Collaborative Testing Servs., Inc., Forensic Testing Program, Latent Prints Examination, Report No. 9508** (1995). As a former editor of the *Journal of Forensic Identification* noted, "reaction to the results of the CTS 1995 Latent Print Proficiency Test within the forensic science community has ranged from shock to disbelief." David L. Grieve, *Possession of Truth*, 46 J. Forensic Identification 521, 524 (1996). Subsequent tests have been less difficult in that they rarely include latents and knowns from such highly related individuals. Importantly, the absence of a clear metric by which to gauge print difficulty is a problem in its own right. Haber & Haber, supra note 27, at 95 ("[T]he profession lacks a quantitative measure of print quality (difficulty). One expert observed that the prints used in the FBI proficiency test are so easy they are a joke.") (citation omitted)).

54. Ideally, the participants should be a random sample of this population.

55. Because the forensic science community is unlikely to embrace and voluntarily participate in tough, realistic proficiency tests, the courts might require such participation from laboratories that wish to offer results in a trial. This point arose in the context of DNA evidence as well. See Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 Cal. L. Rev. 721, 796 (2007) (arguing that laboratory error rate data would have to be provided "if reliability were treated as a threshold question of admissibility"); Barry C. Scheck, *DNA and Daubert*, 15 Cardozo L. Rev. 1959, 1997 (1994) ("Unfortunately, forensic laboratories have historically resisted external blind proficiency testing and other efforts to assess laboratory error rates. ... Unless there are strong decisions on the error rate issue, forensic DNA laboratories and legislators will not voluntarily do what Daubert requires.") (footnote omitted)). For a different view, see Berger, supra note 32, at 1088 (arguing that it is “pointless” to argue that reliable error rates should be admitted because error rates for individual laboratories are “unattainable”), and Edward J. Imwinkelried & D. H. Kaye, *DNA Typing: Emerging or Neglected Issues*, 75 Wash. L. Rev. 413, 460 (arguing that the admissibility “standards relate to the capacity of an analytical procedure to generate accurate results when properly applied,” not to whether the individual using those procedures is prone to err).
a proficiency test. The following section provides a simple blueprint for the production and administration of proficiency tests that are designed to identify the rates at which errors occur under various conditions.

III. PROFICIENCY TEST DESIGN FOR MEASURING ERROR RATES

As noted above, proficiency tests serve a variety of goals. One of those goals—and the central issue in this Article—is the identification of reasonable first pass estimates for error rates in case work under a variety of conditions. This goal cannot be attained under the existing approach to proficiency testing. Instead, a variety of factors related to the production of samples, selection of participants, and administration of the test need to be sharpened before the public can trust that the error rates identified on proficiency tests offer reasonable first-pass estimates for the rate at which fingerprint examiners err in actual cases.

A. WHAT FEATURES SHOULD THE SAMPLE PRINTS HAVE?

The latent prints used in proficiency tests should approximate a random sample of the type of latents found in legal cases. This may be accomplished in different ways. One way is for test administrators to access a database of all cases in a county, state, country, or other population over some time period (e.g., five years), and to note which of these cases included latent prints. A random sample of those cases might then be identified as prototypes for proficiency test samples. The random sample is likely to include samples and cases that vary widely. One case might include two detailed latents, and rolled prints from one suspect and two innocents. Another case might include a single, smudged latent and rolled prints from ten suspects, including a pair of siblings. Once a random sample of fingerprint cases has been identified, proficiency test administrators should write comparable cases and produce sets of latent prints that resemble those in the sample. Materials should not be reused across tests. The newly created latent prints should be rated for

56. The highest quality proficiency testing is “double-blind” administered by an outside agency. In ideal double-blind testing, the test taker and test taker’s supervisor(s) should not be aware that the sample(s) under analysis is a proficiency test. Anything less than double-blind testing administered by an outside agency significantly erodes the effectiveness and value of a proficiency test. Tobin & Thompson, supra note 48, at 21 n.25.


58. See Lyn Haber & Ralph Norman Haber, Letter to the Editor, 56 J. FORENSIC IDENTIFICATION 493, 496 (2006) (arguing that fingerprint error rate studies should adopt “conditions representative of case work,” including varying the difficulty of latents, using latents in which the majority do not match known prints, and using representative examiners).

59. Proficiency test administrators commonly reuse materials in different tests. This is a poor idea because examiners sometimes recognize the materials and because it limits testing materials.
difficulty using an agreed-upon rating scheme\textsuperscript{60} to ensure that they parallel the sample of selected cases and to allow researchers to track the impact of latent difficulty on accuracy. Likewise, administrators should track task features such as whether some latents are from common sources or not, and whether rolled prints that match latents are or are not present. Proficiency tests that are created in this (or a similar) manner should increase public and judicial confidence that test materials reflect those that occur in actual case work.

B. SHOULD THE PROFICIENCY TESTS BE BLIND OR OPEN?

Proficiency tests should be double blind. This means that neither the examiners nor any party that has a direct interest in how the examiners perform should be aware that the proficiency test materials are part of a test rather than part of actual casework.\textsuperscript{61} The reason for this precaution is that examiners' performance is likely to improve when they know they are being tested.\textsuperscript{62} Presumably, knowledge that one is being tested makes examiners more vigilant than they are when performing casework.\textsuperscript{63}

Proficiency tests, like ordinary casework, should also be conducted under conditions of contextual blindness. That is, examiners should not be exposed to anything more than a few essential facts of the target case that are required to complete their examination. They should not be told anything about the beliefs of law enforcement, prosecutors, or other personnel, nor should they know the type or strength of other evidence in the case. Research demonstrates that match reports issued by fingerprint examiners are influenced by contextual factors, including what they believe others think of the evidence.\textsuperscript{64} Contextual blindness is required to ensure that examiners base their conclusions on the forensic science evidence alone.

Blind proficiency testing is used in some forensic science areas\textsuperscript{65} and

\begin{footnotesize}
\begin{enumerate}
\item Development of a print-difficulty rating scheme will require some empirical testing. See Haber & Haber, \textit{supra} note 27 (calling for an experimentally validated quantitative metric of latent print difficulty).
\item In open proficiency testing, examiners know they are being tested. Joseph L. Peterson et al., \textit{The Feasibility of External Blind DNA Proficiency Testing. I. Background and Findings}, 48 J. FORENSIC SCI. 21, 26 (2003).
\item \textit{Id.}
\item \textit{Id.} (stating open proficiency tests sometimes receive "special treatment").
\item Peterson et al., \textit{supra} note 61 ("Currently, some proficiency testing programs, such as the Department of Defense's proficiency testing program for forensic urine drug testing [reference omitted] and HIV testing, are blind."); Steven Rand et al., \textit{The GEDNAP (German DNA Profiling Group) Blind Trial Concept}, 116 INT'L J. LEGAL MED. 199, 201 (2002) (reporting results from DNA
\end{enumerate}
\end{footnotesize}
has broad support in the academic community. Moreover, a detailed pilot investigation with DNA analysts showed that blind testing can be done.

C. WHO SHOULD ADMINISTER THE PROFICIENCY TESTS?

The administrators should be qualified, disinterested parties. By “qualified” I mean people who have expertise across a broad range of subject matters including experimental design, testing, statistics, behavioral sciences, forensic science, and police investigation. If cost is a major concern, then cost control specialists should also be consulted. By “disinterested” I mean that proficiency test administrators should not be affiliated with the examinees or the examinees’ laboratories, nor should they stand to benefit from any particular outcome or set of outcomes on the proficiency tests.

D. WHO SHOULD TAKE THE PROFICIENCY TESTS?

Proficiency test participants should be a random or otherwise representative sample of laboratories and examiners drawn from the population of those who provide fingerprint testimony in court. If we are to obtain reliable data on the accuracy of fingerprint examination, participation should not be a matter of choice. As part of the tests, participants should provide pertinent background information (e.g., training, experience, number of cases, etc.). The policies I recommend

blind trials across 129 laboratories in twenty-eight European countries).

66. Peterson et al., supra note 61, (“[B]lind proficiency testing ... provides a more realistic test of lab performance” than open proficiency testing.); Lempert, supra note 15, at 447 (“The obvious way to estimate laboratory error rates is through blind proficiency testing.”); Risinger et al., supra note 64, at 45 (“The simplest, most powerful, and most useful procedure to protect against the distorting effects of unstated assumptions, collateral information, and improper expectations and motivations is blind testing.”); cf. Alfred Biasotti & John Murdock, The Scientific Basis of Firearms and Toolmark Identification, in 3 Modern Scientific Evidence The Law and Science of Expert Testimony 495, 510 (Faigman et al., eds., 2002) (“It is clear that in such a high stakes game, laboratory administration will do everything possible to ensure that the proficiency tests results are correct before reporting them.”).

67. Joseph L. Peterson et al., The Feasibility of External Blind DNA Proficiency Testing. II. Experience with Actual Blind Tests, 48 J. Forensic Sci. 32, 39 (2003). One examiner was tipped off to the proficiency test by a friendly administrator. Id. at 38. This problem is less likely to arise when the administrators are disinterested parties.

68. Ideally, a large set of proficiency data would be collected from every examiner who wishes to testify in court. However, as others have noted in the context of DNA evidence, this would require an impractical amount of testing. Joseph L. Peterson & R.E. Gaensslen, Developing Criteria for Model External DNA Proficiency Testing 109 (2001), available at http://www.ncjrs.gov/pdffilesi/nij/grants/192282.pdf (“At this stage, though, it is highly unlikely that declared or blind proficiency tests can be administered with sufficient frequency, relative to the quantity of most laboratories’ caseloads, to allow a meaningful calculation of the lab’s or any examiner’s ‘error rate.’”).

69. Peterson et al., supra note 61, at 23 (“In fact, it has been demonstrated that mandated proficiency testing enhances overall quality of clinical laboratory testing, including turnaround time, accuracy of results, and training of laboratories, whereas self regulation has been found less effective in achieving those goals.”).
here represent a significant change from current practices where proficiency test participation is voluntary and little is known about the characteristics of test-taking examiners. If the sampling of examiners is done properly, the resultant performance data may be generalized to the broader population of fingerprint examiners. Furthermore, by tracking examiner characteristics, we will gain insight into the conditions under which performance varies.

E. How Should the Results Be Scored?

After noting the key features of the sample prints and test situation four important rates should be computed: (1) the false positive error rate, (2) the false negative error rate, (3) the false discovery rate, and (4) the inconclusive rate. To avoid an illusion of precision, 95% confidence intervals around those error rate estimates should be presented in reports and explained at trial.

It will be a challenge to convey the significance of the different error rates to jurors. Previous research indicates that people struggle to understand the significance of probabilistic forensic science evidence, including the experts themselves. Further research is needed to identify how best to convey the probative value of fingerprint examination to jurors in light of the reported error rates.

70. Writing in the context of proficiency tests for firearms and toolmark examiners, Professor Adina Schwartz argues that “proficiency testing can (at most) establish an error rate for the particular people tested, not for firearms and toolmark examination as a whole.” Schwartz, supra note 52, at 25. Professor Schwartz is right if her comment is interpreted in the context of current proficiency test practices: if those who take proficiency tests are not representative of the forensic science population, then we cannot be confident that the resultant error rates are valid estimates for the whole industry. However, if participants are randomly sampled, a statistically sound extrapolation can be made.

71. Although the initial studies with well-constructed proficiency tests are likely to produce error rate data that are broadly applicable, the research should not and will not end there. Ultimately, the goal is to understand the conditions under which the risk of error is more and less likely and by how much. Characteristics of the examiner, prints, crime, and technology, are some of the variables that may matter. See Faigman, supra note 27, at 63 n.7 (“Since such errors can occur and have occurred, the questions move on to: what are the probabilities of such errors under varying conditions.”); Lawson, supra note 3, at 44 (“Two critical questions must, however, be tested: 1) what is the probability of an erroneous match, and 2) under what circumstances are errors more likely or less likely to occur. These questions have not been asked of fingerprint identification evidence and the lack of knowledge of the real answers to these questions is what makes the conclusions drawn from fingerprint analysis suspect and arguably unreliable.” (citations omitted)).

72. See supra Part III.A.

73. Some scholars recommend using the upper bound of the error rate’s 95% confidence interval as an estimate of the error rate rather than the estimated error rate itself. Lempert, supra note 15, at 453.

74. See generally Koehler, supra, note 4.

F. Is Not This Type of Proficiency Testing Program Too Expensive?

No. Professor Joseph Peterson and his colleagues provided a series of estimates for the cost of a double-blind proficiency test program for forensic DNA laboratories. They assumed participation from 150 laboratories per year, two tests per year, a 20% fringe benefits rate, and a ten-person oversight committee. The cost estimates ranged from $814,000 to $3,020,000.

Cost estimates associated with a similar program for fingerprint examination would almost surely be lower than those for DNA due to differences in the technologies and sample preparation protocols. Moreover, because many of the testing costs are fixed (a consideration not taken into account in the Peterson et al. DNA estimates), future program costs will likely decrease. A fingerprint proficiency testing pilot study would be useful to help pin down cost estimates for a broader program. But in light of how much is at stake and where there is broad agreement that existing tests are inadequate, it does not appear that a double-blind national proficiency testing program for fingerprint examination is impossible to conduct or too expensive to maintain.

Once we have accumulated a body of proficiency test data that meet the criteria described above, the scientific integrity of the forensic sciences will improve and the resultant error rate data should take on much greater significance in the courtroom.

Conclusion

The reliability and probative value of a reported fingerprint match is inextricably linked to the rate at which fingerprint examiners make errors. For this reason, it is important that legal decision makers have valid information about the risks of false identification and false nonidentification. Examiners' pseudo-scientific claims of absolute certainty must no longer fool us into thinking that the risk of error is negligible or nonexistent.

In discussing errors, error rates, and the design of proficiency tests that can assist with error rate estimates, I considered—and rejected—a variety of skeptical arguments. For example, I reject the claim that information about the training and experience of an examiner provides sufficient information about risk of error. I reject the claim that industrywide error rates obtained from well-constructed proficiency tests

76. Peterson et al., supra note 61, at 28-29.

77. An estimate extrapolated from the authors' project was at the low end, an estimate from a government agency test provider was at the high end, and an estimate from a commercial test provider fell between the other two estimates but was closer to the lower end ($1,050,000). Id. at 29 tbl.3.

78. See generally Thompson, supra note 15.

are irrelevant to estimates of the risk of error in specific cases. I reject the
claim that it is too difficult, expensive, or time consuming to implement a
mandatory program of double-blind proficiency tests administered by
disinterested parties.

And yet there is a kernel of truth in each of the rejected claims. An
examiner's training and experience is relevant to an assessment of the
risk of error. Other things being equal, a poorly trained examiner
probably is at higher risk of making a false identification than a well-
trained examiner. Likewise, a set of industrywide error rates cannot tell
us everything we wish to know about the risk of error in a specific case.
Testimony about contamination problems in a particular laboratory or
about the faintness of certain markings on a latent print may reasonably
inform a juror's subjective estimate of the chance of error in a particular
case. And surely the rigorous proficiency testing system that I envision
will require significant effort and will run into a variety of snags when it
is implemented across the nation.

But there is no contradiction between all of these rather obvious
concessions and the viewpoint put forward in this Article. Though the
idealized goal may be identification of the unique risk of error that
attends each and every case, insistence on achievement of this goal prior
to embracing a methodologically rigorous testing program belies a
misunderstanding of the purpose and value of scientific measurement.

Consider one of the most celebrated medical studies in modern
history, the Physicians' Health Study.80 This randomized, double-blind81
study found that people who took a daily dose of aspirin had a 44%
reduction in the risk of myocardial infarction relative to those who did
not take aspirin. The study was widely hailed as important, well-designed
and a model for future medical studies.83 Thousands of physicians have
recommended a daily aspirin regime to millions of patients and, as a
result, many lives have been saved. Yet it would not be hard for a critic
to poke gaping holes in this celebrated study. Consider the participants.
They were hardly a random sample of the general population: they were
male physicians over the age of forty who were in excellent health.83
Consider the lack of oversight in the administration of the daily
medication. Some participants may have skipped doses or taken some
other unreported medications on their own. Or consider that the study

80. Steering Comm. of the Physicians' Health Study Research Group, Final Report on the Aspirin
81. Neither the participants nor the investigators knew whether a participant was in the
experimental or placebo group. Group assignment was random.
82. See generally Videotape: Against All Odds: Inside Statistics, Program 12: Experimental
Design (The Consortium for Mathematics and its Applications and Chedd-Angier 1989), available at
83. Steering Comm. of the Physicians' Health Study Research Group, supra note 80, at 129.
provided no insight into dose effects: all participants in the experimental group took one 325 milligram tablet every other day. A skeptic could point to one or more of the study's imperfections and draw the mistaken inference that the study is irrelevant to the risk of myocardial infarction of any particular person. Such an inference would be an instance of what might be called the "imperfection fallacy" (i.e., a tendency to treat imperfect information as irrelevant).

Why do I mention this study and the imperfection fallacy here? Because a person who is motivated to dismiss test results—including those from a well-designed proficiency test—can always do so by embracing the test's inevitable imperfections. All proficiency tests will exclude some features that are relevant for assessing the risk of error in a particular case. Despite these imperfections, a well-designed test should still serve as an initial best estimate for the risk of error in a particular case. This case-specific risk can and should be adjusted if case-specific particulars that are not captured by the proficiency test data indicate that such adjustment is appropriate. But the anchor data provided by a properly designed proficiency test should not be discarded by any person, agency, scientific body, or court on account of an imperfect mapping of particulars any more than the results of the Physicians' Health Study should be discarded by those who are not male physicians in excellent health.

In closing, I note that one of the great ironies of the forensic science community's resistance to scientific testing to identify error rates is that identification of those rates may not reduce the perceived probative value of reported fingerprint matches. At present, jurors hear little about error rates, though some may be aware of cases in which examiners erred. If credible data show that all relevant error rates are miniscule, jurors may treat the data as evidence that bolsters the credibility of the reported match. Although such data would force examiners to acknowledge the risk of error, it might also go a long way toward improving the actual and perceived scientific status of fingerprint examination.

84. Id.
86. Another way to impress upon jurors that examiner testimony is opinion rather than fact is to require an instruction from the bench to this effect. Utah v. Quintana, 103 P.3d 168, 170, 171 n.3 (Utah Ct. App. 2004) (Thorne, J., concurring) ("[T]rial courts should be directed to instruct juries about the existing weaknesses of fingerprint examiner training and identification protocol. . . . [T]he jury should have been specifically instructed that the examiner's testimony was opinion and not fact and that the jurors should examine the fingerprint evidence independently. It is vital that we remove the near mystical awe that fingerprints evoke, and replace it with a more cautious regard for forensic evidence and its overall lack of certainty."); FAIGMAN, supra note 27, at 65 (suggesting that one solution to the problem of insufficient data to back fingerprint identification claims is a "judicial instruction of the jury about the limitations of the field"). The sufficiency of such judicial instruction is unknown.