Forensic Science International 321 (2021) 110714

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Judges and forensic science education: A national survey☆

Brandon L. Garrett[a,*], Brett O. Gardner[b], Evan Murphy[c], Patrick Grimes[d]

[a] Duke University School of Law, USA
[b] Institute of Law, Psychiatry, and Public Policy, University of Virginia, USA
[c] University of Nevada, Reno, and The National Judicial College, USA
[d] The National Judicial College, USA

ARTICLE INFO

Article history:
Received 25 September 2020
Accepted 27 January 2021
Available online 30 January 2021

Keywords:
Forensic science
Judges
Education
Reliability
Rule 702
Daubert

ABSTRACT

In criminal cases, forensic science reports and expert testimony play an increasingly important role in adjudication. More states now follow a federal reliability standard, which calls upon judges to assess the reliability and validity of scientific evidence. Little is known about how judges view their own background in forensic scientific evidence, and what types of specialized training they receive on it. In this study, we surveyed 164 judges from 39 different U.S. states, who attended past trainings at the National Judicial College. We asked these judges about their background in forensic science, their views concerning the reliability of common forensic disciplines, and their needs to better evaluate forensic science evidence. We discovered that judges held views regarding the scientific support for different forensic science disciplines that were fairly consistent with available literature; their error rate estimates were more supported by research than many estimates by laypersons, who often assume forensic methods are nearly infallible. We did not find any association between how judges rate forensic reliability and prior training. We did, however, find that training corresponded with judges' views that they should, and do in fact, take on a more active gatekeeping role regarding forensics. Regarding the tools judges need to vet forensic experts and properly evaluate forensic science evidence, they reported having very different backgrounds in relevant scientific concepts and having forensic science education needs. Judges reported needs in accessing better material concerning reliability of forensic science methods. These results support new efforts to expand scientific evidence education in the judiciary.

1. Introduction

In criminal cases, forensic science reports and expert testimony play an increasingly important role in adjudication. All states and the federal government require judges to screen forensic evidence, as with any expert evidence, before admitting it at a trial. Many states now follow standards based on the U.S. Supreme Court's ruling in *Daubert v. Merrell Dow Pharmaceuticals* and its progeny, a standard, which calls upon judges to assess the reliability and validity of scientific evidence [1,2]. Although scholars predicted a "paradigm shift" in judicial approaches towards forensic science following the U.S. Supreme Court's ruling in *Daubert* (Saks & Koehler, 2001), no dramatic shift has occurred in the courts: "at least in criminal cases, forensic science evidence is not routinely scrutinized pursuant to the standard of reliability enunciated in *Daubert*" ([3], p. 106). Indeed, there is evidence that judges do not commonly engage with or rule on reliability questions regarding forensic evidence, forensic science methods (Garrett & Fabricant, 2018), or the proficiency of a given expert [4]. Nevertheless, a more gradual shift in judicial approaches may be forthcoming.

In recent years, more forensic techniques have received scientific scrutiny regarding their reliability, with new research documenting error rates [5], new standards for terminology and conclusions (DOJ, 2019), and critical scientific reports asking that limits be placed on overreaching uses of forensics in criminal courts [3]. Leading reports, such as the influential 2009 National Academy of Sciences Report

☆ This work was partially funded by The National Judicial College and the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement 70NANB20H019 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, Duke University, University of California Irvine, University of Virginia, West Virginia University, University of Pennsylvania, Swarthmore College and University of Nebraska, Lincoln. We would like to thank President Benes Aldana, Provost Joy Lyngar, and The National Judicial College for their support with this research. The data and materials used in this research are publicly available at: https://osf.io/evnzd/

* Correspondence to: Duke University School of Law, 210 Science Drive, Durham, NC 27708, USA.
E-mail address: bgarrett@law.duke.edu (B.L. Garrett).

https://doi.org/10.1016/j.forsciint.2021.110714
0379-0738/

[3], have noted that judges have not taken an active or effective role in accomplishing their gatekeeping responsibilities to review the reliability of forensic evidence in criminal cases. Indeed, while in recent years, forensic practitioners, prosecutors, and forensic professional associations [6] have adopted new limits and standards for their work (DOJ, 2019), in contrast, judges have imposed very few such limits or standards. The challenge for the judiciary is how to take advantage of the work being done in the broader scientific community to improve the use of forensic science in court.

One challenge for efforts to improve judicial use of forensic evidence is the lack of scientific background and education among lawyers generally, and among judges in particular. After all, "scientific evidence tests the abilities of judges, lawyers, and jurors, all of whom may lack the scientific expertise to comprehend the evidence and evaluate it in an informed manner" [7]. The NRC Report noted that, "judges and lawyers…generally lack the scientific expertise necessary to comprehend and evaluate forensic evidence in an informed manner" ([3], p. 78). In its tenth recommendation, the Report noted, the need to "support law school administrators and judicial education organizations in establishing continuing legal education programs [in forensic science] for law students, practitioners, and judges" ([3], p. 239).

New training programs geared towards judges have been developed in response to such recommendations. Some have been hosted by the American Association for the Advancement of Science (AAAS); ABA Judicial Division Forensic Science Committee; Federal Judicial Center (FJC); The National Courts and Sciences Institute (NCSI); and the National Judicial College (NJC). In 2015, the National Commission of Forensic Science recommended that a national curriculum for judges be developed; the Department of Justice did not follow up on that recommendation [8]. There have been training materials developed for judges, most prominently the Reference Manual on Scientific Evidence, developed by the Federal Judicial Center and the National Research Council [9], as well as the Science Bench Book for Judges, developed by The National Judicial College and the Justice Speakers Institute [10]. However, the effectiveness of such training materials—and consequent judge perceptions—is unknown.

In the wake of the *Daubert* court ruling and recent pushes to advance the scientific base for many forensic sciences, there have been studies addressing perceptions of forensic science evidence in court. Studies have examined how laypersons, jurors, and even forensic analysts, assess forensic evidence (e.g., [11,12]; Lieberman et al., 2008; [13]). Taken together, the literature suggests that many individuals are aware that forensic science analyses can be erroneous, but also that perceptions of forensic science evidence vary widely. Far less is known about the state of judicial training in forensic science, how judges view their own background in forensic and scientific evidence, or what additional types of specialized training they would like to receive on forensics, including at the state level. Recruiting lay participants is far more readily managed than recruiting judges. Further, there are reasons to think that judges may have different attitudes towards forensics than lay jurors; for example, one study comparing views of lawyers and jurors found that lawyers were more skeptical of forensic evidence than laypersons [14].

A handful of surveys of judges concerning forensic evidence has been conducted. Gatowski and colleagues [15] surveyed 400 state judges concerning their attitude towards gatekeeping under *Daubert,* and found that they overwhelmingly supported that role and agreed that scientific support is needed for expert methods, but that they were quite unfamiliar with scientific concepts such as error rates. A 2001 survey of Virginia judges inquired into attitudes toward mental health experts, finding that most preferred traditional clinical testimony to research or statistically-based testimony [16]. A 2009 survey of Polish judges found that "a great many judges do not possess adequate and sufficient knowledge for the assessment of scientific evidence" [17]. Judges have also been surveyed regarding their attitudes toward eyewitness evidence in the U.S. [18] and in other countries [19]. One study compared lay and judicial responses to MtDNA evidence [11]. Other surveys of judges involved very small numbers of participants, such as a study of 13 judges regarding their familiarity and views concerning digital evidence [20].

In the current study of a national sample of judges who attended past trainings at the NJC, we surveyed judges concerning their background in forensic science, their views concerning the reliability of common forensic science disciplines, and their needs to help better evaluate forensic science evidence. Moreover, we explored whether judges' histories of forensic science trainings influenced their views of forensic science evidence.

2. Method

2.1. Participants

Participants were recruited using the National Judicial College's (NJC) course enrollment database. We invited judges who had attended a course at the NJC between 2015 and 2020, and who reported overseeing felony, criminal, and jury trials as a judge. Using this criteria, 938 judges were randomly selected and invited to participate in our study.[1]

Participants were 164 judges (17.5% response rate) from 39 different U.S. states, with Kansas (n = 11), Minnesota (n = 8), and New Mexico (n = 7) having the largest representation. Almost all participants held law degrees (n = 138; 97.9%) and most participants were general jurisdictions judges (n = 129; 91.5%). However, other participants indicated that they served as limited jurisdiction judges (n = 4; 2.8%), tribal court judges (n = 4; 2.8%), or in other judicial roles (e.g., military court judge; n = 1; 0.7%). Participants typically had 13 years of judicial experience (SD = 8.5; range = 1–42) and were an average age of 60 years old (SD = 8.5; range = 33–80). Of those who responded to our demographic questions, most were male (n = 100; 73.0%), and Caucasian (n = 118; 71.1%). However, participants also identified as African American (n = 7; 4.2%), Hispanic/Latino (n = 4; 2.4%), American Indian/Alaskan Native (n = 3; 1.8%), Asian (n = 3; 1.8%), and "Other" (n = 3; 1.8%) ethnicities.

2.2. Procedure and materials

Participants who were selected to take our survey were notified through email one week prior to receiving our survey. The notification email informed the participants about the purpose of the survey, the time it would take to complete, and the day they would receive a survey invitation. After a week, participants received a survey link to our online Qualtrics survey. Participants were given two weeks to respond before they were sent a reminder email from the researchers. Two reminders were sent to participants, both two weeks apart. After 6 weeks, data collection was complete, and the survey was closed. The data and materials used in this research are publicly available at: https://osf.io/evnzd/.

2.2.1. Survey

Survey participants completed a three-part survey reporting their background in forensic science, their views concerning the reliability of common forms of forensic evidence, and their needs for evaluating forensic science evidence. Section one of the survey asked participants to describe their past experiences with forensic science

[1] 1000 judges were randomly selected under our study selection criteria. However, 62 judges were unable to receive the email invitation to our study due to email security filters. Since these judges didn't have the opportunity to receive our invitation, they were not counted in our total.

through multiple questions inquiring about the number and percentage of past cases in which forensic science evidence was presented before them.

Section two of the survey focused on training and resources. Participants were asked to report the amount (e.g., "less than one day of training"), content (e.g., "statistical methods"), and context (e.g., "law school") of past trainings on forensic science evidentiary issues. Relatedly, participants were asked to report the ideal context for judicial training on forensic science evidence and which topics they would be interested in receiving specific training. They also described the availability of forensic science education resources in their jurisdiction and identified types of resources they currently use, or would use, to help evaluate forensic science methods.

Section three of the survey asked participants to describe their familiarity with statistical methods and estimate the rate of false positive errors in seven common forensic science disciplines. In other words, we asked judges to estimate the rate in which an experienced forensic scientist would mistakenly conclude a match or identification where none truly exists. In line with previous error rate studies (e.g., [12,13,21]), judges were given a 14-point logarithmic scale to estimate error rates ranging from "*approximately 1 time in 2*," to "*such an error is impossible*." Finally, judges were asked to provide demographic information and details about their position and career as a judge.

3. Results

3.1. Judicial backgrounds in forensic science

3.1.1. Experiences

There was significant variability in the reported percentages of criminal cases that presented forensic science evidence, and in judges' usual responses to such evidence. Judges typically indicated that 37.4% of cases presented forensic science evidence (*Mdn* = 31.0%; *SD* = 23.4), but estimates ranged from 0% to 92%. Judges further indicated that they held a hearing on the admissibility of forensic science evidence in only 14.7% (*Mdn* = 10.0%; *SD* = 19.3) of cases, but again estimates ranged from 0% to 100%. In addition, judges estimated that they ruled forensic science evidence testimony inadmissible in approximately 13.5% of cases (*Mdn* = 5.0%; *SD* = 20.4), with estimates ranging from 0% to 100%. Most judges indicated that their jurisdiction used the *Daubert* standard (*n* = 105; 65.2%), although others endorsed the *Frye* standard (*n* = 30; 18.6%) or another standard (*n* = 26; 16.1%). Interestingly, judges from jurisdictions using the *Daubert* standard (13.9%) did not differ from those in jurisdictions using the *Frye* standard (11.0%) or another standard (11.8%) in the estimated percentage of cases in which they ruled forensic science evidence inadmissible, $F(2, 136) = 0.27$, $p = .76$, $\eta p^2 = .004$. Within the past three years, most judges estimated that 40 of their cases presented forensic science evidence. However, this data was positively skewed in that the median estimate was 10 cases and estimates ranged from 0 to 1000 cases.

3.1.2. Training

Judges endorsed wide-ranging training histories specific to forensic science evidence.[2] Specifically, 29.6% (*n* = 45) of judges reported completing more than one week of training, 27.0% (*n* = 41) completed more than two days of training, 17.1% (*n* = 26) completed one day of training, 16.4% (*n* = 25) completed less than one day of training, and 9.9% (*n* = 15) reported that they had not received any training on the topic.

Regarding particular forensic science training topics, most judges reported that they had received training on the standards for admissibility of forensic evidence (*n* = 128; 77.1%), DNA technology (*n* = 94; 56.6%), and crime lab procedures (*n* = 83; 50.0%). However, less than half of surveyed judges received training on cognitive bias and human factors (*n* = 80; 48.2%), standards and terminology for expressing conclusions (*n* = 46; 27.7%), statistical methods (*n* = 40; 24.1%), cognitive methods (*n* = 28; 16.9%), measurement of error rates (*n* = 28; 16.9%), and privacy issues concerning forensic tests (*n* = 14; 8.4%). Overall, judges reported a moderate level of familiarity with the statistical methods that underlie different types of forensic science evidence (*M* = 4.51; *SD* = 2.31) on a scale of 1–10, with 1 = *Unfamiliar* and 10 = *Very Familiar*, although some judges indicated that they were completely unfamiliar (i.e., endorsed 1) or rather familiar (i.e., endorsed 9).

We also asked judges to describe the context of their forensic science trainings. An overwhelming majority of judges reported that they received training on forensic science evidence through continuing education as a judge (*n* = 130; 78.3%). However, judges also received training through continuing education as a lawyer (*n* = 81; 48.8%), undergraduate studies (*n* = 28; 16.9%), and law school (*n* = 23; 13.9%).

In order to draw comparisons, we also asked judges to report where they *should* receive training about forensic science evidence. As shown in Fig. 1, responses were generally consistent with actual training contexts. Nearly all judges reported that they should receive training on forensic science evidence through continuing education as a judge (*n* = 153; 92.2%). However, responses suggest that judges believe more trainings should occur during law school than actually do occur in law school courses [22].

3.2. Views concerning the reliability of common forensic science disciplines

3.2.1. Error rate estimates

We asked judges to estimate, as noted, how reliable common forensic science methods are. In their responses, judges' false positive error rate estimates varied both within and across different forensic science disciplines (see Fig. 2).

As Table 1 depicts, judges typically believed DNA evidence to be the most reliable form of forensic evidence. The most common error rate estimate for DNA evidence was 1 error in 1000,000 matches, with almost two thirds of judges (63.0%) providing error estimates between 1 in 100,000 and 1 in one billion matches. Judges indicated that the second most reliable form of forensic evidence was toxicology evidence, with 72.4% of judges providing error estimates between 1 in 100 and 1 in 100,000 matches. Conversely, judges estimated that bitemark evidence and shoeprint evidence were the least reliable forms of evidence. Approximately, 85.7% (bitemark evidence) and 81.1% (shoeprint evidence) of judges provided error estimates for these two disciplines between 1 in 2 and 1 in 1000 matches.

3.3. Judicial needs for evaluating forensic science evidence

3.3.1. Forensic science resources

Judges were asked to rate the current availability of resources to help evaluate and understand forensic science evidence presented in their courtroom on a scale from 1 to 10, with 1 = *Nonexistent* and 10 = *Excellent.* On average, judges (*n* = 153) rated the availability of current resources at 5.10 (*SD* = 2.23). However, judges' perceptions of their current resources varied dramatically, with responses ranging from nonexistent to excellent.

[2] Although all judges in our study had participated in education at the NJC previously, our questions about forensic science training were not specific to training at the NJC. Instead, judges were asked to report details about their forensic science training at any institution or organization.

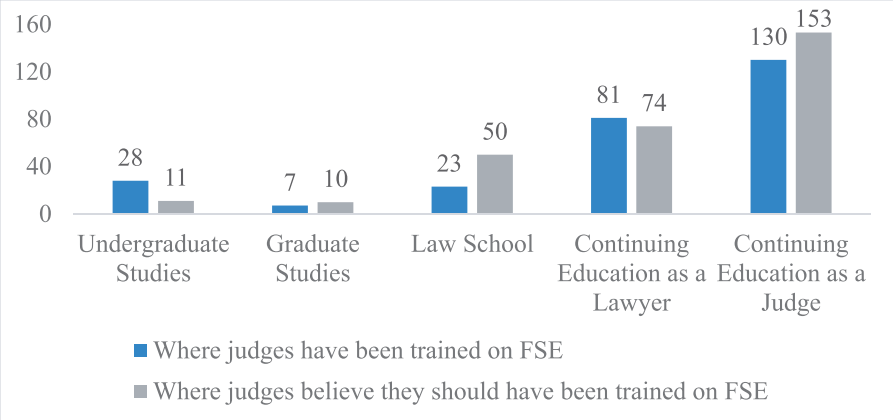

Fig. 1. Judicial training on forensic science evidence (FSE). Note. n = 164.

In addition to commenting on the availability of resources, judges identified the resources they use to evaluate and understand forensic science methods via open-ended questioning. Of the 127 judges who responded, the most common identified resource was legal or scientific research articles ($n = 46$; 36.2%); however, many judges also identified the testimony of experts ($n = 39$; 30.7%), continuing education ($n = 38$; 29.9%) and case law ($n = 27$; 21.2%) as important resources for evaluating and understanding forensics science evidence.

Judges were also asked to report additional resources that they *would* find helpful for evaluating and understanding forensic science methods. Of the 117 judge who responded to the open-ended question, the most common response was judicial training or education ($n = 41$; 39.4%). However, judges also identified online research or reference databases ($n = 25$; 24.0%), and bench books ($n = 11$; 10.6%), as additional resources they would find helpful.

3.3.2. Future training interests

To understand judges' forensic science evidence training needs, we asked judges to rate their interest in receiving training on a variety of forensic evidence topics using a scale ranging from 1 to 10, with 1 = *Not Interested* and 10 = *Very Interested*. Judges reported high levels of interest in training on topics such as digital, DNA, toxicology, firearms, and fingerprint evidence. Judges showed lower levels of training interest in shoeprint and bitemark evidence. See Table 2 for detailed results.

3.4. Impact of past forensic science training

We next explored whether judges' participation in forensic science training was associated with their views of, and experiences with, forensic science evidence by conducting a series of nonparametric analyses. Judicial training on forensic science evidence was associated with several viewpoints measured in the present study. Specifically, judges who reported more extensive training regarding forensic science evidence endorsed greater familiarity with the statistical methods that underlie forensic science evidence, $\rho(149) = 0.21$, $p = .01$, more strongly agreed with the sentiment that it is a judge's responsibility to prevent "junk science" from being presented at trial, $\rho(150) = 0.19$, $p = .02$, and perceived increased availability of resources to evaluate forensic science evidence, $\rho(149) = 0.40$, $p < .001$.

Regarding behavioral practices, judicial training in forensic science evidence was not associated with the percentage of cases in which judges held an admissibility hearing regarding forensic science evidence, $\rho(145) = -0.14$, $p = .10$. However, judicial training was associated with estimated rates of ruling forensic science evidence inadmissible. Specifically, judges who endorsed more extensive training in forensic science evidence indicated that they ruled such evidence to be inadmissible at greater rates, $\rho(133) = -0.21$, $p = .01$.

Finally, to understand the impact of judicial training on perceptions of forensic science evidence reliability, we conducted Spearman's rank-order correlations and found that broad judicial training in forensic science evidence was unrelated to false positive error rate estimates across seven common forensic science disciplines ($ps > 0.05$). We next conducted a series of Mann-Whitney tests to determine whether judges who endorsed forensic training specific to "statistical methods" or "measurement of error rates" provided error rate estimates that differed from judges with no such training. Ultimately, judges who completed specific training in "statistical methods" or "measurement of error rates" did not provide error rate estimates that differed from judges with no specialized training ($ps > 0.05$). Moreover, judges who completed training in "DNA technology" also did not provide error rate estimates for DNA analyses that differed from judges with no such training, $p = .84$.

4. Discussion

As criminal investigations and litigation have become more complex and technical, judges are called upon to weigh in on a range of scientific issues, and as a result, "scientific educational programs for judges are becoming an increasingly important part of continuing judicial education" [23]. Little was known about the extent of judicial forensic science training or needs for such training. At its most basic, the current findings suggest significant variability in judicial practices across the U.S. Among a sample of experienced judges who regularly oversee criminal cases, some indicated that they never hold admissibility hearings for forensic science evidence while others indicated that they always hold admissibility hearings. Further, judges reported that they ruled such evidence to be inadmissible at varying rates that were not associated with their jurisdiction's evidentiary standard. While the participating judges may, in fact, encounter forensic work of varying quality, based on their jurisdictions, the observed variability in routine practices suggests that standardized judicial training may be helpful in reducing any discrepant practices.

Of course, most judges in the current sample (90.1%) have completed training specific to forensic science evidence, which may make them unusual in this respect. However, the content and duration of trainings varied widely. For example, three of four judges (77.1%) received training on the standards for admissibility of

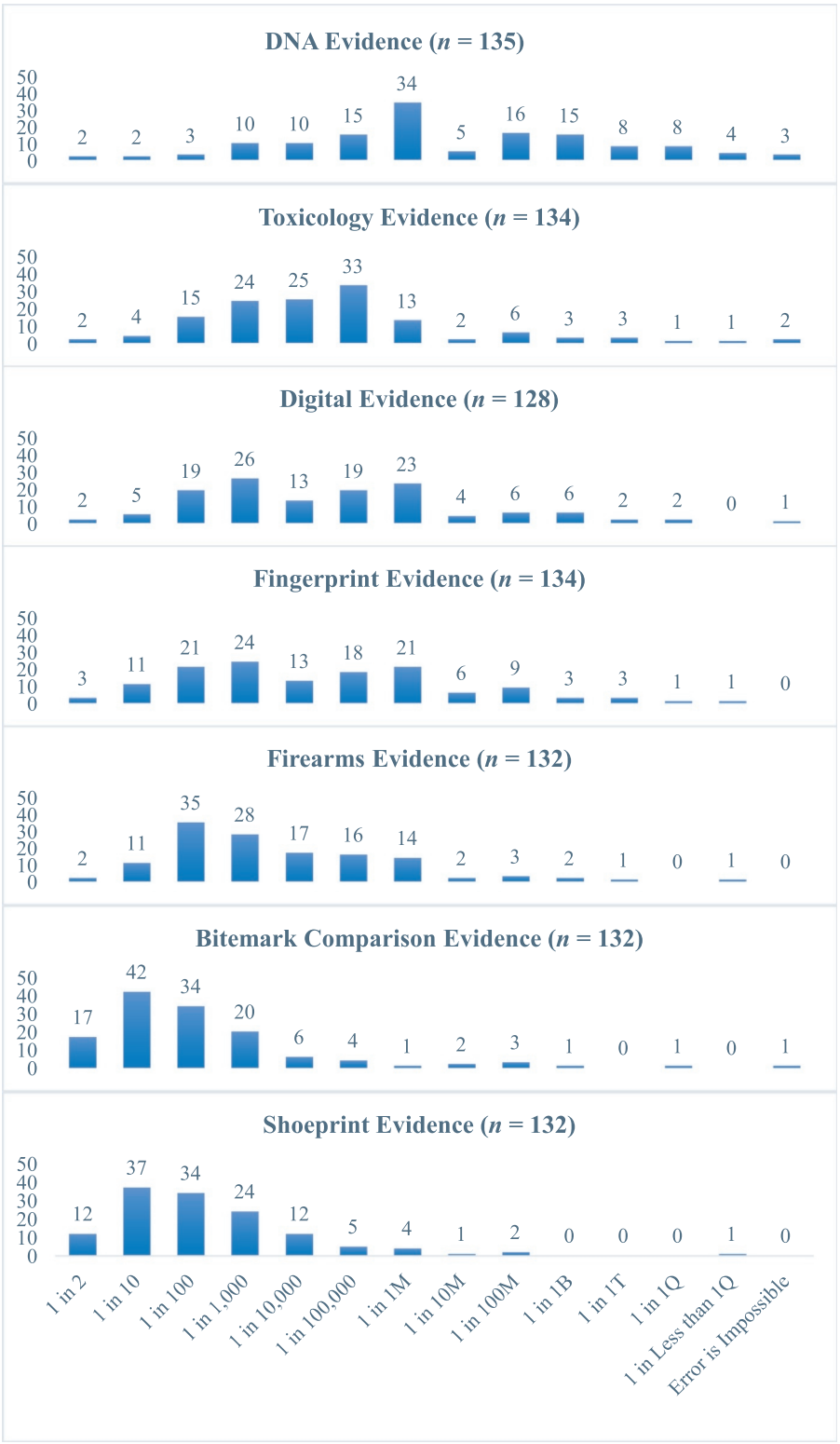

Fig. 2. Judicial false positive error rate estimates for forensic science evidence. Note. M = Million, B = Billion, T = Trillion, Q = Quadrillion.

forensic evidence, but only one in four judges (24.1%) received training in statistical methods. Additionally, the recency of judicial training varied, with some judges indicating that they last completed training on forensic science during undergraduate study. Given the new scientific developments in forensics, including the greater focus on use of statistical methods and standards, there may be a real need for continuing education in this area.

4.1. Perceptions of forensic science evidence

There is very little appropriately designed research examining the scientific reliability of forensic science disciplines and documenting formal error rates [12]. The few existing studies suggest wide-ranging error rates across disciplines. For example, the Netherlands Forensic Institute reported that DNA laboratory errors occurred in less than 1% of

all analyses [24]. In 2016, the PCAST report concluded that the "best

Table 1
Modal Responses in False Positive Error Rates Provided by Judges.

Forensic Evidence Type	False Positive Error Rate Estimate
DNA Evidence	1 time in 1000,000 (25.2%)
Toxicology Evidence	1 time in 100,000 (24.6%)
Digital Evidence	1 time in 1000 (20.3%)
Fingerprint Evidence	1 time in 1000 (17.9%)
Firearms Evidence	1 time in 100 (25.9%)
Bitemark Evidence	1 time in 10 (31.8%)
Shoeprint Evidence	1 time in 10 (28.0%)

Note. Percentages represent the proportion of judges who chose the most common response option.

estimates" of false positive error rates within latent print evidence are currently 1 in 604 cases (0.2%; [25]) and 1 in 24 cases (4.2%; [26]). One study examining firearm evidence estimated the false positive error rate to be 1 in 66 cases (1.5%; [27]). At the other end of the spectrum, bitemark evidence has demonstrated a 64.0% false positive error rate in one study [28] and, to our knowledge, there are no existing studies examining the error rates of shoeprint evidence [5].

Given the lack of established error rates in many forensic science disciplines, and the variability in the research base and reliability across disciplines, judges' perceptions of forensic science evidence appear fairly well-informed. Importantly, judges do appear to distinguish between forensic science disciplines when evaluating the reliability of evidence. Moreover, their general ranking of forensic science disciplines according to false positive error rates appears broadly consistent with the extant (and sparse, for many commonly used disciplines) literature on the topic.

However, judges often underestimated the false positive error rates of forensic science disciplines, using the few available studies as benchmarks. For example, most judges estimated that false positive errors occur in approximately 1 of 1000 cases in latent fingerprint evidence compared to studies suggesting the error rate is between 1 in 604 and 1 in 24 cases [25,26]. Judges similarly estimated that errors occur in approximately 1 of 100 cases in firearm evidence compared to a study providing an error rate estimate of 1 in 66 cases [27]. Indeed, judges were perhaps most skeptical of bitemark evidence, but they still overestimated its reliability, with most judges indicating that false positive errors occur in 1 of every 10 cases in contrast to research suggesting error rates closer to 6 of every 10 cases [28]. More concerning, at least one judge (and sometimes as many as 28% of responding judges) estimated the false positive error rate to be impossible or extremely low (i.e., equal to or less than 1 in 1 billion cases) for each discipline.

This overestimation of the reliability of forensic science evidence is not unique to judges and, in fact, judges may overestimate the reliability to a lesser extent than do laypersons or even forensic analysts. As an example, Koehler [12] found that laypersons provided a median false positive error rate estimate of 1 in 10 million cases for DNA evidence while Murrie and colleagues (2019) found that forensic biology analysts typically provided estimates of 1 in 100 million cases. Judges in this survey provided a higher median estimate of 1 in 1 million cases for DNA evidence. Similarly, laypersons provided median estimates of 1 in 5.5 million cases for latent print evidence while latent print analysts (n = 7) provided a much lower median false positive error rate of 1 in 1 billion. Again, judges in this survey provided a higher median estimate of 1 in 1000 cases; although still overly optimistic, judges' estimates appear to most closely resemble the empirical literature of all surveyed populations. The results also echo findings by Garrett and Mitchell [14] that lawyers are far more skeptical of forensic evidence, such as latent print evidence, than laypersons are, although lawyers are also aware that laypeople place extremely strong weight on such evidence.

Table 2
Judicial Forensic Science Evidence Training Interests.

Forensic Evidence Type	Mean Level of Interest
Digital Evidence ($n = 144$)	8.28 (2.05)
DNA Evidence ($n = 147$)	7.80 (2.43)
Toxicology Evidence ($n = 146$)	7.66 (2.37)
Firearms Evidence ($n = 145$)	7.44 (2.40)
Fingerprint Evidence ($n = 140$)	7.02 (2.50)
Shoeprint Evidence ($n = 136$)	5.06 (2.94)
Bitemark Evidence ($n = 135$)	4.79 (3.12)

Note. Interest ranged from 1–10, where 1 represented *not interested*, and 10 represented *very interested*. Standard deviations are in parentheses.

A recent survey of forensic scientists found that, like judges, practicing analysts often overestimated the accuracy of their discipline when compared to existing error rate estimates [13]. Although the current data cannot speak to this directly, it is possible that these findings are closely related. Indeed, it is not surprising that judges overestimate the reliability of forensic science disciplines. Judicial educators rely on the information communicated by forensic experts, whom also appear to overestimate the reliability of forensic science evidence. Approximately one in three surveyed judges (30.7%) identified expert testimony as a primary resource for evaluating and understanding forensic sciences. Thus, this finding underscores the need for more accurate measurements of error rates by forensic experts. With a more accurate understanding of error rates across forensic disciplines among experts, judicial educators will be better positioned to provide forensic science evidence training that is both accurate and based on scientific research.

4.2. Current judicial needs regarding forensic science

Judges in the current sample indicated that they typically relied upon journal articles, expert testimony, case law, and continuing education to help evaluate and understand forensic science methods. However, they did not describe such resources as particularly accessible, with an average rating of 5.10 on a scale with 1 = *Nonexistent* and 10 = *Excellent*. When asked about additional resources that would increase their ability to evaluate forensic sciences, judges primarily expressed a desire for additional judicial training and online research/reference databases. Judges were specifically interested in receiving training on forensic science disciplines that they perceived to be relatively reliable, such as DNA, digital, and toxicology evidence. They were especially interested in digital evidence, identifying this discipline as the primary method of interest. Finally, judges in the current sample reported that forensic science training should more frequently occur during law school and through continuing education as a judge than is presently the case. We note that law school coverage of forensic science is not common (Garrett, Cooper & Beckham, in draft). Particularly given ongoing and active research and standards-setting, continuing education seems much needed.

4.3. Forensic science training

Although the current research cannot establish causal relationships, results do suggest that forensic science training is associated with some judicial views and behavioral practices. Judges who completed more training specific to forensic science endorsed greater familiarity with related statistical methods. Further, judges with more extensive training were more likely to perceive resources that help judges evaluate forensic science evidence as readily available. In this manner, judicial training may not only help judges, but help them to help themselves. Judicial training may improve judges' statistical fluency and also improve judges' ability to find resources needed to critically evaluate forensic science evidence. Indeed, approximately 16% of the variance in judges' perceptions of resource availability was explained by training history.

Importantly, judicial training was also associated with how judges view their own role in evaluating forensic science evidence. Although *Daubert v. Merrell Dow Pharmaceuticals* and its progeny, in the states that have adopted the federal standard, instruct judges to evaluate the admissibility of forensic science evidence, not all judges agree that this is their duty. Judges with more extensive forensic science training in the current study more strongly agree with the sentiment that it is a judge's responsibility to prevent "junk science" from being presented at trial. Perhaps relatedly, judges with more training indicated that they ruled forensic science evidence to be inadmissible at greater rates.

Thus, comprehensive training may help judges embrace their roles as gatekeeper and be more critical of presented evidence. That said, despite judges' increased familiarity with statistical methods, endorsed training (even training specific to the measurement of error rates) did not influence judges' perceptions regarding the reliability of forensic science disciplines (i.e., false positive error rate estimates). Careful examination of forensic science error rates may represent a worthy focus for future training curriculum. We did not ask judges about other interventions designed to educate jurors, which themselves might require judicial education [31]. For example, jury instructions might convey additional information to jurors about the strength and limitations of forensic evidence [29]. Or judges may instead approve funding for a defense expert, who might call attention to limitations of the relevant discipline.

These results also suggest that further judicial training is important not because judges are unaware of differences in the reliability of forensic evidence, but rather because judges would like to render more informed rulings on these technical subjects. Indeed, past research has found that judges' primary motivation for judicial education is their own professional growth [30]. This suggests, then, that training may further inform judicial decisions, but not necessarily alter the overall priorities and perspectives of judges. To be sure, this study surveyed judicial attitudes towards forensic science disciplines, training, and their gatekeeping role. We did not measure how judges actually rule in cases before them, nor what other non-gatekeeping actions they might take to assure the comprehensibility and the reliability of evidence relied upon in criminal cases. To what degree training and education affect behavior is an important subject for future research; there is little research generally on the connection between legal education and the quality of the work that lawyers and judges do in any setting.

5. Conclusion

We surveyed a national sample of judges regarding their background in forensic science, views concerning the reliability of forensic disciplines, and educational needs regarding forensic science evidence. We found that participant judges held views concerning the scientific support for different forensic disciplines that were fairly consistent with available literature. We did not find any association between prior training and how judges rated the reliability of forensic science disciplines. However, we did find that training corresponded with judges' views that they should, and reports that they do in fact, take on a more active gatekeeping role regarding forensics. Regarding the educational resources that judges need to vet forensic experts and properly evaluate evidence, they reported having very different backgrounds in relevant scientific concepts. These results support new efforts to expand scientific evidence education in the judiciary, particularly as they may inform both greater judicial understanding of key methods and concepts, and have the potential to produce better informed judicial decisions.

Conflict of interest statement

The authors have no conflicts of interest to disclose.

References

[1] Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 1993.
[2] Federal Rules of Evidence, Rule 702.
[3] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington D.C, 2009.
[4] B.L. Garrett, G. Mitchell, The proficiency of experts, Univ. Pa. Law Rev. 166 (2018) 901–960.
[5] 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.
[6] International Association for Identification, IAI Resolution 2010–18, passed July 16, 2010. Retrieved from: http://clpex.com/swgfast/Resources/100716_IAI_Resolution_2010-18.pdf.
[7] Developments in the law – confronting the new challenges of scientific evidence. 108 Harv. L. Rev. 1481, 1484 1995.
[8] National Commission on Forensic Science, Recommendation to the Attorney General forensic science curriculum development 2015. Retrieved from: https://www.justice.gov/ncfs/file/818206/download.
[9] Federal Judicial Center and National Research Center, Reference Manual on Scientific Evidence, 3rd edition., National Academies Press, USA, 2011.
[10] The National Judicial College & Justice Speakers InstituteScience bench book for judges 2019. http://resources.judges.org/downloads/science-bb.pdf.
[11] V.P. Hans, Judges, juries, and scientific evidence, J. Law Policy 16 (2007) 19–46.
[12] J.J. Koehler, Intuitive error rate estimates for the forensic sciences, Jurimetrics 57 (2017) 153–168.
[13] D.C. Murrie, B.O. Gardner, S. Kelley, I.E. Dror, Perceptions and estimates of error rates in forensic science: a survey of forensic analysts, Forensic Sci. Int. 302 (2019) 1–9, https://doi.org/10.1016/j.forsciint.2019.109887
[14] B.L. Garrett, G. Mitchell, Forensics and fallibility: comparing the views of lawyers and jurors, West Va. Law Rev. 119 (2016) 621–637.
[15] S.I. Gatowski, S.A. Dobbin, J.T. Richardson, G.P. Ginsburg, M.L. Merlino, V. Dahir, Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-Daubert world, Law Hum. Behav. 25 (2001) 433–458, https://doi.org/10.1023/A:1012899030937
[16] R.E. Redding, M.Y. Floyd, G.L. Hawk, What judges and lawyers think about the testimony of mental health experts: a survey of the courts and bar, Behav. Sci. Law 19 (2001) 583–594, https://doi.org/10.1002/bsl.455
[17] J. Wójcikiewicz, Judges' attitude towards scientific evidence, Rev. za Krim. Kriminol. 64 (2013) 249–255.
[18] R.A. Wise, M.A. Safer, What US judges know and believe about eyewitness testimony, Appl. Cogn. Psychol. 18 (2004) 427–443, https://doi.org/10.1002/acp.993
[19] L. Jiang, D. Luo, Legal professionals' knowledge of eyewitness testimony in China: a cross-sectional survey, PLoS ONE 11 (2016) 1–8, https://doi.org/10.1371/journal.pone.0148116
[20] G.C. Kessler, Retrieved from NSUWorks, Graduate School of Computer and Information Sciences 2010 (196) https://nsuworks.nova.edu/gscis_etd/196.
[21] W.C. Thompson, S.O. Kaasa, T. Peterson, Do jurors give appropriate weight to forensic identification evidence? J. Empir. Leg. Stud. 10 (2013) 359–397.
[22] B.L. Garrett, G. Cooper, Q. Beckham, Forensic science in legal education (in draft).
[23] J. Cecil, Science education for federal judges, ABA J. 56 (4) (2017) 8–10.
[24] A. Kloosterman, M. Sjerps, A. Quak, Error rates in forensic DNA analysis: definition, numbers, impact and communication, Forensic Sci. Int. 12 (2014) 77–85.
[25] B.T. Ulery, A. Hicklin, J. Buscaglia, M.A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, Proc. Natl. Acad. Sci. 108 (2011) 7733–7738, https://doi.org/10.1073/pnas.1018707108
[26] I. Pacheco, B. Cerchiai, S. Stoiloff. Miami-Dade researchstudy for the reliability of the ACE-V process: Accuracy & precision inlatent fingerprint examinations 2014. Retrieved from www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf.
[27] D. P. Baldwin, S. J. Bajic, M. Morris and D. Zamzow, A study of false-positiveand false-negative error rates in cartridge case comparisons. Ames Laboratory, USDOE, Technical Report#IS-5207 2014. Available at: afte.org/uploads/documents/swggun-false-postive-false-negative-usdoe.pdf.
[28] 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.
[29] B.L. Garrett, W.E. Crozier, R. Grady, Error rates, likelihood ratios, and jury evaluation of forensic evidence, J. Forensic Sci. 65 (2020) 1199–1209, https://doi.org/10.1111/1556-4029.14323
[30] E. Murphy, M. Kemmelmeier, P. Grimes, Motivations, barriers, and impact ofcontinuing judicial education: A survey of U.S. judges. Court Review. National Commissionon Forensic Science 2021.
[31] B.L. Garrett, G. Mitchell, How jurors evaluate fingerprint evidence: the relative importance of match language, method information and error acknowledgement, J. Empir. Leg. Stud. 10 (2013) 484–511.