Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence

Ronald Meester[1] and Marjan Sjerps[2]

June 11, 2003

[1] Divisie Wiskunde, Faculteit der Exacte Wetenschappen, Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands, rmeester@cs.vu.nl;

[2] Nederlands Forensisch Instituut, PO Box 3110, 2280 GC Rijswijk, The Netherlands, m.sjerps@nfi.minjus.nl

Abstract

The introduction of DNA evidence has transformed human individualisation in criminal litigation, but it also introduced daunting statistical, philosophical and practical problems into the process. The current practice in many legal cases is that a forensic expert reports a match probability or a likelihood ratio. However, the value of the likelihood ratio depends on the particular hypotheses used by the expert. Often there are various choices possible for the hypotheses used, and the corresponding likelihood

ratios for different hypotheses are typically different. We therefore argue that the findings of an expert should not be given with a single number, and that any report with a match probability or likelihood ratio should be accompanied by a discussion of the effect of these numbers. We suggest a way to do this, using the so called posterior odds, which are invariant under various sets of hypotheses. Our approach is applicable in legal systems where the judge or jury is not allowed to know that the suspect was in a database. However, juridical problems may arise when courts insist on e.g. only a frequency estimate to be reported.

1 Introduction

DNA evidence plays an increasingly important role in criminal cases. In a typical situation, one identifies a DNA profile (the *crime sample*) via for instance a blood stain found at the scene of a crime, presumably left by the perpetrator. This DNA profile can then be compared to the DNA profile of a suspect, or to a database of DNA profiles. In the first case, the DNA evidence can provide additional evidence against the suspect; in the second case, a match in the database could point to a possible suspect who was up to that stage not under suspicion.

On the one hand, DNA evidence clearly is a powerful tool, and it has already transformed human individualisation in criminal litigation. On the other hand however, it has also introduced daunting statistical, philosophical and practical problems into the process. For instance, there has been a serious controversy about the strength of DNA evidence when a suspect is identified through a database search; this discussion was statistical and philosophical in nature, see e.g. National Research Council (1996), Stockmarr (1999), Donnelly and Friedman (1999), Evett, Foreman and Weir (2000), Dawid (2001), and Balding (2002). One of the main practical problems seems to be the question how to report the DNA evidence to judges and jurors, since it seems that they are easily mislead by the (numerical) statements of the forensic or statistical expert. In this paper, we address statistical, philosophical, as well as practical problems.

Statements about the strength of DNA evidence always are probabilistic or statistical in nature, and based on the assumption that a particular DNA profile occurs with a certain frequency in a population. More precisely, any DNA profile has a certain *match probability*, being the probability that a randomly selected member of the population has this particular profile, given that the suspect has it. In practical cases, this match probability can be as big as 1 in ten, and as small as 1 in a few billion or even smaller, depending on the quality of the crime stain.

In this article, we make two simplifying assumptions. First, we assume that the match probability of a profile A is known exactly, and given to us. Second, we ignore relatedness between individuals. If a certain individual has a profile A, then the probability that his brother also has this profile is not the match probability, but typically much higher. In actual cases, this can be very important, but in this paper this effect is not taken into account. Furthermore, we will use the word 'juror' to mean judge or jury-member, whichever is appropriate in the legal system.

How is DNA evidence reported? There are many ways, and we will focus on two widely used approaches. According to the first approach, one simply reports the *match probability*. So if a certain suspect, John Smith say, has a DNA profile matching the crime sample, the forensic expert would simply state that given that the suspect has the profile, the probability that a randomly selected member of the population has the profile is 1 in R, where $1/R$ is the match probability of the profile. A larger R means stronger evidence against John Smith.

The second approach is based on a so called *likelihood ratio*. A likelihood ratio of a piece of evidence with respect to two hypotheses, is the ratio of the probability that the evidence would arise given the first hypothesis to the probability that the evidence would arise given the second hypothesis; we will give concrete examples below. The idea is that the larger the likelihood ratio, the stronger the piece of evidence is probative of the first hypothesis as compared to the second. A forensic scientist testifying in terms of a likelihood ratio with numerical value R, might say something like the following: 'The evidence is R times more likely to have arisen given the proposition that John Smith was the donor of the crime sample than given the proposition that a person selected at random from the population was the donor of the crime sample.' Again, the larger the likelihood ratio, the stronger the evidence is supportive of the first proposition with respect to the second.

Example. Consider the simplest possible case, in which a certain crime stain is obtained with profile A, and in which a suspect, John Smith say, also has this profile. Suppose further that the match probability of the profile is 1 in R. According to the first approach, the expert simply reports the 1 in R. According to the second approach, we deal with two hypotheses. The first hypothesis says that *John Smith is the crime stain donor*; the second hypothesis says that *someone else is the crime stain donor*. In order to compute the likelihood ratio for these two hypotheses (Evett, 1983), we note that if the first hypothesis is true, then we are *certain* to find profile A in the crime sample: indeed, since this hypothesis assures that John Smith is the crime stain donor, we will certainly find the profile corresponding to John Smith, and this is A (unless an error has been made, which possibility we will ignore). If the second hypothesis is true, then the probability to find profile A is $1/R$, since for all we know, a random member of the population left the crime stain, and the probability that his profile is A is simply $1/R$. Hence the likelihood ratio is $1/(1/R) = R$. A large R reflects the fact that the profile is rare, and therefore a large R means stronger evidence against John Smith. This complies with our intuition.

The first approach using match probabilities is simpler, but often not possible or appropriate. In more complex situations, one can find mixtures of DNA profiles, or several different profiles, or profiles whose relations to each other are not clear. In such cases, there is no one single profile whose match probability can be used directly, and one preferably uses the more complicated likelihood ratios. But even the use of match probabilities is not without problems; the infamous prosecutor's fallacy tends to confuse jurors (Thompson and Schumann 1987).

The second (likelihood) approach is much more general, but also more complicated, and (based on our own experience) tends to confuse jurors even more (for further reading see Taroni and Aitken 1998 and Koehler 1996). This is one reason to be careful with its use.

In this article, we shall argue that there is a much more important reason to be careful with the use the likelihood ratio approach in the above form. We shall explain in the next section that the numerical value of the likelihood ratio depends on the particular choice of the hypotheses used by the expert. Since the choice of the hypotheses is sometimes arbitrary, this implies that reporting only the likelihood ratio (in conjunction with the hypotheses of course) as a measure for the strength of the evidence can lead to misinterpretations. After that, in Section 3, we give a suggestion as to how one could report instead. A central issue here will be the fact that we have to carefully distinguish between the findings of the *expert* and the judgement of the *juror*. We shall argue that these two components - findings of the forensic expert and judgement of the juror - are not very useful in isolation, and become meaningful only in combination with each other. We will also give a number of concrete examples to which we apply our analysis.

In some legal systems, the juror is not supposed to know whether or not the suspect was found through a database search. This is an extra complication of the process, but our approach can deal with this situation. Problems may arise, however, when courts insist on reporting e.g. profile frequency estimates, like the guidelines given in the UK cases R. v. Doheny and Gary Adams (1997) 1 Cr App Rep 369. Another difficulty may arise when courts object to the use of quantitative probability estimates based on non-scientific evidence as in the infamous UK Adams case (R.v. Dennis Adams (1996) 2 Cr App Rep 467; 61 JCL 170).

For a while, we concentrate on likelihood ratios. Match probabilities will return to the scene in Section 4. In Sections 2 and 3 we explain how reporting the likelihood ratio may be improved. We implement our findings in a number of concrete examples in Section 4.

2 Which likelihood ratio should be reported?

We now first explain why reporting a likelihood ratio only is sometimes problematic. We do this by analysing a particular forensic problem, namely the *two-stain problem* (Evett, 1987).

Consider a crime which was committed by two people, each of which left a blood stain at the scene of the crime; one of the stains is on a pillow, the other on a sheet. The DNA profiles of the stains are investigated: the stain on the pillow has profile A and the stain on the sheet has profile B. The match probabilities of these profiles are 1 in R_A and 1 in R_B, respectively.

Suppose that John Smith is arrested for completely different reasons, unrelated to the above crime, but that it turns out that Smith's DNA profile matches with A. As a result, poor John Smith becomes a suspect, and the question arises as to how strong the DNA-evidence against Smith actually is. To perform the likelihood methodology, the expert needs to set up two competing hypotheses. Various natural choices for these hypotheses can now be set up, without a clear a priori preference for either of them.

One can, for instance, compute a likelihood ratio for the hypothesis that *John Smith was one of the crime stain donors* versus the hypothesis that *John Smith* *was not one of the crime stain donors*. This is, in fact, the set of hypotheses that has been advocated and used frequently, see Evett (1987), Aitken (1995) or Stoney (1991). We call this the *first* set of hypotheses.

However, one can also compute the likelihood ratio for the hypothesis that *John Smith was the donor of the pillow stain* versus the hypothesis that *John Smith was not one of the crime stain donors*. We call this the *second* set of hypotheses. Note the difference between the first and second set of hypotheses: in the first set, there is no distinction between the stain on the pillow and the stain on the sheet, wheras in the second set, this distinction is consciencely made.

Finally, there is a third, natural set of hypotheses which completely ignores the stain on the sheet, namely the hypothesis that *John Smith was the donor of the pillow stain* versus the hypothesis that *John Smith was not the donor of the pillow stain*. Given the fact that Smith matches with the pillow stain, this third set of hypotheses might even be the most natural one, see also Meester and Sjerps (2003).

At first sight, it is not clear why it is a problem that various sets of hypotheses are possible. The problem becomes clear though, when we compute the likelihood ratios for the various sets of hypotheses. In general, it is not possible to construct a single likelihood ratio for so called *composite* hypotheses, and one needs additional assumptions. It turns out that in order to compute the likelihood ratio for the third set of hypotheses, we need the probability δ, say, that John Smith was one of the crime stain donors (prior to observing the DNA profiles).

For the first set, one can show (see the Appendix) that the likelihood ratio is

$$\frac{R_A}{2}. \tag{1}$$

For the second set, we find (see the Appendix) that the likelihood ratio is equal to

$$R_A, \tag{2}$$

which is different from the likelihood ratio corresponding to the first set by a factor of 2.

The likelihood ratio in the third set has a slightly more complicated form. It turns out (see again the Appendix) that in this case the likelihood ratio is equal to

$$\frac{R_A(2-\delta)}{2(1-\delta)}, \tag{3}$$

and this is again different from the previous answers. Especially when δ is close to 1, this third likelihood ratio becomes much larger than the previous two. So here is the problem: we have a number of natural sets of hypotheses, and each of these gives rise to a very different likelihood ratio. Yet, it is precisely this likelihood ratio, which is used as a measure for the strength of the evidence against suspect John Smith.

In the above two-stain problem, an expert using the third set of hypotheses might report a much higher likelihood ratio than an expert using the first or second set. This becomes even more confusing when you realise that after the evidence, that is, after finding out that the crime profiles are A and B and that John Smith has profile A, all three sets of hypotheses are equivalent to each other. Indeed if the first hypothesis of one of the sets is true, so is the first hypothesis of the other sets, and the same is true for the second hypotheses. Our problem now is abundantly clear: which set of hypotheses is the 'correct' one? And if we do not know whether this question has a meaningful answer, how do we know which hypotheses (and corresponding likelihood ratio) we have to use? Note that recent developments concerning the distinction between hypotheses at the activity and source level also address the question as to what hypotheses should be used; see for instance Evett, Jackson and Lambert (2000). This is a different issue though, and not of immediate interest to the current discussion.

3 Likelihood ratios and prior odds; the role of expert and juror

In the previous section we explained that the choice of the hypotheses has an effect on the numerical value of the likelihood ratio, and we also saw that the difference can be quite dramatic. Thus one faces a problem when reporting to the juror, since the choice of the hypotheses is not always trivial. But what can one do instead? In this section, we describe a way around the problem, by carefully distinguishing and combining the roles of the expert and the juror.

Consider two competing hypotheses, for instance (in the two-stain problem) the hypotheses that John Smith was one of the crime stain donors, against the hypothesis that he was not. *Before* we confront the juror with the DNA evidence, the juror can, in theory at least, decide which of two hypotheses he or she finds more likely. More precisely, he or she can determine the *odds* of one hypothesis against the other. Formally, the odds equal the probability of the first hypothesis divided by the probability of the other hypothesis. It must be stressed that these odds are *subjective*, and can take into account external knowledge about the suspect or additional evidence not based on DNA, like witness accounts.

Example. As a simple example, consider the case in which a murder is committed on a tiny island, with only 10 potential criminals. If John Smith is one of these 10 people, then most people would say, in the absence of any other evidence, that the probability of the hypothesis *John Smith was the murderer* is 1/10, and the probability of the hypothesis *Someone else was the murderer* is 9/10. This corresponds to the odds being $(1/10)/(9/10) = 1/9$. If one finds a cigarette at the scene of the crime, then the prior odds can change, since one is then perhaps willing to assign higher odds against John Smith, if Smith is the only smoker on the island.

In this example, we initially took *uniform* odds, that is, we considered all possible people equally likely to be the criminal. Without any evidence, this is a reasonable thing to do in many cases. Uniform odds simply means that we have no information whatsoever. After finding the cigarette, we had reasons to believe that John Smith is our man, and as a result, the odds change.

In other, more complicated circumstances, evaluation of the odds is not so easy, and can depend on other evidence, witness reports, personal beliefs, etcetera. Perhaps it is not possible at all to express one's prior belief in the guilt of the suspect numerically, and perhaps the only thing one is willing to say is that the prior is 'high' or 'low'. We shall see that this is not an obstacle for the methodology explained in this article.

The odds as described here, *precede* the presentation of the DNA evidence. This is the reason to call these odds the *prior* odds. The adjective 'prior' is chosen here to distinguish between the odds *before* the presentation of the DNA evidence, and the odds *after* the presentation of the DNA evidence. The latter odds are called *posterior odds*, and represent the ratio of the probabilities of the two hypotheses *after* the presentation of the DNA evidence. It is intuitively clear that if the profile of John Smith is the same as the crime profile, then the posterior odds should be larger than the prior odds.[1] Therefore, it is intuitively clear that a high likelihood ratio should make the posterior odds larger.

But how can the posterior odds be quantified? For the *prior* odds, we saw that these are subjective. Are the posterior odds also subjective? The answer is that

[1]In fact, this is not always true. If the crime profile is very common, then a match might lead to lower posterior odds, depending on the hypotheses chosen. However, this is merely a theoretical situation.

once someone has determined his or her prior odds for a set of two hypotheses, the posterior odds are determined by an elementary mathematical result called *Bayes' theorem*, as follows:

$$\text{posterior odds} \;=\; \text{likelihood ratio} \times \text{prior odds}. \tag{4}$$

This simple result says that once we know our own personal prior odds, the posterior odds can be obtained by multiplication with the likelihood ratio. The personal prior odds in conjuction with the likelihood ratio together give the personal posterior odds. We see that a large likelihood ratio tends to make the posterior odds high, but this can possibly be counterbalanced by very small prior odds. If someone considers a certain hypothesis very unlikely, then a large likelihood ratio can make the hypothesis more probable after presentation of the DNA evidence, although perhaps still very unlikely.

Why are posterior odds useful? We illustrate this with, again, the two-stain problem.

Example. The two-stain problem. Let us now see how we compute the posterior odds in the two-stain problem of the previous section. Recall that we had three different sets of hypotheses, and that each of these sets gave rise to a different likelihood ratio.

The first set, with the hypothesis that John Smith was one of the crime stain donors against the hypothesis that he was not, we computed the likelihood ratio as $R_A/2$. In the Appendix, we show that the prior odds are $\delta/(1-\delta)$. Hence we have from (1) and (4) that

$$\text{posterior odds} = \frac{R_A}{2} \times \frac{\delta}{1-\delta}. \tag{5}$$

For the second set, the likelihood ratio is R_A. In the Appendix we show that the prior odds are

$$\frac{\delta/2}{1-\delta}.$$

Hence, from (2) and (4) we find

$$\text{posterior odds} = R_A \times \frac{\delta}{2(1-\delta)}, \tag{6}$$

which is the same answer as with the first set.

A simple computation of the prior odds in the third set of hypotheses (see again the Appendix) yields that the prior odds are

$$\frac{\delta/2}{1-\delta/2},$$

and with (3) and 4), this gives

$$\text{posterior odds} = \frac{R_A(2-\delta)}{2(1-\delta)} \times \frac{\delta/2}{1-\delta/2} = \frac{R_A\delta}{2(1-\delta)}, \tag{7}$$

the same answer as in (5) and (6).

We see that although the likelihood ratios of the various sets of hypotheses are different, *the prior odds counterbalance this*, giving the *same posterior odds*. Is this a coincidence? No, it is not. The above computations confirm the claim of Dawid (2001) that '... one should avoid talk of 'the likelihood ratio' as if this term designated a well-defined objective measure of evidence: at best, it can only be regarded as such relative to a chosen specification of the hypotheses.' The analysis above shows that there are situations where there is no obvious single 'relevant' pair of hypotheses, and reporting only a likelihood ratio is misleading.

Fortunately, the posterior odds are invariant for pairs of *conditionally equivalent hypotheses*. We call two sets of hypotheses conditionally equivalent if they are equivalent after the DNA evidence, even if they were not before. In the two-stain case, once you know that John Smith has profile A, and that the crime stains on the pillow and sheet are respectively A and B, then all three sets of hypotheses express the same thing. Before we know this, this need not be the case. For instance, the hypothesis that John Smith was one of the crime stain donors, is really different from the hypothesis that he was the donor of the stain on the pillow. But as soon as you know that the profile of John Smith is A, and the profile of the pillow and sheet stain are A and B respectively, then, with this additional information, the two hypotheses say the very same thing. Therefore it is no surprise that if we start off with the same assumptions (that is, there is probability δ that John Smith was one of the crime stain donors) and conditionally equivalent hypotheses (that is, after the DNA evidence the various sets of hypotheses express the same thing), then the posterior odds are the same. Apparently, a high likelihood ratio is counterbalanced by small prior odds and vice versa.

So what can we conclude at this point? We consider it misleading to choose one of the possible hypotheses pairs, and report its likelihood ratio as a measure of evidential strength for these hypotheses. Jurors may tend to focus on this likelihood ratio, and forget about counterbalancing prior odds, since these are not mentioned. We have argued that the likelihood ratio is inappropriate in isolation from the prior odds, but that the posterior odds are invariant under different sets of conditionally equivalent hypotheses. Hence, the posterior odds *are* meaningful. Note that the posterior odds are obtained by *combining* the expert opinion, and the prior odds of the juror. Only the combination (multiplication in this case) of these two quantities is meaningful. This implies that it is impossible for the forensic expert to report the strength of the DNA evidence in one single number, since the expert can deliver only part of a combined quantity. Only the *joint* effort of expert and juror yields something meaningful.

Of course, we still have to deal with the fact that the priors are subjective. In the next section we shall suggest a way for the expert to facilitate the juror in judging the consequences of the DNA evidence.

4 How to report the evidence; examples

As motivated above, we fear that reporting only a likelihood ratio with corresponding hypotheses may lead to misinterpretations in some situations. Serious overestimation of the strength of the case against the suspect may occur, since jurors may not realise that a large likelihood ratio has to be combined with potentially extremely small prior odds. Essentially the same problem arises with match probabilities; indeed, in the Example in Section 1, the match probability was the reciprocal of the likelihood ratio. In addition, reporting only a match probability easily leads to the infamous prosecutor's fallacy.

We conclude that reporting a match probability or likelihood ratio should preferably be accompanied by a description of the effect of prior odds. How can we do this? We think that a table relating a *range* of possible prior odds to their posteriors may be the least misleading way to report the evidence (such a table was suggested before in e.g. NRCII 1996, or as a nomogram in e.g. Riancho and Zarrabeitia 2002). Indeed, the posteriors are invariant under varying conditionally equivalent hypotheses. In such a table, the juror can easily see how the posterior odds would change when varying the prior odds. Since we think that it is in most cases almost impossible to identify well defined numerical prior odds (very few jurors will be willing to state that their prior odds are 14.5, say, or whatever other number; most jurors will instead state that their prior odds are 'high', or 'low', without specifying the exact number), a table can visualise the effect of high ar low prior odds. In fact, since it might even be easier for a juror to say something about the prior *probability* that a suspect left a crime stain, it might be better for the expert to ask the juror for this probability, compute the prior and posterior odds, and report the posterior probability. We now illustrate these ideas with two examples.

Example 1. The two-stain problem. As a first example, consider again the two-stain problem. We have seen that the computation of the posterior odds require the prior probability δ that John Smith was one of the crime stain donors. Hence, we can create a table, relating various values of δ to the corresponding posterior odds. A juror can consult the table, and read off directly how the posterior odds change when his (subjective) prior probability that John Smith is involved varies. As an example, we consider a two-stain problem where the profile A of the suspect and the pillow stain has match probability 1 in 4,000, that is, $R_A = 4,000$. The forensic expert can now report the match probability 1 in 4,000, or the likelihood ratio of his favourite hypotheses, and add to this the effect of the prior probability, like for example the following table:

prior probability	posterior probability
0.00001	0.02
0.001	0.67
0.01	0.95
0.1	0.996
0.2	0.998
0.5	0.9995
0.9	0.9999

Table 1: The relation between prior probability and posterior probability that John Smith was one of the crime stain donors in the two-stain case with $R_A = 4,000$.

A juror can now see that the posterior probabilities are much larger than the prior probabilities; the match of the profile does indeed increase the evidence against John Smith. However, if the prior probability of John Smith being involved is small, and 0.00001 is not even *that* small, the posterior probabilities are larger, but probably still much too small to be useful in the legal case against John Smith.

Example 2. Database search. Consider the situation in which a DNA profile A (with match probability 1 in R_A) is found at the scene of a crime, and in which this profile is compared to profiles in an existing database. Suppose that there is exactly one hit, John Smith say. What is the evidential value of the DNA hit against John Smith? This question has been the subject of an interesting debate in the literature, since various people do not agree on the appropriate set of hypotheses to use. In Balding and Donnelly (1995) and Dawid and Morterra (1996) it was argued that the appropriate hypotheses are *John Smith is the donor of the crime stain*, versus *John Smith is not the donor of the crime stain.* The corresponding likelihood ratio can then be computed to be (see Dawid 2001)

$$\frac{R_A(1-\pi)}{1-\delta}, \tag{8}$$

where δ is the prior probability that John Smith is the donor of the crime stain, and π is the prior probability that the donor is contained in the database. Stockmarr (1999) criticised this; he argued that the hypothesis *John Smith is the donor of the crime stain* is data-dependent, since we only consider this hypothesis *after* searching through the database. He suggested to use the hypothesis *the crime* *stain donor is in the database* versus *the crime stain donor is not in the database.* These hypotheses lead to a likelihood ratio of $\delta R_A/\pi$, which clearly is very different from (8). In a high profile case in The Netherlands, the match probability of the crime profile was about 1 in 4000, and the database contained 600 profiles, leading (under certain assumptions) to likelihood ratios of about 4000 and 7, respectively, an enormous difference.

The analysis in this paper can also be applied in this controversal situation, and then it will be clear that it really does not matter which of the sets of hypotheses one uses. Indeed, the sets are conditionally equivalent, because *after* a single hit in the database, the two hypotheses *John Smith is the donor of the crime stain* and *the crime stain donor is in the database* are equivalent. It is not hard to see that the posterior odds in both cases are $R_A\delta/(1-\pi)$ (Stockmarr 1999, Dawid 2001).

Clearly, to compute the odds, one needs the prior probability δ that John Smith is the donor of the crime stain, and the prior probability π that the donor is in the database. In The Netherlands, the judge is allowed to know how the suspect was found; the judge therefore knows that in the above mentioned case, the suspect was found as the single hit in a database search. In principle, it is therefore possible to ask the judge his or her prior probabilities δ and π and create a table as above, for various combinations of δ and π. This, however, might be confusing for a judge or juror: it might be effectively impossible to make a simultaneous prior statement about δ and π. Therefore we suggest the following:

the expert uses δ, the prior probability that John Smith is the donor of the crime stain, and then *computes* π, the probability that the donor is in the database, by using δ, plus the assumption that all *other* people in the population are equally likely to be the crime stain donor. If there is no other information, this seems a reasonable and conservative assumption. However, if there is other information, it should be made clear that the table does not take this into account. For the computation, the expert needs to know the size n of the database, and the size N of the relevant population.

Here is an example of a table, where $n = 600$, $N = 10,000,000$, and the match probability of the stain is 1 in 4,000.

prior probability	posterior probability
0.00001	0.04
0.001	0.8
0.01	0.98
0.1	0.998
0.2	0.999
0.5	0.999
0.9	0.9999

Table 2: The relation between prior probability and posterior probability of the event that John Smith left the stain in the database case, with $n = 600$, $N = 10,000,000$ and $R_A = 4,000$.

We note a few things:

1. This solution is also possible if the juror is not supposed to know that the suspect was found through a database search; only the prior probability that John Smith is the crime stain donor is needed, and the number N. This means that the tables can also be used in legal systems where the juror is not allowed to know about the database search.

2. The good thing about this method is that the fact that John Smith was the only hit is really used. That is, we do take into account the fact that all other people in the database did not match and can not be one of the culprits.

3. This solution is not perfect, but a perfect and completely honest solution does not exist in a situation where the database search affects the outcome, while at the same time, the juror is not supposed to know about the database search.

4. Drawback is that the size of the relevant population, N, is needed in the computations, and this number might be difficult to choose for the forensic expert. Perhaps the expert can discuss this issue with the juror, or present a table where also N varies.

5 Final conclusion

We have argued that reporting only a match probability or likelihood ratio may sometimes be misleading and confusing, and we suggest to accompany the likelihood ratio by a table which shows the effect of various prior odds or probabilities on the match probability or likelihood ratio. Using this table, a juror can immediately see how prior probabilities or prior odds transform into posterior probabilities and odds. These posterior odds are the quantities of interest, and turn out to be invariant for conditionally invariant hypotheses.

Appendix: the computation of the likelihood ratios and prior odds in the two-stain problem

We sketch the computation of the likelihood ratios in the two-stain problem. For more details, see Meester and Sjerps (2003).

For the first set, we first compute the numerator of the likelihood ratio. This is the probability that we find the profiles A and B on the pillow and sheet respectively, under the assumption that John Smith was one of the crime stain donors and has profile A. If John Smith was one of the donors, he can have left the stain on the pillow, or on the sheet. Since we have no reason to consider one of these options more likely than the other (before the DNA evidence), we give them equal probability 1/2. If he was the donor of the pillow stain, the probability that the profile of the stain on the sheet is B, is then simply 1 in R_B, since we have no information whatsoever about the identity of the donor of the second stain. If he was the donor of the stain on the sheet, the probability that its profile is B is zero, since John has profile A. Hence the numerator of the likelihood ratio is $1/2R_B$. For the denominator, we need to compute the probability that two randomly selected individuals would be donor of profile A on the pillow and profile B on the sheet: $1/R_A \times 1/R_B$. Hence the likelihood ratio is $R_A/2$.

What are appropriate prior odds for the first set? We already assumed that we have prior belief that Smith was one of the crime stain donors with probability δ. This implies that the probability that he was not one of the crime stain donors is equal to $1-\delta$. Hence the prior odds are $\delta/(1-\delta)$.

For the second set, the denominator of the likelihood ratio is as before, but the numerator is different, since we specify in the hypothesis that John was the donor of the stain *on the pillow*. The probability of finding profile A on the pillow and profile B on the sheet, if John Smith was the donor of the pillow stain and has profile A, is $1/R_B$, and the likelihood ratio now becomes R_A, which is different from the likelihood ratio corresponding to the first set.

What are the prior odds? Well, we deal with the hypotheses that John Smith was the donor of the stain on the pillow, versus the hypothesis that two unknowns were the donors of the two stains. Since our prior belief that Smith was involved is δ, the prior probability that he was the donor of the stain *on the pillow* is $\delta/2$, as before. The probability of the second hypothesis is $1-\delta$, and we find that the prior odds are $(\delta/2)/(1-\delta)$.

The computation of the likelihood ratio in the third set is slightly more complicated. It turns out that in order to compute this likelihood ratio, we need the probability that John Smith was one of the stain donors. Suppose that this probability is δ. The numerator of the likelihood ratio is the same as in the above case, namely $1/R_B$. The denominator is slightly more complicated now. The hypothesis is that John Smith was not the donor of the *pillow* stain, and this does not rule out the possibility that he was the donor of the *other* stain. The probability that Smith was not one of the crime stain donors, *given* that he was not the donor of the pillow stain is the ratio of the probability that he was not one of the stain donors and the probability that he was not the donor of the pillow stain. The first probability is $1-\delta$, the second is $\delta\times 1/2+(1-\delta) = 1-\delta/2$, where the first term comes from the possibility that Smith was donor of the other stain, and the second comes from the possibility that he was not one of the donors. Hence, under the assumption that Smith was not the donor of the pillow stain, the probability that he was not one of the donors is $(1-\delta)/(1-\delta/2)$. Now we can compute the denominator of the likelihood ratio. Given that Smith was not the donor of the pillow stain, the probability that he was not one of the stain donors is $(1-\delta)/(1-\delta/2)$ and then the probability to find the two profiles A and

B is $1/R_A \times 1/R_B$. If Smith was donor of the sheet stain, then the probability to find B on the sheet is zero, since Smith has profile A himself. We conclude that the likelihood ratio is equal to

$$\frac{R_A(2-\delta)}{2(1-\delta)}.$$

The prior odds in the third set of hypotheses is easy now. The probability of the first hypothesis equals $\delta/2$, and the probability of the second hypothesis equals $1-\delta/2$, as explained before. Hence the prior odds are $(\delta/2)/(1-\delta/2)$.

References

Aitken C.G.G. (1995) Statistics and the evaluation of evidence for forensic scientists, Wiley, Chichester, UK.

Balding D.J. (2002) The DNA Database controversy, Biometrics **58**, 241-244.

Balding D.J. and Donnelly P. (1995) Inference in forensic identification, Journal of the Royal Statistical Society Series A **158**, 21-53.

Dawid A.P. (2001) Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search", Biometrics **57**, 976-978.

Dawid A.P. and Morterra J. (1996) Coherent analysis of forensic identification evidence, Journal of the Royal Statistical Society Series B **58**, 425-443.

Donnelly P. and Friedman R.D. (1999) DNA database searches and the legal consumption of the scientific evidence, Michigan Law Review **97**, 931-984.

Evett I.W. (1983) What is the probability that this blood came from that person? A meaningful question?, Journal of Forensic Science Society **23**, 35-39.

Evett I.W. (1987) On meaningful questions: a two-trace transfer problem, Journal of Forensic Science Society **27**, 375-381.

Evett I.W., Foreman L.A. and Weir B.S. (2000) Correspondence, Biometrics **56**, 1274-1277.

Evett I.W., Jackson, G. and Lambert J.A. (2000) More on the hierarchy of propositions: exploring the distinction between explanations and propositions, Sience & Justice **40**(1), 3-10.

Koehler J.J. (1996) On conveying the probabitive value of DNA evidence: Frequencies, likelihood ratios and error rates, University of Colorado Law Review, **67**, 859-886.

Meester R. and Sjerps, M. (2003) The evidential value in the DNA database search controversy and the two stain problem, to appear in Biometrics.

NRCII (1996) National Research Council, The evaluation of forensic DNA evidence, National Academy Press, Washington DC.

Riancho J.A. and Zarrabeitia M.T. (2002) The prosecutor's and defendant's Bayesian nomograms, International Journal of Legal Medicine **116**, 312-313.

Stockmarr A. (1999) Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search, Biometrics **55**, 671-677.

Stoney D.A. (1991) in: The use of statistics in forensic science, Aitken C.G.G. and Stoney D.A. (eds.), Ellis Horwood Limited, Chichester, UK.

Taroni F and Aitken C.G.G. (1998) Probabilistic reasonig in the law Part I: assessment of probabilities and explanation of the value of DNA evidence, Science & Justice **38**, 165-177.

Thompson W.C. and Schumann E.L. (1987) Interpretation of statistical evidence in criminal trials-the prosecutors fallacy and the defence attorneys fallacy, Law and Human Behavior **11**, 167-187.