Forensic Science International 192 (2009) 56–61

Contents lists available at ScienceDirect

Forensic Science International

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

Selection effects and database screening in forensic science

Marjan Sjerps [a,*], Ronald Meester [b,1]

[a] Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB DEN HAAG, Postbus 24044, 2490 AA DEN HAAG, The Netherlands
[b] Vrije Universiteit Amsterdam, Department of mathematics, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 21 November 2008
Received in revised form 2 July 2009
Accepted 28 July 2009
Available online 10 September 2009

Keywords:
Selection effects
Database search
Evaluation of evidence
Interpretation
Post-hoc correction
Selection bias

ABSTRACT

We argue that it is, in principle, not difficult to deal with selection effects in forensic science. If a suspect is selected through a process that is related to the forensic evidence, then the strength of the evidence will be compensated by very small prior odds. No further correction is necessary. The same is true for so-called data-dependent hypotheses. These are allowed, since if the hypothesis is really "tailored around" the evidence, the evidential value will be high but the prior odds will compensate for that. The assessment of the prior odds is outside the scope of the forensic scientist, but he should make lawmakers, judges and juries aware of the phenomenon. This discussion applies to many situations—we discuss four concrete examples.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The main objective in forensic science is to produce reliable evidence and to report this in a clear and unequivocal way. This evidence may be used to identify a suspect, but it may also be used in court either in favour or against a suspect. Usually, the reasons to select a certain item for submission to a forensic laboratory have nothing to do with the subsequent forensic analysis. For example, one submits a reference blood sample of a suspect and a sample from a bloodstain found at the scene of a crime. The bloodstain was selected because it was found on a broken window where the perpetrator is believed to have entered the house, and the suspect was identified through a witness who said she recognised him running from the house. The forensic scientist compares the DNA profiles of the two samples and reports "match" or "no match", and an estimate of the random match probability. The crime stain and the suspect in such a case already were the focus of police attention before their DNA profiles were known. The features that are used to select the samples for comparison (the witness statement and the location of the stain) are thus completely independent of the features that the forensic scientist compares (the DNA profiles). The interpretation of the evidential value in such cases is extensively discussed in the literature (see for instance [1–4]).

However, there are also situations where the items were selected in a special way, for instance by searching a large number of items and selecting those items that satisfy a criterion that is *not* independent of the features used in the forensic analysis. The evaluation of the value of the forensic evidence in such cases is potentially problematic because of so-called *selection effects*. For example, consider a situation in which a crime stain is submitted, and the forensic scientist compares its DNA profile to a database of DNA profiles. When a match is found, the matching person automatically becomes a suspect in the case. Obviously, the reason for selecting this person as a suspect is not independent of the outcome of the forensic DNA analysis. In such cases, it is not straightforward at all to derive the evidential value of the forensic evidence.

In fact, this example has been the subject of a considerable debate. The issues that seem important in this debate are

(a) *Double-counting*: Balding [20] notes that "the notion that evidence that has led to the identification of the suspect should not be subsequently used as evidence in court is analogous with some modes of statistical reasoning. But it is inconsistent with legal practice and would, I believe, be regarded as absurd by legal commentators".
(b) *Data-dependent hypotheses*: Stockmarr [5] objects to the use of "data-dependent hypotheses", that is, hypotheses that can be set up only after seeing the data, and insists we should only

* Corresponding author. Tel.: +31 70 888 6666/6122; fax: +31 70 888 6565.
E-mail addresses: m.sjerps@nfi.minjus.nl (M. Sjerps), rmeester@few.vu.nl (R. Meester).
1 Tel.: +31 20 5987711; fax: +31 20 5987653.

0379-0738/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2009.07.022

consider hypotheses that are data-independent. According to this point of view, it is misleading to consider the hypothesis that the one person in the database who matched was the actual criminal, since this hypothesis depends on the outcome of the search and is therefore clearly data-dependent.

(c) *Correcting for selection bias*: The first report of the National Research Council [6] is very much concerned with the danger of "multiple hypothesis testing". According to this report, the best way to correct for the "selection bias" when a DNA match is found between, e.g. a crime stain sample and a suspect's sample in the database, is to ignore the DNA loci used in the search and use only additionally typed loci for evidence in court. The second report of the National Research Council [7] abandons this view and suggests a "correction to account for the database search", in the sense that the match probability should be multiplied by the number n of comparisons made.

(d) *Reporting*: Others argue that this correction is unnecessary and "that it is conservative to just report the random match probability" [8] and "... it would usually seem preferable to neglect the observed non-matching profiles when calculating match probabilities. This practice is slightly favourable to defendants and therefore need not be reported to the court, nor any adjustment need to be made to the likelihood ratio because of it" [3]. We have argued that reporting only a match probability is misleading because lay people should be made aware that this probability may be compensated by the prior odds that the suspect is the DNA-donor ([9,10]; see also below).

Most of the discussion around the DNA evidential value is meaningful in a much more general context. Indeed, in forensic casework, there are many situations concerning evidence other than DNA, where an item or person is selected from a large pool of items or persons by some criterion, and subsequently the same information that is used to select the item or person is used as evidence in court. Questions concerning selection effects apply to these situations just as well. Here are four concrete examples. Examples 1, 2 and 4 are based on questions raised by employees of the Netherlands Forensic Institute concerning their casework. One of us (RM) acted as an expert witness in a situation as described in Example 3.

1.1. Fibre comparison

Purple polyester fibres are found on the white clothing of a murder victim dumped in the woods. One identifies a suspect and searches his house, his garage, his car, his caravan, and his office for similar fibres. They find a small rug in his caravan with similar fibres. The forensic scientist compares the rug with the fibres found on the victim and concludes they are both purple polyester. However, he considers this not very surprising given the way that the rug was selected.[1] He wonders how to assess the evidential value of his observations, and how to report it. Should he somehow correct for the large number of items that were compared to the fibres?

1.2. Crime Watch

A robber is filmed by a surveillance camera. The tape is shown on "Crime Watch", a popular TV-show, and the audience is requested to contact the police in case they recognise the robber. Millions of people watch the show and the police receives 40 tips. One of them is from a woman who fears she recognised her son-in-law. The police obtains good photographs of the son-in-law and submits them to a forensic expert for comparison with the tape. The expert compares the photos to the robber on the tape and observes many similarities. He wonders how he should take into account that the suspect was selected precisely because he resembles the robber. He is also worried because in many other cases he is not told how the suspect was selected. Should he make sure that he is always informed about this or should the selection procedure be irrelevant to him?

1.3. Angel of death

A nurse is being prosecuted for murdering several of her patients. Part of the evidence is the report of a statistician. He calculated the probability that she would be present by so many medically unexplained incidents if in fact she was innocent and it was all mere coincidence. However, he thinks it is necessary to apply a "post-hoc correction" to this probability because the nurse became a suspect in the first place because she was present at so many incidents. The post hoc correction is disputed in court.

1.4. Spider in the web

A crime analyst constructs a social network of a group of persons suspected of criminal activity. One of these persons appears to be the "spider in the web", having links with many persons. This person subsequently becomes a suspect. (This way of interpreting such graphs is in fact dangerous since it is based on the premise that available information is equivalently complete for all the individuals appearing in the graph, which is definitely not the case in most investigations or intelligence activities—we ignore this practical difficulty here.) Other evidence is gathered against him and he is finally accused of leading a criminal organisation. The analyst wonders whether this analysis can be used to first identify the suspect, and subsequently as evidence against him. He is worried that the same information is used twice against the suspect.

There is surprisingly little literature about how to deal with selection effects in these kinds of situations. One is occasionally warned against selection effects, but unfortunately, most literature is vague on how we should take these effects into account and especially on how to report the evidence.

Early papers on this issue seem to just note the problems without offering a solution. For example, Aitken [11] mentions several situations with a selection effect. He warns that "The number of comparisons made before considering a match, or similarity, in transfer evidence has to be taken into account when addressing the value of the evidence". Furthermore, "The reason for apprehension of a suspect has also to be considered when assessing the weight of the evidence". Stoney [12] states that "The use of evidence for investigative screening of suspects is in conflict with its subsequent use to evaluate the suspect". He considers an example where a suspect's jacket is found stained with blood of the same blood type as the victim. He suggests that the evidential value of this observation depends on the way that the suspect was identified: was he selected because he had a bloodstained jacket, or was he selected on the basis of other evidence.

However, Robertson and Vignaux [13] point out that this is in fact an error of thinking. We find their reasoning very instructive. "The power of the evidence is still determined by the ratio of the two probabilities of the accused having a bloodstained shirt if guilty and if not guilty. It is just that there happens to be less evidence in one case than the other. When the suspect is stopped because of a bloodstained shirt there may be no other evidence. When the suspect is arrested on the basis of other evidence and then found to have a bloodstained shirt, the likelihood ratio for the bloodstained shirt is to be combined with a prior which has already

[1] Throughout this paper one should read "he or she" in phrases like this.

Download English Version:

https://daneshyari.com/en/article/97278

Download Persian Version:

https://daneshyari.com/article/97278

Daneshyari.com