Probabilistic reporting in criminal cases in the United States: A baseline study

Simon A. Cole Funding acquisition Supervision Writing - review & editing Writing - original draft Data curation Resources Investigation Formal analysis Methodology Conceptualization * scole@uci.edu, Matt Barno Formal analysis Investigation Resources Data curation Writing - original draft Writing - review & editing mbarno@uci.edu

Department of Criminology, Law & Society, University of California, Irvine, CA 92697-7080, USA

*Corresponding author.

1 Introduction

Over a decade ago, the National Research Council (NRC) Report, *Strengthening Forensic Science in the United States*, highlighted the lack of standards with regard to reporting of evidence in forensic science:

> [M]any terms are used by forensic examiners in reports and in court testimony to describe findings, conclusions, and the degrees of association between evidentiary material (e.g., hairs, fingerprints, fibers) and particular people or objects. Such terms include but are not limited to "match," "consistent with," "identical," "similar in all respects tested," and "cannot be excluded as the source of."...Yet the forensic science disciplines have not reached agreement or consensus on the precise meaning of any of these terms. Although some disciplines have developed vocabulary and scales to be used in reporting results, they have not become standard practice [1].

In subsequent years, working groups in a number of forensic disciplines promulgated guidelines in an effort to standardize~~d~~ reporting among practitioners [2–7]. As the above quotation~~e~~ illustrates, many forensic disciplines historically reported in what can reasonably be characterized as a "non-statistical" manner, using verbal formulations that were categorical rather than continuous, reflecting certainty rather than uncertainty (e.g., [8,9]). The standards issued following the NRC report often continued to promote these non-statistical forms of forensic evidence reporting, albeit it in a more systematized manner.

At the same time, recent years have seen increasing efforts to promote the application of statistics to forensic evidence in the United States (e.g., [8,9]). Within this burgeoning field of forensic statistics, a vigorous scholarly debate has developed over how forensic scientists *should* report results. Although all forensic statisticians urge that results should be reported in a statistically defensible manner, they differ regarding the precise tools and frameworks for such reporting. For example, some forensic statisticians have promoted the likelihood ratio framework as an all-encompassing solution to forensic reporting, applicable to all, or nearly all, situations (e.g., [10,11]). Yet, even among those who advocate for likelihood ratios, debates remain over myriad technical issues in formulating those ratios. Some forensic statisticians argue that, in order to adequately address underlying model assumptions and sampling uncertainty, forensic reports should be expressed in terms of a range of plausible likelihood ratios rather than a single likelihood ratio (e.g., [8,9]). Likewise, there have been strong debates over what we might call "reporting standards": documents that instruct forensic practitioners how they *should* report results.

But what if practitioners do not actually report in the way that disciplinary standards say they should? In contrast to the robust scholarship on how forensic scientists *should* report, less attention has been paid to how they *do* report. Accordingly, our study focuses on the following key research questions:

1. To what extent are forensic reports in these disciplines consistent with published standards?
2. To what extent are forensic reports in these disciplines probabilistic, and, if so, how is probability expressed?

2 Literature review

Researchers have increasingly looked to the content of forensic reports and courtroom testimony as data. For example, a team of Australian researchers have published a number of studies on the "readability" of forensic reports, defined as "the ease with which something can be read and understood due to the style of writing" [12–14].

However, fewer studies have focused on forensic reports and testimony in relation to either disciplinary standards or probabilistic reporting frameworks. With regard to the former, our earlier study of U.S. trial transcripts on friction ridge (fingerprint) evidence found a low degree of adherence to disciplinary standards [15]. Siegel, King, and Reed of the Laboratory Report Project examined forensic laboratory reports (but not courtroom testimony) in relation to both generic scientific reports from outside forensics as well as numerous extant disciplinary standards. While they found that most reports were fairly consistent with disciplinary standards, they also found that reports tended to deviate from disciplinary standards in two key areas: description of methods and discussion. In addition, the disciplinary standards themselves tended not to prescribe the inclusion of detailed procedures, data, or limitations [16]. Another recent study of forensic biological evidence in ten high-profile cases in Tasmania found a high degree of adherence to standards in courtroom testimony, but less adherence in written laboratory reports [17].

With regard to probabilistic reporting, our earlier study of U.S. trial transcripts on friction ridge evidence found that all testimony was reported categorically, rather than probabilistically [15]. While not a formal scientific study, a 2015 comprehensive review of trial transcripts from U.S. Federal Bureau of Investigation (FBI) expert testimony~~s~~ on microscopic hair comparison attracted a great deal of attention after revealing a shockingly high (96%) tendency to inappropriately overstate the probative value of the evidence ~~(96%)~~. The findings of the review were neither presented nor interpreted in terms of categorical versus probabilistic reporting; rather, they were presented in terms of "exceeding the limits of the science" [18–20]. Nonetheless, these findings may be interpreted as relevant to probabilistic reporting, insofar as what the FBI called "exceeding the limits of the science" essentially consisted of reporting what should have been an uncertain finding ("the hair is consistent with that of the defendant and with an unknown number of other potential sources") as a certain one ("the hair came from the defendant"). An Interpol

study of expert reports on voice identification also found a preference for categorical reports over probabilistic ones [21]. In contrast, Reid and Howes in their study of high-profile cases in Tasmania found high use of probabilistic reporting on forensic biology [15].

Most recently, Bali *et al.* analyzed a random sample of simulated laboratory reports from proficiency tests in eight forensic disciplines: fibres analysis, firearms examination, glass analysis, handwriting examination, paint analysis, questioned documents, shoeprints, and toolmarks [22]. The researchers found an overwhelming tendency toward categorical reporting in these proficiency test responses: "We found that the conclusion type which has received the most criticism in recent years (categorical statements) was the preferred expression in the overwhelming majority of responses." Indeed, they found probabilistic reporting was almost nonexistent. The study found no uses of random match probabilities or source probabilities. Likelihood ratios appeared in only half of one percent of all reports, were confined to only 1 of the 8 analyzed disciplines (glass analysis), and were rare even in that discipline. Likewise, likelihoods of observed similarity appeared in fewer than 1% of all reports, were confined to only 3 of the 8 analyzed disciplines (glass, handwriting, and firearms), and were rare even in those disciplines. In at least one of those three disciplines, the use of probability may have been an anomaly: a likelihood of observed similarity was used in only 0.8% of firearms reports, whereas a categorical conclusion was rendered in 96% of firearms reports [22].

The present study adds to this nascent body of research by measuring the current state of forensic evidence reporting across four of the most prominent forensic disciplines in the United States: friction ridge prints, firearms and toolmarks,[1] questioned documents, and shoeprints. In doing so, the study offers three main research contributions. First, the study is among the first to analyze the relationship between forensic disciplinary standards and courtroom testimony, in addition to laboratory reports of the type analyzed by the Laboratory Report Project. Second, in contrast to Bali *et al.* [22], the study includes reports and testimony on friction ridge prints, one of the most frequently used forensic evidence disciplines in the U.S. [23]. Lastly, for the disciplines included in the sample, the study is among the first to analyze the frequency of probabilistic reporting in actual forensic reports and testimony, in addition to simulated reports from proficiency tests.

The current study also contributes to the ongoing normative discussion among academics and statisticians regarding how forensic results *should* be reported to lawyers, judges, and juries. By providing a baseline understanding of how forensic experts are *currently* reporting their results, scholars can get a better sense of the degree to which contemporary reporting deviates from the ideal of probabilistic reporting, and what it might take to bring forensic evidence reporting closer to this ideal.

3 Materials & methods

3.1 Specifying disciplines

This project was sponsored by the Center for Statistics and Applications in Forensic Evidence (CSAFE). CSAFE has a mandate from the U.S. National Institute of Standards and Technology to advance statistical applications in forensic evidence in specific disciplines, defined by CSAFE as "in-scope." Therefore, this study focused only on four pattern disciplines defined as "in-scope" by CSAFE: friction ridge prints ~~("fingerprints")~~, firearms and toolmarks, questioned documents, and shoeprints. Although blood pattern evidence is also considered in-scope, we were unable to obtain sufficient data to include that discipline in this study. Our disciplinary coverage is, therefore, narrower than that of Bali *et al.* [22], who were able to cover eight disciplines. However, our study includes the widely used discipline of friction ridge prints, which their study omits.

3.2 Identifying disciplinary standards

Next, we identified a reporting standard for each discipline. For friction ridge prints we used the Scientific Working Group for Friction Ridge Study Analysis and Technology's (SWGFAST) Document #10, "Standards for Examining Friction Ridge Impressions and Resulting Conclusions." This standard allowed for three conclusions: Exclusion, Individualization, and Inconclusive [24]. For firearms and toolmarks, we used the Association of Firearm and Tool Mark Examiners (AFTE) Range of Conclusions. This standard allowed for four conclusions: Identification, Inconclusive, Elimination, and Unsuitable (we did not include the last conclusion in our analysis) [25]. For questioned documents, we used the Scientific Working Group for Forensic Document Examination (SWGDOC) Standard Terminology for Expressing Conclusions of Forensic Document Examiners. This standard allowed for nine conclusions: Identification, Strong probability, Probable, Indications, No conclusion, Indications did not, Probably did not, Strong probability did not, and Elimination [26]. For shoeprints, we used the Scientific Working Group for Shoeprint and Tire Tread Evidence (SWGTREAD) Range of Conclusions. This standard allowed for seven conclusions: Lacks sufficient detail, Exclusion, Indications of non-Association, Limited association of class characteristics, Association of class characteristics, High degree of association, and Identification (we did not include the first conclusion in our analysis) [27]. Appendix A provides a summary table with descriptions of the conclusion types from each of the standards documents used in the study.

It should be apparent that all of the prevailing standards were~~, in fact,~~ quite literally categorical. A statistically purist position would hold that all of the above categories should be done away with, and scientists should report along a continuum of probability (~~f~~e.g., [[Instruction: please hyperlink]28]). However, some forensic statisticians have also embraced the notion of "verbal scales" by which findings along a continuum of probability might be "translated" onto a "verbal scale" for communication with courts and juries [29]. The questioned document and shoeprint standards can be interpreted as enactments of just such a verbal scale.

3.3 Obtaining data sources

In order to ascertain the current state of forensic reporting in the United States, we would ideally have liked to obtain the following materials:

- A representative sample of forensic reports filed in all criminal cases in the United States
- A representative sample of transcripts of the testimony of forensic analysts in every criminal trial in the United States

Unfortunately, these materials are neither preserved nor collected in such a way that makes them easily obtainable as a single data set. Trial testimony in the United States is not even always transcribed. When it is transcribed, the transcript usually must be purchased from the court reporter who recorded it. Even when transcripts exist, they are not systematically archived, indexed, or made publicly available, making it difficult to obtain a sample of transcripts, let alone a representative sample, from a particular forensic discipline [30,31]. Laboratory reports, meanwhile, are probably even more difficult to obtain than transcripts (~~f~~but see [[Instruction: please hyperlink

]16]). They, likewise, are not systematically archived, indexed, or made publicly available. While forensic laboratories may possess copies of their own reports, they have no mandate or process to make them publicly available.

Under these circumstances, we employed a heterogeneous, opportunistic data collection approach. Our preference was to obtain transcripts and reports from the opportunistic collection of documents in the Westlaw database "Expert Materials." This is a small collection of expert trial transcripts, depositions, affidavits, and laboratory reports included in the Westlaw legal database. The materials are classified by discipline and include the following disciplines: DNA, fingerprint, firearms, handwriting, pathology, and voice. For this study, we used the materials on fingerprint, firearms, and handwriting. We eliminated civil cases and cases in which the source of the forensic evidence was not at issue in the case, leaving 118 reports[2] for inclusion in the sample. The materials appear to overrepresent the federal jurisdiction and some particular states (e.g., Michigan and California). Westlaw did not appear to follow any sort of systematic data collection strategy in compiling this database [30]. It must, therefore, be regarded as a non-random, opportunistic data set. Nonetheless this source was able to provide an adequate amount of data in three of our four disciplines.

In addition, we were able to supplement this collection strategy with an opportunistic sample of 73 transcripts and reports obtained from one consultant on friction ridge prints (an author of this paper) and two consultants on firearms and toolmarks. We are fairly confident that our sample is not skewed toward poorly formulated expert evidence. Our Westlaw source oversamples federal cases which tend to be, if anything, better litigated. Cases in which consultants were involved are likewise, if anything, likely to be better litigated.

These sources still left us without any data on the fourth discipline, shoeprints. For the shoeprint discipline, the Collaborative Testing Services (CTS) proficiency tests contain a report-writing exercise. The CTS reporting exercise has also been used as data on reporting by Howes *et al.* for glass analysis [14] and by Bali *et al.* for eight disciplines [22]. We used these proficiency test responses as our shoeprint data [32]. In contrast to our other data, the shoeprint data are~~is~~ simulated, rather than from real cases. Our shoeprint data does not overlap with Bali *et al.*'s because we used the 2017 CTS test, whereas they used the 2016 test. In addition, Bali *et al.* sampled the data, whereas we used the complete set. Therefore, our sample size for shoeprints (381) is larger than theirs (64).

Table 1 summarizes the sample sizes and sources of the data used in this study and gives the median year of the reports used in each discipline. One limitation of our data is that ~~it~~ some of it is not current. Our data will not necessarily capture changes in practice that occurred very recently.

Table 1

ⓘ The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Descriptive statistics of data sources.

Discipline	Report sample size	From Westlaw	From consultants	From CTS	Median year
Friction ridge Prints	91	41	50	0	2009
Firearm and Toolmarks	48	25	23	0	2003
Questioned Document	52	52	0	0	2006
Shoeprints	381	0	0	381	2017
Total	**572**	**118**	**73**	**381**	

To be sure, the data used here are~~is~~ heterogeneous and far from ideal. Whilst Bali *et al.* analyzed CTS simulated data across disciplines, allowing for homogeneity of data source across eight different disciplines [22], our study, in contrast, combined CTS simulated data with "real-world" data across a smaller number of disciplines. While the heterogeneity of our data is a liability and our opportunistic sampling methods limit the generalizability of our findings, the authenticity of our "real-world" data in three disciplines is an asset.

3.4 Analysis

We initially coded the "type" of each report within the framework of the disciplinary standard. For example, if the disciplinary standard states that the only possible conclusions are "Identification," "Inconclusive," and "Exclusion," then we coded our reports as to whether they were attempts to report "Identification," "Inconclusive," or "Exclusion." It should be noted that our data are~~is~~ quite heterogeneous in terms of the distribution of report types across disciplines. For example, we have almost no "Inconclusive" reports for friction ridge prints.

Next, we coded the words actually used in the report. These words were then aggregated into broader categories. For example, the statements "I identified the print to the defendant" and "It was an identification" were aggregated into a category called "Identification." We then coded whether the language used by the examiner to express the report adhered to the language used in the disciplinary standard for that report type.

Finally, we coded whether the report wa~~i~~s probabilistic. Our definition of "probabilistic" was simple: We asked whether the expert's report assigned any probability greater than zero but less than 1 to the alternate hypothesis that the source of the evidence was someone other than the suspect. This meant that all reports from the intermediate steps of the above conclusion scales (e.g., "Inconclusive" or "Strong probability") were coded as "probabilistic" despite being literally categorical. This definition provided a bright-line rule that facilitated easy classification and minimized the possibility of coding errors. The definition is also quite generous and therefore more likely to overstate than understate the degree to which probability factors into the decision-making of forensic practitioners. Nevertheless, as the results below demonstrate, few reports included in the sample could be considered probabilistic even under this more generous definition.

Finally, if reporting was probabilistic, we coded the form in which the probability was reported. For this step, we used a typology of probabilistic reporting found in the Guidance on Probability and Statistical Evidence for Judges, Lawyers, Forensic Scientists and Expert Witnesses published by the Royal Statistical Society in the United Kingdom [33]. It proposed the following types of forensic reports about evidence:

- Likelihood Ratio
- Random Match Probability
- Probability Inclusion
- Compound Probability of Exclusion
- Consistent with; Match
- Subjective Posterior Probability, Verbal Statement
- Subjective Posterior Probability, Numerical Statement
- Objective Posterior Probability, Verbal Statement
- Objective Posterior Probability, Numerical Statement
- Categorical Conclusion

It should be noted that Bali *et al.*'s study [22] employed a similar, though slightly different, typology derived from Thompson and Newman [34]. One curious attribute of the scale we used is the lumping together of "Consistent with" and "Match." From our perspective, these two types of reports are quite different. While testimony about a "Match" might be intended or interpreted to mean something equivalent to "Consistent with," it might also be intended or interpreted to means something much closer to what the scale calls a "Categorical Conclusion."

4 Results

4.1 Friction ridge

Table 2 shows the results for the friction ridge discipline. At the time of data collection, SWGFAST advocated a three-conclusion scale: "Individualization," "Inconclusive," and "Exclusion." The data set, which was approximately evenly split between Westlaw- and consultant-sourced reports, was overwhelmingly skewed toward the "Individualization" conclusion. Around three quarters of the reports adhered to the SWGFAST conclusions. The most common language used was "identified," although a small number of reports used the more technical term "individualized." The second most common language was to state that two impressions derived from the same source or the same person, or—somewhat nonsensically—that the two impressions themselves were "the same."

Table 2

Friction ridge (n = 91).

Report type according to standard	n	Actual language used	n	Adherence to standard	Probabilistic	Statistical Report Type	
						Consistent/match	Categorical
Exclusion	4	Exclusion	4	4	0	2	2
Inconclusive	1	Inconclusive	1	1	1	1	0
Individualization	86	Identified/Individualized	42	42	0	0	41
		Same source/person	27	27	0	0	27
		Match	8	0	0	7	1
		Possession	9	0	0	0	9
Total (%)	**91**		**91**	**74 (81%)**	**1 (1%)**	**10 (11%)**	**81 (89%)**

Smaller numbers of reports used the vague and disfavored term "match" or said that a questioned impression was possessed by the person of interest (e.g., "the print belonged to the defendant"). If taken literally, the latter is an odd claim for a forensic scientist to make, but we suggest that those scientists—whether consciously or not—assume that listeners will understand it to mean more or less the same thing that is conveyed by the standardized language of "Individualization": that the two impressions derive from the same source.

No "Individualization" reports were probabilistic. The only friction ridge report that was probabilistic was an "Inconclusive" report, which, as noted above, is inherently probabilistic. The reports were overwhelmingly categorical. An example is the following report dating from 2012:

> A: That print was identified to [person of interest (POI)] [35].

4.2 Firearm and toolmark

The leading standard in the firearm and toolmark discipline used the same three-conclusion scale as in the friction ridge discipline. The data set, which was smaller than the friction ridge data set and also approximately evenly split between Westlaw- and consultant-sourced reports, was more evenly distributed among the three

reports, though still skewed toward “Individualization” conclusions (Table 3).

Table 3

The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Firearms and toolmarks (n = 48).

Report type according to standard	n	Actual language used	n	Adherence to standard	Statistical Report Type				
					Probabilistic	Consistent/match	Subjective posterior verbal	Subjective posterior numerical	Categorical
Exclusion	6	Exclusion	6	6	0	0	0	0	6
Inconclusive	14	Inconclusive	14	14	14	5	8	1	0
Individualization	28	Identified	12	12	2	0	2	0	10
		Same source/person	16	16	0	0	0	0	16
Total (%)	**48**		**48**	**48 (100%)**	**16 (33%)**	**4 (8%)**	**10 (21%)**	**1 (2%)**	**32 (67%)**

The firearm and toolmark discipline showed perfect adherence to the standard. “Individualization” reports were nearly evenly split between statements about identification and statements about “same source.” Two-thirds of reports were categorical. An example, dating from 2016, is:

> Q: And, within a reasonable degree of ballistic certainty, what can you tell us about those shell casings?
>
> A: In my opinion, all of these shell casings were fired by the same firearm [35].

All of the probabilistic reports pertained to “Inconclusive” reports, which, as noted above, are inherently probabilistic. Most of these “Inconclusive” reports assigned subjective verbal posterior probabilities to the likelihood that the source of the evidence was someone other than the suspect. A small number of “Inconclusive” reports used the language of “consistency.”

Firearms and toolmarks also yielded the only numerical probability in the entire study for an “Inconclusive” report:

> I was not able to assign it a hundred percent certainty, but a, what we call an ‘entirely consistent,’ 95 percent certainty [36].

Some readers might question our coding of the above statement as “Inconclusive.” It might be argued that the expert was trying to render a probabilistic report of “Identification.” As noted above, our approach bound us to the prevailing disciplinary standards that we used. In this case, our task was made more difficult by the fact that the expert implied that the discipline provided for a report called “entirely consistent,” when, in fact, it does not.

However, given the definitions contained in the AFTE Range of Conclusions, we feel that the above statement must be considered “Inconclusive.” Although the language in the Range of Conclusions is not a model of clarity, we do not see how the above statement can be considered an “Individualization” report because that conclusion requires that “the extent of agreement exceed[~~d~~] that which can occur in the comparison of toolmarks made by different tools.” A report of “95 percent certainty” explicitly acknowledges a 5% probability that the observed agreement exists despite originating from *different* tools. But the definition of “Individualization” states the agreement observed in the instant case cannot exist between different tools.

To add to the confusion, the usage of the term “Inconclusive” in forensic science is wildly inconsistent [21]. While some understand the term to mean that the expert is in equipoise between “Identification” and “Elimination” or that the evidence has no probative value in either direction, the AFTE Range of Conclusions makes clear that its definition of “Inconclusive” is not that narrow. In particular, the Range specifies three subtypes of “I~~i~~nconclusive.” The first of those subtypes pertains to “some agreement ... but insufficient for an identification.” Our view, therefore, is that, according to the Range, the above statement must be coded as “I~~i~~nconclusive,” not “I~~i~~dentification.”

Two “Individualization” reports were coded as probabilistic. The first occurred in a 2011 case, where the expert testified:

> The probability [that the forensic evidence was produced by another source] is very very low. Which was stated in when I make identification [sic]. Anything is possible, anything. But the probability, due to all the variables that I mentioned before, is close to zero [37].

The second probabilistic “Individualization” report occurred in a 2016 case, where the expert testified:

> In this particular case, I examined eight cartridge cases. And they were--I identified them has [sic] having been fired from the same firearm.

Later, however, the expert elaborated:

> Q. Okay. So again, in response to my question, what are the chances you've made an error in the source attribution in this case?
>
> A. I would have to -- that would have to be a very subjective answer for myself. In my opinion, on this case, it is very low to a zero chance on this case [37].

In our effort to be as generous as possible in coding statements as probabilistic, we so coded the above statements. We assume, however, that many readers would have qualms about characterizing these statements as probabilistic. Both statements adopt the tactic of using a verbal characterization of probability as “very close to

zero," without actually quantifying precisely how far from zero this purported probability actually is (see [38]). And the second statement has already undermined the probability by stating that the two cartridges were fired from the same firearm.

4.3 Questioned documents

In contrast to friction ridge and firearms and toolmarks, the SWGDOC standard advocated a broader 9-conclusion scale at the time of data collection. The data were sourced entirely from Westlaw, and the reports were fairly well distributed across the 9-conclusion scale (Table 4). Adherence to the SWGDOC standard was almost perfect.

Table 4

ⓘ The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Questioned documents (n = 52).

Report type according to standard	n	Actual language used	n	Adherence to standard	Probabilistic	Statistical Report Type		
						Consistent/match	Subjective posterior verbal	Categorical
Elimination	10	Elimination	1	1	0			1
		Not same	9	9	0			9
Strong probability did not	1	May have	1	0	1		1	
Probably did not	2	Probably not	2	2	2		2	
Indications did not	2	Indications did not	2	2	2		2	
No conclusion	5	No conclusion	3	3	3		3	
		Could not determine	2	2	2		2	
Indications	1	May have	1	0	1	1		
Probable	7	Probably	7	7	7		7	
Strong probability	7	Strong/highly probable~~ly~~/likely	7	7	7		7	
Individualization	17	Identified	6	6	0			6
		Same source/person	11	11	1		1	10
Total (%)	**52**		**52**	**50 (96%)**	**26 (50%)**	**1 (2%)**	**25 (48%)**	**26 (50%)**

Precisely half of the reports were probabilistic. This result was largely a function of the fact that 7 of the 9 reports on the scale were considered inherently probabilistic, as noted above. Almost half of the reports assigned subjective verbal posterior probabilities, such as the following report:

> An examination was then performed using the known handwriting of [POI] to the signatures contained in Q-1 and Q-2 and in as much as possible to identify from copies, it is the opinion of this examiner that the written signature, [POI] where it appears on Q-1 and Q-2 demonstrates a strong probability that the signatures were written by [POI] [39].

There was one use of the language of "consistency." There was also one probabilistic expression of an "Individualization" report:

> Based on my knowledge, experience, and training and taking into consideration any strengths or weaknesses of my analysis, I find that the Questioned Document was produced by the same writer of the Known samples for [POI]. It is with a high degree of confidence that I state [POI] produced the signature on the Questioned Document [40].

The remaining half of the questioned document reports were categorical, and all of these categorical reports were either "Eliminations" or "Individualizations." An example, dating from 2011, is:

> It was determined that the non-overwritten questioned writing on specimen Q1 was prepared by [POI] known writer of specimen K1 [41].

4.4 Shoeprints

The shoeprint discipline advocated a 7-conclusion scale at the time of data collection. The data were sourced entirely from CTS tests, and consequently the sample size was much larger than for the other three disciplines. The data were overwhelmingly skewed toward the extremes of the scale: "Exclusion" and "Individualization" (Table 5). Adherence to the standard was high, but examiners used a wide variety of language in their reports. Very few reports (13%) were probabilistic. What probabilistic reports there were derived entirely from the intermediate 4 reports on the scale, which, as noted above, are inherently probabilistic.

Table 5

The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Shoeprints (n = 381).

Report type according to standard	n	Actual language used	n	Adherence to standard	Probabilistic	Statistical Report Type			
						Consistent/match	Likelihood ratio	Subjective verbal probability	Categorical
Exclusion	168	Exclusion	46	46	0	0	0	0	46
		Not same source	23	17	0	1	0	0	22
		Not made	75	70	0	0	0	0	75
		Eliminated	24	4	0	0	0	0	24
Indications of non-association	2	Indications of non-association	2	2	2	2	0	0	0
Limited association	6	Limited association	1	1	1	1	0	0	0
		Similar	4	0	4	2	0	2	0
		Consistent/correspondence	1	0	1	1	0	0	0
Association	4	May have	2	0	2	0	0	2	0
		Similar	2	0	2	1	0	1	0
High degree of association	36	High degree of association	17	15	17	15	0	2	0
		Extremely/very strong support	6	0	6	0	6	0	0
		Consistent/Correspondence	7	0	7	7	0	0	0
		(Very) likely/probably	4	0	4	0	0	4	0
		Similar	2	0	2	2	0	0	0
Identification	165	Identification	62	61	0	0	0	0	62
		Same source	24	24	0	0	0	0	24
		Made	79	74	0	0	0	0	79
Total (%)	**381**		**381**	**314 (82%)**	**48 (13%)**	**32 (8%)**	**6 (2%)**	**11 (3%)**	**332 (87%)**

Reports were overwhelmingly categorical. An example is:

> The questioned imprints, identified “Q2, Q5″ were made by the right suspect shoes, identified"-"R"“. The questioned imprint, identified Q7 was made by the left suspect shoe, identified"-"L"“ (Response #47HPNG).

A small number of reports used the language of “consistency.” The shoeprint data also contained the only uses of the likelihood ratio in the entire study. A small pocket (2%) of shoeprint reports~~examiners~~ used likelihood ratio language (“extremely/very strong support”) for the “high degree of association” conclusion, although these reports did not involve numerical estimations of this likelihood. The following is a representative example of such reports:

> The above findings provide extremely strong support for the view that the right runner K1, rather than other footwear, made the impression Q2 from the scene. The findings provide very strong support for the view that the left runner K1, rather than other footwear, made the impression Q7 from the scene. They also provide strong support for the view that the right runner K1, rather than another runner, made the impression Q5 from the scene. Taken together these findings provide extremely strong support for the view that the runners K1 made some of the impressions at the scene (Response #MW7ZWD).

4.5 Summary

Table 6 summarizes all four disciplines. As noted above, adherence to standards was generally high. Probabilistic reporting was quite rare, except for reports for the intermediate conclusions on the conclusion scales, which, as noted above, were inherently probabilistic. Reports were overwhelmingly categorical, except in the questioned documents discipline. The next most commonly used language types were “consistency” and subjective verbal posterior probabilities. Numerical probabilities and likelihood ratios, whether numerical or verbal, were extremely rare. Random Match Probabilities, probabilities of inclusion, compound probabilities of exclusion, objective (that is, data-based) probabilities of any kind, and numerical likelihood ratios were entirely absent from the ~~entire~~ study sample.

Table 6

The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Summary.

Discipline	n	Adherence to standard	Probabilistic	Statistical report type				
				Consistent/match	Likelihood ratio	Subjective verbal	Subjective numerical	Categorical
Friction Ridge	91	74%	1%	11%				89%
Firearm/toolmark	48	100%	33%	8%		21%	2%	67%
Questioned documents	52	96%	50%	2%		48%		50%
Shoeprints	381	82%	13%	32%	2%	3%		87%

5 Discussion

5.1 Adherence to standards

Most reports generally adhered to standards. This was especially true in the firearms and toolmark discipline, in which all reports used the language mandated by the standard. But the finding also holds for friction ridge prints, questioned documents, and shoeprints. Over 80% of reports in each of these disciplines adhered to prevailing practitioner standards.

There was, however, great variety in terms of the actual words used in testimony and reports. Experts frequently used casual, vernacular terms and phrases. Common ones were:

- X “was identified to” Y.
- X “was identical to” Y.
- X “made” Y

And, of course, the notorious term “match” and phrase “consistent with” appeared frequently. This is consistent with Reid and Howes [17], which found high use of the term “match” in testimony about forensic biology in Tasmania.

5.2 Probabilistic reporting

Probabilistic reporting remains rare in the United States in these four disciplines. We found relatively few probabilistic reports. The probabilistic reports that we did find were almost entirely “Inconclusive” reports which, by their very nature, were always coded as probabilistic. As noted above, there were three probabilistic “Individualization” reports, two for firearms and toolmarks and one for questioned documents. In all three cases, the expert combined a non-probabilistic statement of certainty with a probabilistic statement. Hence, our results suggest that practitioners in these disciplines are generally unwilling to issue probabilistic reports for results other than “Inconclusive.” This finding is consistent with that of Bali *et al.* [28], who found probabilistic reports virtually nonexistent in the CTS tests, and actually nonexistent in 5 of the 8 disciplines analyzed (the exceptions being glass, handwriting, and firearms).

Indeed, we found that on the rare occasions in which probability was discussed, it was as often to dismiss or denigrate probability as it was to employ it. For example, in a 2011 trial ~~in Louisiana~~, a firearm and toolmark examiner testified as follows:

> Q: Yes, Ma’am. So if we’re doing – Do you deal with statistics?
>
> A: Absolutely not.

The witness followed that comment up by saying

> A: They match – No. This is important, because you’re putting statistics where it doesn’t belong [42].

Despite having stated that she does not use statistics, the same witness then goes on to assign a probability to the hypothesis that the known tool is *not* the source of the unknown mark:

> THE WITNESS: The probability is very very low. Which was stated in when I make identification. Anything is possible, anything. But the probability, due to all the variables that I mentioned before, is close to zero [42].

Consider as well the following testimony by a friction ridge print examiner ~~in Washington State~~ in 2009:

> We don’t come in with okay, there’s a probability of, you know, 100,000 to 1 or whatever. It’s either an exclusion, an individualization, or there’s not enough information in the print to include or exclude. There’s no probability [43].

5.3 Types of probability

The Royal Statistical Society’s elaborate typology of ways of reporting probabilities in forensic science [33] speaks to its expectation that forensic scientists have a varied toolkit available to them for reporting results probabilistically. Anyone expecting to see forensic practitioners in the studied disciplines drawing on these variety of tools, however, will be disappointed. Likelihood ratios, and numerical probabilities went largely unused. Forensic statisticians will be particularly disappointed by the scant use of likelihood ratios. The exception for likelihood ratios was shoeprints, in which likelihood ratios were used in a small minority (2%) of reports. Match probabilities, probabilities of inclusion, compound probabilities of exclusion, and objective probabilities of any kind went entirely unused.

Instead, reports were usually categorical or invoked the language “consistent with/match.” When probability was used, it overwhelmingly referred to posterior probabilities, rather than the likelihood of the evidence ~~ratios~~. Those posterior probabilities, meanwhile, were overwhelmingly subjective, rather than objective. Those subjective posterior probabilities, in turn, were overwhelmingly verbal, rather than numerical. This finding is consistent with that of Bali *et al.* [28], who found that

probability statements on proficiency exams came either in the form of likelihood ratios or likelihoods of observed similarity and were exceedingly rare, appearing in less than 2% of reports in their sample.

There was only one numerical probability in the entire study--for an “Inconclusive” firearms and toolmarks report in which the examiner reported “95 percent certainty.” It should be noted that while the “95 percent” figure is indeed a number, we are dubious that it represents the output of any sort of calculation. Indeed, the testimony did~~oes~~ not offer any suggestion that this number hads any formal basis other than ~~as~~ an expression of the expert witness’s subjective state of confidence. This is why we coded it as a “subjective,” rather than an “objective” posterior probability.

Although the above statement was the only quantitative one, several other “Inconclusive” firearm and toolmark reports in our data set were similarly expressions of near-certainty. For example:

> I couldn't see enough individual markings to say positively, although there were several markings that were close enough that it took quite awhile [sic] to make a determination. I couldn't say a hundred percent, but it was very close [44].

To be sure, the confusion that arises regarding near-certainty reports is a liability of the kind of three-conclusion scale embodied by the AFTE Range of Conclusions. Had the above statement pertained to shoeprints, for example, it could have easily been accommodated by the “High degree of association” report. To be fair, the Firearms and Toolmarks Subcommittee of the National Institute of Standards and Technology Organization of Scientific Area Committees has now promulgated a proposed reporting standard consisting of a five-conclusion scale [45].

6 Conclusion

This study sought to measure the extent of probabilistic reporting in forensic pattern disciplines in the United States to establish a baseline from which progress could be measured. Probabilistic reporting of the type advocated by forensic statisticians appeared very infrequently in the four disciplines studied. When it comes to probabilistic reporting, there is almost nowhere to go but up. While there is a great deal of discussion of probabilistic reporting in the scholarly literature, on governmental and scientific commissions, and in standards developing organizations, there is a long way to go before probabilistic reporting becomes common in US courtrooms and laboratory reports.

Funding

This research was funded in part by the Center for Statistical Applications in Forensic Evidence (CSAFE), sponsored by the National Institute of Standards and Technology (NIST).

Cr~~R~~edit~~T~~ authorship contribution statement

Simon A. Cole: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Matt Barno:** Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing.

Appendix A . Standards Table.[Instruction: There are indents at the start of the text in each cell for the Friction Ridge Prints, Firearms and Toolmarks, and Questioned Documents sections of the Appendix (curiously, all of the cells for the Shoeprints section were fully left-aligned). Those indents should be removed, and we could not figure out how to do that in the table editor. Also, it would be great if the lefthand column of the table could be widened so that words are not split across two verticle lines in the final proof.]

ⓘ The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

Friction Ridge Prints (Fingerprints): Scientific Working Group for Friction Ridge Study Analysis and Technology, “Document 10: Standards for Examining Friction Ridge Impressions and Resulting Conclusions”	
Individualization	“Individualization of an impression to one source is the decision that the likelihood the impression was made by another (different) source is so remote that it is considered as a practical impossibility.”
Inconclusive	“[T]he inconclusive conclusion means that the impression needs to be reexamined using clearly and completely recorded known impressions...[or] the inconclusive conclusion means that the unknown impression was neither individualized nor excluded as originating from the same source.”
Exclusion	“Exclusion is the decision by an examiner that there are sufficient features in disagreement to conclude that two areas of friction ridge impressions did not originate from the same source.”
Firearms and Toolmarks: Association of Firearm and Tool Mark Examiners, “Range of Conclusions”	
Identification	“Agreement of a combination of individual characteristics and all discernible class characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool.”
Inconclusive	“a. Some agreement of individual characteristics and all discernible class characteristics, but insufficient for an identification. b. Agreement of all discernible class characteristics without agreement or disagreement of individual characteristics due to an absence, insufficiency, or lack of reproducibility.

	c. Agreement of all discernible class characteristics and disagreement of individual characteristics, but insufficient for an elimination.”
Elimination	“Significant disagreement of discernible class characteristics and/or individual characteristics.”
Unsuitable	“Unsuitable for examination.”
Questioned Documents: Scientific Working Group for Forensic Document Examination, “SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners”	
Identification	“The examiner has no reservations whatever, and although prohibited from using the word ‘fact,’ the examiner is certain, based on evidence contained in the handwriting, that the writer of the known material actually wrote the writing in question.”
Strong Probability	“[T]he evidence is very persuasive, yet some critical feature or quality is missing so that an identification is not in order; however, the examiner is virtually certain that the questioned and known writings were written by the same individual.”
Probable	“[T]he evidence contained in the handwriting points rather strongly toward the questioned and known writings having been written by the same individual; however, it falls short of the“ virtually certain” degree of confidence.
Indications (Evidence to Suggest)	“[A] body of writing has few features which are of significance for handwriting comparison purposes, but those features are in agreement with another body of writing.”
No Conclusion (Totally Inconclusive, Indeterminable)	“This is the zero point of the confidence scale. It is used when there are significantly limiting factors, such as disguise in the questioned and/or known writing or a lack of comparable writing, and the examiner does not have even a leaning one way or another.”
Indications Did Not	“[T]his carries the same weight as the indications term that is, it is a very weak opinion.”
Probably Did Not	“[T]he evidence points rather strongly against the questioned and known writings having been written by the same individual, but, as in the probable range above, the evidence is not quite up to the ‘virtually certain’ range.”
Strong Probability Did Not	“[T]his carries the same weight as strong probability on the identification side of the scale; that is, the examiner is virtually certain that the questioned and known writings were not written by the same individual.”
Elimination	“[T]his, like the definite conclusion of identity, is the highest degree of confidence expressed by the document examiner in handwriting comparisons. By using this expression the examiner denotes no doubt in his opinion that the questioned and known writings were not written by the same individual.”
Shoeprints: Scientific Working Group for Shoeprint and Tire Tread, “Evidence Range of Conclusions Standard for Footwear and Tire Impression Examinations (03/2013)”	
Identification	“In the opinion of the examiner, the particular known footwear or tire was the source of, and made, the questioned impression. Another item of footwear or tire being the source of the impression is considered a practical impossibility.”
High Degree of Association	“In the opinion of the examiner, the characteristics observed exhibit strong associations between the questioned impression and known footwear or tire; however, the quality and/or quantity were insufficient for an identification.”
Association of Class Characteristics	“In the opinion of the examiner, the known footwear or tire is a possible source of the questioned impression and therefore could have produced the impression. Other footwear or tires with the same class characteristics observed in the impression are included in the population of possible sources.”
Limited Association of Class Characteristics	“Some similar class characteristics were present; however, there were significant limiting factors in the questioned impression that did not permit a stronger association between the questioned impression and the known footwear or tire.”
Indications of Non-Association	“The questioned impression exhibits dissimilarities when compared to the known footwear or tire; however, the details or features were not sufficiently clear to permit an exclusion.”
Exclusion	“Sufficient differences were noted in the comparison of class and/or randomly acquired characteristics between the questioned impression and the known footwear or tire. In the opinion of the examiner, the particular known footwear or tire was not the source of, and did not make, the impression.”
Lacks Sufficient Detail	“No comparison was conducted: the examiner determined there were no discernible questioned footwear/tire impressions or features present. This opinion applies when there is insufficient detail to conduct any comparison.”

References

The corrections made in this section will be reviewed and approved by a journal production editor. The newly added/removed references and its citations will be reordered and rearranged by the production team.

[1] NRC, Strengthening Forensic Science in the United States: A Path Forward, National Research Council, National Academies Press, Washington, Committee on Identifying the Needs of the Forensic Science Community, 2009.

[2] ~~S.A. Cole, Forensics without Uniqueness, Conclusions without Individualization: The New Epistemology of Forensic Identification, Law, Probability and Risk, 8 (2009) 233–255.~~Cole S.A., Forensics without Uniqueness, Conclusions without Individualization: The New Epistemology of Forensic Identification, Law, Probability and Risk 8 (2009) 233–255.

[3] Koehler J.J., Saks M.J., Individualization claims in forensic science: still unwarranted, Brooklyn Law Rev. 75 (2010) 1187–1208.

[4] Biedermann A., Bozza S., Taroni F., Decision theoretic properties of forensic identification: underlying logic and argumentative implications, For. Sci. Int. 177 (2008) 120–132.

[5] Thompson W.C., How should forensic scientists present source conclusions?, Seton Hall Law Rev. 48 (2018) 773–814.

[6] Broeders A.P.A., Of earprints, fingerprints, scent dogs, cot deaths and cognitive contamination–a brief look at the present state of play in the forensic arena, For. Sci. Int. 159 (2006) 148–157.

[7] Mnookin J.L., Cole S.A., Dror I.E., Fisher B., Houck M.M., Inman K., Kaye D.H., Koehler J.J., Langenburg G., Risinger D.M., Rudin N., Siegel J.A., Stoney D.A., The need for a research culture in the forensic sciences, UCLA L. Rev. 58 (2011) 725–780.

[8] ~~S.P. Lund, H. Iyer, Likelihood Ratio as Weight of Forensic Evidence: A Metrological Perspective, arXiv:1608.07598v2 [stat.AP], (2016).~~Lund, S.P., Iyer, H., Likelihood Ratio as Weight of Forensic Evidence: A Metrological Perspective, arXiv:1608.07598v2 [stat.AP], (2016).

[9] Morrison G.S., Enzinger E., What should a forensic practitioner's likelihood ratio be?, Sci. Justice 56 (2016) 374–379.

[10] Aitken C.G.G., Taroni F., Statistics and the Evaluation of Evidence for Forensic Scientsts, second ed., Wiley, Chichester, 2004.

[11] Biedermann A., Bozza S., Taroni F., Aitken C.G.G., Reframing the debate: a question of probability, not of likelihood ratio, Sci. Justice 56 (2016) 392–396.

[12] Howes L.M., Julian R., Kelty S.F., Kemp N., Kirkbride K.P., The readability of expert reports for non-scientist report-users: report of DNA analysis, For. Sci. Int. 237 (2014) 7–18.

[13] Howes L.M., Kirkbride K.P., Kelty S.F., Julian R., Kemp N., The readability of expert reports for non-scientist report-users: reports of forensic comparison of glass, For. Sci. Int. 236 (2014) 54–66.

[14] Howes L.M., Kirkbride K.P., Kelty S.F., Julian R.D., Kemp N., Forensic scientists' conclusions: how readable are they for non-scientist report-users?, For. Sci. Int. 231 (2013) 102–112.

[15] Cole S.A., Where the rubber meets the road: thinking about expert evidence as expert testimony, Villanova Law Rev. 52 (2007) 803–842.

[16] Siegel J.A., King M., Reed W., The laboratory report project, Forensic Sci. Policy Manage. 4 (2013) 68–78.

[17] Reid C.A., Howes L.M., Communicating forensic scientific expertise: an analysis of expert reports and corresponding testimony in tasmanian courts, Sci. Justice 60 (2020) 108–119.

[18] Reimer N.L., The hair microscopy review project: an historic breakthrough for law enforcement and a daunting challenge for the defense bar, The Champion (2013) 16.

[19] Cole S.A., Duster T., Microscopic hair comparison and the sociology of science, Contexts 15 (1) (2016) 28–35.

[20] ~~S.S. Hsu, FBI Admits Flaws in Hair Analysis over Decades, in: Post, Washington, 2015.~~Hsu, S.S., FBI Admits Flaws in Hair Analysis over Decades, Washington Post (2015).

[21] Morrison G.S., Sahito F.H., Jardine G., Djokic D., Clavet S., Berghs S., Dorny C.G., INTERPOL survey of the use of speaker identification by law enforcement agencies, For. Sci. Int. 263 (2016) 92–100.

[22] Bali A.S., Edmond G., Ballantyne K.N., Kemp R.I., Martire K.A., Communicating forensic science opinion: an examination of expert reporting practices, Sci. Justice 60 (2020) 216–224.

[23] ~~M.R. Durose, Census of Publicly Funded Forensic Crime Laboratories, 2005, Department of Justice, Bureau of Justice Statistics Bulletin, Office of Justice Programs, NCJ 222181, 2008.~~Durose, M.R., Census of Publicly Funded Forensic Crime Laboratories, 2005, Department of Justice, Bureau of Justice Statistics Bulletin, Office of Justice Programs, NCJ 222181, 2008.

[24] SWGFAST, Standards for Examining Friction Ridge Impressions and Resulting Conclusions, Scientific Working Group on Friction Ridge Analysis Study and Technology, ver. 1.0, Sept. 13, 2011, http://www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf.

[25] Association of Firearm and Tool Mark Examiners, Range of Conclusions, AFTE J., 43 (2011).

[26] SWGDOC, Standard Terminology for Expressing Conclusions of Forensic Document Examiners, S.W.G.f.F.D. Examination, 2013, http://www.swgdoc.org/images/documents/standards/SWGDOC%20Standard%20Terminology%20for%20Expressing%20Conclusions%20of%20Forensic%20Document%20Examiners%20150114.pdf.[Instruction: Please replace "S.W.G.f.F.D. Examination" with "Scientific Working Group for Forensic Document Examination"
]

[27] SWGTREAD, Range of Conclusions, S.W.G.f.S.a.T.T. Evidence, 2013, http://www.swgtread.org/images/documents/standards/published/swgtread_10_conclusions_range_201303.pdf.[Instruction: Please replace "S.W.G.f.S.A.T.T. Evidence" with "Scientific Working Group for Shoeprint and Tire Tread Evidence"
]

[28] Champod C., Evett I.W., A probabilistic approach to fingerprint evidence, J. Foren. Identification 51 (2001) 101–122.

[29] Martire K.A., Watkins I., Perception problems of the verbal scale: a reanalysis and application of a membership function approach, Sci. Justice 55 (2015) 264–273.

[30] S.A. Cole, W.C. Thompson, B. Valazquez, Assessing the Feasibility of Building a Database of Trial Transcripts Containing Scientific Testimony, Project on Scientific Knowledge and Public Policy, Report, 2007.[Instruction: Please reformat author names with surname first.
]

[31] Koppel G., A tale of two counties: divergent responses in Los Angeles and orange county superior courts to the ban on electronic recording in California court reporters association v. Judicial Council, San Diego Law Review 37 (2000) 47–100.

[32] Collaborative Testing Services, Footwear Imprint Evidence Test No. 17-5331/2/5 Summary Report, 2017.

[33] ~~C. Aitken, P. Roberts, G. Jackson, Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society, London, Communicating and Interpreting Statistical Evidence in the Administration of Justice, 2010.~~Aitken C., Roberts P., Jackson G., Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society, London, Communicating and Interpreting Statistical Evidence in the Administration of Justice (2010).

[34] Thompson W.C., Newman E.J., Lay understanding of forensic statistics: evaluation of random match probabilities, likelihood ratios, and verbal equivalents, Law Human Behavior 39 (2015) 332–349.

[35] United States v. Harris (2016) 2:15-cr-80(2)

[36] United States v. Andres Espinoza-Torres (2001) 2001 WL 35807683.

[37] People v. Cedric Poore (2016) Testimony of Jeffrey Jerek Brown.

[38] ~~S.A. Cole, Indvidiualization Is Dead, Long Live Individualization! Reforms of Reporting Practices for Fingerprint Analysis in the United States, Law, Probability and Risk, 13 (2014) 117–150.~~Cole S.A., Indvidiualization Is Dead, Long Live Individualization! Reforms of Reporting Practices for Fingerprint Analysis in the United States, Law, Probability and Risk 13 (2014) 117–150.

[39] United States v. Wingrove Edward Michael (2007) 2007 WL 7228639.

[40] People v. Sherman Buggs (2007) 2007 WL 7266599.

[41] United States v. Ihenacho (2011) 2011 WL 9154359.

[42] State v. Bryant Boudoin (2011) 11-KA-967.

[43] State v. Kenneth Sims (2009) No. 08-1-06912-1 KNT

[44] United States v. Jesse Leahy (2004) 2004 WL 5653985.

[45] Organization for Scientific Area Committees (OSAC) for Forensic Science, Standard Scale of Source Conclusions and Criteria for Toolmark Examinations, National Institute of Standards and Technology, 1.0, n.d., https://www.nist.gov/document/range-source-conclusions-and-criteria-toolmark-examinations.

Footnotes

Text Footnotes

[1] Bali *et al.* treat firearms and toolmarks analysis as two separate disciplines, presumably because the distributor of the proficiency tests used in their sample offers separate tests in each of these two areas. However, because practitioners view these two forms of analysis as falling under the same disciplinary umbrella (see https://afte.org/), we treat firearms and toolmarks as a single discipline.

[2] We use the term “report” ~~to describe opinions delivered in both laboratory reports and courtroom testimony included in our sample. It has increasingly become standard in forensics to use the term~~ to describe opinions delivered in both laboratory reports and courtroom testimony. It has increasingly become standard in forensics to use the term “report” to describe opinions rendered by forensic examiners regardless of format. We also find it among the most neutral of competing terms like “conclusion,” “opinion,” “decision,” “determination,” “result,” and so on.

Highlights

- Forensic evidence reporting shows a high degree of adherence to prevailing disciplinary standards.
- Probabilistic reporting of forensic results remains rare.
- Probabilistic reports were mostly subjective verbal assignments of posterior probabilities.

Queries and Answers

Query: Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact PA.Crabtr@elsevier.com immediately prior to returning your corrections.

Answer: Not special issue

Query: The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.

Answer: Yes

Query: Please check whether the affiliation is set correctly, and correct if necessary.

Answer: Correct

Query: As per journal instructions, KEYWORD is necessary. Kindly provide the same.

Answer: This is not where keywords belong. Keywords were provided through online submission system

Query: Have we correctly interpreted the following funding source(s) and country names you cited in your article: Center for Statistical Applications in Forensic Evidence (CSAFE), United States? National Institute of Standards and Technology (NIST), United States? /

Answer: Yes