Science and Justice 56 (2016) 371–373

Contents lists available at ScienceDirect

Science and Justice

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

Special issue on measuring and reporting the precision of forensic likelihood ratios: Introduction to the debate☆

Geoffrey Stewart Morrison

Morrison & Enzinger, Independent Forensic Consultants, Vancouver, British Columbia, Canada, and Corvallis, Oregon, United States of America
Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Article history:
Received 17 April 2016
Received in revised form 3 May 2016
Accepted 4 May 2016

Keywords:
Forensic evidence
Likelihood ratio
Precision
Reliability
Accuracy
Validity

ABSTRACT

The present paper introduces the *Science & Justice* virtual special issue on measuring and reporting the precision of forensic likelihood ratios — whether this should be done, and if so how. The focus is on precision (aka reliability) as opposed to accuracy (aka validity). The topic is controversial and different authors are expected to express a range of nuanced opinions. The present paper frames the debate, explaining the underlying problem and referencing classes of solutions proposed in the existing literature. The special issue will consist of a number of position papers, responses to those position papers, and replies to the responses.

1. Introduction

The papers gathered in this special issue of *Science & Justice* represent an open debate on the topic of whether the precision of forensic likelihood ratios should be measured and reported to the courts, and if so how. This is a controversial issue, with authors expressing contrary opinions. Some believe that the very concept of the precision of likelihood ratios is not appropriate, whereas others believe that measuring and reporting the precision of likelihood ratios is essential. In addition, even authors who mostly agree with each other differ on details. The debate has been ongoing for a number of years, but much of it has been hidden: Reviewers of submitted manuscripts complaining that precision has or has not been measured, and authors justifying their choices in rebuttal letters.

The present special issue aims to bring the debate out into the open, to lead to a better understanding of the reasons behind different authors' opinions, and potentially even to lead to some degree of consensus. A precursor to the present journal special issue took place at the European Academy of Forensic Sciences (EAFS) Conference in Prague in September 2015, where several presenters (Geoffrey Stewart Morrison; James M Curran; Colin GG Aitken; Peter Vergeer with Ivo Alberink, Annabel Bolck, Marjan Sjerps, and Reinoud Stoel; and Didier Meuwly) gave a series of brief presentations on this topic. This virtual special issue (VSI) of *Science & Justice* had an open call for papers which welcomed and encouraged submissions both from authors who did not present at the EAFS conference and from those who did.

The present introductory paper is intended to be neutral, framing the debate but not taking a position in the debate. An initial draft of this introduction was written before the initial call for papers went out, and (except for corrections of any typographical errors, etc.) the final version was written before any of the submitted papers were received.

The initial position papers will be published online as a single block, at which point a second open call for papers will go out requesting responses to those position papers. Authors of the position papers will subsequently be given the opportunity to reply to the response papers. All the position, response, and reply papers will be collected together online.

2. Precision versus accuracy

First, we wish to make it clear that all contributors to the current debate agree that the likelihood ratio framework is the logically correct framework for the evaluation of evidence. Next, we wish to make clear what it is we mean by precision. Fig. 1 shows four targets that four archers have shot at. Two archers each have a tight grouping of arrows — these archers are precise. Two other archers have a wide spread in the location of their arrows — these archers are not precise. In each of these pairs, one archer is also accurate, the average location of their arrows is near the centre of the target, and the other is not accurate, the average

☆ This paper is part of the Virtual Special Issue entitled: Measuring and Reporting the Precision of Forensic Likelihood Ratios, [http://www.sciencedirect.com/science/journal/13550306/vsi], Guest Edited by G. S. Morrison.

E-mail address: geoff-morrison@forensic-evaluation.net.

http://dx.doi.org/10.1016/j.scijus.2016.05.002
1355-0306/

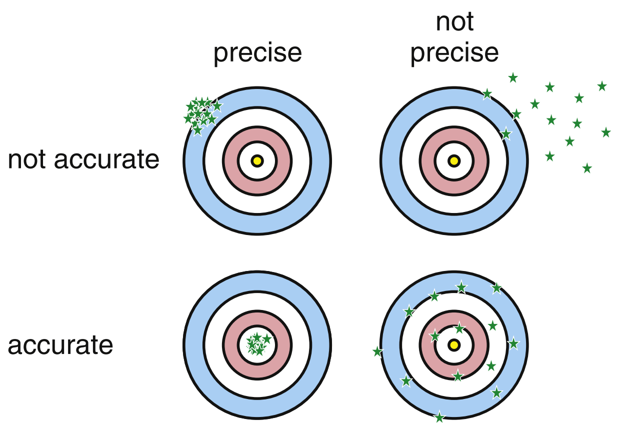

Fig. 1. Illustration of the difference between accuracy and precision.

location of their arrows is far from the centre of the target. Although the concept of accuracy requires comparison with a reference value, the concept of precision does not. Precision quantifies the degree of closeness of the results to each other irrespective of whether they are close to a reference value.

That one should assess the validity of forensic systems that output likelihood ratios does not appear to be controversial and we expect that all contributors to the debate will agree on the need to do this. The present debate is restricted to whether one should assess the precision of forensic systems that output likelihood ratios, and if so how that precision should be assessed and reported.

Although the terms validity and reliability may be used with a wider range of meanings, they are often used as synonyms of accuracy and precision respectively. Debaters may choose to use either the terms accuracy and precision or validity and reliability.

3. The underlying problem

We expect that all debaters will agree on the underlying problem, and it is with respect to the appropriate solution where they will potentially disagree with one another. What is the underlying problem? Sources of potential variability in the output of a forensic analysis system include:

- intrinsic variability at the source (e.g., [1,2])
- variability in the transfer process (e.g., [3])
- variability in sampling the relevant population (e.g., [4–6])
- variability in sampling the known source (e.g., [7])
- variability in the measurement technique employed (e.g., [8,9])
- variability in the statistical modelling technique employed (e.g., [9–11])
- variability due to modelling assumptions (e.g., [12]).

Here we illustrate the effect of the third and fourth items from the list above: sampling variability for the relevant population and the known source. Fig. 2 contains several panels each showing the results of fitting univariate Gaussian distributions to data which are pseudo-random samples of specified Gaussian distributions. One underlying distribution is broad and represents a population distribution, and the other is narrow and represents a single known source. Each underlying distribution has a fixed specified mean and variance. In each panel we take a pseudo-random sample of 4 points from the population distribution and a pseudo-random sample of 2 points from the known-source distribution. We then calculate sample means and sample variances, and plot the sample distributions. In each panel, we also have a fixed-value specimen of questioned source. Given the sample distributions we calculate a likelihood ratio for the probability of obtaining the value of the questioned specimen had it come from the known source versus the probability of obtaining the value of the questioned specimen had it come from the population (technically, we evaluate the likelihood of the models at the value of the questioned specimen). Each panel represents the results from a different set of pseudo-random samples, and, as can be seen in the different panels, the value of the calculated likelihood ratio varies considerably across sample sets. Fig. 3 shows a box plot of the likelihood ratio values resulting from 100 sample sets. This example is obviously extreme, a very small sample size results in high variability in parameter estimates, but to some extent sample variability will always affect our parameter estimates and therefore always affect our calculations of likelihood ratios. The effect may be small or large, we will not know until we look at it, and it may be due to some combination of any of the potential sources of variability listed above. This is the underlying problem, the debate is about what to do about it.

4. Proposed solutions

Multiple solutions have been proposed including:

- Ignore the problem. Not actually a solution, and probably never explicitly proposed as a solution, but widely practised. Rather than being aware of the problem and deliberately ignoring it, the vast majority of those who ignore the problem likely do so because they are not aware of the problem.
- Calculate an empirical estimate of imprecision using test data. Generally this will be done under case-specific conditions, and a numeric value representing imprecision will be reported. For example, a numerically calculated likelihood ratio could be reported as 1000 with a 98% credible interval of plus or minus one order of magnitude, or it could be reported that it is 99% certain that the likelihood ratio is at least 100. For examples, see Morrison [7], Hancock et al. [6], and Stoel and Sjerps [13].
- Report a level from an ordinal scale in which each level on the scale covers a relatively wide range of likelihood ratio values (Evett et al. [14]). Such scales usually have verbal expressions associated with each level and the range of values covered by each level is often an order of magnitude. For recently published examples of such scales,

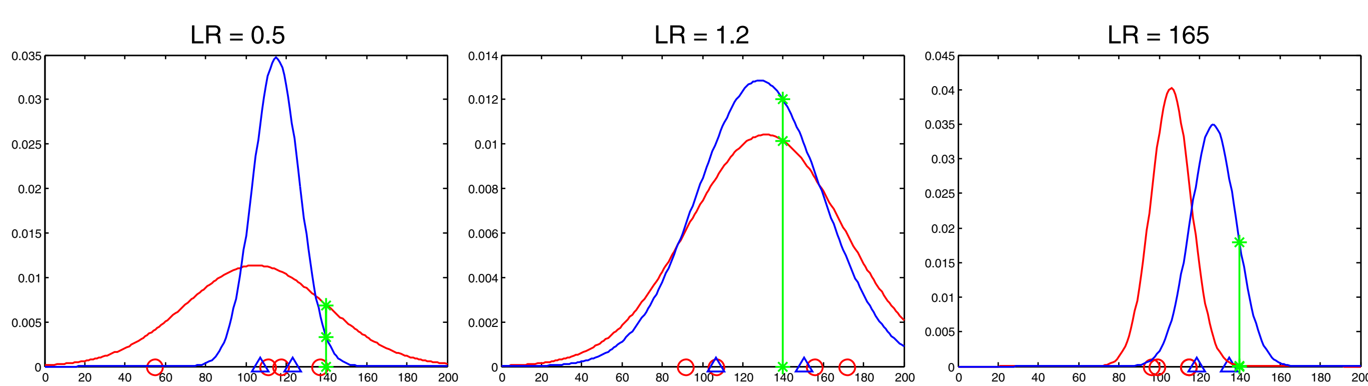

Fig. 2. Example of variability in likelihood ratio calculation due to sampling variability. Each panel represents a different sample set.

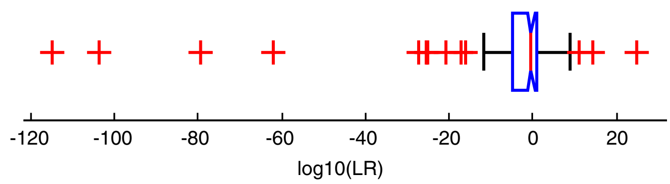

Fig. 3. Example of variability in likelihood ratio calculation due to sampling variability in 100 different sample sets.

see Association of Forensic Science Providers [15], Nordgaard et al. [16], and Willis et al. [17]. These scales are predefined and the imprecision of the particular system under the conditions of the particular case is not empirically assessed. These scales are also used to express strengths of evidence which are assigned directly on the basis of expert subjective judgement, rather than calculated using quantitative measurements and statistical models.

- The underlying problem as described in §3 above is misstated. The problem is not estimation of true but unknown parameter values, probability is a state of belief. The likelihood ratio itself is a measure of uncertainty and it is not appropriate to attempt to quantify the uncertainty of the uncertainty. The uncertainly due to the sources of variability should be directly incorporated into the calculation of the likelihood ratio value. Use Bayesian techniques which combine prior probability density distributions with sample data, and which integrate out nuisance variables to arrive at posterior predictive probability distributions. These are used to calculate likelihood ratios (technically, Bayes factors) which are point values but which are closer to a neutral likelihood ratio of 1 than would otherwise be the case. For examples, see Brümmer and Swart [18], and Taroni et al. [19].

With the exception of "ignore", we expect that the special issue will include submissions from proponents of variants of each of these solutions. We suspect that some of the differences between positions leading to these different solutions are attributable to underlying differences in the conception of what a forensic likelihood ratio actually is, and we hope that this will be explored as part of the debate.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] P. Rose, Forensic Speaker Identification, Taylor & Francis, London, UK, 2002.
[2] G.S. Morrison, Forensic voice comparison, in: I. Freckelton, H. Selby (Eds.), Expert Evidence, Thomson Reuters, Sydney, Australia, 2010 (ch. 99).
[3] E. Enzinger, G.S. Morrison, F. Ochoa, A demonstration of the application of the new paradigm for the evaluation of forensic evidence under conditions reflecting those of a real forensic-voice-comparison case, Sci. Justice 56 (2016) 42–57, http://dx.doi.org/10.1016/j.scijus.2015.06.005.
[4] J.M. Curran, J.S. Buckleton, C.M. Triggs, B.S. Weir, Assessing uncertainty in DNA evidence caused by sampling effects, Sci. Justice 42 (2002) 29–37, http://dx.doi.org/10.1016/S1355-0306(02)71794-2.
[5] J.M. Curran, An introduction to Bayesian credible intervals for sampling error in DNA profiles, Law Probab. Risk 4 (2005) 115–126, http://dx.doi.org/10.1093/lpr/mgi009.
[6] S. Hancock, R. Morgan-Smith, J.S. Buckleton, The interpretation of shoeprint comparison class correspondences, Sci. Justice 52 (2012) 243–248, http://dx.doi.org/10.1016/j.scijus.2012.06.002.
[7] G.S. Morrison, Measuring the validity and reliability of forensic likelihood-ratio systems, Sci. Justice 51 (2011) 91–98, http://dx.doi.org/10.1016/j.scijus.2011.03.002.
[8] C. Zhang, G.S. Morrison, F. Ochoa, E. Enzinger, Reliability of human-supervised formant-trajectory measurement for forensic voice comparison, J. Acoust. Soc. Am. 133 (2013) EL54–EL60, http://dx.doi.org/10.1121/1.4773223.
[9] J.-A. Bright, K.E. Stevenson, J.M. Curran, J.S. Buckleton, The variability in likelihood ratios due to different mechanisms, Forensic Sci. Int. Genet. 14 (2015) 187–190, http://dx.doi.org/10.1016/j.fsigen.2014.10.013.
[10] G.S. Morrison, J. Lindh, J.M. Curran, Likelihood ratio calculation for a disputed-utterance analysis with limited available data, Speech Comm. 58 (2014) 81–90, http://dx.doi.org/10.1016/j.specom.2013.11.004.
[11] E. Enzinger, G.S. Morrison, Mismatched distances from speakers to telephone in a forensic-voice-comparison case, Speech Comm. 70 (2015) 28–41, http://dx.doi.org/10.1016/j.specom.2015.03.001.
[12] D. Taylor, J.-A. Bright, J.S. Buckleton, J.M. Curran, An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations, Forensic Sci. Int. Genet. 11 (2014) 56–63, http://dx.doi.org/10.1016/j.fsigen.2014.02.003.
[13] R.D. Stoel, M.J. Sjerps, Interpretation of forensic evidence, in: S. Roeser, R. Hillerbrand, P. Sandin, M. Peterson (Eds.), Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk, Springer Netherlands, Dordrecht, The Netherlands 2012, pp. 135–158, http://dx.doi.org/10.1007/978-94-007-1433-5_6.
[14] I.W. Evett, G. Jackson, J.A. Lambert, S. McCrossan, The impact of the principle of evidence interpretation on the structure and content of statements, Sci. Justice 40 (2000) 233–239, http://dx.doi.org/10.1016/S1355-0306(00)71993-9.
[15] Association of Forensic Science Providers, Standards for the formulation of evaluative forensic science expert opinion, Sci. Justice 49 (2009) 161–164, http://dx.doi.org/10.1016/j.scijus.2009.07.004.
[16] A. Nordgaard, R. Ansell, W. Drotz, L. Jaeger, Scale of conclusions for the value of evidence, Law Probab. Risk 11 (2012) 1–24, http://dx.doi.org/10.1093/lpr/mgr020.
[17] S.M. Willis, L. McKenna, S. McDermott, G. O'Donell, A. Barrett, B. Rasmusson, A. Nordgaard, C.E.H. Berger, M.J. Sjerps, J.J. Lucena-Molina, G. Zadora, C.G.G. Aitken, L. Lunt, C. Champod, A. Biedermann, T.N. Hicks, F. Taroni, ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes, Wiesbaden, Germany, http://enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf, 2015.
[18] N. Brümmer, A. Swart, Bayesian calibration for forensic evidence reporting, Proceedings of Interspeech 2014, pp. 388–392 (http://www.isca-speech.org/archive/interspeech_2014/i14_0388.html).
[19] F. Taroni, S. Bozza, A. Biedermann, C.G.G. Aitken, Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio, Law Probab. Risk 15 (2015) 1–16, http://dx.doi.org/10.1093/lpr/mgv008.