A comment on the PCAST report: Skip the "match"/"non-match" stage

Abstract

This letter comments on the report "Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods" recently released by the President's Council of Advisors on Science and Technology (PCAST). The report advocates a procedure for evaluation of forensic evidence that is a two-stage procedure in which the first stage is "match"/"non-match" and the second stage is empirical assessment of sensitivity (correct acceptance) and false alarm (false acceptance) rates. Almost always, quantitative data from feature-comparison methods are continuously-valued and have within-source variability. We explain why a two-stage procedure is not appropriate for this type of data, and recommend use of statistical procedures which are appropriate.

Keywords: PCAST report; forensic statistics; match/non-match; sensitivity; false alarm; likelihood ratio

Highlights

- Feature-comparison methods produce continuously-valued data.
- The PCAST report advocates a two-stage procedure:
- (1) Dichotomise the data into "match" or "non-match".
- (2) If "match", assess correct acceptance and false acceptance rates.
- A better procedure would directly statistically model the continuously-valued data.

On September 20, 2016, the President's Council of Advisors on Science and Technology (PCAST) released their report: *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods* [1]. The report is rightly critical of "heterodox" "non-empirical" views ([1] §4.7, pp 59–63), and we wholeheartedly endorse the report's call for forensic analysis methods to be empirically validated under casework conditions. We see the report as an important contribution to improving forensic science practice, and implementation of the report's recommendations would constitute a major step forward. Our intention in this letter is to encourage an additional step forward.

The PCAST report advocates a procedure for evaluating strength of evidence that is a substantial improvement over historical (and in many places current) practice, but further improvement is desirable and achievable. The report describes a procedure for quantifying "probative value"[1] in which if a forensic practitioner declares a "match",[2] they also report the results of an empirical assessment of the probability of declaring a "match" if the questioned-source specimen came from the known source[3] and the probability of declaring a "match" if the questioned-source specimen came from some other source.[4] "The forensic examiner should report the overall false positive rate and

---

[1] what we call "strength of evidence"

[2] also called "proposed identification" in the report ([1] p 46)

[3] this would be the numerator for a likelihood ratio

[4] this would be the denominator for a likelihood ratio

sensitivity for the method established in the [empirical] studies of foundational validity and should demonstrate that the samples used in the foundational studies are relevant to the facts of the case." ([1] p 56). This is not an inappropriate procedure for quantifying strength of evidence if the data are discrete and have no within-source variability.

If, however, the data are continuously-valued and have within-source variability,[5] then such a procedure discards information that could be exploited by more appropriate statistical procedures. Almost always, measurements made in "feature-comparison methods" will naturally result in continuously-valued data with within-source variability. For this type of data it is generally inappropriate to use a procedure which includes a stage that assesses "whether the features in an evidentiary sample and the features in a sample from a suspected source lie within a pre-specified measurement tolerance" ([1] p 48). Such a procedure suffers from a cliff-edge effect: A questioned-source specimen which falls just above the threshold for "match" with the known-source sample and a questioned-source specimen which falls just below the threshold will result in very different conclusions as to the strength of the evidence, even though the difference between the two is negligible (the two specimens could in fact be from the same source, with the difference between them due to within-source variability). Also, a procedure that includes a "match"/"non-match" stage limits the strength-of-evidence conclusion to one of two possible values: A questioned-source specimen which vastly exceeds the threshold will be assessed as having exactly the same strength of evidence as a questioned-source specimen which just exceeds the threshold, even if the former should in theory constitute much stronger evidence than the latter. *Mutatis mutandis* for a specimen which falls just short of the threshold and one which falls far below the threshold.

A more appropriate procedure would not include a "match"/"non-match" stage, would not use a threshold, and would instead directly assess two probabilities[6] based on the continuously-valued data: (1) The probability of obtaining the measured properties of the questioned-source specimen had it come from the known source; versus (2) the probability of obtaining the measured properties of the questioned-source specimen had it come not from the known source but from some other source in the relevant population. The former is the numerator and the latter is the denominator of a likelihood ratio.[7] There is a substantial body of literature describing and validating statistical procedures which work directly with continuously-valued data. Such statistical procedures would *a priori* be expected to have higher degrees of validity than those that include a "match"/"non-match" stage, and for particular applications actual performance can be compared via empirical tests (e.g., [2], [3]).

---

[5] Within-source variability could be intrinsic or due to transfer or measurement processes.

[6] technically "likelihoods"

[7] There are multiple appropriate statistical procedures for calculating forensically interpretable likelihood ratios for continuously-valued data. The choice of the best procedure to use will depend on the details of the particular question to be answered in the case, the structure of the data to be analysed, and empirical testing of performance. To accurately represent many of these procedures would require more complex descriptions than the verbal description of a likelihood ratio just given in the main text. For simplicity, we only include the latter description in the present letter.

The PCAST report does not exhibit familiarity with the extensive existing literature on forensic inference and statistics, very little is referenced. The history of forensic science includes multiple examples in which procedures including a "match"/"non-match" stage were advocated and used, but which were subsequently replaced by procedures that more directly exploit continuously-valued measurements. Aitken & Taroni ([4] pp 10–11) and Foreman et al. ([5] pp 474–476) discuss examples from glass and DNA respectively. Additional publications critical of two-stage procedures in which the first stage is "match"/"non-match" include [6]–[16]. Progress toward "ensuring scientific validity of feature-comparison methods" will be quicker if forensic practitioners skip the "match"/"non-match" procedure advocated in the PCAST report and move directly to using validated statistical procedures which are appropriate for continuously-valued data.

Acknowledgments

This work was funded in-part by a fellowship awarded to Morrison by the Simons Foundation. Morrison, Balding, Dawid, Aitken, Robertson, Pope, Neil, Martire, Gill, Jamieson, de Zoete, and Caliebe would like to thank the Isaac Newton Institute for Mathematical Sciences for its hospitality during the program Probability and Statistics in Forensic Science which was supported by EPSRC Grant Number EP/K032208/1. All opinions expressed are those of the authors/signatories and do not necessarily represent the opinions or policies of any funding agencies or organizations with which they are affiliated.

References

[1] President's Council of Advisors on Science and Technology (2016) *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Washington DC: Executive Office of The President's Council of Advisors on Science and Technology. https://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports

[2] Berry DA, Evett IW, Pinchin R (1992) Statistical inference in crime investigations using deoxyribonucleic acid profiling. Appl Stat 41:499-531 http://www.jstor.org/stable/2348086

[3] I.W. Evett, J. Scranage, R. Pinchin, An illustration of efficient statistical methods for RFLP analysis in forensic science, American Journal of Human Genetics 52 (1993) 498–505.

[4] C.G.G. Aitken, F., Taroni F., Statistics and the Evaluation of Forensic Evidence for Forensic Scientist (2nd Ed.), Wiley, Chichester, UK, 2004. http://dx.doi.org/10.1002/0470011238

[5] L.A. Foreman, C. Champod, I.W. Evett, J.A. Lambert, S. Pope, Interpreting DNA evidence: A review, International Statistical Review 71 (2003) 473–495. http://dx.doi.org/10.1111/j.1751-5823.2003.tb00207.x

[6] C.G.G. Aitken, Statistics in forensic science. Part II. An aid to evaluation of evidence. Problems of Forensic Sciences 65 (2006) 68–81. http://www.forensicscience.pl/component/option,com_jbook/task,view/Itemid,2/catid,15/id,101/lang,en/

[7] D.A. Berry, Inferences using DNA profiling in forensic identification and paternity cases. Statistical Science 6 (1991) 175–205.

[8] J.M. Curran, T.N. Hicks, J.S. Buckleton, Forensic Interpretation of Glass Evidence, CRC Press, Boca Raton, FL, 2000.

[9] I.W. Evett, Interpretation: a personal odyssey, in C.G.G. Aitken, D.A. Stoney (Eds.), The Use of Statistics in Forensic Science, Ellis Horwood, Chichester, UK, 1991, pp 9–22.

[10] D.H. Kaye, D.E. Bernstein, J.L. Mnookin, The New Wigmore: A Treatise on Evidence: Expert Evidence (2nd Ed.), Aspen Publishing, New York, 2011, §14.2.1.

[11] D.V. Lindley, A problem in forensic science, Biometrika 64 (1977) 207–213. http://links.jstor.org/sici?sici=0006-3444%28197708%2964%3A2%3C207%3AAPIFS%3E2.0.CO%3B2-E

[12] G.S. Morrison, Forensic voice comparison, in I. Freckelton, H. Selby (Eds.), Expert Evidence, Thomson Reuters, Sydney, Australia, 2010, ch. 99. http://expert-evidence.forensic-voice-comparison.net/

[13] B. Robertson, G.A. Vignaux, DNA evidence: Wrong answers or wrong questions? in B.S. Weir (Ed.), Human Identification: The Use of DNA Markers, Kluwer, Dordrecht, The Netherlands, 1995, pp 145–152. http://dx.doi.org/10.1007/978-0-306-46851-3_16

[14] K. Roeder, DNA fingerprinting: A review of the controversy, Statistical Science 9 (1994) 222–247.

[15] P. Rose, G.S. Morrison, A response to the UK position statement on forensic speaker comparison, International Journal of Speech, Language and the Law 16 (2009) 139–163. http://dx.doi.org/10.1558/ijsll.v16i1.139

[16] G. Zadora, A. Martyna, D. Ramos, C.G.G. Aitken, Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data, Wiley, Chichester, UK, 2014, http://dx.doi.org/10.1002/9781118763155

Geoffrey Stewart Morrison *

Independent Forensic Consultant, Vancouver, British Columbia, Canada

Adjunct Professor, Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada

Simons Foundation Visiting Fellow, Isaac Newton Institute for Mathematical Sciences, Cambridge, England, United Kingdom

David H. Kaye

Distinguished Professor and Weiss Family Scholar, Penn State Law, Pennsylvania State University, University Park, Pennsylvania, United States of America

Regents' Professor Emeritus, Arizona State University College of Law and Department of Life Sciences, Tempe, Arizona, United States of America

David J. Balding

Professor of Statistical Genetics, Centre for Systems Genomics, School of Biomedical Sciences, and School of Mathematics & Statistics, University of Melbourne, Melbourne, Victoria, Australia

Duncan Taylor

Principal Scientist of Forensic Statistics, Forensic Science South Australia, Adelaide, South Australia, Australia

Associate Professor of Biology, School of Biological Sciences, Flinders University, Adelaide, South Australia, Australia

Philip Dawid

Emeritus Professor of Statistics, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, England, United Kingdom

Colin GG Aitken

Professor of Forensic Statistics, School of Mathematics, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Simone Gittelson

Forensic Statistician, Statistical Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland, United States of America

Grzegorz Zadora

Associate Professor, Institute of Forensic Research, Krakow, Poland

Senior Lecturer, Chemometric Research Group, Institute of Chemistry, University of Silesia in Katowice, Poland

Bernard Robertson

Barrister, Wellington, New Zealand

Sheila Willis

Director General, Forensic Science Ireland, Dublin, Ireland

Susan Pope

Independent Consultant, DNA Principal Forensics Ltd., Reading, England, United Kingdom

Martin Neil

Professor of Computer Science and Statistics, Risk Assessment and Decision Analysis Research group, Department of Computer Science, Queen Mary, University of London, London, England, United Kingdom

Kristy A. Martire

ARC DECRA Fellow & Senior Lecturer, School of Psychology, University of New South Wales, Sydney, New South Wales, Australia

Amanda Hepler

Senior Analyst, Innovative Decisions, Inc., Vienna, Virginia, United States of America

Richard D. Gill

Professor of Mathematical Probability, Department of Mathematics, Leiden University, Leiden, The Netherlands

Allan Jamieson

Director and Consultant Scientist, The Forensic Institute, Glasgow, Scotland, United Kingdom

Jacob de Zoete

PhD Candidate (ABD), Department of Mathematics, University of Amsterdam, Amsterdam, The Netherlands

R. Brent Ostrum

Senior Forensic Document Examiner, Forensic Document Examination Section, Science and Engineering Directorate, Canada Border Services Agency, Ottawa, Ontario, Canada

Amke Caliebe

Forensic Statistician, Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany

[*] Corresponding author. *E-mail address:* geoff-morrison@forensic-evaluation.net (G.S. Morrison).