



Perceptions and estimates of error rates in forensic science: A survey of forensic analysts

Daniel C. Murrie^{a,*}, Brett O. Gardner^a, Sharon Kelley^a, Itiel E. Dror^b

^a Institute of Law, Psychiatry, and Public Policy, University of Virginia, United States

^b Center for the Forensic Sciences, University College London, United Kingdom

ARTICLE INFO

Article history:

Received 19 February 2019

Received in revised form 4 July 2019

Accepted 20 July 2019

Available online 29 July 2019

Keywords:

Forensic science

Error rates

Bias

Forensic decision making

Cognitive and human factors

Expert judgment

ABSTRACT

Every scientific technique features some error, and legal standards for the admissibility of scientific evidence (e.g., *Daubert v. Merrill Dow Pharmaceuticals, Inc.*, 1993; *Kumho Tire Co v. Carmichael*, 1999) guide trial courts to consider known error rates. However, recent reviews of forensic science conclude that error rates for some common techniques are not well-documented or even established (e.g., NAS, 2009; PCAST, 2016). Furthermore, many forensic analysts have historically denied the presence of error in their field. Therefore, it is important to establish what forensic scientists actually know or believe about errors rates in their disciplines. We surveyed 183 practicing forensic analysts to examine what they think and estimate about error rates in their various disciplines. Results revealed that analysts perceive all types of errors to be rare, with false positive errors even more rare than false negatives. Likewise, analysts typically reported that they prefer to minimize the risk of false positives over false negatives. Most analysts could not specify where error rates for their discipline were documented or published. Their estimates of error in their fields were widely divergent – with some estimates unrealistically low.

© 2019 Elsevier B.V. All rights reserved.

All scientific procedures inevitably involve some error, particularly when the procedures rely on subjective human judgment. Understanding the nature and extent of error is crucial to understanding the meaning of any particular piece of scientific evidence. Thus, most scientific disciplines work to document the reliability (i.e., consistency, reproducibility) and validity (i.e., accuracy) of their scientific procedures.

For scientific evidence to be admitted in court, data regarding reliability and validity are crucial. In the United States, landmark Supreme Court decisions *Daubert v. Merrill Dow Pharmaceuticals, Inc.* [1] and *Kumho Tire Co. v. Carmichael* [2], direct federal trial courts (and state trial courts that adopted this framework) to evaluate whether “an expert’s testimony pertain[s] to ‘scientific knowledge’” as articulated in the Federal Rule of Evidence 702 ([1], p. 590). The *Daubert* Court acknowledged that many factors are relevant in determining whether evidence derived from a particular technique qualifies as “scientific knowledge,” but offered criteria that courts could consider. One enumerated criterion is the “known or potential rate of error” (p. 594).

Error rates are a crucial component in assessing the reliability and validity of a method purporting to produce scientific findings.

A fact-finder’s ability to properly interpret the conclusions of any scientific analysis depends upon the ability to know the chance that an error occurred. And, although the discussion above concerns admissibility criteria in the United States, other countries have similar guidelines for admitting scientific evidence [5,6]. Errors are especially important in the criminal justice system because they are not isolated within an investigation. Indeed, they often *cascade* and *snowball* to influence numerous other aspects of the investigative and legal proceedings [7].

1. What do we know about error rates in forensic science?

What do we know about reliability and error rates of popular forensic science techniques admitted in court? Historically, many forensic scientists adamantly denied that there are any errors in their field (e.g., [8,9]). In a well-known example, in 2004, the FBI laboratory supported their “conclusion of a 100% positive identification” that Brandon Mayfield was the source of latent prints implicating him as the Madrid train bomber [10]. However, this identification by multiple FBI latent fingerprint experts was erroneous [11]. Landmark reviews (e.g., National Academy of Sciences [3] and President’s Council of Advisors on Science and Technology [PCAST] [4]) have since noted that such statements implying 100% certainty in forensic conclusions are not scientifically defensible. Other reviews have documented that errors in forensic science analyses clearly exist and have significant

* Corresponding author at: Institute of Law, Psychiatry, and Public Policy, University of Virginia, Box 800660, Charlottesville, VA 22908, United States.
E-mail address: Murrie@virginia.edu (D.C. Murrie).

consequences. For example, a review of DNA exoneration cases found that 63% of erroneous guilty verdicts were caused, in part, by forensic science testing errors [9]. But acknowledging that errors occur is not the same as documenting rates of error, or reliability of techniques.

Indeed, a primary critique from the 2009 National Academy of Sciences [NAS] report was that the field of forensic science must better study, document, and acknowledge the reliability (including error rates) of forensic science techniques. Specifically, the report noted that “wide variability is found across forensic science disciplines not only with regard to techniques and methodologies, but also with regard to reliability, error rates, reporting, research foundations, general acceptability, and published material” (p. 188). Thus, the authors concluded that “research is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines” (p. 190).

Perhaps due to the NAS report and other recent reform efforts, some forensic science disciplines have begun to publish research that sheds light on error rates. A primary example is latent print analysis. The 2016 PCAST report noted that, “In response to the 2009 NRC report, the latent print analysis field has made progress in recognizing the need to perform empirical studies to assess foundational validity and measure reliability” (p. 87–88). However, despite the progress toward establishing error rates for latent print analysis, the report also said,

Remarkably, there have been only two black-box studies that were intentionally and appropriately designed to assess validity and reliability—the first published by the FBI Laboratory in 2011 [Ulery, Hicklin, Buscaglia, & Roberts, 2011]; the second completed in 2014 but not yet published [Pacheco, Cerchiai, & Stoiloff [12]]. Conclusions about foundational validity thus must rest on these two recent studies. (p. 91)¹

After reviewing all available literature, the PCAST report ultimately concluded that the “best estimates” of false positive error rates in latent print examinations are 1 in 604 cases (with an upper bound of 1 in 306) and 1 in 24 (with an upper bound of 1 in 18), derived from the 2011 FBI study and the 2014 Miami-Dade study, respectively.

For the most part, there has been similarly scarce published research on error rates in other forensic science disciplines. For example, the 2016 PCAST report identified only one appropriately designed study regarding firearm analysis (i.e., an unpublished report to the Federal government completed by the Ames Laboratory; [13]), noting that the rate of false positive errors was estimated to be 1 in 66, with an upper bound of 1 in 46. Studies examining error rates in handwriting analysis have typically suggested that analysts offer erroneous conclusions approximately 40% of the time, although error rates vary widely [14]. Few studies have examined the accuracy of bitemark analysis but the scant research shows similarly high error rates, with 64% of analysts making a false positive error and 22% of examiners making a false negative error across four cases in one study [14]. As a result, several experts have concluded that analysts are currently unable to reliably identify or exclude a suspect based on bitemark analysis [15]. Finally, the error rate among identification conclusions in microscopic hair comparisons was 11% when using DNA analysis results to determine accuracy [16]. The above studies were ultimately published, though other studies, with less favorable

results, have not made their way to the public domain and remain unpublished.

In summary, few published studies have attempted to estimate error rates in various forensic science disciplines and, among those that do, poor ecological validity severely limits the generalizability of results because such studies do not typically resemble “real-world” casework. Nevertheless, this slim published research base indicates that many forensic analysts could, at least in theory, identify at least one published study in their discipline that presents some estimate of error rates. Error rates depend on the validity (i.e., accuracy) of their scientific procedures, but more basically they depend on the reliability (i.e., consistency, reproducibility) of their results [17]. With respect to reliability, two of the more robust forensic domains show concerning data. Fingerprint data suggests a lack of reliability in 10% of conclusions. In other words, the same examiner, looking at the same pair of fingerprints, will reach a different conclusion 10% of the time [18]. A past study examining the reliability of DNA analyses showed some inconsistency in conclusions [19], and this has recently been replicated by a number of studies showing great variability in DNA results when analyzing the same sample (e.g., some included alleles that were absent, others neglected alleles that were present) further indicating that errors in DNA do exist [20–22]. Although DNA analysis is widely regarded as the prevailing standard of forensic science [3], “we do not know any more about human or laboratory error rates in DNA cases than we know about human or laboratory error rates in cases that employ other forensic science techniques” ([23], p. 3). Sweden’s forensic laboratory described general rates of “contamination” and DNA laboratory failures such as lost samples or instrument-related faults, but their data do not speak to error rates in erroneous conclusions [24]. The Netherlands Forensic Institute also reported general DNA laboratory error rates (e.g., contamination, technical errors, sample mix-ups) and found that such errors were infrequent (i.e., less than 1% of all analyses) and often involved human error [25]. Furthermore, analysts may overstate the certainty of their findings or the rarity of error. Nevertheless, there continues to be a lack of published data detailing the rate of erroneous conclusions in DNA analyses or any other forensic science analyses.

2. Types of error

When discussing error in forensic science conclusions, it is important to distinguish between three types of error. Analysts can ultimately conclude that the results of their analysis suggest an identification or match between two sources (e.g., a latent print originated from same source as a known print), that the results suggest an exclusion (e.g., a latent print did *not* originate from same source as a known print), or that the results are inconclusive (e.g., that the clarity of the fingerprints, the quantity and quality of information, does not permit an identification or an exclusion decision). As depicted in Table 1, a true positive occurs when analysts correctly offer an identification, a true negative occurs when analysts correctly exclude a source, and a true inconclusive occurs when there is not enough information to draw any other conclusion. A false positive error occurs if an analyst concludes a match between two items when, in fact, they originated from different sources. Conversely, a false negative error occurs if an analyst concludes that the fingerprints cannot have come from the same source, when, in fact, they came from the same source. Lastly, a false inconclusive error occurs if an analyst concludes inconclusive, when there is actually sufficient information to conclude an identification or exclusion [26].

The consequences of false positive and false negative errors vary significantly according to the context of the decision-making process, a calculation labeled *the payoff matrix* [27]. For example, in

¹ “By a “black-box study,” we mean an empirical study that assesses a subjective method by having examiners analyze samples and render opinions about the origin or similarity of samples” (PCAST [4], p. 48). For further detail on black-box studies and the reasons these are crucial to gauging validity, see the PCAST [4] report in full.

Table 1Error types in forensic science conclusions.^a

Analyst's conclusion	Ground truth	
	Identification (Same source)	Exclusion (Different source)
Identification	Correct identification (True Positive)	False Positive
Exclusion	False Negative	Correct Exclusion (True Negative)
Inconclusive	False Inconclusive	False Inconclusive

^a Assuming that there is sufficient information in the fingerprints to make an identification or exclusion decision. If there is insufficient information to reach such conclusions, then all decisions, except inconclusive, are erroneous.

smoke detectors, false positive errors (i.e., an alarm sounding when there is no smoke) are associated with more minor consequences than are false negative errors (i.e., an alarm does not sound in the presence of smoke). In the criminal investigation process, the reverse is often true. A false positive error in forensic science conclusions typically results in criminal charges against an innocent individual whereas false negative errors result in guilty individuals avoiding legal charges. Determining which type of error is more problematic presents a moral judgment with no objective answer. However, most analysts work in an environment that considers false positive errors worse than false negative errors [28]. Indeed, some scholars assert that analysts should explicitly lean towards minimizing false positive errors as the United States criminal justice system values human life and freedom [29]. Within many forensic domains, inconclusive decisions are not even regarded as errors, and therefore are a very attractive option for many forensic analysts [26].

3. Do proficiency tests provide “known error rates”?

Forensic analysts' performance on proficiency tests are sometimes cited as revealing error rates in routine casework, perhaps because they can be examined for agreement or disagreement. For example, approximately 12% of latent print examiners who completed and submitted one of the seven prior Collaborative Testing Services (CTS) proficiency tests dating back to 2014 did not correctly identify all latent prints on the exam.² These data suggest that examiners do not universally agree on all conclusions, contrary to some examiners' assertions [30]. However, the providers of proficiency testing emphasize that test results should not be considered “error rates.” For example, on the front page of all summary reports, CTS [31] states:

This report contains the data received from the participants in this test. Since these participants are located in many countries around the world, and it is their option how the samples are to be used (e.g., training exercise, known or blind proficiency testing, research and development of new techniques, etc.), the results compiled in the Summary Report are not intended to be an overview of the quality of work performed in the profession and cannot be interpreted as such . . . These comments are not intended to reflect the general state of the art within the profession. (p. 1)

Furthermore, as currently implemented, there are several important differences between proficiency testing and real-world casework. Forensic analysts are distinctly aware that they are being tested; it is unclear whether most respondents work alone, in groups, or under close supervision while completing proficiency

tests. Further, some have argued that proficiency testing in latent print examination may be unrealistically simple and may not adequately simulate actual casework (e.g., [8,23]). Some scholars have thus called for a revision of the proficiency testing process to allow for calculation of actual error rates [32]. Currently, however, it is likely that proficiency testing only informs estimates of the lower bounds of error rates in routine casework because “laboratory technicians can take special care when they know they are dealing with test samples” ([8,33], p. 447).

4. Public perception of forensic science accuracy and error rates

Of course, the sparse empirical data examining forensic science error rates does not preclude *opinions* regarding the accuracy of forensic science techniques. Several studies have explored public perception of errors and reliability in forensic science, with varied results. Lieberman et al. [34] found that undergraduate students and former jurors provided accuracy estimates for various forensic science techniques ranging from 78% (“alcohol and drug tests”) to 95% (DNA evidence). In contrast, a recent survey of Australian community members revealed high estimates of the chances of an error occurring during forensic science testing ($M=39\%$ chance of error) and analysis ($M=45\%$; [35]). Regarding latent print analyses specifically, another study found that 20% of laypersons believed fingerprint evidence was “somewhat reliable,” 51% believed such evidence was “reliable,” and 26% believed such evidence was “very reliable” [36]. The same study also surveyed defense attorneys, who were more skeptical: only 39% believed fingerprint evidence to be “somewhat reliable,” 23% viewed such evidence as “reliable,” and only 6% viewed such evidence as “very reliable.”

There is little research examining beliefs about how often false positive and false negative errors occur within forensic science analyses. Members of a jury pool estimated the false positive error rate in DNA analyses to be one in one million, but the researchers did not ask participants to estimate the prevalence of false negative errors [37]. Similarly, Koehler [23] asked jury-eligible community members to estimate the false positive rate in five forensic science disciplines using a 14-level logarithmic scale. Participants' median estimates were all very low, with average responses ranging from 1 in 10 million for DNA analyses, 1 in 5.5 million for latent print analyses, 1 in 1 million for bitmark analysis and microscopic hair analysis, and 1 in 100,000 for handwriting analysis.

Taken together, the limited literature suggests that laypersons and lawyers perceive that forensic science analyses are prone to error and that error rates vary by analysis type, but also that people vary in their perceptions regarding how often such errors occur. Moreover, there is very little information regarding perceptions of false negative error rates in forensic science analyses. Most importantly, none of the existing research includes the forensic examiners themselves.

5. Statement of purpose

We do not know how often forensic science analysts arrive at erroneous conclusions in their casework. Furthermore, we know very little about how forensic analysts approach and estimate error in their discipline. Research reveals that laypersons and lawyers hold varied beliefs regarding the accuracy of forensic science techniques, but no study has examined the opinions of those who actually conduct such analyses and communicate the results: i.e., the forensic analysts.

It is especially important to explore what forensic analysts think about error rates. First, the forensic analysts are those who communicate their results to the fact finder (e.g., in court to the jurors or judge, or earlier in the process, to investigators and attorneys). Hence, it is important to understand what analysts

² Data available at: <https://cts-forensics.com/program-4.php>. CTS proficiency exams typically comprise 11 or 12 latent prints.

consider their error rates to be, as this clearly determines how they communicate their results. Second, changes and improvements in forensic science can only be motivated by acknowledging errors; if a domain is considered error-free, then the motivation to change or improve is minimal. The present study is the first to examine the perception of error rates among forensic analysts themselves.

The present study explored forensic analysts' estimates of error rates in their disciplines and whether demographic characteristics (e.g., age, gender) influenced such perceptions. This study also investigated how forensic analysts weigh the risk of false positive and false negative errors in their work. Understanding whether examiners adopt a more conservative or liberal decision criterion when making conclusions, and whether this is consistent across examiners and laboratories, has the potential to inform training and promote reliability across examiners. Finally, this study asked forensic analysts how they knew the "known error rates" in their disciplines, that is, where they believed error rates to be published or otherwise available. This final question is important because legal standards of admissibility (e.g., [1]) advise judges to consider the "known or potential rate of an error" and forensic analysts are typically responsible for communicating conclusions and information about error rates to the court.

6. Method

6.1. Participants

Participants were 189 practicing forensic analysts within the United States. We removed six participants from analyses due to either unintelligible and/or incomplete responses on the survey ($n=2$), or because they identified a discipline outside of forensic science (e.g., administration, "police officer"; $n=4$). Thus, our final sample comprised 183 forensic analysts. Most participants were women ($n=118$, 65%) and their average age was 39 years ($SD=8.9$; range=24 to 74). Most participants described themselves as Caucasian ($n=96$; 53%) or Asian/Pacific Islander ($n=44$; 24%), with fewer endorsing Hispanic/Latino ($n=19$; 10%), African American ($n=4$; 2%), or unspecified ($n=2$; 1%) ethnicities. Other participants identified as Multiracial ($n=6$; 3%) or declined to provide an ethnicity ($n=12$; 6%).

Of participants who responded to the demographics survey, the vast majority held a Master's degree ($n=86$; 49%) or Bachelor's degree ($n=76$; 43%). Few participants held a doctoral degree ($n=9$; 5%) or completed less than three years of college ($n=4$; 2%). Regarding work experience, participants typically reported working 9.9 years ($Mdn=8$; $SD=7.4$; range=0.5 to 34) in their primary discipline.

Participants endorsed a variety of forensic science disciplines. To facilitate comparisons among disciplines, we categorized participants according to four broad disciplines that were included in analyses: Biology ($n=84$; 46%), Pattern Evidence (i.e., latent print analysis, firearm and tool mark analysis, questioned documents, and trace evidence; $n=43$; 24%), Chemistry (i.e., forensic, drug, and environmental chemistry, toxicology; $n=32$; 18%), and Crime Scene Investigation ($n=12$; 7%).³ We categorized a small subset of participants' primary discipline as other (e.g., environmental forensics, anthropology; $n=6$; 3%) and some participants did not identify a primary discipline ($n=6$; 3%). More than one third of participants ($n=73$; 40%) endorsed multiple disciplines

within forensic science and, of these participants, most identified crime scene investigation ($n=38$; 52%) as an ancillary discipline.

6.2. Measures

6.2.1. Procedure

Participants completed a three-part survey addressing topics relating to (1) attitudes regarding the acceptability of error types, (2) estimates of error rates within forensic science, and (3) perceptions of task relevance regarding available case information. For the current study, we present results from the first two parts of the survey (for additional detail regarding part three, see Ref. [42]). Before asking questions about perceptions or estimates, the survey provided a brief education on false positive and false negative errors. We defined false positive errors as "the error of determining that evidence does support a match or identification, when a match or identification does not truly exist," and false negative errors as "the error of determining that evidence does not support a match or identification, when a match or identification truly does exist."

6.2.2. Survey

In part one of the survey, participants indicated how they "weigh and prioritize the risk of each type of error" on an 11-point scale ranging from 0=*I minimize the risk of false positives, and therefore tolerate greater risk of false negatives*, to 10=*I minimize the risk of false negative, and therefore tolerate greater risk of false positives*. The midpoint of the scale (i.e., 5) noted, *I accept equal risk of false positives and false negatives*. Participants then completed similar questions indicating how they believed their laboratory, and discipline as a whole, weigh and prioritize the risk of each type of error.

In part two of the survey, we asked participants to provide separate estimates of the false positive and false negative error rates in their discipline. We used the logarithmic scale first published by Thompson and colleagues [37] and later used by other researchers (e.g., [23]) to provide participants with 14 possible error rates ranging from "approximately 1 time in 2," to "such an error is impossible." After estimating the error rates in their discipline, participants were asked where the error rates for their discipline were published or documented. They could either identify a source for their error rate estimates or indicate that they did not know of any source documenting known error rates in their discipline.

6.2.3. Demographics form

A standard demographics form asked for participant age, gender, and ethnicity. We also asked participants to identify their primary discipline, as well as any ancillary disciplines. Additionally, the form asked participants to indicate how many years they had worked in their primary discipline and their highest educational degree.

7. Results

7.1. Analysis strategy

We used SPSS (version 24) to perform all analyses. We assessed participants' perceptions of error rates using a 14-point scale with responses ranging from "approximately 1 time in 2" (coded as 1 for analyses) to "such an error is impossible" (coded as 14 for analyses). Thus, higher scores on this variable indicated lower estimates of error rates. We used non-parametric analyses (i.e., Spearman's ρ) to treat this variable as ordinal because the data were not equally distributed. We similarly used non-parametric analyses (i.e., chi square test of independence) to evaluate potential differences

³ We could not obtain comparison data detailing the disciplines or demographics of forensic analysts within crime laboratories. However, the most recent census of publicly funded crime laboratories in the United States [41] indicates that analyses relating to controlled substances (81% of all laboratories), latent prints (63%), and biology (62%) are the most common among crime laboratories.

between participants who provided a source for their error rate estimates and those who did not on categorical variables (e.g., discipline). Finally, we used parametric analyses to evaluate the association between participants' ratings of the acceptability of error types and other variables of interest. Specifically, we conducted a mixed-factor analysis of variance (ANOVA) to assess whether participant ratings differed according to primary discipline and ratings subject (i.e., self, workplace, discipline). We also conducted Pearson product-moment correlations to assess the association between participant ratings and other continuous variables such as years of work experience.

7.2. Estimated error rates

Overall, participants estimated that false positive errors were less frequent than false negative errors in their respective disciplines, although estimates of false positive and negative errors were highly correlated (Spearman's ρ [116] = .67, $p < .001$). Analysts most commonly estimated that false positive errors occurred at a rate of 1 in 10,000 and that false negative errors occurred at a rate of 1 in 100. Approximately half of analysts ($n=60$; 50%) indicated that the rate of false positive errors fell between 1 in 1,000 and 1 in 1,000,000. However, 38% of analysts

($n=46$) indicated that the rate of false positive errors was equal to, or less than, 1 in one billion. Moreover, 10% of analysts ($n=12$) stated that false positive errors were not possible.

Regarding false negative errors, half of analysts ($n=60$; 50%) indicated that the rate of false positive errors fell between 1 in 100 and 1 in 10,000. Almost one quarter of analysts ($n=26$; 22%) indicated that the rate of false negative errors was equal to, or less than, 1 in one billion. Additionally, 6% of analysts ($n=7$) stated that false negative errors were not possible.

Participant estimates of error rates varied according to their primary discipline. As shown in Fig. 1, analysts in forensic biology generally provided lower errors rates than did analysts in other disciplines, such as pattern evidence. However, there was significant variability within all disciplines, with estimates for identical error types sometimes ranging from as high as "1 in 2," to as low as "less than 1 in quadrillion."

Interestingly, older analysts estimated that both false positive (ρ [111] = -.28, $p = .003$) and false negative (ρ [110] = -.28, $p = .003$) errors occurred more frequently than did younger analysts. This pattern may be partially explained by age differences among forensic disciplines. Specifically, the average age of forensic biologists (who generally provided lower error estimates) was 37.4 ($SD=6.7$) whereas pattern evidence analysts were typically 41.2

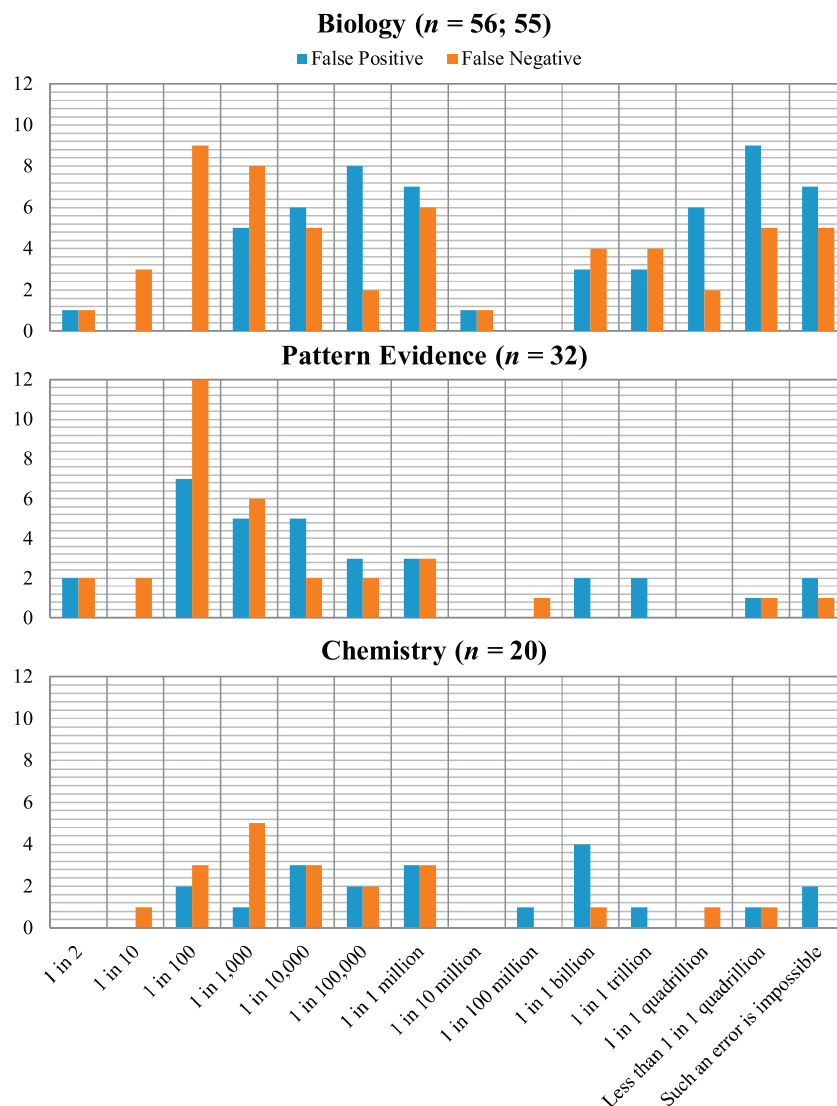


Fig. 1. Estimated error rates within primary discipline.

years old ($SD=9.9$). Although the trend persisted, the association between analyst age and error rate estimates was not significant solely among forensic biologists or chemists. Older analysts in pattern evidence provided higher estimates of false negative errors ($\rho[30] = -.42, p=.02$) but not false positive errors ($\rho[30] = -.30, p=.10$).

7.2.1. Source of error rate estimates

When asked to provide a source⁴ for their estimated error rates, most analysts ($n=144$; 78.7%) were not able to provide a source. Among the minority who provided a source, 46.2% ($n=18$) of analysts cited their personal work or training experiences as evidence for their estimated error rates. Fewer analysts ($n=12$; 30.8% of those who provided any source) referenced a scientific journal or study, whereas 10.3% ($n=4$) cited proficiency testing data (e.g., CTS examinations). An additional 12.8% ($n=5$) of analysts reported a personal belief as justification for their error rate estimations (e.g., "I just believe not enough labs take appropriate QA measures when it comes to exclusions/inconclusives") or described laboratory procedures. In total, three quarters (75.3%) of analysts who directly responded to our source questions, and two-thirds (66.7%) of all surveyed analysts, explicitly indicated that they did not know of a source for error rates in their discipline.

The proportion of analysts who explicitly indicated that they were unaware of an error rate source varied significantly by discipline, $\chi^2(3, N=153)=9.13, p=.03$. As demonstrated in Fig. 2, of analysts who responded, more biologists (85.7%) indicated that they were unaware of a source for error rates than did other disciplines (63.2–77.8%). Moreover, female analysts (81.0%) were more likely to indicate that they were unaware of an error rate source than were male analysts (64.2%), $\chi^2(1, N=158)=5.35, p=.02$, though (as detailed earlier) the sources male analysts referenced were often nonspecific.

7.3. Attitudes regarding the acceptability of error types

Overall, analysts indicated that they ($M=3.48$; $SD=2.22$), their workplace ($M=3.57$; $SD=2.33$), and their discipline ($M=3.62$; $SD=2.22$) preferred to minimize the risk of false positive errors and therefore tolerated a greater risk of false negative errors. However, there was substantial variability in attitudes toward error types as evidenced by the relatively high standard deviations and responses ranging from 0 = *I minimize the risk of false positives, and therefore tolerate greater risk of false negatives*, to 10 = *I minimize the risk of false negative, and therefore tolerate greater risk of false positives*.

We conducted a 3 (Subject: self, workplace, discipline) \times 4 (Primary discipline: biology, crime scene investigation, chemistry, pattern evidence) mixed-factor ANOVA to assess (1) whether analysts described their personal attitudes as different from the attitudes held by their workplace and discipline, and (2) whether analysts' attitudes regarding error types varied according to their primary discipline. Analysts did not describe their attitudes as different than the attitudes held by their workplace or discipline, $F(2, 310)=0.30, p=.75, \eta_p^2=.002$. However, the between-subjects effect of primary discipline on analyst attitudes approached significance, $F(3, 155)=2.53, p=.06, \eta_p^2=.05$. More specifically, crime scene investigators endorsed a stronger preference to minimize the risk of false negative errors than did analysts from all other disciplines, $F(3, 162)=3.73, p=.01, \eta_p^2=.07$ (a finding that

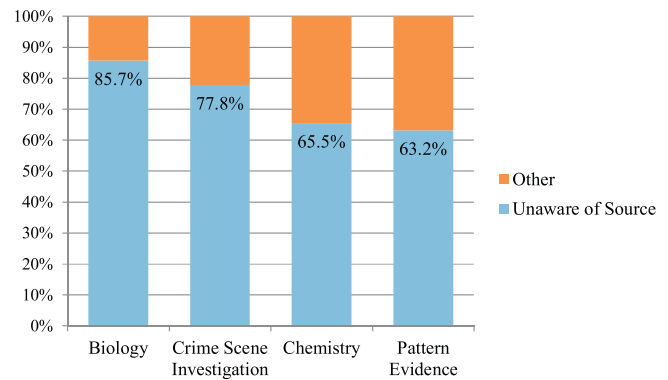


Fig. 2. Percentage of analysts explicitly indicating unawareness of error rate source within primary discipline.

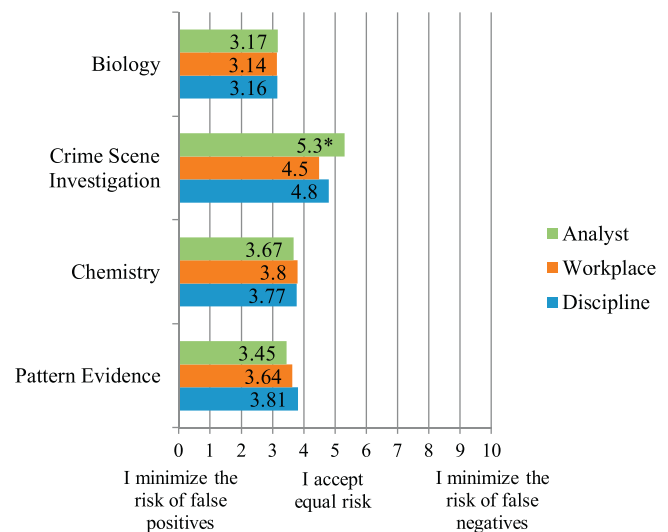


Fig. 3. Attitudes regarding the acceptability of error types. * denotes value is significantly different from other values depicted in same-colored bars at $p < .05$.

seems to reflect the different duties of each discipline). Thus, although analysts from different disciplines did not perceive their workplaces or disciplines as having contrasting attitudes regarding the acceptability of error types, crime scene investigators endorsed a stronger personal preference to minimize false negative errors and therefore tolerated a greater risk of false positive errors than did other disciplines. Fig. 3 reports analysts' average attitudes regarding error acceptability according to their primary discipline.

There were no significant attitudinal differences between male and female analysts, and analyst age was unrelated to reported attitudes.⁵ Analyst age and work experience are highly correlated constructs but are not redundant; almost half of the variance in one variable is not explained by the other ($r[165]=.74, p<.001$). Further, unlike age, work experience was unrelated to analyst gender or primary discipline in the current sample. Analysts with greater work experience expressed a greater preference to minimize false positive errors and therefore tolerate a greater

⁴ The exact question was: "What is the basis of your reported error rates in questions 1 and 2? Please be as specific as possible (i.e., if your response comes from a specific publication or document, please identify that publication or document). Imagine that you were being asked this question in court and had to convey to the court exactly what source of information you used."

⁵ Although not significant, male analysts reported a stronger preference to minimize the risk of false positives ($M=3.0$; $SD=2.2$) than did female analysts ($M=3.7$; $SD=2.2$; $p=.053$; $d=.32$). They did not differ in their perceptions of their workplace and discipline's efforts to prevent false positive/negatives. Perhaps relatedly (as men were older than women in the current sample), an insignificant trend emerged in which older analysts emphasized the minimization of false positive errors, $r(164)=-.14, p=.07$.

risk of false negative errors in their work ($r[164] = -.23, p = .003$). Those with more work experience also believed that their lab ($r[167] = -.19, p = .01$) and discipline ($r[161] = -.19, p = .02$) placed a greater emphasis on minimizing the risk of false positive errors.

8. Discussion

Although forensic science disciplines do not have well-established error rates [23], the few studies that have been published present widely divergent error rates across forensic science disciplines, with false positive error rates ranging from 0.1% in latent fingerprint analyses [38] to 64.0% in bitemark analyses [14]. False negative error rates in these studies are similarly wide-ranging, with estimates ranging from 7.5% (latent fingerprint analyses) to 22% (bitemark analyses). The current findings are the first to explore *what the forensic analysts themselves believe* about error rates in their field. Analysts most commonly estimated that false positive errors occurred at a rate of 1 in 10,000 (i.e., a 0.01% error rate) and that false negative errors occurred at a rate of 1 in 100 (i.e., a 1.0% error rate). These rates appear lower than even the smallest error rates reported in the literature. Further, more than one in five examiners estimated the risk of each type of error to be impossible or extremely low (i.e., equal to or less than 1 in one billion). This is concerning in light of (a) the inevitability of error in human decision making, (b) human vulnerability to cognitive bias, and (c) the small body of available empirical research suggesting actual error rates are likely much higher (all factors emphasized in the landmark reviews of forensic sciences in the U.S. [3,4]).

Daubert et al. [1] governs admissibility decisions in many jurisdictions and explicitly advises judges to consider the “known or potential rate of an error” when admitting a particular technique as evidence. However, analysts provided widely divergent estimates of error in their fields – with some estimates impossibly low – and most analysts could not identify where error rates for their discipline were published or otherwise available. Indeed, among the minority of analysts who did identify a source (21.3% of analysts surveyed), most cited personal experiences or beliefs (59.0%) as opposed to scientific publications or studies. Put differently, of all analysts that were asked to provide a source for their estimated error rates, only 6.6% made any reference to a scientific journal or study. Forensic analysts are responsible for conducting scientific analyses and communicating subsequent conclusions; they are usually the expert expected to convey information about error rates to the court. But in our sample, less than 7% could provide the kind of error rate information that many standards of legal admissibility (e.g., *Daubert*) require.

8.1. Analyst vs. layperson error estimates

Fig. 1 suggests that analyst opinions varied widely, such that they do not hold a consensus view of error rates in their disciplines. Koehler [23] used an identical response scale to ask jury-eligible laypersons what they believed about the false positive error rates in various forensic science techniques. Whereas he found that laypersons gave a median estimate of a 1-in-10-million false positive error rate for DNA analyses, the forensic biology analysts in the present study typically estimated a 1-in-100-million false positive error rate. Similarly, laypersons estimated a 1-in-5.5-million error rate for latent print analyses while the seven latent print analysts who provided responses in this study gave a much lower median false positive error rate: 1 in 1 billion. Koehler's [23] findings suggested that laypersons were quite optimistic in estimating the accuracy of forensic science techniques. But, by our comparisons, those laypersons shared less optimistic (and more accurate, according to the limited data available) perceptions of error rates than the analysts we studied, who actually perform

the techniques and conveyed even more optimistic estimates of exceedingly low error rates.

8.2. Attitudes regarding error types

In general, analysts reported that they, their workplace, and their discipline prefer to minimize the risk of false positive errors and thus tolerate a greater risk of false negative errors. Similarly, analysts' error rate estimates also indicated a belief that false positive errors are less prevalent than false negative errors across disciplines. These findings are consistent with scholars asserting that analysts tend to minimize false positive errors in their work [28,29]. Ulery and colleagues [38] also noted that more information is typically required to conclude that two items do originate from the same source than to conclude two items do *not* originate from the same source, thus making false positive conclusions rarer than false negative conclusions.

More experienced analysts reported an even greater preference to minimize false positive errors. Approximately 5% of the variance in personal attitudes regarding error acceptability can be explained by years of work experience, a small-to-moderate effect. Perhaps extended exposure to work environments that warn false positive errors are more serious than false negative errors [28] influence analyst attitudes over time. Additionally, or alternatively, perhaps extended experience with a criminal justice system that expresses an explicit preference to avoid false positive errors pulls analysts to minimize such errors. In any case, it appears that analysts become slightly more conservative in their approach to error with experience.

Finally, crime scene investigators appeared distinct from other disciplines in their attitudes toward the acceptability of error types. They held a more balanced view regarding errors than other disciplines and their responses typically indicated that they accepted equal risk of false positive and false negative errors. This distinction does not appear surprising given crime scene investigators' tendency to consider more information relevant to their analyses than other disciplines [42]. Moreover, they often conduct analyses earlier in the investigation process, when identifying potential suspects may be equally important to excluding innocent individuals. Crime scene investigators may also rely on different decision rules than other analysts, or have access to additional information that differentially informs similar decision rules. For example, where some analysts may prefer to make decisions that minimize the maximum loss scenario (e.g., avoid incriminating the innocent), others may rely more heavily on the probabilities associated with specific outcomes (for a detailed discussion of forensic identification decision making, see [39]). Alternatively, two analysts may hold similar decision rules, but one may have access to additional information (i.e., less uncertainty) in the decision making process that influences behavioral tendencies. In sum, crime scene investigators may hold different beliefs regarding the perceived consequences and probabilities of erroneous decisions, but they may also have additional information that informs their decision-making compared to other analysts.

8.3. Limitations

There are several limitations when considering results. As Koehler [8] stated, “there is no agreed-upon answer to what counts as an error” (p. 1080). Although we provided explicit definitions of false positive and false negative errors, it is possible that some amount of the variability in analyst opinions can be explained by differences in what they consider an error. For example, some analysts may disagree regarding whether incorrect inconclusive opinions (e.g., an analyst gives an inconclusive determination

when the two items actually did result from the same source) constitute formal errors. There are also multiple methods to calculate and report error rates; thus, it is important to ensure results describe comparable statistics procedures before contrasting error rates reported in the literature.

A strength of this study was that participants were actual forensic examiners and that they came from several different disciplines. However, a weakness of the study was that we did not have particularly large samples from any single discipline, making it difficult to assess differences among disciplines, or to generalize results to a particular sub-discipline within forensic science. Many analysts in the current study indicated that they worked in forensic science disciplines beyond their primary discipline. It is possible that participation in multiple disciplines influences analyst beliefs and attitudes such that those with varied backgrounds approach error differently than those with experience only in one discipline. Finally, accurately describing attitudes and decision-making criteria can be difficult because these processes often occur without our complete awareness (e.g., [40]). Analysts who express a preference to minimize false positive errors may, in practice, make more balanced conclusions. Follow-up research examining analyst beliefs, and patterns in actual conclusions, across additional forensic science disciplines would shed further light on this important area of study.

9. Conclusion

No forensic science discipline has well-established error rates (disciplines are beginning to examine error rates, but some of these data are not published). Nevertheless, people hold intuitive opinions about the accuracy of forensic techniques, and these opinions are used to make important legal decisions. Most analysts expressed a preference to minimize false positive errors and, likewise, estimated false positive errors to be less common than false negative errors. However, they varied considerably in their estimates. A significant number of analysts provided unrealistically low estimates, and the vast majority of analysts could not identify a source from which they *knew* estimated rates of error. Overall, results highlight the need for further work investigating and disseminating error rates as a step toward improving forensic science, as well as informing policy and practice, guarding against overconfidence, and complying with legal standards for the admissibility of scientific evidence.

Funding

This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.forsciint.2019.109887>.

References

- [1] Daubert v. Merrill Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
- [2] Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
- [3] National Academy of Sciences, Strengthening Forensic Science in the United States: A Path Forward, National Academies Press, Washington, DC, 2009.
- [4] President's Council of Advisors on Science and Technology, Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods, Executive Office of the President of the United States, Washington, DC, 2016.
- [5] G. Edmond, The admissibility of forensic science and medicine evidence under the Uniform Evidence Law, *Crim. Law J.* 38 (2014) 136.
- [6] G. Edmond, J. Tangen, R. Searston, I.E. Dror, Contextual bias and cross-contamination in the forensic sciences: the corrosive implications for investigations, plea bargains, trials and appeals, *Law Prob. Risk* 14 (2015) 1–25.
- [7] I.E. Dror, Biases in forensic experts, *Science* 360 (6386) (2018) 243, doi:<http://dx.doi.org/10.1126/science.aat8443>.
- [8] J.J. Koehler, Fingerprint error rates and proficiency tests: what they are and why they matter, *Hastings Law J.* 59 (5) (2008) 1077–1099.
- [9] M.J. Saks, J.J. Koehler, The coming paradigm shift in forensic identification science, *Science* 309 (2005) 892–895.
- [10] Affidavit of Rickard K. Werder, In Re: Federal Grand Jury 03-01, No. 04-MC-9071 (USDC WD OR), May 6, (2004).
- [11] Officer of the Inspector General, A Review of the FBI's Handling of the Brandon Mayfield Case, Office of the Inspector General, Oversight & Review Division, US Department of Justice, 2006.
- [12] I. Pacheco, B. Cerchiai, S. Stoiloff, Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Examinations, (2014) Retrieved from www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf.
- [13] D.P. Baldwin, S.J. Bajic, M. Morris, D. Zamzow, A study of false-positive and false-negative error rates in cartridge case comparisons, Ames Laboratory, USDOE, Technical Report #IS-5207 (2014) Available at: afte.org/uploads/documents/swgun-false-positive-false-negative-usdoe.pdf.
- [14] D.L. Faigman, E.K. Cheng, J.L. Mnookin, E.E. Murphy, J. Sanders, C. Slobogin, *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Thomas Reuters, St. Paul, MN, 2016.
- [15] M.A. Bush, H.I. Cooper, R.B. Dorion, Inquiry into the scientific basis for bitemark profiling and arbitrary distortion compensation, *J. Forensic Sci.* 55 (2010) 976–983.
- [16] M.M. Houck, B. Budowle, Correlation of microscopic and mitochondrial DNA hair comparisons, *J. Forensic Sci.* 47 (2002) 1–4.
- [17] I.E. Dror, A hierarchy of expert performance (HEP), *J. Appl. Res. Mem. Cogn.* 5 (2) (2016) 121–127, doi:<http://dx.doi.org/10.1016/j.jarmac.2016.03.001>.
- [18] B.T. Ulerly, R.A. Hicklin, J. Buscaglia, M.A. Roberts, Repeatability and reproducibility of decisions by latent fingerprint examiners, *PLoS ONE* 7 (3) (2012) e32800.
- [19] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, *Sci. Just.* 51 (4) (2011) 204–208, doi:<http://dx.doi.org/10.1016/j.scijus.2011.08.004>.
- [20] P.A. Barrio, M. Crespillo, J.A. Luque, M. Aler, C. Baeza-Richer, L. Baldassarri, . . . , S. Vannelli, GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation, *Forensic Sci. Int. Genet.* 35 (2018) 156–163, doi:<http://dx.doi.org/10.1016/j.fsi-gen.2018.05.005>.
- [21] J. Bright, K. Cheng, Z. Kerr, C. McGovern, H. Kelly, T.R. Morretti, . . . , J. Buckleton, STRmix™ collaborative exercise on DNA mixture interpretation, *Forensic Sci. Int. Genet.* 40 (2019) 1–8, doi:<http://dx.doi.org/10.1016/j.fsi-gen.2019.01.006>.
- [22] J.M. Butler, M.C. Kline, M.D. Coble, IST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned, *Forensic Sci. Int. Genet.* 37 (2018) 81–94, doi:<http://dx.doi.org/10.1016/j.fsi-gen.2018.07.024>.
- [23] J.J. Koehler, Intuitive error rate estimates for the forensic sciences, *Jurimetrics* 57 (2017) 153–168.
- [24] R. Ansell, Internal quality control in forensic DNA analysis, *Accred. Qual. Assur.* 18 (2013) 279–2898, doi:<http://dx.doi.org/10.1007/s00769-013-0968-9>.
- [25] A. Kloosterman, M. Sjerps, A. Quak, Error rates in forensic DNA analysis: definition, numbers, impact and communication, *Forensic Sci. Int.* 12 (2014) 77–85.
- [26] I.E. Dror, G. Langenburg, “Cannot Decide”: the fine line between appropriate inconclusive determinations vs. unjustifiably deciding not to decide, *J. Forensic Sci.* 64 (2019) 1–15, doi:<http://dx.doi.org/10.1111/1556-4029.13854>.
- [27] D.E. Kornbrot, Random walk models of binary choice: the effect of deadlines in the presence of asymmetric payoffs, *Acta Psychol.* 72 (1989) 103.
- [28] G. Langenburg, A performance study of the ACE-V process: a pilot study to measure the accuracy, precision, reproducibility, repeatability, and bias ability of conclusions resulting from the ACE-V process, *J. Forensic Identif.* 59 (2009) 219–257.
- [29] M. Du, Analysis of errors in forensic science, *J. Forensic Sci. Med.* 3 (2017) 139–143, doi:http://dx.doi.org/10.4103/jfsm.jfsm_8_17.
- [30] R. Epstein, Fingerprints meet Daubert: the myth of fingerprint “science” is revealed, *Southern California Law Rev.* 75 (2002) 605–658.
- [31] Collaborative Testing Services, Inc, Latent Print Examination Test No. 18-5171/2/5 Summary Report, (2018) Retrieved from https://cts-forensics.com/reports/38171_Web.pdf. 1–31.
- [32] J.J. Koehler, Proficiency tests to estimate error rates in the forensic sciences, *Law Prob. Risk* 12 (2013) 89–98, doi:<http://dx.doi.org/10.1093/lpr/mgs013>.
- [33] R. Lempert, After the DNA wars: skirmishing with NRC II, *Jurimetrics* 37 (1997) 439–468.
- [34] J.D. Lieberman, C.A. Carrell, T.D. Miethe, D.A. Krauss, Gold versus platinum: do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychol. Public Policy Law* 14 (2008) 27–62, doi:<http://dx.doi.org/10.1037/1076-8971.14.1.27>.

- [35] G. Ribeiro, J.M. Tangen, B.M. McKimmie, Beliefs about Error Rates and Human Judgment in Forensic Science, (2019), doi:<http://dx.doi.org/10.31234/osf.io/td8fx> (in press).
- [36] B.L. Garrett, G. Mitchell, Forensics and fallibility: comparing the views of lawyers and judges, *West Virginia Law Rev.* 119 (2016) 621–637.
- [37] W.T. Thompson, S.O. Kaasa, T. Peterson, Do jurors give appropriate weight to forensic identification evidence? *J. Empirical Legal Stud.* 10 (2013) 359–397.
- [38] B.T. Ulery, A. Hicklin, J. Buscaglia, M.A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 7733–7738, doi:<http://dx.doi.org/10.1073/pnas.1018707108>.
- [39] A. Biedermann, S. Bozza, F. Taroni, P. Garbolino, A formal approach to qualifying and quantifying the “goodness” of forensic identification decisions, *Law Prob. Risk* 17 (2018) 295–310, doi:<http://dx.doi.org/10.1093/lpr/mgy016>.
- [40] R.E. Nisbett, T.D. Wilson, Telling more than we can know: verbal reports on mental process, *Psychol. Rev.* 84 (1977) 231–259.
- [41] M.R. Durose, A.M. Burch, K. Walsh, E. Tiry, Publically funded forensic crime laboratories: resources and services, 2014, Bureau Just. Stat. (2016) 1–11.
- [42] B.O. Gardner, S. Kelley, D.C. Murrie, I.E. Dror, What do forensic analysts consider relevant to their decision making? *Sci. Justice.* <https://doi.org/10.1016/j.scijus.2019.04.005> (in press).