

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 (01 Dec 2023): 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*, 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>



© 2023 The Author(s). Published with license by Taylor and Francis Group, LLC



Accepted author version posted online: 01 Dec 2023.



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



Article views: 253



View related articles [↗](#)



View Crossmark data [↗](#)

Letter to the Editor

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

Alex Biedermann^{a,*}, Kyriakos N. Kotsoglou^b

^aUniversity of Lausanne, School of Criminal Justice, 1015 Lausanne-Dorigny (Switzerland)

^bUniversity of Northumbria, School of Law, Newcastle Upon Tyne, NE1 8ST (United Kingdom)

*Corresponding author. Email: alex.biedermann@unil.ch

Abstract

This letter to the Editor comments on the paper “Three-Way ROCs for Forensic Decision Making” by Nicholas Scurich and Richard S. John, published in *Statistics and Public Policy* (2023, 10, 1–10).

Accepted Manuscript

Dear Editor,

In their recent article, Scurich and John (hereinafter SJ) [14] discuss methods for evaluating the performance of forensic examiners based on ideas previously presented in Dror and Scurich [7]. We are writing to reiterate our concern that the analyses presented by SJ [14] promote illogical and retrograde thinking in forensic science and thus constitute an obstacle to long-awaited fundamental progress in the field [4]. Specifically, as we will briefly outline below, SJ [14] 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 [14] and others before them [e.g., 15] 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. [8], a match refers to a descriptive summary of the findings, i.e. “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” [8, 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 [14], 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 authorise 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 [14] 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 [1]. Instead, they are continuously valued. For this reason, as explained in detail in [12], 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., 5, 6, 17], forensic examiners are not in a position to make “decisions” except at the cost of becoming unscientific [10, 16].

There is a further instance where SJ [14] 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 [14] 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 [14] 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 [14]. 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 [4]. 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. [2] 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 [14] that the three-way ROC “approach is (...) appropriate for other forensic domains that permit inconclusive decisions”. Instead, we argue that it is *in*appropriate across all domains where the term “inconclusive” is being used.

Furthermore, the analyses in SJ [14] amount to an attempt to treat a symptom, rather than the root cause. Here, the symptom is the problem of how to summarise “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 (S.A.D.), also known as identification or individualisation. This is the assertion that two compared items come from the same source. This reporting scheme has already been exposed as unscientific [13] because of the exaggerations and overstatements to which it amounts. Therefore, rather than trying to handle a consequence of this reporting scheme, i.e. the conclusion category “inconclusive”, it would make more sense to abandon the reporting scheme altogether in favour of a more methodologically defensible reporting format that focuses on assessing and reporting the value of findings only [e.g., 9, 11, 12]. Such statements about the value of evidence avoid scientifically unfounded, indeed forbidden, opinions about (source) propositions. Progress towards 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 [3]. To conclude on a positive note, the analyses proposed by SJ [14], 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”.

Conflict of Interest 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.

References

- [1] Aitken, C. G. G., F. Taroni, and S. Bozza (2020). *Statistics and the Evaluation of Evidence for Forensic Scientists* (Third ed.). Chichester: John Wiley & Sons.
- [2] Arkes, H. R. and J. J. Koehler (2022). Inconclusives and error rates in forensic science: a signal detection theory approach. *Law, Probability and Risk* 20, 153–168.
- [3] Biedermann, A. (2022). The strange persistence of (source) “identification” claims in forensic literature through descriptivism, diagnosticism and machinism. *Forensic Science International: Synergy* 4, 100222.
- [4] Biedermann, A. and K. Kotsoglou (2021). Forensic science and the principle of excluded middle: “inconclusive” decisions and the structure of error rate studies. *Forensic Science International: Synergy* 3, 100147.
- [5] Biedermann, A. and J. Vuille (2018). Understanding the logic of forensic identification decisions (without numbers). *sui-generis* 5, 397–413.
- [6] Cole, S. A. and A. Biedermann (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.
- [7] Dror, I. E. and N. Scurich (2020). (Mis)use of scientific measurements in forensic science. *Forensic Science International: Synergy* 2, 333–338.
- [8] Evett, I. W., C. E. H. Berger, J. S. Buckleton, C. Champod, and G. Jackson (2017). Finding the way forward for forensic science in the US – A commentary on the PCAST report. *Forensic Science International* 278, 16–23.
- [9] Koehler, J. J., J. L. Mnookin, and M. J. Saks (2023). The scientific reinvention of forensic science. *Proceedings of the National Academy of Sciences* 120, 1–10.

[10] Kotsoglou, K. N. and A. Biedermann (2022). Inroads into the ultimate issue rule? structural elements of communication between experts and fact finders. *The Journal of Criminal Law* 86, 223–240.

[11] 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.

[12] Morrison, G. S., D. H. Kaye, D. J. Balding, D. Taylor, P. Dawid, C. G. G. Aitken, S. Gittelsohn, G. Zadora, B. Robertson, S. Willis, S. Pope, M. Neil, K. A. Martire, A. Hepler, R. D. Gill, A. Jamieson, J. de Zoete, R. B. Ostrum, and A. Caliebe (2017). A comment on the PCAST report: Skip the “match”/“non-match” stage. *Forensic Science International* 272, e7–e9.

[13] Saks, M. J. and J. J. Koehler (2005). The coming paradigm shift in forensic identification science. *Science* 309, 892–895.

[14] Scurich, N. and R. S. John (2023). Three-way ROCs for forensic decision making. *Statistics and Public Policy* 10, 1–10.

[15] Smith, A. M. and T. M. S. Neal (2021). The distinction between discriminability and reliability in forensic science. *Science & Justice* 61, 319–331.

[16] Stoney, D. A. (2012). Discussion on the paper by Neumann, Evett and Skerrett. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175, 399–400.

[17] Taroni, F., S. Bozza, and A. Biedermann (2021). Decision theory. In D. L. Banks, K. Kafadar, D. H. Kaye, and M. Tackett (Eds.), *Handbook of Forensic Statistics*, pp. 103–130. Boca Raton: CRC Press.