

<p>Institution: The University of Edinburgh/Heriot-Watt University (Maxwell Institute)</p>
<p>Unit of Assessment: B10, Mathematical Sciences</p>
<p>Title of case study: Bayesian statistical methods applied to the quantification of forensic evidence</p>
<p>1. Summary of the impact</p> <p>In a series of papers published from 1999 on, Aitken (Maxwell Institute) and collaborators applied Bayesian statistics to develop a methodology for the quantification of judicial evidence derived from forensic analyses. They proposed and implemented procedures for (i) determining the optimal size of samples that should be taken from potentially incriminating material (such as drugs seized); and (ii) the estimation of likelihood ratios characterising evidence provided by multivariate hierarchical data (such as the chemical composition of crime-scene samples). Their procedures have been recommended in international guideline documents (including a 2009 publication by the United Nations Office on Drugs and Crime) and have been routinely used by forensic science laboratories worldwide since 2008. The research has therefore had an impact on the administration of justice, leading to a better use of evidence and accompanying judicial and economic benefits. Examples are given from laboratories in Australia, Sweden and The Netherlands.</p>
<p>2. Underpinning research</p> <p>Our judicial system increasingly relies on the quantification of the value of evidence presented in court. As a result, advanced statistical methods have a strong impact on the administration of justice. The key research insight in this area is the recognition that the Bayesian framework provides the tools needed for the interpretation of forensic evidence. This has led to the development of increasingly sophisticated statistical analyses driven by new measuring equipment for the examination of trace evidence and by the increase in computing power that enables the lengthy calculations required to be performed efficiently. In papers published from 1999 on Aitken (Maxwell Institute, MI) and co-workers have contributed to this development and tackled two important problems: the determination of the optimal size of samples to be taken from seized material and the treatment of multivariate, hierarchical evidence data. The methodology issued from his research has since been adopted by forensic laboratories worldwide.</p> <p>Optimal sample size. When large quantities of potentially incriminating material are seized, it is difficult to determine what fraction should be used for forensic testing: small samples are open to challenge as providing too little information; large samples are costly. A procedure that determines optimal sample sizes in terms of clearly expressed criteria is therefore of obvious benefit for the administration of justice. This led the Scottish Forensic Science Liaison Group (SFSLG) to approach Aitken in the late 1990s to resolve the problem of the lack of criteria for the choice of sample size. Motivated by this, Aitken developed a Bayesian procedure and published the underpinning statistical research in 1999 [1]. The theory applies when the sampling unit may be classified into two or more possible categories (e.g., licit or illicit). As examples we cite cases about which Aitken was directly consulted: (a) the sampling of drug tablets from consignments; (b) the sampling of computer files for evidence of child pornography; and (c) the sampling of CDs for evidence of piracy. In such cases, the theory provides an estimate of the number of tablets, computer files or CDs that need to be inspected to obtain reliable evidence, potentially sufficient for a prosecution. Further work by Aitken and collaborators (see [2] and references therein) considered the estimation of the quantity of drugs in a consignment and provided the probability distribution for the amount of illicit material as a function of the sample size.</p>

Likelihood ratios for multivariate hierarchical data. When samples of material obtained from a crime scene are compared with those obtained from a suspect, it is necessary to quantify the support for the proposition that they come from the same source. In many cases the data characterising the material is multivariate, continuous and hierarchical. Examples include the composition of glass taken from fragments of windows, or the composition of drugs. The hierarchical nature then arises because variations within-source and between-source differ (variation of glass composition in a single window pane versus variation between different panes, or variation of composition within a drug batch versus variations between batches). Research in the MI developed a Bayesian methodology to quantify the value of the evidence derived from such multivariate and hierarchical data. This overcame the drawbacks of earlier methodologies (which often incorrectly assumed the independence of the different variables) by providing a likelihood ratio (LR) that can be combined with other forms of evidence in an integrated analysis and leads to readily interpretable conclusions. The initial work by Lucy and Aitken [3] considering a two-level hierarchy of data was extended to a three-level hierarchy in [4-5]. The paper [4] also developed an implementation based on graphical modelling techniques which is adapted to multivariate data.

Dissemination. The methodology developed by Aitken and collaborators and published in [1-3] has been further disseminated through its inclusion in the book [6], a well-cited authority on the role of statistics in the evaluation of evidence in forensic science (1740 sales to 31st August 2013).

Software implementing the sampling method of [1-2] has been developed and is available on the website <http://www.enfsi.org> (see [9]). A R package ‘comparison’ computing LRs following [3] has been developed by Lucy and is freely available at <http://cran.r-project.org/web/packages/comparison/index.html>

Attribution. C. G. G. Aitken has been with the Maxwell Institute since 1979. D. Lucy was a PDRA at the Maxwell Institute from 2001 and joined the University of Lancaster in 2006. G. Zadora is at the Institute for Forensic Research in Krakow (Poland), J.M. Curran at the University of Auckland, New Zealand and F. Taroni at the Institute of Forensic Science at the University of Lausanne.

3. References to the research

Those marked with a * best indicate the quality of the research

- [1]* Aitken, C.G.G., Sampling – how big a sample? *Journal of Forensic Sciences*, **44**, 750-760 (1999). <http://www.maths.ed.ac.uk/~mthdat25/forensic/Sampling-how-big-a-sample>
- [2] Aitken, C. G. G. and Lucy, D., Estimation of the quantity of a drug in a consignment from measurements on a sample, *Journal of Forensic Sciences*, **47**, 968-975 (2002). <http://www.maths.ed.ac.uk/~mthdat25/forensic/Estimation-of-the-quantity-of-a-drug-in-a-consignment-from-measurements-on-a-sample>
- [3]* Aitken, C.G.G. and Lucy, D., Evaluation of trace evidence in the form of multivariate data. *Applied Statistics*, **53**, 109-122, with corrigendum 665-666 (2004). <http://www.maths.ed.ac.uk/~mthdat25/forensic/Evaluation-of-trace-evidence-in-the-form-of-multivariate-data-plus-corrigendum>
- [4] Aitken, C.G.G., Zadora, G. and Lucy, D., A two-level model for evidence evaluation. *Journal of Forensic Sciences*, **52**, 412-419 (2007). <http://dx.doi.org/10.1111/j.1556-4029.2006.00358.x>
- [5] Aitken, C.G.G., Lucy, D., Zadora, G. and Curran, J.M., Evaluation of trace evidence for three-level multivariate data with the use of graphical models, *Computational Statistics and Data Analysis*, **50**, 2571-2588 (2006). <http://dx.doi.org/10.1016/j.csda.2005.04.005>
- [6]* Aitken, C.G.G. and Taroni, F., *Statistics and the evaluation of evidence for forensic scientists*, John Wiley and Sons Ltd (2004, 2nd edition).

Grants. Aitken's research on been funded by a series of research grants: SHEFC (01.03.01–31.07.04), value: £338,366. ESRC RES-000-23-0729 (01.10.04–31.03.08), value: £205,292. EPSRC GR/S98603/01 (01.12.04–31.03.07), value: £90,598. EPSRC EP/C532627 (01.08.2006–31.07.2008), value: £95,538.

4. Details of the impact

Beneficiaries. The beneficiaries of the research are forensic science services and law-enforcement agencies worldwide. They can now optimise the size of the samples they test and quantify in precise Bayesian terms the weight of evidence. This impact on professional practice in turn improves the judicial system of the countries relying on these services and agencies by enabling the best use of the evidence available and ultimately leading to safer verdicts.

Impact on beneficiaries. The impact started in the late 1990s with the initial work leading up to [1]: the procedure was referred by the SFSLG to the Crown Office in Scotland which approved the ideas and issued guidance to the Scottish forensic science laboratories for the procedure to be used in cases in which sampling was desirable [9]. Cases (a)-(c) are examples of this early impact which led to cost savings and, in the case (b) of sampling of pornographic files, to a reduction of stress-related illnesses amongst the law enforcement agents examining the files (prior to Aitken's involvement, out of four officers of the Strathclyde Police Force who examined all files on certain seized computers in a particular case, three had to take sick leave on stress-related grounds).

The impact of [1-2] has considerably extended since 2008, due in part to the publication of high-profile guidance documents published by crime enforcement agencies that refer to the work; these include the 'Guidance for best practice sampling in forensic science' published in 2007 by the European Network of Forensic Science Institutes (ENFSI, which represents forensic science laboratories throughout Europe including Russia, also Turkey and some trans-Caucasian countries), and the 'Guidelines on representative drug sampling' [7] published in 2009 jointly by the United Nations Office on Drugs and Crime and ENSFI. The software implementing the sampling method of [1-2] is available on the ENFSI website: <http://www.enfsi.org> (see [8]). It is used widely in Europe (including Sweden, The Netherlands, Poland, Switzerland, UK) and is disseminated world-wide.

We document the adoption of Aitken's methodology for both sample-size determination [1-2] and LR for multivariate hierarchical data [3-5] by describing three specific examples of applications in laboratories in Australia, Sweden, and the Netherlands.

Australian National University. [Text removed for publication]. The method of [3] was applied by ANU consultants to a high-profile court case in Australia to estimate the strength of the evidence of a telephone conversation. [Text removed for publication].

Since this case, the LR derived in [3] has been used more broadly in cases involving voice comparison. A senior staff member of the Forensic Voice Comparison Laboratory (University of New South Wales, Australia) has commented that *'the work on statistical modelling for numerical calculation of the strength of forensic evidence [3] has become a standard tool in the field of forensic voice comparison'* [11].

Statens Kriminaltekniska Laboratorium (SKL, Swedish National Laboratory of Forensic Science). SKL, practices a framework for sampling of drug units that is built on [1]. The paper [1] gave rise to a research project within SKL, that led to general rules for sampling of pills; according to senior SKL staff, the process *'has substantially reduced the amount of material that needs to be analysed, still preserving the precision needed for legal purposes, and has hence increased cost-*

Impact case study (REF3b)

efficiency' [12]. SKL are in the process of implementing the approach described in [3] for the comparison of amphetamine seizures and for the strengthening of glass evidence by the use of composition measurements.

Netherlands Forensic Institute. The glass experts at the Netherlands Forensic Institute now use the method developed in [3] in every case as a support to earlier analyses. The verbal statements of the value of the evidence that they issue to the court are on both methods and on graphical displays. A senior forensic statistician at the Netherlands Forensic Institute has commented that *'the ground breaking work of Aitken and others has transformed the way we evaluate forensic evidence'* and *'the LR method is the next step in the evolution from forensic craft to forensic science* [13]'.

5. Sources to corroborate the impact

[7] United Nations Office on Drugs and Crime Guidelines on representative drug sampling. *UNITED NATIONS PUBLICATION*; Sales No. E.09.XI.13 ISBN 978-92-1-148241-6 (2009).

See <http://www.maths.ed.ac.uk/~mthdat25/forensic/UN-Office-on-Drugs-and-Crime-Drugs-Sampling-Guidelines> or [8].

[8] ENFSI publications may be found on the website: <http://www.enfsi.org>

Click on 'Documents' then 'External Publications'. Three are of relevance:

- a. Validation of the 'Guidelines on representative sampling_DWG-SLG-001-vers002.
- b. Drugs Sampling Guideline UNODC-ENFSI.
- c. ENFSI DWG Calculator for Qualitative Sampling of seized drugs (2012) (Software).

Confirmation of the benefits of the research to forensic science can be obtained from:

[9] Senior manager of the Forensic Science Services, Scottish Police Services Authority.

[10] Senior member of the Forensic Speech Science Committee, Australasian Speech Science.

[11] Senior member of the Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney, New South Wales, Australia.

[12] Senior statistician at SKL.

[13] Senior statistician the Netherlands Forensic Institute.

Note: should links to web pages be broken, please use the website

<http://www.maths.ed.ac.uk/~mthdat25> to access pdf versions of the pages