Evaluating DNA profiles in a case where the defence is "It was my brother"

IW EVETT

Central Research and Support Establishment, Forensic Science Service, Aldermaston, Reading, Berkshire, United Kingdom RG7 4PN

The widespread application of DNA profiling has brought to some prominence the issue of close relatives sharing body fluid types. In a particular crime case it may be suggested that a close relative of the suspect was the person responsible. This paper explains how an assessment may be made in the context of single locus profiling and where the suggestion is that a full sibling is involved. A formula is established for the likelihood ratio which is then worked out in detail for the particular case where the suspect profile has two distinct bands. Generalization to other types of case is briefly explained and there is a short discussion of the impact of the evidence in the context of a trial. For all parties, justice will best be done if the scientist has ample notice, before the trial, of relationship issues such as this one.

Key Words: Bayes; DNA; Interpretation; SLP; Genetics.

Journal of the Forensic Science Society 1992; 32: 5–14
Received 15 October 1990; accepted 27 March 1991

Introduction

Consider a crime case in which a body fluid stain has been left at the scene by the perpetrator; a suspect has been apprehended and a blood sample from him is found to be of the same type as that of the crime stain. The probative value of this evidence is, in most cases, a function of the relative frequency with which that blood type occurs in the population. Whereas this assertion is intuitively accepted apparently universally in courts of law, the logical foundation of that relationship is often misunderstood. Clarification can be achieved by adopting a Bayesian view, along the following lines.

Within the adversary system there are two competing explanations for the evidence.

C: the suspect left the crime stain

$\bar{C}$: someone else left the crime stain

It is necessary to evaluate the probability of the evidence given each of these alternatives. In the present context, as has been explained elsewhere [1], the strength of the evidence is given by the likelihood ratio.

$$\frac{\text{Probability of the crime stain type given } C}{\text{Probability of the crime stain type given } \bar{C}}$$

In the simplest situation, the body fluid types are determined without error so the numerator is 1. What is the denominator? This is the probability of the observed crime stain type if someone else had left it. The problem is what we mean by "someone else". It is recognized that there is no burden on the defence to provide any details of this alternative scenario and so, in the absence of any other information, it is universal convention to consider this hypothetical alternative person as one selected at random from the population. Just what the population is in an individual case may not be a straightforward matter but, for this discussion, assume that the crime has been committed under such circumstances that there is an identifiable population and that there are data from a sample which is representative of that population. In those circumstances, the denominator of the likelihood ratio is simply the probability that a person selected at random would be of the same type as the crime stain; and the best estimate of that is the relative frequency of that type in the sample.

Given these conditions, the likelihood ratio is therefore the inverse of the observed relative frequency of the type of the crime stain. The relationship in this case is a simple one, though it is important to remember that in other situations it may not be nearly so simple.

This paper is concerned with an aspect of interpretation which does arise from time to time in this sort of case. In the foregoing, it has been taken that there is no particular reason to address any particular "other man" when addressing the probability of the evidence given $\bar{C}$. However, it might happen that a suggestion is made that a sibling of the suspect committed the crime. If this suggestion arises early enough in the course of the investigation, it may be possible to address it directly by interview, or, if necessary, by means of a blood sample from any proposed alternative sibling. Whereas this is the desirable procedure, there are situations in which it may not be feasible: the suggestion may not come until the trial itself and no time is available; the sibling (or siblings) concerned may refuse to co-operate or may not be approachable for a range of reasons, including death.

The problem for the forensic scientist, then, is to estimate the likelihood ratio given the two alternatives

C: the suspect left the crime stain

$\bar{C}$: a sibling of the suspect left the crime stain

The necessary calculations are explained, for reasons of topicality, from the perspective of DNA single locus profiling.

The model

The treatment of measurement error in SLPs is not considered here. It is assumed that there is a reliable matching rule and that the probability of a non-match between separate samples from one person is effectively zero. It is also assumed that the relative frequency of a given SLP band weight can be reliably estimated from a suitable population sample.

Consider the case in which the crime and suspect profiles both contain bands a and b. The extension to cases involving the use of more than one probe is obvious.

First, given no specific alternative explanation we address

C: the suspect left the crime stain

$\bar{C}$: someone else left the crime stain

Because we assume no error in the matching process the probability of the crime stain bands given C is simply one; the bands are exactly what we would expect if the suspect had left the crime stain. For the evidence given $\bar{C}$ we require the probability that a person selected at random would have a profile containing bands a and b; the best estimate of this is $2f_a f_b$, where f_a and f_b are the respective relative frequencies in the population sample. The likelihood ratio is $1/2f_a f_b$. If, for example, $f_a = f_b = 0{\cdot}05$, then the likelihood ratio is 200.

Now consider the evidence when the alternative explanation is that a brother of the suspect left the crime stain.

C: the suspect left the crime stain

$\bar{C}$: a brother of the suspect left the crime stain

The numerator of the likelihood ratio is still one. The evaluation of the denominator is by no means so obvious but it is fairly simple to establish the order of magnitude. This can be done by visualizing a case in which bands a and b are both very rare in the population. Then it is almost certain that one of the parents of the suspect has a profile of the kind ax and the other parent has a profile of the kind bx, where x denotes any band other than a or b. Because a and b are both very rare, we are ignoring the possibility of parental configurations such as aa/bx, ab/bx, etc.

If the parental band configuration is ax/bx, the probability that a brother of the suspect would have bands a and b is simply 1/4. So when the alternative explanation was that some random person from the population left the stain, the likelihood ratio was very large—in the thousands, or tens of thousands—because a and b are both very rare. When the alternative explanation is that a brother left the crime stain, it falls to approximately 4 (in fact, slightly less than 4).

Non-scientists might not be aware that such changes are possible. But they are no more than a reflection of the fact that different questions lead to different answers! This issue will be discussed a little more, later in the paper. In the next section, however, the ground work is laid for calculating the probability of the brother's type more precisely.

Theory

To determine the probability that a sibling will have bands a and b in his or her profile it is necessary first to establish the range of combinations which are possible for the parental profiles and their probabilities. Let ϕ_i denote a parental configuration. Examples are: ab/ab, aa/bb and so on. Note that it is not necessary to distinguish between mother and father, so $\phi = aa/bb$ is shorthand for saying "one parent has two copies of band a, the other parent has two copies of band b". It is worth noting also that the word "homozygous" is not used in this paper. Saying that a profile has two copies of band a is merely saying that there are two bands which are so close in size that the profiling system does not resolve them. It has been argued convincingly by Gill (personal communication) that at hypervariable mini-satellite loci, true homozygosity is most unlikely in a randomly mating population.

Let α denote the suspect's profile. Given α, we can itemize a range of parental configurations ϕ_i, $i = 1, 2, \ldots$. Some of these are more probable than others and it is necessary to establish the probabilities before we can establish the probability of a sibling's profile being the same as α.

Ignoring α it is a simple matter to write out probabilities for the ϕ_i based purely on random sampling from the population. For example, the probability of aa/bb is $2f_a^2f_b^2$. These can be termed prior probabilities $p(\phi_i)$. Taking account of α will modify these probabilities to posterior probabilities $p(\phi_i \mid \alpha)$ in a manner which is given by Bayes' theorem.

$$p(\phi_i \mid \alpha) = \frac{p(\alpha \mid \phi_i)p(\phi_i)}{p(\alpha)} \tag{1}$$

where $p(\alpha)$ is the unconditional probability of α, in this case $2f_af_b$, and $p(\alpha \mid \phi_i)$ is the probability of α given the parental configuration ϕ_i.

Given any one of the configurations ϕ_i, the probability that a full sibling will have a profile α is $p(\alpha \mid \phi_i)$. So the unconditional probability is obtained by weighting this by the probability of ϕ_i and summing over all values of i. Using equation (1), the probability that a sibling's profile will be α is, therefore,

$$\frac{1}{p(\alpha)} \sum_i p^2(\alpha \mid \phi_i) p(\phi_i) \tag{2}$$

This is a general expression which will apply whatever the form of α.

Calculation

It has already been explained that an upper limit for the likelihood ratio is 4 in the given scenario, where the suspect has bands a and b. It will provide a useful, and relatively simple, application of the formula (2) to establish a lower limit for the likelihood ratio; this occurs when bands a and b are very common. For illustration we consider what happens when $f_a = f_b = 0{\cdot}5$. This is, of course, unrealistic for SLPs but it is feasible in a conventional typing system with two alleles.

The calculation is carried out in Table 1. There are four possible configurations for the parental profiles and these are shown in the second column. There are two other configurations—*aa*/*aa* and *bb*/*bb*—but neither of these could result in a child with the suspect's profile, so they have been omitted from the table. The second column shows the prior probability of each configuration, that is, the probability that two people selected at random would exhibit the given configuration. The unconditional probability of the suspect's profile is 1/2, so the probability that a sibling will have the same profile, using the formula (2), is

$$2\{1/8 + 1/4\,.\,1/4 + 1/4\,.\,1/4 + 1/4\,.\,1/4\} = 5/8$$

In this case, the likelihood ratio is 8/5 = 1.6.

The more general calculation will next be carried out for all values of f_a and f_b. However, it is now clear that whatever their values the likelihood ratio must fall in the range 1·6 to 4.

Table 2 has a similar design to Table 1 but it is more complicated because it deals with the general case. Again, the possible parental configurations are

TABLE 1. Calculation of the probability that a person will have a profile containing bands *a* and *b* given that a sibling has that profile, in a special case where $f_a = f_b = 0{\cdot}5$

i	*Parental configuration* (ϕ_i)	*Prior probability* $p(\phi_i)$	*Probability of suspect's profile given* ϕ_i $p(\alpha \mid \phi_i)$
1	*aa*/*bb*	1/8	1
2	*aa*/*ab*	1/4	1/2
3	*ab*/*bb*	1/4	1/2
4	*ab*/*ab*	1/4	1/2

TABLE 2. Calculation in a general case of the probability that a person will have a profile containing bands *a* and *b* given that a sibling has that profile

i	*Parental configuration* (ϕ_i)	*Prior probability* $p(\phi_i)$	*Probability of suspect's profile given* ϕ_i $p(\alpha \mid \phi_i q)$
1	*aa/bb*	$2f_a^2 f_b^2$	1
2	*aa/ab*	$4f_a^3 f_b$	1/2
3	*aa/bx*	$4f_a^2 f_b f_x$	1/2
4	*ab/bb*	$4f_a f_b^3$	1/2
5	*ab/ab*	$4f_a^2 f_b^2$	1/2
6	*ab/bx*	$8f_a f_b^2 f_x$	1/4
7	*ax/bb*	$4f_a f_b^2 f_x$	1/2
8	*ax/ab*	$8f_a^2 f_b f_x$	1/4
9	*ax/bx*	$8f_a f_b f_x^2$	1/4

listed in column 2. The symbol x has been used to denote any band other than band a or band b. The meaning of f_x in the third column should then be obvious.

$$f_x = 1 - f_a - f_b$$

The probabilities in the last two columns are calculated in the same way as those in Table 1. Applying formula (2) gives the probability that a sibling has bands a and b (from Table 2)

$$\{2f_a^2 f_b^2 + f_a^3 f_b + f_a^2 f_b f_x + f_a f_b^3 + f_a^2 f_b^2 + f_a f_b^2 f_x/2 + f_a f_b^2 f_x + f_a^2 f_b f_x/2 + f_a f_b f_x^2/2\}/(2f_a f_b)$$

This can be simplified to

$$\{1 + f_a + f_b + 2f_a f_b\}/4$$

The likelihood ratio is the inverse of this. To confirm the first calculation note that as f_a and f_b become small the probability decreases towards 1/4, so the likelihood ratio increases to 4.

Dealing with other scenarios

The foregoing has dealt with one particular situation—when the suspect has a two-banded profile, and the alternative explanation is that the crime sample was left by a sibling. There are obvious ways in which different situations may arise.

First, there is the case where the suspect has just one band in his profile. In such a case, it would be necessary to start again from formula (2). However, there is at present a complication relating to one-banded profiles because there are two reasons why only one band may occur: either there are two bands present which are unresolved; or there is another band which, for one reason or another, has not been detected by the profiling system. It is possible to work through the calculation using appropriate weighting terms for these two eventualities but this will depend on local practice. The basic methodology has been established so this case is not discussed further here.

Next, there is the case where some relative other than a sibling is proposed as an alternative. The simplest case is where the suspect's father is suggested; it is a simple matter to show that, in that case, the likelihood ratio is $2/(f_a + f_b)$. Other, more distant relationships, such as uncles, half-brothers and cousins become more complicated but the general approach adopted here should always be possible.

Discussion

To discuss the impact that such considerations may have on the conduct of a case in court it is necessary to understand the crucial role of the likelihood ratio, which is determined by Bayes' theorem. If there are two alternatives C and $\bar{C}$, then whatever the odds are in favour of C before the scientific evidence (the prior odds), the odds in favour of C after the scientific evidence (the posterior odds) are the prior odds multiplied by the likelihood ratio. This simple result will first be illustrated with two hypothetical scenarios; but to explore the "brothers" issue it is necessary later to deal with the more complex case where there are not two, but three (or more) alternative explanations for the evidence.

For the first two scenarios let $f_a = f_b = 0{\cdot}01$.

Case Scenario 1

The defendant has no brothers. There is no suggestion of any of his other relatives being involved. In this case, the alternative $\bar{C}$ is based on a random selection from the population and the likelihood ratio is 5000. This must be regarded as very strong evidence to support C. Whatever the odds based on the other evidence they are multiplied by 5000. If, for example, the prior odds were 10 to 1 against, they become 500 to 1 on. If they were evens, they become 5000 to 1 on.

Case Scenario 2

The defendant has one brother. The circumstances of the case are such that one of these two men must have left the crime stain. However, no sample is available from the brother. In this case the likelihood ratio is approximately 4. If the other evidence gives no cause for favouring either of the brothers—i.e., prior odds are evens—then the posterior odds are 4 to 1 on.

It is recognized that this scenario is bizarre in the extreme but the author has found in conversations with both lawyers and scientists that it helps to present the extreme purely for illustrative purposes.

Case Scenario 3

In this case the defendant has a number of brothers. There is no particular item of evidence to implicate any one of them, neither have they been eliminated from suspicion. No samples have been taken from any of them. The proposal that a brother left the stain is put forward by defence counsel as a potential explanation for the scientist to consider.

This is the scenario which is most likely to occur. It will be most helpful to work it through in fairly general terms so the numerical values used in the present scenario are replaced for the time being with symbols. Let q denote the probability of the crime stain type given that a random member of the population left it; and let Q denote the probability given that a sibling of the defendant left it. In this kind of problem Q is much larger than q.

There are three explanations to be considered.

C: the defendant left the crime stain

$\bar{C}_1$: a random member of the population left the crime stain

$\bar{C}_2$: a brother of the defendant left the crime stain

Let $p(C)$, $p(\bar{C}_1)$ and $p(\bar{C}_2)$ denote the prior probabilities of these three explanations. Let E denote the crime stain evidence. Then $p(C \mid E)$, $p(\bar{C}_1 \mid E)$ and $p(\bar{C}_2 \mid E)$ are the posterior probabilities. Bayes' theorem gives

$$p(C \mid E) = \frac{p(E \mid C)p(C)}{p(E \mid C)p(C) + p(E \mid \bar{C}_1)p(\bar{C}_1) + p(E \mid \bar{C}_2)p(\bar{C}_2)}$$

i.e.

$$p(C \mid E) = \frac{p(C)}{p(C) + qp(\bar{C}_1) + Qp(\bar{C}_2)}$$

Simple manipulation shows that the posterior odds on C are given by

$$\frac{p(C)}{qp(\bar{C}_1) + Qp(\bar{C}_2)} \tag{3}$$

At the two values $p(\bar{C}_2) = 0$ and $p(\bar{C}_1) = 0$ this expression leads to the results already stated for Case Scenario 1 and Case Scenario 2, respectively.

In general, however, the expert is going to be in a difficult position. In the simpler kind of case where there are only two alternatives, the likelihood ratio gives a measure of evidence strength which is quite independent of the prior odds. In the present kind of situation, however, the import of the scientific evidence cannot be detached neatly from the other evidence. Like it or not, to assess his evidence it is necessary for the scientist to gain some idea of how the Court is thinking in relation to $\bar{C}_1$ and $\bar{C}_2$. The difficulties of doing this should not be underestimated.

The common courtroom approach to this problem is to talk in terms of population sizes. This is fraught with difficulty but it is generally necessary, if only for illustrative purposes.

Let us adopt the concept of a pool of possible suspects of size N—including the defendant himself. Let the defendant be one of n brothers all included in the suspect population.

If there is no evidence to favour the defendant or any other member of the suspect population, it may be argued that the prior probability for each and every member of the suspect population is $1/N$. Then this approach assigns

$$p(C) = \frac{1}{N}$$

$$p(\bar{C}_1) = \frac{(N-n)}{N}$$

$$p(\bar{C}_2) = \frac{(n-1)}{N}$$

The posterior odds on C can, using expression (3), be shown, if $N \gg n$, to be approximately

$$\frac{1}{qN + Q(n-1)}$$

In this framework, in the absence of the "brother" explanation, the posterior odds on C are $1/qN$.

The effects of the "brother" explanation will be illustrated by two numerical examples.

Example 1 Two SLPs are used, each giving two bands which match the suspect. Each of the four bands is rare, with a frequency of 0·01, so $q = 1/25{,}000{,}000$. The circumstances of the case are such that the defendant is one of a suspect population of $N = 100{,}000$ men. Then within the given framework of an equal prior for every one of the suspect population, the posterior odds are 250. However, if the suspect is known to have a brother then $Q = 1/16$, $n = 2$ and the posterior odds are $1/(0{\cdot}004 + 0{\cdot}063) \cong 15$. The "brother" alternative has a considerable impact on the overall assessment.

Example 2 One MLP has been used and there is a 14 band match. If the band matching probability is 0·26, the probability that a band in the profile of a brother of the defendant would match a band in the crime stain is 0·62. Let the overall pool of suspects be 100,000. Then $q = 1/155$ million, $Q = 1/800$. The defendant has 4 brothers.

If there were no "brother" suggestion the posterior odds would have been 1,550. Given the "brother" suggestion the posterior odds are 177.

Conclusion

To deal with the "it was my brother" defence (or, for that matter, the citation of any other relative as an offender) in court effectively and objectively is an extreme test of the forensic scientist's powers of reasoning and communication. It also involves considerations which, strictly speaking, are outside the scientist's recognized domain. It is difficult to imagine that the cause of justice can be served by a court room discussion around the treatments which have been presented in this paper. The messages are clear. To the prosecution, it is essential that the investigator ensures that any relatives who might credibly be considered suspects be eliminated—by taking samples if necessary. To the defence, that if there is a genuine concern about a relative's involvement then, for the scientist to address it in a manner which will best serve the cause of justice, it must be brought to his or her notice at the earliest opportunity.

References

1. Evett IW. What is the probability that this blood came from that person? A meaningful question? Journal of the Foresic Science Society 1983; 23: 35–39.