

Commentary on “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John (in: *Statistics and Public Policy*)

Alex Biedermann & Kyriakos N. Kotsoglou

To cite this article: Alex Biedermann & Kyriakos N. Kotsoglou (2024) Commentary on “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John (in: *Statistics and Public Policy*), *Statistics and Public Policy*, 11:1, 2288166, DOI: [10.1080/2330443X.2023.2288166](https://doi.org/10.1080/2330443X.2023.2288166)

To link to this article: <https://doi.org/10.1080/2330443X.2023.2288166>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 12 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 650




View related articles [↗](#)



View Crossmark data [↗](#)

Commentary on “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John (in: *Statistics and Public Policy*)

Alex Biedermann^a  and Kyriakos N. Kotsoglou^b 

^aSchool of Criminal Justice, University of Lausanne, Lausanne, Switzerland; ^bSchool of Law, University of Northumbria, Newcastle Upon Tyne, UK

Dear Editor,

In their recent article, Scurich and John (hereinafter SJ) (2023) discuss methods for evaluating the performance of forensic examiners based on ideas previously presented in Dror and Scurich (2020). We are writing to reiterate our concern that the analyses presented by SJ (2023) promote illogical and retrograde thinking in forensic science and thus constitute an obstacle to long-awaited fundamental progress in the field (Biedermann and Kotsoglou 2021). Specifically, as we will briefly outline below, SJ (2023) use convoluted terminology, make assumptions that are known to be incorrect, and rely on a reporting format that is ubiquitous but that has long since been exposed as unscientific.

As regards terminology, SJ (2023) and others before them (e.g., Smith and Neal 2021) use the terms “matching” and “non-matching” to denote propositions representing ground truth states. In the present context, ground truth means whether a pair of compared items (e.g., material from a crime scene and a sample from a known source) come from the same source or from different sources, respectively. However, this is not how the terms “matching” and “non-matching” are used in forensic science. As explained by Evett et al. (2017), a match refers to a descriptive summary of the findings, that is “a judgement, by the scientist, as to whether or not the two sets of observations agree within the range of what would be expected if the questioned sample had come from the same origin as the reference sample” (Evett et al. 2017, at p. 18, *emphasis added*). Here, the highlighted “if” shows that a (non-)match is distinct from and *conditioned* by ground truth, not the reverse. This is not just semantic pedantry. Using terminology reserved for describing observations, to denote propositions, as is done in SJ (2023), is highly problematic because findings (observations) are *not* the same as propositions. The two should not be confused. More specifically, when an examiner describes observations made during a comparison between two items in terms of a match, this does not logically imply or authorize anyone to conclude, without further consideration, that the items compared are from the same source. But the problems do not end there. Even if SJ (2023) had used the terms “matching” and “non-matching” to denote findings, this would have been

inappropriate in that findings in forensic science comparisons, especially in the so-called pattern-based disciplines, are *not* discrete, as the terms “matching” and “non-matching” might suggest (Aitken, Taroni, and Bozza 2020). Instead, they are continuously valued. For this reason, as explained in detail in Morrison et al. (2017), the terms “matching” and “non-matching” should no longer be used. As an aside, it is unsound to describe the process of drawing conclusions in forensic comparisons as “decision making” because, as decision-theoretic analysis shows (e.g., Biedermann and Vuille 2018; Cole and Biedermann 2020; Taroni, Bozza, and Biedermann 2021), forensic examiners are not in a position to make “decisions” except at the cost of becoming unscientific (Kotsoglou and Biedermann 2022; Stoney 2012).

There is a further instance where SJ (2023) carry a descriptor for findings over to a ground truth state, but in a way that raises fundamental conceptual issues that undermine the rationality of the analyses proposed by the authors. In fact, one of the key assumptions made by SJ (2023) is that the category of results termed “inconclusive” could serve as a ground truth state, thus enabling a three-way ROC. As an analogy, SJ (2023) mention medical applications, such as when “deciding whether a film displays malignant lesions, benign lesions, or no lesions.” Clearly, the latter categories of medical conditions denote real(istic) ground truth states, but this cannot serve as an analogy in the way suggested by SJ (2023). Why not? In forensic science comparison contexts, there are simply *no* ground truth states other than “same source” and “different sources.” To claim or suggest otherwise would violate the principle of the excluded middle (Biedermann and Kotsoglou 2021). But even if one were to adopt such a conceptually flawed mindset, by making assumptions *known* to be incorrect, the practical application is far from obvious (see e.g., Arkes and Koehler 2022 for a critical discussion). It would just shift the problem to a new one. For all of these reasons, we disagree with the view of SJ (2023) that the three-way ROC “approach is (...) appropriate for other forensic domains that permit inconclusive decisions.” Instead, we argue that it is *inappropriate* across all domains where the term “inconclusive” is being used.

Furthermore, the analyses in SJ (2023) amount to an attempt to treat a symptom, rather than the root cause. Here, the symptom is the problem of how to summarize “inconclusive” statements made by examiners during feature-comparison work. The root cause is the traditional reporting scheme, in which forensic examiners directly express opinions about ground truth conditions. The most prominent conclusion category of this reporting scheme is the source attribution determination (SAD), also known as identification or individualization. This is the assertion that two compared items come from the same source. This reporting scheme has already been exposed as unscientific (Saks and Koehler 2005) because of the exaggerations and overstatements to which it amounts. Therefore, rather than trying to handle a consequence of this reporting scheme, that is, the conclusion category “inconclusive,” it would make more sense to abandon the reporting scheme altogether in favor of a methodologically more defensible reporting format that focuses on assessing and reporting the value of findings only (e.g., Morrison et al. 2017; Morrison 2022; Koehler, Mnookin, and Saks 2023). Such statements about the value of evidence avoid scientifically unfounded, indeed forbidden, opinions about (source) propositions. Progress toward this post-identification era will be hampered, however, as long as we continue to see, and tolerate, studies that opt for the traditional reporting scheme of “identification – inconclusive – exclusion” as the object of study. These studies may cover themselves superficially in scientificity, but beneath the surface they violate fundamental methodological principles, including the metaphysical substrate of the empirical sciences (Biedermann 2022). To conclude on a positive note, the analyses proposed by SJ (2023), through their need to resort to assumptions that openly violate logical principles, do have value in the sense that they provide us with yet another argument in support of the call to abandon the traditional reporting format of “identification – inconclusive – exclusion.”

Disclosure Statement

The authors declare that this letter was written in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

ORCID

Alex Biedermann  <http://orcid.org/0000-0002-0271-5152>
Kyriakos N. Kotsoglou  <http://orcid.org/0000-0003-3458-4182>

References

- Aitken, C. G. G., Taroni, F., and Bozza, S. (2020), *Statistics and the Evaluation of Evidence for Forensic Scientists* (3rd ed.), Chichester: Wiley. [1]
- Arkes, H. R., and Koehler, J. J. (2022), “Inconclusives and Error Rates in Forensic Science: A Signal Detection Theory Approach,” *Law, Probability and Risk*, 20, 153–168. [1]
- Biedermann, A. (2022), “The Strange Persistence of (Source) “Identification” Claims in Forensic Literature through Descriptivism, Diagnosticism and Machinism,” *Forensic Science International: Synergy*, 4, 100222. [2]
- Biedermann, A., and Kotsoglou, K. (2021), “Forensic Science and the Principle of Excluded Middle: “inconclusive” Decisions and the Structure of Error Rate Studies,” *Forensic Science International: Synergy*, 3, 100147. [1]
- Biedermann, A., and Vuille, J. (2018), “Understanding the Logic of Forensic Identification Decisions (Without Numbers),” *sui-generis*, 5, 397–413. [1]
- Cole, S. A., and Biedermann, A. (2020), “How Can a Forensic Result be a “Decision”? A Critical Analysis of Ongoing Reforms of Forensic Reporting Formats for Federal Examiners,” *Houston Law Review*, 57, 551–592. [1]
- Dror, I. E., and Scurich, N. (2020), “(Mis)use of Scientific Measurements in Forensic Science,” *Forensic Science International: Synergy*, 2, 333–338. [1]
- Evet, I. W., Berger, C. E. H., Buckleton, J. S., Champod, C., and Jackson, G. (2017), “Finding the Way Forward for Forensic Science in the US – A Commentary on the PCAST Report,” *Forensic Science International*, 278, 16–23. [1]
- Koehler, J. J., Mnookin, J. L., and Saks, M. J. (2023), “The Scientific Reinvention of Forensic Science,” *Proceedings of the National Academy of Sciences*, 120, 1–10. [2]
- Kotsoglou, K. N., and Biedermann, A. (2022), “Inroads into the Ultimate Issue Rule? Structural Elements of Communication between Experts and Fact Finders,” *The Journal of Criminal Law*, 86, 223–240. [1]
- 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*, 21, 127–129. [2]
- Morrison, G. S., Kaye, D. H., Balding, D. J., Taylor, D., Dawid, P., Aitken, C. G. G., Gittelsohn, S., Zadora, G., Robertson, B., Willis, S., Pope, S., Neil, M., Martire, K. A., Hepler, A., Gill, R. D., Jamieson, A., de Zoete, J., Ostrum, R. B., and Caliebe, A. (2017), “A Comment on the PCAST Report: Skip the “match”/“non-match” Stage,” *Forensic Science International*, 272, e7–e9. [1,2]
- Saks, M. J., and Koehler, J. J. (2005), “The Coming Paradigm Shift in Forensic Identification Science,” *Science* 309, 892–895. [2]
- Scurich, N., and John, R. S. (2023), “Three-Way ROCs for Forensic Decision Making,” *Statistics and Public Policy*, 10, 1–10. [1,2]
- Smith, A. M., and Neal, T. M. S. (2021), “The Distinction between Discriminability and Reliability in Forensic Science,” *Science & Justice*, 61, 319–331. [1]
- Stoney, D. A. (2012), “Discussion on the Paper by Neumann, Evett and Skerrett,” *Journal of the Royal Statistical Society, Series A*, 175, 399–400. [1]
- Taroni, F., Bozza, S., and Biedermann, A. (2021), “Decision Theory,” in *Handbook of Forensic Statistics*, eds. D. L. Banks, K. Kafadar, D. H. Kaye, and M. Tackett, pp. 103–130, Boca Raton, FL: CRC Press. [1]