Preprint of:
Morrison G.S. (2022). A plague on both your houses: The debate about how to deal with “inconclusive” conclusions when calculating error rates. Law, Probability and Risk. https://doi.org/10.1093/lpr/mgac015

A plague on both your houses: The debate about how to deal with “inconclusive” conclusions when calculating error rates

Author and affiliations:

Geoffrey Stewart Morrison *

Forensic Data Science Laboratory, Aston University, Birmingham, UK

Forensic Evaluation Ltd, Birmingham, UK

* e-mail: geoff-morrison@forensic-evaluation.net

ORCID:

0000-0001-8608-8207

Disclaimer:

All opinions expressed in the present paper are those of the author, and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the author is associated.

Declaration of competing interest:

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding:

This work was supported by Research England's Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2023.

A plague on both your houses: The debate about how to deal with "inconclusive" conclusions when calculating error rates

Abstract:

There is an ongoing debate about how to deal with "inconclusive" conclusions when calculating error rates. The appropriate solution to the problem is to abandon reporting of conclusions as "identification", "inconclusive", and "exclusion", and to adopt the likelihood-ratio framework for interpretation and adopt the log-likelihood-ratio cost (C_{llr}) as a validation metric.

Keywords:

Bayes factor; Error rate; Inconclusive; Likelihood ratio; Validation

Letter to the Editor:

There is an ongoing debate about how to deal with "inconclusive" conclusions when calculating error rates. Recent contributions to this debate include, in chronological order, Dror & Scurich (2020), Weller & Morris (2020), Biedermann & Kotsoglou (2021), Arkes & Koehler (2021), Dror (2022), and Arkes & Koehler (2022). All the proposed solutions are inappropriate because they do not address the real problem. The real problem is not what to do with "inconclusive" conclusions but the fact that forensic practitioners report conclusions as "identification", "inconclusive", and "exclusion" at all. The real solution is for practitioners to abandon this practice and, instead, adopt the logically correct framework for interpretation of forensic evidence, the likelihood-ratio

framework. Rather than reporting each of their conclusions as one of three categories, practitioners should report them as continuously-valued likelihood-ratio values.[1]

When conducting tests, if the test pair is a same-source pair, then a good output by the forensic-evaluation system would be a large positive log-likelihood-ratio value, a poorer output would be a small positive log-likelihood-ratio value, a worse output would be a small negative log-likelihood-ratio value, and a much worse output would be a large negative log-likelihood-ratio value. *Mutatis mutandis*, for a different-source pair, for which a good output would be a large negative log-likelihood-ratio value.[2] When calculating error rates in a categorical framework, a penalty value of 1 is assigned to each response that is the incorrect category and a penalty value of 0 is assigned to each response that is the correct category, then the penalty values are averaged. In contrast, an appropriate validation metric for a system that outputs likelihood ratios assigns a small penalty value to each log-likelihood-ratio value that is large in the correct direction, a less small penalty value to each log-likelihood-ratio value that is small in the correct direction, a large penalty value to each log-likelihood-ratio value that is small in the incorrect direction, and an even larger penalty value to each log-likelihood-ratio value that is large in the incorrect direction, then the penalty values are averaged. The penalty values are assigned using a continuously-valued function of the continuously-valued log-likelihood-ratio outputs of the forensic-evaluation system. A well-established validation metric that has these properties is the log-likelihood-ratio cost (C_{llr}; Brümmer & du Preez, 2006). It has been in use in forensic science since at least 2007 (González-Rodríguez et al., 2007), and its use is recommended in the recently-published *Consensus on validation of forensic voice comparison* (Morrison et al., 2021).[3] An appendix in the latter publication provides an

[1] Biedermann & Kotsoglou (2021) hints at this solution, but then dives into the debate.

[2] Likelihood-ratio values in the range 0 to 1 correspond to log-likelihood-ratio values in the range $-\infty$ to 0, and likelihood-ratio values in the range 1 to $+\infty$ correspond to log-likelihood-ratio values in the range 0 to $+\infty$.

[3] With minor wording changes, the recommendations in the *Consensus on validation of forensic voice comparison* become applicable across other branches of forensic science.

introduction to C_{llr} and its interpretation.

The appropriate solution to the problem of how to deal with "inconclusive" conclusions when calculating error rates is to abandon reporting of conclusions as "identification", "inconclusive", and "exclusion", and to adopt the likelihood-ratio framework for interpretation and adopt C_{llr} as a validation metric.[4]

References

Arkes H.R., Koehler J.J. (2021). Inconclusives and error rates in forensic science: A signal detection theory approach. *Law, Probability and Risk*, 20, 153–168. https://doi.org/10.1093/lpr/mgac005

Arkes H.R., Koehler J.J. (2022). Inconclusives are not errors: a rejoinder to Dror. *Law, Probability and Risk*. https://doi.org/10.1093/lpr/mgac009

Biedermann A., Kotsoglou K.N. (2021). Forensic science and the principle of excluded middle: "Inconclusive" decisions and the structure of error rate studies. *Forensic Science International: Synergy*, 3, 100147. https://doi.org/10.1016/j.fsisyn.2021.100147

Brümmer N., du Preez J. (2006). Application independent evaluation of speaker detection. *Computer Speech and Language*, 20, 230–275. https://doi.org/10.1016/j.csl.2005.08.001

Dror I.E. (2022). The use and abuse of the elusive construct of inconclusive

[4] Although, ideally, this would involve replacement of procedures based on human perception and subjective judgement with procedures based on relevant data, quantitative measurement, and statistical models (Morrison, 2022a), a preliminary step in the right direction could be to use statistical models to convert practitioners' "identification", "inconclusive", and "exclusion" conclusions into well-calibrated likelihood-ratio values (or into well-calibrated Bayes-factor values; Morrison, 2022b).

decisions. *Law, Probability and Risk*. https://doi.org/10.1093/lpr/mgac008

Dror I.E., Scurich N. (2020). (Mis)use of scientific measurements in forensic science. *Forensic Science International: Synergy*, 2, 333–338. https://doi.org/10.1016/j.fsisyn.2020.08.006

González-Rodríguez J., Rose P., Ramos D., Toledano D.T., Ortega-García, J. (2007). Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Transactions on Speech and Audio Processing*, 15, 2104–2115. https://doi.org/10.1109/TASL.2007.902747

Morrison G.S. (2022a). Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science. *Forensic Science International: Synergy*, 5, 100270. https://doi.org/10.1016/j.fsisyn.2022.100270

Morrison G.S. (2022b). A method to convert traditional fingerprint ACE / ACE-V outputs (“identification”, “inconclusive”, “exclusion”) to Bayes factors. Manuscript submitted for publication. Preprint at https://geoff-morrison.net/#ID_IN_EX_to_BF

Morrison G.S., Enzinger E., Hughes V., Jessen M., Meuwly D., Neumann C., Planting S., Thompson W.C., van der Vloed D., Ypma R.J.F., Zhang C., Anonymous A., Anonymous B. (2021). Consensus on validation of forensic voice comparison. *Science & Justice*, 61, 229–309. https://doi.org/10.1016/j.scijus.2021.02.002

Weller T.J., Morris M.D. (2020). Commentary on: I. Dror, N Scurich “(Mis)use of scientific measurements in forensic science” Forensic Science International: Synergy 2020 https://doi.org/10.1016/j.fsisyn.2020.08.006. *Forensic Science International: Synergy*, 2, 701–702. https://doi.org/10.1016/j.fsisyn.2020.10.004