LETTER TO THE EDITOR

Authors' response

See Original Article here.
See Commentary on here.

Editor,

AIM AND MOTIVATION

We would like to begin by describing the motivation and aim of the current recommendations [1]: Within the German judicial system—which is inquisitorial in nature—forensic scientists do not act as “expert witnesses” as they do in adversarial court systems, but as experts of the court. The expert, in fact, is part of the court and not a “witness.” The main task of the expert is to provide the court with an expertise. This means that there is usually only one expert to give evidence and the parties usually do not have their own experts as “advisors.” A lack of cross examination means that it remains the expert's task to reveal and explain any uncertainties associated with the evidence presented. Furthermore, it is the general idea within the German legal system that experts are replaceable, that is, a second expert—if presented with the same facts—is expected to reach the same conclusions.

These underlying principles influence the way we express an “expert's opinion.” In the current situation, most German experts still use binary models to calculate likelihood ratios (LRs), which means that many trace DNA profiles are not subjected to statistical evaluation at all. In Germany, laboratories are free to decide which fully continuous model (FCM) software to use. It is therefore to be expected that various FCM programs will be implemented. The assumption that experts should be replaceable leads to the expectation that different calculation models used by different experts should yield at least very similar results. The fact that this cannot be expected from FCMs, currently hinders the implementation of such programs. Some colleagues would even go as far as claiming that FCM therefore cannot and must not be used in Germany. Providing a reporting framework by implementing reporting thresholds is currently a necessary step to overcome this obstacle and enable laboratories to start validating and implementing FCM.

CALIBRATION

Indeed, the four programs described by Templin et al. [2] were not fully calibrated according to manufacturers’ recommendations.

We agree that observed differences are not related to an error in calculations but due to slightly differing parameters and modeling choices. Different calculation results are inherent in this method due to different mathematical models. And it can be assumed that some variation might be the result of incomplete implementation.

It was, however, not our aim to evaluate different FCM regarding their performance, but rather to learn, to what extent differences between FCM have to be expected. While we agree that from a purely theoretical point of view, a direct comparison of different FCM might not be meaningful, understanding differences is necessary from the German's user’s point of view. Since there is no “true” LR, we agree that evaluating the performance of the models would be meaningless.

The importance of calibrating FCM programs is explicitly highlighted several times throughout the recommendations. Each laboratory must determine for itself, based on the results of validation studies, a range of application for the program used (e.g., see the "Conclusion" section):

"Die Vorgaben der Programmhersteller in Bezug auf die für die Auswertung relevanten Laborparameter einschließlich experimenteller Analysen von Spuren bekannter Zusammensetzung im Rahmen einer Validierung bzw. Verifizierung sind zu beachten und zu dokumentieren."

LOWER THRESHOLD

Regarding the recommendation of a lower threshold for reporting LRs, we certainly agree that a positive LR < $10^6$ may still provide some evidential value. We would like to draw attention to the verbalization we suggest, clearly stating that there is some support for H1. We observed in our own work [2], however, that multiple calculations with different FCM using the same raw data, lead to differing LRs. Alladio et al. [3] recently compared Lab Retriever, LRmix Studio, DNA-VIEW, EuroForMix, and STRmix. In general, the quantitative models used in DNA-VIEW, EuroForMix, and STRmix performed similarly while the qualitative models used in Lab Retriever...
and LRmix Studio also performed similarly to each other, but differed from the quantitative methods. They concluded that "results provided by fully continuous models proved similar and convergent to one another, with slightly higher within-software differences (i.e., approximatively 3–4 degrees of magnitude)." This shows the need to understand differences between models and how they might affect how an LR is perceived.

When different models used on the same datasets produce LRs that provide extremely strong support for the same hypotheses over the alternative, such differences usually do not change the way the LR is perceived. When different models produce LR values closer to 1, however, this might change the way they are perceived. The case described by Gill et al. [4] shows that there is a real risk of achieving negative LRs with a second calculation method, if the first calculation yielded such a "low" LR. We are aware of the discussion on the reliability of low LRs within one model. There is, however, currently very limited data available describing the extent of differences between models. This is why we recommend for the time being and in line with others (e.g., [5]), a very careful interpretation of LRs with the necessary distance to the "prosecutor’s fallacy" and leaves the "gray zone," however, our opinion differs from that expressed by the authors: If we cannot be satisfied that LR values within the "gray zone" are reliable and reproducible, we believe that giving the numerical value of such LRs is of very limited use to the court and potentially misleading. In our opinion, such results should therefore only be evaluated and reported in a verbal statement as before.

Thus, we recommend a more cautious assessment of LR values that fall into the so-called "gray zone." Our recommendations are adapted to the German legal framework. Furthermore, as mentioned, we observe direct support for our approach from the calibration data shown in table 1 of [6]. The data clearly show that below LR of $1.1 \times 10^6$ the number of non-contributors rises significantly.

**FINAL REMARKS/ON THE ALLEGATION OF WITHHOLDING IMPORTANT INFORMATION**

At this point we would like to clarify that we are not withholding information from the court. We agree with the authors that misleading LR values might occur in DNA analysis for reasons already discussed that cannot be avoided. A detailed implementation and validation is necessary to ensure that such misleading LR values occur as rarely as possible and to estimate the frequency with which misleading LRs occur in specific scenarios.

Regarding the evaluation and presentation of LR values within the so-called "gray zone," however, our opinion differs from that expressed by the authors: If we cannot be satisfied that LR values within the "gray zone" are reliable and reproducible, we believe that giving the numerical value of such LRs is of very limited use to the court and potentially misleading. In our opinion, such results should therefore only be evaluated and reported in a verbal statement as before.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Meinhard Hahn PhD
Cornelius Courts PhD
Martin Eckert PhD
Rolf Fimmers PhD
Stefanie Grethe PhD
Sebastian Kranz PhD
Christoph Leuker PhD
Claus Oppelt PhD
LETTER TO THE EDITOR

Sven Razbin PhD
Michael Templin PhD
Marielle Vennemann PhD
Peter Zimmermann PhD
Katja Anslinger PhD

1 State Criminal Police Office of Lower Saxony, Hanover, Germany
2 Institute of Legal Medicine, University Hospital of Cologne, Cologne, Germany
3 Federal Criminal Police Office, Wiesbaden, Germany
4 Institute for Forensic Statistics and Quality Assurance, St. Augustin, Germany
5 State Criminal Police Office of Rhineland-Palatinate, Mainz, Germany
6 State Criminal Police Office of Hamburg, Hamburg, Germany
7 State Criminal Police Office of North Rhine-Westphalia, Dusseldorf, Germany
8 State Criminal Police Office of Bremen, Bremen, Germany
9 Institute of Legal Medicine, University of Munster, Munster, Germany
10 State Criminal Police Office of Baden-Wuerttemberg, Stuttgart, Germany
11 Institute of Legal Medicine, Ludwig Maximilian University, Munich, Germany

Correspondence
Katja Anslinger, Institute of Legal Medicine, Ludwig Maximilian University, Munich, Germany.
Email: katja.anslinger@med.uni-muenchen.de

REFERENCES


6. Berger C, Kruijver M, Hicks T, Champod C, Buckleton J. Reaction to the recommendations by Hahn et al. regarding the interpretation and reporting of LRs from fully continuous models. JOFS In press.
