Forensic Science International xxx (2009) xxx–xxx

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 xxx

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.

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 Published by Elsevier Ireland Ltd.
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.

been raised by the other evidence. Once again the power of an item of evidence is being confused with the strength of the evidence as a whole". Aitken and Taroni [14] confirm this view.

There are some papers that consider the evidential value of searching a database containing features other than DNA. For example, Robertson and Vignaux [13] consider database searches in general and conclude among others that, contrary to intuition, a larger database increases the probability that the "match" (or one of the matches) is indeed the perpetrator. Evett et al. [15] consider a case example where a suspect's shoe is compared to a database of shoe marks left at crime scenes, and appears to "match" one of these. They state that we need not be concerned that the fact that the mark was retrieved as a result of a database search, in some way weakens the evidence: this fact is relevant to the forming of prior odds, but not for the likelihood ratio. Kaye [22] shows a flaw in an analogy to DNA database search used by the Court in the case People v. Johnson, where a suspect is identified by comparing the image of a robber from a surveillance camera to a database of driver's license photographs.

In this paper, we extend some arguments that were developed for the interpretation of a DNA database match to much more general situations with selection effects. In most cases, the mathematical models used in the analysis are probably too simple to be of direct practical use. However, they will enable us to decide which issues are potentially problematic, and which issues can be dealt with without any problem. We will see that the conclusions drawn in the DNA-debate actually do generalise to other situations with selection. In particular, this implies that we do not need to worry about, and understand how to deal with, the double use of information, data-dependent hypotheses and corrections for selection effects. However, in order to do that, we do need to emphasise to the readers of forensic reports the role of "prior odds" and report either a table relating priors to posteriors, or a warning.

Interpretation of evidence is a complicated matter. This paper only deals with one aspect of it. Other aspects, like formulating hypotheses at the "activity level" [16], evaluating a combination of different pieces of evidence, or avoiding errors of thinking, are explicitly not considered here. In fact, our first example (fibre comparison) focuses on the "source level" hypotheses, whereas the "activity level" may in real cases be more appropriate to address.

The paper is built up as follows. First we review in some detail the discussion about the DNA database match mentioned above. After that we argue that similar considerations can be used in more general models.

2. The DNA database search

A crime is committed by an unknown criminal C. The evidence in the case consists of two things. First of all, C left (via a blood stain) a DNA profile d at the crime scene. Second, we run this profile d through a database of DNA profiles, and find that among all people in the database, only a single person s also has profile d; one could say that s is selected via the database search.

We now wish to compute the probability, conditioned on this twofold information, that s is the criminal C. To this end, consider the following computation, based on Balding [3], where we use the notation "$C \equiv d$" to mean "C has profile d", and "$s \sim d$" to mean "s is the only person in the database with profile d, the other persons are [vector of names] and their DNA profiles are [vector of DNA profiles]". The computation is elementary, using no more than Bayes' theorem. We use the symbol i as an index denoting a general individual in the population. In the computation below, a quantity like $P(C = i)$ simply refers to the probability that the unknown criminal is individual i without looking at the evidence in question, while $P(C = i|s \sim d;\ C \equiv d)$ is the *conditional* probability that the unknown criminal is individual i, given the evidence that the unknown criminal has profile d and person s is the only person in the database with this profile. In the model, we assume that the profile of any given individual does not depend a priori on the identity of the criminal C. That is, before observing the DNA profiles, the probability that an individual has a certain profile is not affected by knowing who the criminal is.

We are interested in the question whether or not $C = s$, given the evidence that $C \equiv d$ and $s \sim d$, and we would like to know on which quantities the final answer depends. Following the literature, we in fact compute the posterior odds: the ratio of the probability that s is or is not the criminal, given the evidence. Some steps in the computation are just rearrangements, others possibly require some explanation, which we will provide after the computation.

$$\frac{P(C = s|s \sim d; C \equiv d)}{P(C \neq s|s \sim d; C \equiv d)} = \frac{P(C = s|s \sim d; C \equiv d)}{\sum_{i \neq s} P(C = i|s \sim d; C \equiv d)} = \frac{P(C = s; s \sim d; C \equiv d)/P(s \sim d; C \equiv d)}{\sum_{i \neq s} P(C = i; s \sim d; C \equiv d)/P(s \sim d; C \equiv d)} =$$
$$\frac{P(C = s; s \sim d; C \equiv d)}{\sum_{i \neq s} P(C = i; s \sim d; C \equiv d)} = \frac{P(C = s) \cdot P(s \sim d; C \equiv d|C = s)}{\sum_{i \neq s} P(C = i) \cdot P(s \sim d; C \equiv d|C = i)} = \frac{P(C = s) \cdot P(s \sim d; s \equiv d|C = s)}{\sum_{i \neq s} P(C = i) \cdot P(s \sim d; i \equiv d|C = i)} =$$
$$\frac{P(C = s) \cdot P(s \sim d)}{\sum_{i \neq s} P(C = i) \cdot P(s \sim d; i \equiv d)} = \frac{P(C = s)}{\sum_{i \neq s} P(C = i) \cdot P(i \equiv d|s \sim d)} = \frac{1}{\sum_{i\ \text{not in database}} (P(C = i)/P(C = s)) \cdot P(i \equiv d|s \sim d)}.$$

The 6th equality is true because the events $C = i$ and $C = s$ are independent of the DNA profiles of i and s, and the last equality is true because for individual i in the database, it cannot be the case that i has profile d and at the same time s is the only match in the database.

The first ratio in the displayed equations above is known as the "posterior odds", that is, the ratio of the probabilities that C is or is not the perpetrator *after* seeing the evidence. The ratio $P(C = i)/P(C = s)$ is known as the "prior odds", referring to (unconditional) probabilities *before* seeing the DNA evidence.

A number of observations are important. These observations are based on contributions of many authors, including Robertson and Vignaux [13], Balding and Donnelly [17], Donnelly and Friedman [21], Dawid and Mortera [18], Evett and Weir [1] and Dawid [19].

First of all, we see that the probability that s is the perpetrator C increases with the size of the database; indeed, when the database grows, the final sum in the denominator is over fewer elements and therefore the denominator becomes smaller. In other words, the larger the database, the more people are excluded from being the perpetrator (assuming that s is the only match), so the larger the probability that s is the perpetrator C.

More importantly, we see that given the evidence, the probability that s is the perpetrator C, depends on just two things, namely

1. The conditional probabilities $P(i \equiv d | s \sim d)$, and
2. The prior odds $P(C = i)/P(C = s)$ of all individuals i who are *not* in the database.

The conditional probability $P(i \equiv d | s \sim d)$ is the probability that a certain individual i has profile d, given that s is the only person in the database with profile d, and given the names and profiles of the other persons in the database. Ignoring the fact that the non-matching persons in the database may be related to i, or belong to the same sub-population, this is about the same as the *match probability* $P(i \equiv d | s \equiv d)$. Ignoring again relatedness between people and complications arising from the existence of sub-populations, this is about the same as the unconditional probability $P(i \equiv d)$, that is, the relative frequency of the profile d in the population of interest.

The prior odds $P(C = i)/P(C = s)$ are obviously *not* for the DNA expert to assess; indeed they have nothing to do with DNA at all! The expert should restrict to conveying information about the match probabilities. However, when there is little or no evidence besides the DNA match against s, the prior probability that s is the perpetrator may be very small. This may counterbalance strong DNA evidence resulting in a modest posterior probability that s is the perpetrator.

The point of the current article, now, is to realize that the above computation simply asserts that after seeing the evidence, the posterior probabilities are what they are and that it is not necessary to "correct" for anything at all. The "selection effect" of the database search *must* have been taken into account by the prior odds, and the idea that anything needs correction is simply false. The formula gives the correct answer, and in fact that is all there is to say, counterintuitive as this might be.

It is then also clear that there is no need to avoid the use of "data-dependent" hypotheses: although one might be suspicious about using a hypothesis that has been set up after seeing the data, the above computation shows that for the posterior odds, this issue is not relevant. We can illustrate this in the database case. The data-dependent hypotheses set "$C = s$" versus "$C \neq s$" can be replaced by the data-independent hypotheses set "C is in the database" versus "C is not in the database". However, given the evidence that s is the only match in the database, the two sets are equivalent: "$C = s$" implies "C is in the database" and vice versa, and "$C \neq s$" implies "C is not in the database" and vice versa. Hence, the posterior odds on "$C = s$" must be exactly the same as the posterior odds on "C is in the database"; this was shown by Dawid [19]. What happens when we move from the "data-dependent" set to the "data-independent" set is that the prior odds increase, but simultaneously the likelihood ratio decreases. The result, the posterior odds, is the same.

Similarly, there is no need to worry about using the same information to first identify the suspect, and subsequently as evidence against him. We return to this in the more general context below.

We have argued in Meester and Sjerps [10] that a lay person may not realise the counterbalancing effect of the prior odds, and may conclude from a tiny random match probability that s is highly probably the perpetrator. Hence on the one hand, the expert is not allowed to address the prior odds in his report, on the other hand, he is afraid that lay persons may come to the wrong conclusion when he does not mention the prior odds. One way to overcome this dilemma is to report a table relating prior to posterior odds, as we suggested in Meester and Sjerps [10]. Another way may be to put a warning in the report explaining that the match probability should be weighted with the other evidence in the case. Such a warning is reported by experts of the Netherlands Forensic Institute.

3. Extension to more general models

We distinguish between two situations. In Situation A, we know for sure that a crime was committed. In situation B, we only know that a crime may have been committed, or may have been planned. The latter situation applies for instance to prevention activities.

3.1. Screening for identifying the perpetrator of a crime or the crime-related object

A crime has been committed. We consider a population consisting of a large number of items that could be related to the crime, including the single item C that truly is related to the crime. A subset X consisting of $n + 1$ items is screened with some selection criterion f, which is a function $f: X \to \{0,1\}$. We know that f is 1 for item C, and is equal to 1 for other items i with probability p_i. The value of f turns out to be 0 for all items in the subset, except for item s, for which f turns out to be 1. This makes item $s \in X$ the focus of our attention. We consider this as the evidence in the case, that is, the evidence is $f(C) = 1$, $f(s) = 1$, and $f(i) = 0$ for all $i \neq s$, $i \in X$. This evidence is to be evaluated in the light of the hypotheses:

H1 : $s = C$
H2 : $s \neq C$

Note that the selection criterion f could be interpreted as indicating whether some continuous measure exceeds a threshold. Examples of screenings of type A are DNA database search, and Examples 1 and 2 above:

- *DNA database search*: We consider a population of persons, including the perpetrator C. We have a database X consisting of DNA profiles of $n + 1$ individuals. Furthermore, the selection criterion f is the indicator of an individual having a DNA profile that matches the DNA profile d of the crime stain. The evidence is that our database yields a single match s.
- Example 1. *Fibre comparison*: We consider a population of items containing textile that could be the source of the purple polyester fibres found on the murder victim. A single item C in this population is truly the source. A subset X, all items containing textile that were searched in the suspect's house, etc., is screened by the police. They use as a selection criterion f whether or not the item contains purple polyester fibres. A single item s, the small rug in the caravan, is found to satisfy this criterion. The fibre expert confirms that $f(s) = 1$. We assume that item C, if it were screened by the police, would also be selected: $f(C) = 1$.
- Example 2. *Crime Watch*: We consider a large population of persons that could be the robber C on the surveillance tape. The audience of Crime Watch compare everyone they know to the robber on tape, and a number of them will actually call the police when they think they recognise someone. Thus, the set of all persons known to one or more of these Crime Watch viewers is the subset X that is screened by the audience. In this first selection step several members of X are selected by the audience because they resemble the robber. The police carries out a second selection on these individuals, and selects those individuals that show a striking resemblance for a photographic comparison by an expert. Here, we have assumed that only one individual s satisfies their criterion. The expert confirms the striking resemblance of s. Hence, the selection criterion $f(x)$ for any member x of X is striking resemblance to the robber. $f(x) = 1$ if both the audience and the police think that x resembles the person on tape sufficiently well to respectively call the police and submit the photo to the expert, and the expert confirms the resemblance; $f(x) = 0$ otherwise. We assume that $f(C) = 1$, that is, the surveillance tape is of such good quality that C would be

recognised if he were a member of X, and he would pass the selection criterion.

We will now extend the ideas developed in the DNA context to the general situation A. We consider the hypotheses

H1 : $C \neq s$
H2 : $C = s$,

and want to derive the evidential value of the evidence that only items C and s pass the selection criterion f and all other items in the subset X do not. We use the notation, like before, "$C \equiv 1$" to mean "the value of f is 1 for item C", and "$s \sim 1$" to mean "The value of f is 1 for item s, and 0 for all other items in the subset X". We perform a similar computation as before (simply replace "d" with "1", and "database" with "X"), leading to

$$\frac{P(C = s|s \sim 1; C \equiv 1)}{P(C \neq s|s \sim 1; C \equiv 1)} = \frac{1}{\sum_{j\,not\,in\,X}(P(C = j)/P(C = s))P(j \equiv 1|s \sim 1)}.$$

The conclusions for DNA database search generalise to this situation. Given the evidence, the probability that our suspect item s is indeed the crime-related item C increases with the size of the screened population X, and depends only on the match probabilities $P(j \equiv 1|s \sim 1)$ and the prior probabilities $P(C = j)/P(C = s)$ of all items j which are not in X. The "selection effect" is again taken into account by considering the prior odds: it is not necessary to "correct" the match probability.

The prior probabilities are again not for the expert to assess. He should restrict to conveying information about the match probabilities. However, when there is little or no other evidence, the prior probability that s is C may be very small. This may counterbalance the possibly strong evidence we consider with the selection criterion f, resulting in a modest posterior probability that s is indeed C. A lay person may not realise that this is possible, and may conclude from a tiny match probability (a rare fibre type, or a particularly striking resemblance to the robber) that s highly probably is C. Hence, the expert should make clear in his report that the evidential value of the information has to be weighted with the prior odds.

For the fibre example, this means that the probability that the fibres found on the murder victim indeed originate from the suspect's purple rug does not decrease but slightly increases with the number of items that are screened by the police. This probability only depends on the "match" probability of unscreened items that may be the source of the fibres found on the victim, and the prior odds P(item j is the source)/P(suspect's rug is the source) of each of these items. The expert may thus report as in a usual case, for instance his estimate of the rarity of the fibre type, and does not need to correct this estimate for selection bias. However, he should also get the message through that a rare fibre type not necessarily means a high probability that the rug is indeed the source: this depends on the other evidence in the case via the prior odds.

For the Crime Watch example, we conclude that the probability that the suspect is indeed the robber increases with the number of persons that are screened by the audience and the police. This probability only depends on the "match" probability of unscreened persons who may be the robber, and the prior odds P(person j is the robber)/P(suspect is robber). The expert may thus report as in a usual case, for instance his estimate of the likelihood ratio (the ratio of the probability of observing this degree of similarity when the suspect is indeed the robber and when an unrelated person is the robber) and does not need to correct for selection bias. However, he should also get the message through that a large likelihood ratio not necessarily means a high probability that the suspect is indeed the robber: this depends (through the prior odds) on the other evidence in the case (which may be completely lacking).

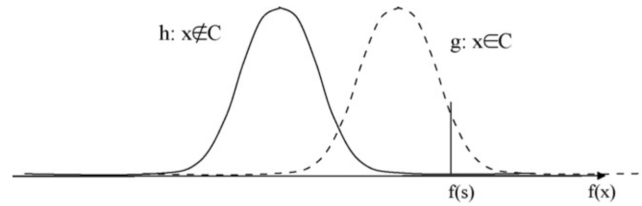

Fig. 1. Illustration of probability densities g and h.

3.2. Screening for identifying unusual persons or objects that may be involved in criminal activity

We consider a certain type of criminal activity, and a population consisting of a large number of items, of which a subset C is involved in this activity (possibly the activity is absent in the population so that $C = \emptyset$). A subset X consisting of $n + 1$ items is screened with some measure f, which is a function f: $X \rightarrow IR$. We expect that items involved in the criminal activity score relatively higher on f than other items: denoting the probability density function of the f-scores of items in C by g, and that of items not in C by h, we expect the location of g to be much larger than that of h (Fig. 1). We observe that item $s \in X$ scores high compared to the other items: $f(s) \gg f(x)$ for all $x \neq s$, $x \in X$. This makes item s the focus of our attention. We consider this as the evidence in the case, and want to evaluate this in the light of the hypotheses

- H1: $s \in C$
- H2: $s \notin C$

Examples of such screenings are Examples 3 and 4 above.

- Example 3. *Angel of death*: We consider a population of medical staff, and are interested to know if any of them are killing or seriously harming their patients (either intentionally or unintentionally). The subset C is formed by these individuals. Medically unexplainable death or incidents are being recorded routinely now in many hospitals, and so it is noticed when a cluster of such deaths or incidents occurs. The rosters of the medical staff in such a case may be screened to see if the shifts of any person coincide with the incidents. The measure $f(x)$ could be some statistic measuring the relationship between the occurrence of an incident and the shifts of a particular person. We expect that nurses and doctors who harm their patients score relatively high on f. One nurse s scores much higher on f than the all the others.
- Example 4. *Spider in the web*: The crime analyst has to prevent (a certain type of) crime in a country. The criminals in the country form a subset C of the country's population. The analyst has data from a number of persons suspected of criminal activity, forming the subset X. He performs a social network analysis on this subset, and expects that criminals will have more "links" to other members of X than non-criminals. Thus f is the number of links to persons in X (or some other criterion indicating the number of contacts or status of this person within X). One individual s appears to be a true "spider in the web".

It follows from Bayes' theorem that

$$\frac{P(s \in C|\,f(s))}{P(s \notin C|\,f(s))} = \frac{g(\,f(s))}{h(\,f(s))} \cdot \frac{P(s \in C)}{P(s \notin C)}$$

We see that some of the conclusions for DNA database search also generalise to this situation. Given the evidence, the probability that our suspect item s is indeed the crime-related item C depends only on the likelihood ratio $g(f(s))/h(f(s))$ and the prior odds $P(s \in C)/P(s \notin C)$. The "selection effect" is taken into account by considering the prior odds: it is not necessary to correct the match probability, since the probabilities are just what they are.

Again, the prior probabilities are not for the expert to assess. He should restrict to conveying information about the match probabilities. So, when there is little or no other evidence, the prior probability that $s \in C$ may be very small. This may counterbalance the possibly strong evidence we consider with the selection criterion f, resulting in a modest posterior probability that s is indeed a member of C. A lay person may not realise that this is possible, and may conclude from a large likelihood ratio (an exceptional high number of incidents/contacts) that s highly probably is a member of C. Hence, the expert should make clear in his report that the evidential value of the information has to be weighted with the prior odds.

For the angel of death example this means that the probability that the suspected nurse indeed harmed her patients depends only on the likelihood ratio of her score $f(s)$ and on the other evidence in the case that determine the prior odds. A "post-hoc" correction is unnecessary; however the expert should get the message through that a large likelihood ratio not necessarily means a high probability that the nurse indeed harmed her patients. This depends on the other evidence in the case (which may be completely lacking).

For the spider in the web example, we conclude analogously that the probability that the suspected person is indeed a criminal depends only on the likelihood ratio of his score $f(s)$ and on the other evidence in the case that determine the prior odds. A correction for a selection effect is unnecessary. However, the expert should get the message through that a large likelihood ratio not necessarily means a high probability that the suspect is indeed a criminal: this depends on the other evidence in the case (which may be completely lacking). Moreover, there is no need to worry about the fact that the same information is first used to identify the suspect, and subsequently as evidence against him.

4. Conclusion

Although the mathematics above is very elementary, it has important practical and philosophical consequences. It shows that the conclusions from the DNA database search debate generalise to very general situations in which selection plays a role. The returning theme of all computations, both in the original and more general models, is that the posterior odds depend on

1. The prior odds,
2. A number measuring the strength of the evidence.

As mentioned before here and elsewhere, the prior odds have nothing to do with the actual data or evidence, and should therefore not be discussed by the forensic expert. Indeed, the forensic expert tries to compute the strength of the evidence, and nothing more than that. How does this work in practice? For concreteness, let us concentrate on the Crime Watch example. One wonders whether (and how) one should take into account that a person was selected just because he looked very much like the robber on the tape. The answer is as simple as elegant: the strength of the evidence is what it is, but even if this strength is very high, the prior odds may be very small. Hence, if an individual is selected just because he resembled the robber on the tape, and for no other reason, and no other evidence is found afterwards, then the prior odds are very low, and therefore the case as a whole is not very convincing. Or, to put it in other words, small prior odds will compensate to a large extent the possibly strong evidence.

Similar conclusions can be drawn for the other examples. In all cases, it follows straightforwardly from the model and the mathematics (and, in retrospection, also from common sense) that once prior odds are taken into account, there is no need to make any corrections for selection effects, multiple testing or for data-dependent hypotheses. This does not mean that it is straightforward (or even meaningful) to compute the posterior odds. Indeed, in most practical cases, it will be difficult or impossible to estimate any prior odds. This being said (and true), we think it is very important to point out that these difficulties cannot really be avoided, and fall outside the scope of the forensic expert. It seems to us that the best a forensic expert can do, is to provide a table, relating the various prior odds to the final posterior odds, or to put a warning in the report that strong evidence not necessarily implies a large probability that the suspect (item) is indeed the offender (item).

Acknowledgement

Part of the research of M.S. was funded by the Netherlands VIA programme (Improving Security through Information Awareness).

References

[1] I.W. Evett, B.S. Weir, Interpreting DNA Evidence, Sinauer, Sunderland, MA, 1998.
[2] J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, 2005.
[3] D.J. Balding, Weight-of-Evidence for Forensic DNA Profiles, Wiley & Sons, Chichester, UK, 2005.
[4] W.K. Fung, Y.-Q. Hu, Statistical DNA Forensics—Theory, Methods and Computation, Wiley & Sons, Chichester UK, 2008.
[5] A. Stockmarr, Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search, Biometrics 55 (1999) 671–677.
[6] National Research Council, DNA technology in Forensic Science, National Academy Press, Washington, DC, 1992.
[7] National Research Council, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, DC, 1996.
[8] S. Walsh, B.J. Buckleton, DNA intelligence databases, in: J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evidence Interpretation, CRC Press, Boca Raton, 2005.
[9] R. Meester, M. Sjerps, The evidential value in the DNA database search controversy and the two-stain problem, Biometrics 59 (2003) 727–732.
[10] R. Meester, M. Sjerps, Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence, Law, Probability and Risk 3 (2004) 51–62 (see also the discussion and response, p. 63–86).
[11] C.G.G. Aitken, Population and samples, in: C.G.G. Aitken, D.A. Stoney (Eds.), The Use of Statistics in Forensic Science, Ellis Horwood Lmt, Chichester, 1991.
[12] D.A. Stoney, Transfer evidence, in: C.G.G. Aitken, D.A. Stoney (Eds.), The Use of Statistics in Forensic Science, Ellis Horwood Lmt, Chichester, UK, 1991.
[13] B. Robertson, G.A. Vignaux, Interpreting Evidence-evaluating Forensic Science in the Courtroom, Wiley & Sons, Chichester, UK, 1995.
[14] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd ed., Wiley and Sons, Chichester, 2004.
[15] I.W. Evett, J.A. Lambert, J.S. Buckleton, A Bayesian approach to interpreting footwear marks in forensic science, Science & Justice 38 (1998) 241–247.
[16] R. Cook, I.W. Evett, G. Jackson, P.J. Jones, J.A. Lambert, A hierarchy of propositions: deciding which level to address in casework, Science and Justice 38 (1998) 231–239.
[17] D.J. Balding, P.J. Donnelly, DNA profile evidence when the suspect is identified through a database search, Journal of Forensic Science 41 (1996) 603–607.
[18] A.P. Dawid, J. Mortera, Coherent analysis of forensic identification evidence, Journal of the Royal Statistical Society Series B 58 (1996) 425–443.
[19] A.P. Dawid, Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search", Biometrics (2001) 976–980.
[20] D.J. Balding, The DNA database search controversy, Biometrics 58 (2002) 241–244.
[21] P. Donnelly, R.D. Friedman, DNA database searches and the legal consumption of scientific evidence, Michigan Law Review 97 (1999) 931–984.
[22] D.H. Kaye, Rounding up the usual suspects: a legal and logical analysis of DNA trawling cases, North Carolina Law Review (2009) 87 (2).

Please cite this article in press as: M. Sjerps, R. Meester, Selection effects and database screening in forensic science, Forensic Sci. Int. (2009), doi:10.1016/j.forsciint.2009.07.022