Letter to the Editor in response to FSI 314 (2020) 110388

Forensic Science International

Letter to the Editor

--Manuscript Draft--

Manuscript Number:	
Article Type:	Letter to the Editor
Keywords:	Calibration; Evidence evaluation; Functional data analysis; Likelihood ratio
Corresponding Author:	Colin Aitken University of Edinburgh Edinburgh, UNITED KINGDOM
First Author:	Colin Aitken
Order of Authors:	Colin Aitken
	Ya-Ting Chang
	Patrick Buzzini
	Grzegorz Zadora
	Genevieve Massonet
Suggested Reviewers:	Anders Nordgaard anders.nordgaard@liu.se Expert on interpretation and evaluation of evidence in forensic science, from a separate institution to those of the people involved in the two papers which are the subject of the correspondence.
Opposed Reviewers:	

To the Editor, Forensic Science International.

Please find attached a letter written in response to

Vergeer,P., Alberink,I. Sjerps,M., and Ypma,R. (2020) Why calibrating LR-systems is best practice. A reaction to "The evaluation of evidence for microspectrophotometry data using functional data analysis", in FSI 305.

Please consider it for publication in Forensic Science International.

Yours sincerely,

Colin Aitken

School of Mathematics, The University of Edinburgh.

cgga@ed.ac.uk

23rd July 2020

Letter to the Editor in response to FSI 314 (2020) 110388

Colin Aitken[a], Ya-Ting Chang[a], Patrick Buzzini[b], Grzegorz Zadora[c,d], Genevieve Massonnet[e]

a School of Mathematics and Maxwell Institute, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK

b Department of Forensic Science, Sam Houston State University, Huntsville, TX, USA

c Institute of Forensic Research, Krakow, Poland

d Institute of Chemistry, The University of Silesia in Katowice, Katowice, Poland

e School of Criminal Justice, University of Lausanne, Switzerland

Dear Sir,

We read with interest the criticisms of our paper, Aitken et al (2019), contained in Vergeer et al (2020). The purpose of Aitken et al (2019) is to bring to the attention of forensic scientists what we believe is the first development of a statistical model for a likelihood ratio for functional data. The example to illustrate the approach uses microspectrophotometry data but there are many other data types to which the method could be applied. Exciting new possibilities are opened up for the evaluation and interpretation of functional data. As with other first developments, the model can be improved. It is to be hoped that this development will encourage the development of other, improved, models for the analysis of functional data in forensic science. This model was part of the doctoral work of one of us (Y-T C) in which several other models were examined. The model described in Aitken et al (2019) is the one we felt was most ready for dissemination to the forensic science community. It is with regret we note that Vergeer et al (2020) ignore the potential of this work and confine their remarks to the criticism of one paragraph in the conclusion, a paragraph the text of which adds nothing to the message of Aitken et al (2019).

All we can do is reiterate here the comments in Aitken et al (2019). The main thrust of the criticism appears to be that our results do not satisfy the general result that the likelihood ratio of the likelihood ratio is the likelihood ratio; our model is not well-calibrated. We accept this criticism. As written above, this is the first model for the development of the likelihood ratio for functional data. It would be surprising if we obtained a very good model at the first time of asking. Also, the results we obtain are the best assessments of the value of the evidence for the circumstances of our analysis.

Vergeer et al (2020) use the likelihood ratio as a score. They argue that "LR-values coming from assumed statistical model families … often cannot be interpreted as such and require a so-called post-hoc calibrating step." We disagree. Consider the standard approach for the statistical evaluation of evidence. Statistical models, either feature-based or score-based, are developed based on training data. Likelihood ratios follow. Their performance is assessed, ideally with validation data or, failing their availability, a cross-validation analysis of the training data. These likelihood ratios are likelihood ratios by definition. The post-hoc calibrating step does not then produce a likelihood ratio in the sense that evaluation of evidence defines it. The resultant statistic is not the ratio of the probabilities (loosely defined) of the evidence given the prosecution proposition and the evidence given the defence proposition. One can, of course, define the original likelihood ratio as a score as Vergeer et al (2020) suggest but this is an artificial construct based on a post-hoc desire to obtain a better result. Once a likelihood ratio is determined, it cannot be adjusted in the manner suggested by calibration.

We cannot accept that because a model is not well-calibrated the results should then be adjusted to obtain something that is well-calibrated. This is an adjustment made after the analysis to obtain a result which looks better in some sense. The correct response to poor calibration is to look for a better model. Consider the weather forecaster whose 90% predictions tend only to be right 70% of the time. The correct response is not an automatic adjustment of 90% to 70%. It is to obtain a better model for forecasting.

We comment that "calibration destroys the updating process of Bayes' theorem whereby the posterior odds of one piece of evidence becomes the prior odds for the next piece of evidence". Vergeer et al (2020) respond with an explanation of how a calibrated likelihood ratio can be updated. This is done by treating the original likelihood ratio as a score and working with the adjusted likelihood ratio. Of course, this treatment enables the updating process of Bayes' theorem to be implemented. However, as we do not accept that calibration is a valid approach for the evaluation of evidence, we do not then accept that their procedure for updating is valid.

Calibration is a measure for the assessment of performance of a model. It is not a method for the evaluation of evidence. Some further comments about the role of calibration in the evaluation and interpretation of evidence are contained in a forthcoming book, Aitken et al (2020).

Incidentally, we find it strange that a paper criticising another paper should use in its title the title of the other paper and yet not include the other paper in the list of references.

References:

Aitken,C.G.G., Chang,Y.-T., Buzzini,P., Zadora,G. and Massonnet,G. (2019) The evaluation of evidence for microspectrophotometry data using functional data analysis. *Forensic Science International*, 305, 10.1016/j.forsciint.2019.110007.

Aitken,C.G.G., Taroni,F. and Bozza,S. (2020) *Statistics and the evaluation of evidence for forensic scientists*, Third edition, John Wiley and Sons Ltd.

Vergeer,P., Alberink,I. Sjerps,M., and Ypma,R. (2020) Why calibrating LR-systems is best practice. A reaction to "The evaluation of evidence for microspectrophotometry data using functional data analysis", in FSI 305.