RTICLE IN PRES

Forensic Science International: Synergy xxx (xxxx) xxx



Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: https://www.journals.elsevier.com/ forensic-science-international-synergy/



(Mis)use of scientific measurements in forensic science

Itiel E. Dror^{a,*}, Nicholas Scurich^b

^a University College London (UCL), 35 Tavistock Square, London, WC1H 9EZ, USA ^b University of California, Irvine, 4312 Social and Behavioral Sciences Gateway, Irvine, CA, 92697, USA

ARTICLE INFO

Article history: Received 9 July 2020 Received in revised form 21 August 2020 Accepted 22 August 2020 Available online xxx

Keywords: Error rates Daubert Forensic science Inconclusive decisions Expert decision making

ABSTRACT

Forensic science error rate studies have not given sufficient attention or weight to inconclusive evidence and inconclusive decisions. Inconclusive decisions can be correct decisions, but they can also be incorrect decisions. Errors can occur when inconclusive evidence is determined as an identification or exclusion, or conversely, when same- or different-source evidence is incorrectly determined as inconclusive. We present four common flaws in error rate studies: 1. Not including test items which are more prone to error; 2. Excluding inconclusive decisions from error rate calculations; 3. Counting inconclusive decisions as correct in error rate calculations; and 4. Examiners resorting to more inconclusive decisions during error rate studies than they do in casework. These flaws seriously undermine the credibility and accuracy of error rates reported in studies. To remedy these shortcomings, we present the problems and show the way forward by providing a corrected experimental design that quantifies error rates more accurately. © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Sciences are based on measurement of data, that are observable, quantifiable, and replicable. If the data are not quantifiable, then the ability to conduct experimentation, find regularities, falsify hypotheses, and make predictions is undermined [1]. Without quantification, science is restricted, perhaps even non-existent, as science requires quantification to measure the phenomena it is investigating. Not only is quantification a basic requirement to conduct scientific inquiry, but it is also critical for communicating the findings. This is especially important in a domain such as forensic science, where science is used as evidence in court [2]. The fair administration of justice requires that science is accurately and effectively communicated to the fact finders (e.g. Ref. [3]).

One critical measurement metric in all sciences, and in forensic science in particular, are error rates -the topic of this article. Knowing the error rates in a particular forensic domain is a vital measurement needed to ascertain the weight of the evidence. The appropriate weight of the evidence cannot be known without some sense of the rates at which the technique errs [4]. Despite that, most forensic domains do not have properly established error rates [5–9]. As noted by the National Academy of Sciences [10] "The assessment of the accuracy of the conclusions from forensic

errors [13].

of judges find that they overwhelmingly lack an understanding of

Corresponding author. E-mail address: i.dror@ucl.ac.uk (I.E. Dror).

https://doi.org/10.1016/i.fsisyn.2020.08.006

Please cite this article as: I.E. Dror, N. Scurich, (Mis)use of scientific measurements in forensic science, Forensic Science International: Synergy, https://doi.org/10.1016/j.fsisyn.2020.08.006

analyses and the estimation of relevant error rates are key components of the mission of forensic science" (p. 122), but, "In most areas of forensic science, no well-defined system exists for determining error rates" (p. 188). Hence, when forensic experts claim that a fingerprint found at the crime scene is that of the suspect, or that the gun found in the defendant's home is the gun which fired the cartridge case found at the crime scene, there is a problem in assessing the probative value of that conclusion.

Not only do properly established error rates not exist in many forensic domains, but forensic experts have appeared in court claiming their error rates were zero, and that they and their techniques are infallible [11]. Such claims are not only unfounded, but also, as stated by the National Academy of Sciences, "claims that these analyses have zero error rates are not scientifically plausible" ([10]; p. 142). Indeed, there have been documented cases of erroneous identifications in fingerprinting as well as in firearms, and more than half of the known wrongful conviction cases have included flawed forensic science evidence [12]. Forensic experts have unsubstantiated, unrealistic and false beliefs regarding their

The "known or potential rate of error" is important information for considering the weight of the evidence, and it is one of the factors judges need to consider for legal admissibility of expert testimony under Daubert v. Merrill Dow Pharmaceuticals ([14], p. 594). Although some analyses suggest that judges do spend time analysing error rates in making admissibility decisions [15], surveys

²⁵⁸⁹⁻⁸⁷¹X/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4. 0/).

ARTICLE IN PRESS

error rates and scientific testing [16]. Forensic science techniques have repeatedly passed the Daubert standard for admissibility, even when they have no properly established error rates and even when experts have implausibly claimed that the error rate is zero (e.g., in Ref. [17]; see Refs. [18]).

In contrast to DNA that uses probabilities, or handwriting and other forensic domains that use verbal scales with various degrees (e.g., the items 'probably' or 'very probably' match), fingerprinting and firearms experts most often testify in categorical terms: They either conclude that there is a definite identification, a definite exclusion, or, that they cannot determine ('inconclusive' –see Ref. [19]). Even with expansion of categorical decisions to a verbal scale with various degrees, the probative weight of these conclusions cannot be understood without some sense of how often the conclusions are erroneous. The lack of appropriately established error rates, coupled with experts testifying in highly-confident terms about their conclusions, can cause jurors to be misled about the strength of the forensic evidence [20].

Establishing error rates is not a simple matter [21]. There are many complex practical and theoretical challenges, from knowing the ground truth and establishing appropriate representative databases (e.g. Ref. [22]), and which ways to calculate error rates [23], to issues of ecological validity and fortuitous factors that contribute to errors (e.g. Refs. [24,25]). Furthermore, error rates may give insights into forensic domains in general, but may say very little about a specific examiner's decision in a particular case [21] –a specific examiner, in a specific case, may have a potential error that is much lower (or much higher) than the general error rate in their field.

When considering a general error rate, there are many issues that may impact error rates in casework that are often not accounted for in error rate studies. As a result, the reported error rates in such studies may be misleading. These may include: 1. Dismissing and not counting errors because they are regarded as 'clerical errors,' 2. Selectively publishing only studies that reveal low error rates, 3. Not mimicking the realities of casework which can further increase errors, such as stress and bias [26,27], 4. Not including inconclusive evidence as test items, and 5. Never counting inconclusive decisions as potential errors (either excluding the inconclusive decisions from calculations, or scoring them as correct decisions). In this article we focus on how the latter two factors have been misused in establishing error rates in the two major forensic domains of fingerprint and firearms.

2. Inconclusives

A central factor in determining an error rate is what counts as an error. In the widely used forensic domains of fingerprinting and firearms, forensic experts can reach a conclusion of 'inconclusive' [28]. Clearly, some cases warrant such an inconclusive decision: when the quantity or quality of the information in the evidence is insufficient to reach a conclusion of identification or exclusion, an inconclusive decision is correct [21]. In fact, such correct inconclusive decisions reflect good meta-cognitive abilities, and even modesty, on the part of the forensic experts. However, an inconclusive decision can also be an erroneous decision. This is the case when the evidence does contain sufficient quantity and quality of information but is nevertheless determined as inconclusive rather than an identification or an exclusion (see Ref. [19]).

It is important to realize that even though each latent forensic evidence found at the crime scene either was, or was not, left by the suspect (i.e., the ground truth of who deposited the mark), that does not necessarily mean that there is sufficient information in the latent evidence to justify reaching a definite conclusion (identification or exclusion). For example, when the evidence is degraded or damaged to a degree that there is very little-to-no information. In such cases an inconclusive decision is the only correct decision (see detailed discussion in Ref. [21]). Forensic examiners should be able to reach the appropriate and correct decision given the quality and quantity of information that is in the evidence. If not, then they are making an error in judgment, an error, which needs to be counted when measuring error rates.

These errors include reaching an identification (or exclusion) decision where there is insufficient information to justify such a decision; or conversely, reaching an inconclusive decision when there is sufficient information to reach an identification (or exclusion) decision. From a practical point of view, imagine a guilty person not being prosecuted and sent to jail because the examiner failed to make the identification decision and incorrectly concluded an inconclusive (when there was sufficient information to justifiably make an identification) –a clear error that should be counted in error rates studies. Conversely, imagine an innocent person, not being dismissed as a suspect because the examiner doing the comparison failed to make the exclusion decision and incorrectly concluded an inconclusive -again, a clear error that should to be counted as an error. Hence, the question is not only about the ground truth of who left the mark, but more about what is the correct conclusion given the information available in the evidence.

Nevertheless, researchers have not properly considered the possibilities of error related to inconclusives. For example, one study of fingerprint experts showed that about 10% of the time the same expert, examining the same pair of fingerprints, reached different conclusions [29]. Despite this obvious error (when an examiner looking at the same fingerprints reached different conclusions at different times, then at least one of their conclusions is erroneous –the different conclusions cannot all be correct), inconclusive decisions in this study were never considered as error; errors were "only in reference to false positive and false negative conclusions ..." (p. 2). In another fingerprint study it is stated that "we have chosen to define an error for the present study, as a definitive opinion (exclusion or individualization) that did not reflect the ground truth" and "Thus, we have chosen not to categorize inconclusive opinions as errors per se" ([30]; p. 577 & 578).

Similarly, in the domain of firearms a study [31], found that expert forensic examiners reached inconclusive decisions on certain cartridge case comparisons, while other expert forensic examiners reached definitive decisions on the same comparisons, but *both* decisions were scored as correct —producing an "overall error rate of 0%" (p. 56). Another firearms study [32] clearly stated that "Since an inconclusive response is not an incorrect response it was totalled [sic] with the correct response [sic] and figured into the error rate as such" (p. 255).

A priori presuming that inconclusive decisions can never be an error is problematic. If some examiners conclude an identification (or exclusion) whereas other examiners conclude an inconclusive, then at least some of the examiners are mistaken [33]. They noted that "no such absolute criteria exist for judging whether the evidence is sufficient to reach a conclusion as opposed to making an inconclusive or no-value decision" (p. 7733) [33]). They are correct that it can be hard, maybe sometimes impossible, to know who is making the mistake, but it is obvious they cannot all be correct when examiners reach different conclusions on identical comparisons. It is therefore a problem when Ulery et al. find that "It was not unusual for one examiner to render an inconclusive decision while another made an individualization decision on the same comparison" (p. 7737). Even more so, when it is not different examiners (who may differ in their education, training or ability, or use different thresholds), but as in their study, the same examiner reached different conclusions on the same evidence about 10% of the time [29]; see review and conceptual hierarchical model at [34].

This is a critical issue in determining error rates: If one refuses a

ARTICLE IN PRESS

priori to count inconclusive decisions as errors, then error rates may be artificially and falsely reduced by making inconclusive decisions. In fact, zero error rates are possible with such an approach: regardless of anything, just reach inconclusive decisions for every comparison and you will have a perfect score! As stated earlier, inconclusive decisions can be appropriate and correct, but they can also be erroneous. It depends on whether or not there is sufficient quality and quantity of information to reach an identification or an exclusion decision. Determining that is not a simple matter (see details in Ref. [21]), but not ever counting inconclusive decisions as error is conceptually flawed and has practical negative consequences, such as misrepresenting error rate estimates in court which are artificially low and inaccurate. Furthermore, not counting inconclusive decisions as potential errors can lead examiners to resort to inconclusive decisions more often during error rate studies than they do in casework. Both of these factors seriously call into question the accuracy of the error rates reported in existing studies.

For establishing accurate error rates, one needs to use an appropriate study design that allows researchers to disentangle correct and incorrect inconclusive decisions, as well as correct and incorrect identification/exclusion decisions when the evidence is inconclusive. This point is developed in detail below. We first deal with 'Classifying Inconclusive Errors', i.e., how to correctly classify and collect data to estimate error rates, specifically dealing with the inconclusives as potential errors. Then, we deal with the 'Implications for Error Rates', i.e., how the framework for establishing error rate differs from actual casework, and the consequences of that.

3. Classifying Inconclusive Errors

Existing error rate studies have two categories into which they classify the *evidence*: either the test items come from the same source, or they come from different sources. Then, there are three possible *decision options* for the human examiner: identification, exclusion, or inconclusive. Decisions are scored as correct or erroneous by their correspondence to the evidence (see Fig. 1, left panel). Inconclusive decisions in existing studies are either always counted as correct and thus added to the 'correct decision tally' (e.g. Ref. [32]), or they are just not considered as either correct or erroneous and thus excluded from any tally (e.g., Ref. [29]). Either way, they are never considered or counted as erroneous.

A more appropriate study design would include a third category of inconclusive evidence (See Fig. 1, right panel). This design includes cases where the evidence is inconclusive, a reality in casework, in which evidence can be, and sometimes is, inconclusive (because the quantity and quality of information is not sufficient to allow any other conclusion —see discussion, above).

Including inconclusive evidence would allow researchers to test whether and to what extent participants correctly or erroneously make inconclusive decisions. Hence, in the proposed study design there are two kinds of errors relating to inconclusives: First, an inconclusive decision is reached when there is sufficient information to decide on an identification or exclusion (the red cells in the bottom row of Fig. 1, right panel); and second, when an identification or an exclusion is reached when there is insufficient information to justify such a decision (the red cells in the right column of Figure, right panel). Establishing that inconclusive decisions can be errors is theoretically and conceptually justified and clear, and is also applicable to casework.

However, as a practical matter, determining which evidence falls within this category is complicated. Ideally, determining when evidence is inconclusive should be done using objective criteria that ascertains whether the quantity and quality of the evidence is "sufficient" to reach an identification or exclusion decision. Unfortunately, most forensic domains currently lack such objective criteria. Given that there is currently no objective way to determine when evidence is inconclusive, we propose two different practical and feasible ways to determine when evidence is inconclusive:

The first option is that the test items would be piloted by a panel of independent experts who will be tasked with determining whether there is insufficient quantity or quality of information to make a source determination for each comparison. Comparisons deemed by this group to lack sufficient quality or quantity of markings would be classified as inconclusive evidence. Of course, this raises the question (and concerns) about who will be the independent experts and how will they carry out this task. What is clear is that this group will consist of established experts, that they will determine which evidence is inconclusive prior to the actual test study taking place, and that they will not participate in the test study itself.

The second option is that the data from the actual test study be used to make the determination of which evidence should be deemed as inconclusive. The responses to each set of test items would be examined to see what percentage of decisions were inconclusive and what percentage were not inconclusive. If most examiners report the comparison as inconclusive, then that evidence would be classified as inconclusive (and hence an inconclusive decision would be deemed correct and any other decision would be an error). However, if most examiners deem a given comparison an identification or exclusion,



Fig. 1. The left panel is the widely used, and misleading, study design for establishing error rates. The evidence is either same- or different-source, and inconclusive decisions are never counted as error. The right panel is the suggested and correct design for studying error rates, whereby evidence can be inconclusive. There are two kinds of errors relating to inconclusive decisions: First, an inconclusive decision is reached when there is sufficient information to decide on an identification or exclusion (see red cells in the bottom row); the second type of error is when an identification or an exclusion decision is reached when there is insufficient information to justify such a decision (see red cells in the right column). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Please cite this article as: I.E. Dror, N. Scurich, (Mis)use of scientific measurements in forensic science, Forensic Science International: Synergy, https://doi.org/10.1016/j.fsisyn.2020.08.006

4

then an inconclusive decision would be deemed an error. This approach for determining which evidence is inconclusive is based on the test data, and assumes that most examiners make the correct decision.

In summary, including the category of inconclusive evidence (see right panel of Fig. 1) is critical in error rate studies. It allows researchers to determine when inconclusive decisions are correct and when they are erroneous. This enables a more accurate estimation of error rates because it does not simply drop -or unjustifiably treat as correct- inconclusive decisions, nor does it ignore the existence of inconclusive evidence in casework. Whether a decision in an error rate study is correct or incorrect must be made in reference to the information contained in the evidence. It is a challenge to classify which evidence falls in an inconclusive category, but we have suggested a couple of practical ways this can be achieved. Failing to include inconclusive evidence, and/or never counting inconclusive decisions as potentially an error (or just not allowing to make inconclusive decisions), yields misleading error rates.

4. Implications for error rates

Existing studies that fail to use a proper design are not able to accurately estimate error rates because they fail to address inconclusive evidence and decisions. These studies just ignore the very existence of inconclusive *evidence*, count automatically inconclusive *decisions* as correct or just exclude them, and a whole myriad of other flaws we specify below (e.g., Refs. [31,33,35]).

First flaw: Avoiding difficult and ambiguous cases where errors are more likely to occur. The difficulty of test items in error rates studies must reflect the distribution of difficulty in casework [21]. This is especially important as difficult cases are more prone to error. Not having sufficiently difficult cases will artificially reduce the error rate; conversely, including too many difficult cases will artificially increase the error rate.

Inconclusive evidence is challenging, because it may contain some similarities, but just not enough to reach an identification decision, or, alternatively, may contain dissimilarities, but just not enough to reach an exclusion decision. Although this makes the decision-making process difficult and prone to error, it also reflects the reality of casework. Hence, studies that do not include inconclusive evidence, by design, fall short of reflecting casework and thus provide inaccurate error rate.

As detailed above, although the ground truth is that the evidence either comes from the same source or it comes from a different source, that does not mean that the available information in the evidence is sufficient to justify reaching a conclusion that the item is or is not from the same source. Not including such inconclusive items in error rate studies is a major problem, as it misrepresents the reality of evidence in casework, and thus will provide a misleading error rate. As noted, all of the existing studies in the domains of fingerprints and firearms suffer from this flaw.

Second flaw: Excluding inconclusive decisions from analysis and error rate calculations. Since inconclusive decisions can be errors, when there is sufficient information to reach an identification (or exclusion) decision (see Fig. 1, and earlier discussion), excluding them may well mean that errors that should be counted within the error rate calculation are ignored. This artificially reduces the error rates and causes them to be misleading.

Excluding inconclusive decisions from error rate calculations is especially problematic when inconclusive decisions are made in the difficult comparisons, which are more prone to error. Indeed, the findings in a fingerprint error rate study showed that "participants reported more inconclusive decisions than correct identifications for latent trials that were rated the most difficult to compare" ([36]; p. 66). This means that if inconclusive decisions are excluded from calculating the error rate, then the error rates are based more on the easy comparisons, since inconclusive decisions are more often made in difficult cases.

Third flaw: Counting inconclusive decisions as correct decisions. Even worse, rather than just excluding the inconclusive decisions from error rate calculations, is including them and counting them as correct decisions. In this case, potentially erroneous decisions are not just ignored and not counted as error, but potential errors are being counted as correct decisions.

Although some studies are explicit about counting inconclusive decisions as correct (e.g. Ref. [32]), other studies implicitly count inconclusive decisions as correct by the virtue of how they calculate error rates (e.g., Ref. [35]). Calculating error rates involves dividing the number of incorrect decisions by the total number of decisions (both correct and incorrect). If we label the incorrect decisions as X and the correct decisions as Y, the error rate = X/(X + Y). Including inconclusive decisions (labelling them as Z) as correct decisions, means that error rate is actually calculated as = X/(X + Y + Z), thus reducing the error rate.

Fourth flaw: Examiners resort to making more inconclusive decisions during error rate studies than they do in casework. The problem with error rates not reflecting casework is further multiplied because examiners resort to inconclusive decisions in error rate studies more often than they do in actual casework. Indeed, the prevalence of inconclusive decisions in error rate studies is quite remarkable. For example, in the domain of firearms, the percentage of inconclusive decisions in some studies is almost a quarter of the total decisions (e.g., [35], and in other studies almost half the decisions are inconclusive, leaving a maximum ceiling of only 1.7% as potentially the highest possible error rate [38].

It is important to remember that in most error rate studies examiners know they are participating in error rate studies and therefore might modify their decision making, particularly if they know that inconclusive decisions never counted as an error. There can be additional reasons for the discrepancy in the use of inconclusive decisions between case work and error rate studies, for example in casework examiners may be pressured to reach a definite decision because "an examiner who is frequently inconclusive is ineffective and thereby fails to serve justice" ([33]; p. 7737). In contrast, in error rate studies, examiners might have pressure to reach inconclusive decisions. Examiners not only know that they are taking part in a study (which already impacts performance, see Ref. [9], but that it is an error rate study, and that errors they make in the study may cast doubt on the field they have been working and invested in, e.g., "The examiners understood they were participating in a blind validation study, and that an incorrect response could adversely affect the theory of firearms identification" ([38]; p.132). Furthermore, the results may be used in court to undermine them and their profession. Because they know that any inconclusive decision they make will not be counted as an error (and may even be counted as correct), it is not surprising that examiners resort in error rate studies to inconclusive decisions.

Regardless of why there is a difference between casework and error rate studies, if examiners resort to inconclusive decisions more often in error rate studies than they do in casework, then the error rates observed in studies do not accurately reflect casework.

We listed four flaws regarding the design and interpretation of existing error rates and how they deal with inconclusive decisions and evidence. These flaws have real implications for how error rates are calculated and (mis)used (we do not make a claim that the misuse is intentional). For example, in the domain of firearms [31], report "an overall error rate of 0%" (p. 57). This error rate counts 207 inconclusive decisions as correct. However, some of these inconclusive decisions could be errors, potentially giving an overall error

rate of up to 8.2%. So, the actual overall error rate within the parameters of this study is between 0% and 8.2%, depending how many of the inconclusive decisions were incorrect. Either of the two extremities of the range, 0% and 8.2%, could be correct –if all the inconclusive decisions were correct or if all the inconclusive decisions were incorrect, respectively.

Similarly, the 746 inconclusive decisions in [35] could potentially include errors, which means that the overall error rate can be as high as 22.8% (the extremity of the range, if all inconclusive decisions are incorrect) –drastically higher than the other extremity of the range (if all inconclusive decisions are correct), the ~1% error rate reported in the study.

Fingerprint research mirrors that found in firearms. For example [33], provide data on false identification and false exclusions, but never consider any of the many inconclusive decisions found in the study as errors (over a third of all decisions), nor do they seem to include inconclusive evidence in their study. Interestingly they define inconclusive decisions as "The comparison/evaluation decision that neither individualization nor exclusion is possible" (p. 24 of Appendix). Hence, by their own definition: A. An inconclusive decision is incorrect when an identification or exclusion is possible. And, B. An identification or exclusion are incorrect when evidence does not allow identification or exclusion (i.e., when inconclusive is correct). However, these were not included in their calculations of error rates.

5. Summary and conclusions

Error rates are a critical measure of performance and are especially important in forensic science. However, error rates are a complex construct which have a whole host of challenges and difficulties that stand in the way of establishing an accurate measurement [21]. One set of issues surrounds inconclusives (other issues, beyond the scope of this paper, include selectively publishing only studies that reveal low error rates and not mimicking the realities of casework which can further increase errors, such as stress and bias).

Inconclusive evidence and inconclusive decisions have received very little attention, including examining the circumstances for when inconclusive decisions are correct and when they are erroneous [19]. Although inconclusive evidence and decisions are less likely to be presented in court, they are indirectly presented (and misused) in court as part of error rate data. Erroneous decisions, including erroneous inconclusives, and inconclusive evidence should be accounted for in error rate studies, and failing to do so results in measurements that are misleading and inaccurate.

In this paper we showed that existing error rate studies have failed to address the issues of inconclusives in two distinct ways, both in terms of the evidence (not including inconclusive evidence), as well as in terms of decisions (not considering inconclusive decisions as potential error). We furthermore identified four specific flaws in existing error rate studies, and we proposed an experimental design that will address these problems and provide a more accurate estimate of error rates. We further proposed possible ways to determine how to classify evidence into the category of inconclusive, as well as how to determine when decisions are incorrect (both when inconclusive evidence is incorrectly determined as an identification or exclusion, and when inconclusive decisions are incorrect).

Without addressing and fixing these issues, error rate studies fall short, and produce inaccurate and misleading error rate estimates. Beyond error rates, thinking about when evidence is inconclusive, and when inconclusive decisions are correct or incorrect will benefit forensic science in its own right.

Declaration of competing interest

No conflict of interest.

Acknowledgements

This work was partially funded by 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.

References

- [1] K. Popper, The Logic of Scientific Discovery, Routledge, New York, 1959.
- [2] Garrett, B.L. (in press). Autopsy of a Crime Lab. University of California Press..
 [3] L.M. Howes, The communication of forensic science in the criminal justice system: a review of theory and proposed directions for research, Sci. Justice 55 (2) (2015) 145–154.
- [4] T.D. Lyon, J.J. Koehler, Relevance ratio: evaluating the probative value of expert testimony in child sexual abuse cases, Cornell Law Rev. 82 (1996) 43–76.
- [5] PCAST, President's Council of Advisors on science and Technology report on forensic science, in: Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods, 2016. Available at: https://obamawhitehouse.archives. gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_ final.pdf.
- [6] B. Sangero, Safety from flawed forensic sciences evidence, Ga. State Univ. Law Rev. 34 (2018) 1129–1220.
- [7] R.N. Haber, L. Haber, Experimental results of fingerprint comparison validity and reliability: a review and critical analysis, Sci. Justice 54 (2014) 375–389.
- [8] M.J. Saks, J.J. Koehler, The coming paradigm shift in forensic identification science, Science 309 (5736) (2005) 892–895.
- [9] J.J. Koehler, Proficiency tests to estimate error rates in the forensic sciences, Law Probab. Risk 12 (1) (2013) 89–98.
- [10] NAS, Strengthening Forensic Science in the United States: A Path Forward. National Academy of Sciences, National Academies Press, Washington, DC, 2009.
- [11] S.A. Cole, More than zero: accounting for error in latent fingerprint identification, J. Crim. Law Criminol. 95 (2005) 985–1078.
- [12] B.L. Garrett, P.J. Neufeld, Invalid forensic science testimony and wrongful convictions, Va. Law Rev. 95 (2009) 1–97.
- [13] D.C. Murrie, B.O. Gardner, S. Kelley, I.E. Dror, Perceptions and estimates of error rates in forensic science, Forensic Sci. Int. 302 (2019).
- [14] Daubert, Merrill Dow Pharmaceuticals, Inc, vol. 509, U.S, 1993, p. 579.
- [15] J.B. Meixner, S.S. Diamond, The hidden Daubert factor: how judges use error rates in assessing scientific evidence, Wis. Law Rev. (2014) 1063–1134.
- [16] 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 (5) (2001) 433–458.
- [17] United States v, Havvard 260 (7th Cir. 2001). F.3d 597.
- [18] D.J. Capra, Symposium on forensic expert testimony, Daubert, and Rule 702, Fordham Law Rev. 86 (4) (2018) 1463–1550.
- [19] I.E. Dror, G. Langenburg, Cannot Decide": the fine line between appropriate inconclusive determinations VS. unjustifiably deciding not to decide, J. Forensic Sci. 64 (1) (2019) 1–15, https://doi.org/10.1111/1556-4029.13854.
- [20] D. McQuiston-Surrett, M.J. Saks, The testimony of forensic identification science: what expert witnesses say and what factfinders hear, Law Hum. Behav. 33 (5) (2009) 436–453.
- [21] I.E. Dror, The Error in 'Error Rate': why error rates are so needed, yet so elusive, J. Forensic Sci. 65 (4) (2020a) 1034-1039, https://doi.org/10.1111/ 1556-4029.14435.
- [22] S. Kelley, B.O. Gardner, D.C. Murrie, K. Pan, K. Kafadar, How do latent print examiners perceive proficiency testing? An analysis of examiner perceptions, performance, and print quality, Sci. Justice 60 (2) (2020) 120–127.
- [23] J.J. Koehler, Fingerprint error rates and proficiency tests: what they are and why they matter, Hastings Law J. 59 (2008) 1077–1100, 59.
- [24] M.L. Pierce, LJ. Cook, Development and implementation of an effective blind proficiency testing program, J. Forensic Sci. 65 (2020) 809–814, https:// doi.org/10.1111/1556-4029.14269.
- [25] M.T. Orne, On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications, Am. Psychol. 17 (11) (1962) 776–783.
- [26] A.M. Jeanguenat, I.E. Dror, Human factors effecting forensic decision making: workplace stress and wellbeing, J. Forensic Sci. 63 (1) (2018) 258–261, https://doi.org/10.1111/1556-4029.13533.
- [27] I.E. Dror, Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias, Anal. Chem. 92 (12) (2020b) 7998–8004, https:// doi.org/10.1021/acs.analchem.0c00704.

Please cite this article as: I.E. Dror, N. Scurich, (Mis)use of scientific measurements in forensic science, Forensic Science International: Synergy, https://doi.org/10.1016/j.fsisyn.2020.08.006

6

ARTICLE IN PRESS

I.E. Dror, N. Scurich / Forensic Science International: Synergy xxx (xxxx) xxx

- [28] A.S. Cole, B.C. Scheck, Fingerprints and miscarriages of justice: "other" types of error and a post-conviction right to database searching, Albany Law Rev. 81 (2018) 807–850.
- [29] B.T. Ulery, R.A. Hicklin, J. Buscaglia, M.A. Roberts, Repeatability and reproducibility of decisions by latent fingerprint examiners, PLoS 7 (2012), e32800.
- [30] G. Langenburg, C. Champod, P. Wertheim, Testing for Potential contextual bias effects during the verification stage of the ACEV methodology when conducting fingerprint comparisons, J. Forensic Sci. 54 (3) (2009) 571–582.
- [31] M.A. Keisler, S. Hartman, A. Kilmon, M. Oberg, M. Templeton, Isolated pairs research study, AFTE Journal 50 (1) (2018) 56–58.
- [32] D.J. Lyons, The identification of consecutively manufactured extractors, AFTE Journal 41 (3) (2009) 246–256.
- [33] B.T. Ulery, A.R. Hicklin, J. Buscaglia, M.A. Roberts, Accuracy and reliability of forensic latent fingerprint decisions, Proc. Natl. Acad. Sci. Unit. States Am. 108 (19) (2011) 7733–7738.
- [34] I.E. Dror, A hierarchy of expert performance (HEP), Journal of Applied Research in Memory and Cognition 5 (2) (2016) 121–127, https://doi.org/ 10.1016/j.jarmac.2016.03.001.
- [35] D.P. Baldwin, S.J. Bajic, M. Morris, D. Zamzow, A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons, Ames Laboratory, Defense Forensic Science Center, 2014. Technical Report # IS-5207, available at: https://www.ncjrs.gov/pdffiles1/nij/249874.pdf.
- [36] I. Pacheco, B. Cerchiai, B. Stoiloff, Miami-dade Research Study for the Reliability of the Ace-V Process: Accuracy & Precision in Latent Fingerprint Examinations, Department of Justice, 2014. Technical Report # 248534, available at: https://www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf.
 [37] S.G. Bunch, D.P. Murphy, A comprehensive validity study for the forensic
- [37] S.G. Bunch, D.P. Murphy, A comprehensive validity study for the forensic examination of cartridge cases, AFTE Journal 35 (2) (2003) 201–203.
 [38] E.D. Smith, Cartridge case and bullet comparison validation study with fire-
- [38] E.D. Smith, Cartridge case and bullet comparison validation study with firearms submitted in casework, AFTE Journal 37 (2) (2005) 130–135.