Crises of uncertainty, reproducibility, and replicability in forensic comparison

Forensic validation tends to focus on the overall performance of methods under casework conditions. This implicitly focuses the expert's attention on discriminability (see Smith & Neal 2021), with different methods chosen, or decisions made, based low values for the validity metric used. In this paper, we argue against this view. Rather, we believe that **the expert's primary concern should be to reduce uncertainty** (Morrison & Enzinger 2016, Ramos et al 2021), rather than maximising discriminability. This is because uncertainty is directly related to the probability of a miscarriage of justice, which should be the expert's ultimate concern.

In this paper, we present results from a score simulation study demonstrating the variability in validation results as a function of sampling variability and demonstrate how Bayesian calibration (Brümmer & Swart 2014) may reduce such uncertainty. We also highlight how this problem applies to all forms of forensic comparison (including those based on entirely human-based, subjective methods), and present a series of specific recommendations. Experts should:

1. Recognise that forensic comparison is a process involving numerous decisions which introduce uncertainty via both systematic and random factors
2. Be explicit about the decisions made at each stage of the process and the implications of such decisions for uncertainty in terms of the results LRs **and** overall method validity
3. Take steps to measure and minimise uncertainty

The focus on uncertainty also directly relates to issues of reproducibility and replicability. In this paper, we also consider the specific challenges these concepts pose for forensic sciences.

References:

Brümmer, N. and Swart, A. (2014) Bayesian calibration for forensic evidence reporting. *Proceedings of Interspeech*, Singapore, 14-18 September. pp. 388-392.

Morrison, G. S. and Enzinger, E. (2016) What should a forensic practitioner's likelihood ratio be? *Science and Justice* (Virtual Special Issue on measuring and reporting the precision of forensic likelihood ratios), 56: 374-379.

Morrison, G. S. and Poh, N. (2018) Avoiding overstating the strength of forensic evidence: Shrunk likelihood ratios/ Bayes factors. *Science and Justice*, 58: 200-218.

Smith, A. M. and Neal, T. M. S. (2021) The distinction between discriminability and reliability in forensic science. *Science and Justice*, 61: 319-331.