Science and Justice 56 (2016) 402–410

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Position paper

Using sensitivity analyses in Bayesian Networks to highlight the impact of data paucity and direct future analyses: a contribution to the debate on measuring and reporting the precision of likelihood ratios

Duncan Taylor [a,b,*], Tacha Hicks [c], Christophe Champod [d]

[a] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia
[b] School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia
[c] Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice and Fondation pour la formation continue UNIL-EPFL, University of Lausanne, Lausanne-Dorigny, Switzerland
[d] Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, University of Lausanne, Lausanne-Dorigny, Switzerland

ARTICLE INFO

Keywords:
Sensitivity analysis
Bayesian networks
Likelihood ratio
Data
Source level propositions

ABSTRACT

Bayesian networks are being increasingly used to address complex questions of forensic interest. Like all probabilities, those that underlie the nodes within a network rely on structured data and knowledge. Obviously, the more structured data we have, the better. But, in real life, the numbers of experiments that can be carried out are limited. It is thus important to know if/when our knowledge is sufficient and when one needs to perform further experiments to be in a position to report the value of the observations made. To explore the impact of the amount of data that are available for assessing results, we have constructed Bayesian Networks and explored the sensitivity of the likelihood ratios to changes to the data that underlie each node. Bayesian networks are constructed and sensitivity analyses performed using freely available R libraries (gRain and BNlearn). We demonstrate how the analyses can be used to yield information about the robustness provided by the data used to inform the conditional probability table, and also how they can be used to direct further research for maximum effect. By maximum effect, we mean to contribute with the least investment to an increased robustness. In addition, the paper investigates the consequences of the sensitivity analysis to the discussion on how the evidence shall be reported for a given state of knowledge in terms of underpinning data.

1. Introduction

This paper aims to contribute to the ongoing debate regarding the relevancy of reporting the precision associated with a likelihood ratio (*LR*). The case of the determination of the nature of body fluids will be used to illustrate the argument and explore practical implications. For the sake of this introduction, we reduce the problem to a potential bloodstain where a stain is observed on the garment of a person of interest. One single presumptive test for human blood is carried out and gives a positive result. The test is known to have a false positive rate of 0.01 and a false negative rate of 0.1. If the propositions of prosecution and defence are respectively 'The stain is human blood' (H_p) and 'The stain is not human blood' (H_d), the *LR* associated with the positive result can be written as: Pr(a positive test result| H_p, I)/Pr(a positive test result| H_d, I) = 0.9/0.01. The information 'I' represents what is known, told and assumed, here the data associated with the test. In this case, our *LR* is assigned as 90. The typical questions that will be explored in this paper are:

- How sensitive is our *LR* of 90 to the data that underpin the rates of false positive and false negative?
- Should this sensitivity be reflected in the reporting of the *LR* by the introduction, for example, of a confidence (or credible) interval associated with our *LR*?

The argument we will try to convey is that, in the above case, there is no such thing as a "true value" for the likelihood ratio, and that the *LR* of 90 conveys in itself all that needs to be known about the weight to be assigned to the forensic results. However, we do not wish to imply that measuring the variability of likelihood ratios and their dependency on the data is useless for forensic scientists. It is useful to decide whether knowledge is sufficient for robust reporting. However, this is a different question (a question about data) and cannot be answered by giving the value of the results in the case at hand.

* Corresponding author at: Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia.
E-mail address: Duncan.Taylor@sa.gov.au (D. Taylor).

http://dx.doi.org/10.1016/j.scijus.2016.06.010
1355-0306/

These questions are currently debated in the literature [1–3]. For example, Ali et al. [4], following their analysis of sampling variability in the training sets used in biometrics, suggest that "*a range of LRs should be reported which incorporates the sampling variability instead of reporting a single value of the LR.*" Sjerps et al. [2] advocate the need for a full transparency on the statistical analysis and not depriving the fact finder from any information that may help them to assess the trustworthy of the reported *LR*. They also refer to forensic publications introducing variability measures on the *LR* in areas such DNA, traces, drugs or speaker recognition. Taroni et al. [3] argue that the very nature of the *LR* encapsulates all required uncertainty and does not need any complementary measures. Our own thinking aligns with that of Taroni et al., who in [3], eloquently wrote that probability is a state of mind, and to present multiple values for a probability (such as a point estimate and a probability interval) is akin to having two different states of mind, and is hence logically flawed.

Body fluid attribution in forensic science is more complex than the above example because the types of fluids encountered in forensic science may be more varied than human blood. Moreover forensic observations can be multiple and dependent (e.g., visual observations, presumptive and confirmation tests). To deal with this complexity, we will use Bayesian networks (*BN*), a tool that can be used to graphically display the dependency relationships and interactions between different elements within a dataset. There have been numerous applications of *BN* within forensic science: quality control monitoring [5], preparation for legal challenges [6], complex pedigree evaluation [7], DNA profile mixture evaluation [8] helping to address activity level propositions [9], as well as numerous applications outside legal or forensic applications (refer to [10] for a review of Bayesian Networks). We direct the reader to [11] for explanations of the structure and terminology of *BN*s. Recently, the authors published a paper that used *BN* to combine DNA profiling results with the results of body fluid tests in order to help address propositions at the source level [12]. When assessing such results, the data used to inform probabilities can be limited to a few experiments. Because the conclusion of the forensic scientist depends on these data, it is important to know whether their knowledge is sufficient to ensure robust reporting, and when it is necessary to perform further research. This aim of the present contribution is to show how *BN*s can help us in this task. The basic structure of the *BN* network that achieved this can be seen in Fig. 1.

The definition of each node is:

Profile matches – This node has states 'yes' and 'no' and is the node that is instantiated when a DNA profile obtained from a recovered trace possesses the same alleles as the reference of the POI.
POI DNA present – This node has states 'yes' and 'no' and is the node that specifies whether the DNA of the POI is the source of the stain.
Hp/Hd – This node has states 'Hp' and 'Hd' and marries the DNA results with the results of tests for body fluid identification (in this particular network, the body fluid is blood, however it could be configured for questions regarding any body fluid by changing prior probabilities in the 'Nature of Stain' node; see appendix table A5).
Nature of stain – This node has states 'blood', 'semen', 'saliva', 'trace' and 'none'.
Quant – Each numerical category represents the concentration of DNA detected per mm^2 of sampled area. Categories are '0', '0 to 50', '50 to 500', '500 to 5000' and '5000+'.
Visual – This node has states 'red/brown', 'white/yellow' and 'none/other' to indicate the presence (or absence) of a visual stain.
HemaStix result – This node has states 'positive' and 'negative' that correspond to the test result.
HemaTrace result – This node has states 'positive' and 'negative'.

Within the work by Taylor et al. [12] a theoretical series of court questions were used to drive the work forward from a simple *BN* to one which could consider a complex mixed DNA profile scenario. The work in [12] provides a useful starting point for the evaluation of evidence when biological source is in question; however there is a common line of questioning in court to which the scenario could be extended.

Q: What are the sample sizes on which you are basing your calculations and is that big enough?

This question is founded in traditional frequentist thinking, where the scientist may want to consider all possible datasets that could have been obtained (given different experiments, or alternative data) but were not. This is then commonly referred to as 'sampling variation' and can be taken into account by producing a distribution of the *LR*s and reporting a confidence or probability interval. On the other hand, Bayesian inference makes probability statements posterior to the data, i.e. it is inferentially complete. Indeed, Bayesian inference provides the conditional probability distribution of the next observation given prior belief and all the data observed thus far. There is therefore no need to worry about the data that could exist but have not been yet obtained, as this is encapsulated in our probabilities.

Still, underlying the question is an important concept, whilst the *LR* being provided utilises the prior beliefs of the scientist and the available data, is this accumulated knowledge enough to provide a robust opinion? By robust in the forensic context, we mean a piece of information that has limited opportunity to mislead the court. In such cases, scientists, in order to decide whether or not the data available are sufficient to warrant a robust opinion, can explore the impact of the size of the dataset on their evaluation. Because *LR*s depend on the data used to inform probabilities, it goes without saying that using different data, will lead to different *LR*s. However, ideally, if the data sufficiently reflect the phenomenon we want to account for, sensitivity analyses should not lead to *LR*s that are

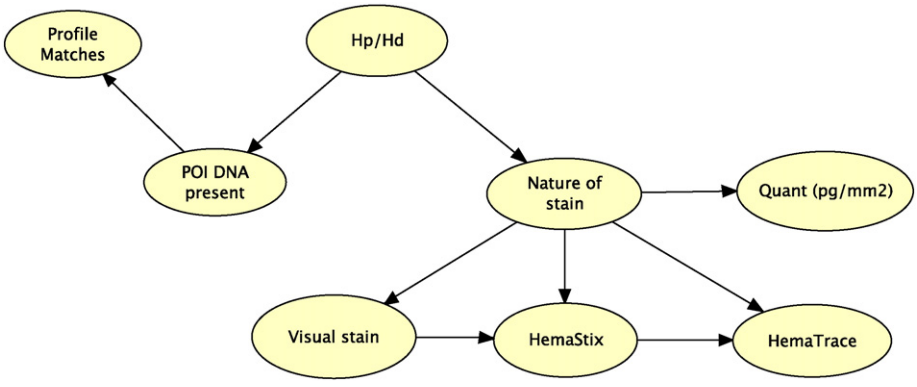

Fig. 1. Bayesian network for the presence of human blood in a sample and incorporating DNA profiling results [8].