Lies, Damned Lies and Expert Witnesses

COLIN AITKEN
The University of Edinburgh

The value of scientific evidence presented by an expert witness may be misrepresented by both prosecution and defence. Consider evidence that a trace of a genetic marker, which occurs with low probability in the population, found at a crime scene has also been found in a suspect. The prosecution may use this to infer a low probability that the suspect is innocent of the crime. Conversely, the defence may argue that, even though the marker is rare, the size of the population from which the criminal may have come is so large that there may be several hundred people in it with such a marker. The suspect is one of these but there are many others with the marker also. Hence the evidence is not relevant for assessing the guilt of the suspect. These conflicting arguments are reconciled through consideration of the issues involved in evaluating evidence.

Every contact leaves a trace. This is a famous principle in forensic science known as Locard's principle. A crime is committed and the criminal makes contact with the scene of the crime. The traces which may be left are, for example, fibres from the criminal's clothing at the scene and fragments of glass from a broken window on the criminal's clothing. Evidence may be transferred in two directions, from the criminal to the crime scene and from the crime scene to the criminal. Such evidence is known as transfer evidence, for obvious reasons, or trace evidence, since it is concerned with traces of material. The purpose of this paper is to describe aspects of the evaluation of such evidence. First, two well-known fallacies are illustrated. Then a formal approach to evaluating evidence, based on the likelihood ratio, is described with an illustration based on evidence of blood grouping. The next section considers DNA profiling. This is an extremely powerful tool in the investigation of crime, but is also the source of much controversy. It is not possible in the space available to give more than a brief description of some of the issues involved. In particular, it is not possible to give a definitive approach based on likelihood ratios as this does not yet exist. Fuller details of all the issues outlined here are available in [1].

FALLACIES

There are two fallacies in the evaluation of evidence which have achieved a certain notoriety. These are known as the prosecutor's and defender's fallacies. The former is also known as the fallacy of the transposed conditional. Suppose a crime has been committed. Blood is found at the scene for which there is no innocent explanation. It is of a type which is present in 1% of the population. The prosecutor may then state

> "There is a 1% chance that the defendant would have the crime blood type if he were innocent. Thus, there is a 99% chance that he is guilty."

Alternatively, the defender may state

> "This crime occurred in a city of 800,000 people. This blood type would be found in approximately 8,000

people. The evidence has provided a probability of 1 in 8,000 that the defendant is guilty and thus has no relevance."

Both these statements are wrong. In order to understand why, it is necessary to consider a mathematical formulation of the evaluation of evidence.

LIKELIHOOD RATIOS

First, some notation has to be introduced. In a trial, there are two hypotheses under consideration: that of the prosecutor, which will be denoted H_0, and that of the defender, which will be denoted H_1. For the statistically minded, note that H_0 is not used here to denote the null hypothesis; the dictum that a defendant is innocent until proven guilty prevents such an interpretation. Nor do H_0 and H_1 necessarily need to refer to guilt and innocence. Another possibility is that H_0 is the hypothesis that the suspect was at the crime scene and H_1 is the hypothesis that the suspect was not at the crime scene; these hypotheses are the ones that are implied in any discussion of transfer evidence. Obviously, presence at the crime scene is not synonymous with guilt. It will be assumed, further, that H_0 and H_1 are mutually exclusive and exhaustive so that $Pr(H_0) + Pr(H_1) = 1$. Let the evidence to be evaluated be denoted by E. All other information, which may include other evidence, is denoted by I.

A fundamental assumption is then made. It is assumed that in evaluating evidence, only two probabilities are of relevance, namely the probability of the evidence given the truth of the prosecutor's hypothesis and all other information, I, denoted $Pr(E \mid H_0, I)$, and the probability of the evidence given the truth of the defender's hypothesis and I, denoted $Pr(E \mid H_1, I)$. Even such a seemingly obvious assumption is novel since it is often assumed that if $Pr(E \mid H_1, I)$ is small then the suspect is guilty; this is the prosecutor's fallacy. The value of $Pr(E \mid H_0, I)$ is not considered. Nor is it recognised that $Pr(E \mid H_1, I)$ is not the same as $Pr(H_1 \mid E, I)$; this is the origin of the term "the fallacy of the transposed conditional".

It can also be shown that the value, V say, of the evidence is a function of the ratio of the two probabilities, that is, the likelihood ratio. A simple proof based on one given in [2] is given here. The I is dropped for convenience.

Let $x = Pr(E \mid H_0)$ and $y = Pr(E \mid H_1)$. The assumption above states that

$$V = f(x, y)$$

for some function f.

Consider another piece of evidence, T say, which is irrelevant to, and hence independent of, E, H_0 and H_1, and which is such that $Pr(T) = \phi$, say. Then

$$\begin{aligned} Pr(E, T \mid H_0) &= Pr(E \mid H_0)Pr(T \mid H_0) \\ &\quad \text{by the independence of } E \text{ and } T \\ &= Pr(E \mid H_0)Pr(T) \\ &\quad \text{by the independence of } T \text{ and } H_0 \\ &= \phi x. \end{aligned}$$

Similarly,

$$Pr(E, T \mid H_1) = \phi y.$$

The value of the combined evidence (E, T) is equal to the value of E, since T has been assumed irrelevant. The value of (E, T) is $f(\phi x, \phi y)$ and the value of $E = V = f(x, y)$. Thus,

$$f(\phi x, \phi y) = f(x, y)$$

for all ϕ in the interval $[0, 1]$ of all possible values of $Pr(T)$. It follows that f is a function of x/y alone and hence that V is a function of

$$Pr(E \mid H_0)/Pr(E \mid H_1),$$

namely, the likelihood ratio.

Thus, replacing I for completeness,

$$V = \frac{Pr(E \mid H_0, I)}{Pr(E \mid H_1, I)}.$$

From the odds version of Bayes' Theorem, V is the factor which converts the prior odds in favour of H_0 into the posterior odds in favour of H_0, namely,

$$\frac{Pr(H_0 \mid E, I)}{Pr(H_1 \mid E, I)} = V \times \frac{Pr(H_0 \mid I)}{Pr(H_1 \mid I)}. \qquad (1)$$

Such a representation neatly separates out the roles of the scientist, or expert witness, and the investigator. The expert witness is concerned with the value of the evidence, which is represented by V, the likelihood ratio. The investigator is concerned with the prior and posterior odds in favour of H_0 and the effect that V has in converting the former into the latter.

A value of V greater than 1 lends support to the prosecutor's hypothesis, a value less than 1 lends support to the defender's hypothesis. If logarithms are used, then the effect of E is an additive one. This has an intuitively attractive analogy with the scales of justice and the weighing of evidence. Also, the posterior odds for one piece of evidence, E say, may be taken as the prior odds for the next piece of evidence, F say. Care has to be taken then when evaluating F that due attention is paid to the dependencies there may be between E and F.

EXAMPLE 1

Consider an example involving a blood typing system. A crime has been committed and blood of type Γ has been found at the scene of the crime. It has been ascertained that this blood can only have come from the criminal. There is eyewitness evidence that the criminal is Caucasian and the frequency of blood of type Γ in the relevant Caucasian population is γ. A suspect has been identified who is of blood type Γ. The value of the evidence may be determined as follows. In the above notation:

- I: The criminal is the source of the crime stain and there is eyewitness evidence that the criminal is Caucasian;
- H_0: Suspect was the source of the stain at the crime scene;
- H_1: Suspect was not the source of the stain at the crime scene;
- E: (E_1, E_2) where:
- E_1: the suspect is of type Γ;
- E_2: the stain at the crime scene is of type Γ.

For the moment I is omitted for clarity. Then

$$V = \frac{Pr(E \mid H_0)}{Pr(E \mid H_1)}$$
$$= \frac{Pr(E_1, E_2 \mid H_0)}{Pr(E_1, E_2 \mid H_1)}$$
$$= \frac{Pr(E_2 \mid E_1, H_0)}{Pr(E_1 \mid E_1, H_1)} \times \frac{Pr(E_1 \mid H_0)}{Pr(E_1 \mid H_1)}.$$

The suspect's blood type is independent of whether he was or was not at the crime scene. Therefore $Pr(E_1 \mid H_0) = Pr(E_1 \mid H_1)$. Also, if H_1 is true and the suspect was not at the crime scene then the blood type of the stain at the crime scene (E_2) is independent of the blood type of the suspect (E_1). Therefore, $Pr(E_2 \mid E_1, H_1) = Pr(E_2 \mid H_1)$. Thus

$$V = \frac{Pr(E_2 \mid E_1, H_0)}{Pr(E_2 \mid H_1)}. \qquad (2)$$

Consider the numerator of (2). If the suspect was the source of the crime stain (H_0 is true) and the suspect was of type Γ (E_1) then the crime stain is of type Γ with probability 1; thus $Pr(E_2 \mid E_1, H_0) = 1$. Consider the denominator of (2). The criminal is Caucasian and is the source of the crime stain (I), which is of type Γ. The relative frequency of blood of this type in the relevant population is γ; thus $Pr(E_2 \mid H_1) = \gamma$. Hence

$$V = 1/\gamma.$$

As a numerical example, consider Γ to be type AB in the ABO system and the relative frequency γ in the relevant Caucasian population to be equal to 0.031. Then

$$V = 1/\gamma = 1/0.031 \simeq 32.$$

The presentation of this result is very simple and this simplicity provides a very attractive feature of this approach. The result may be presented by saying that

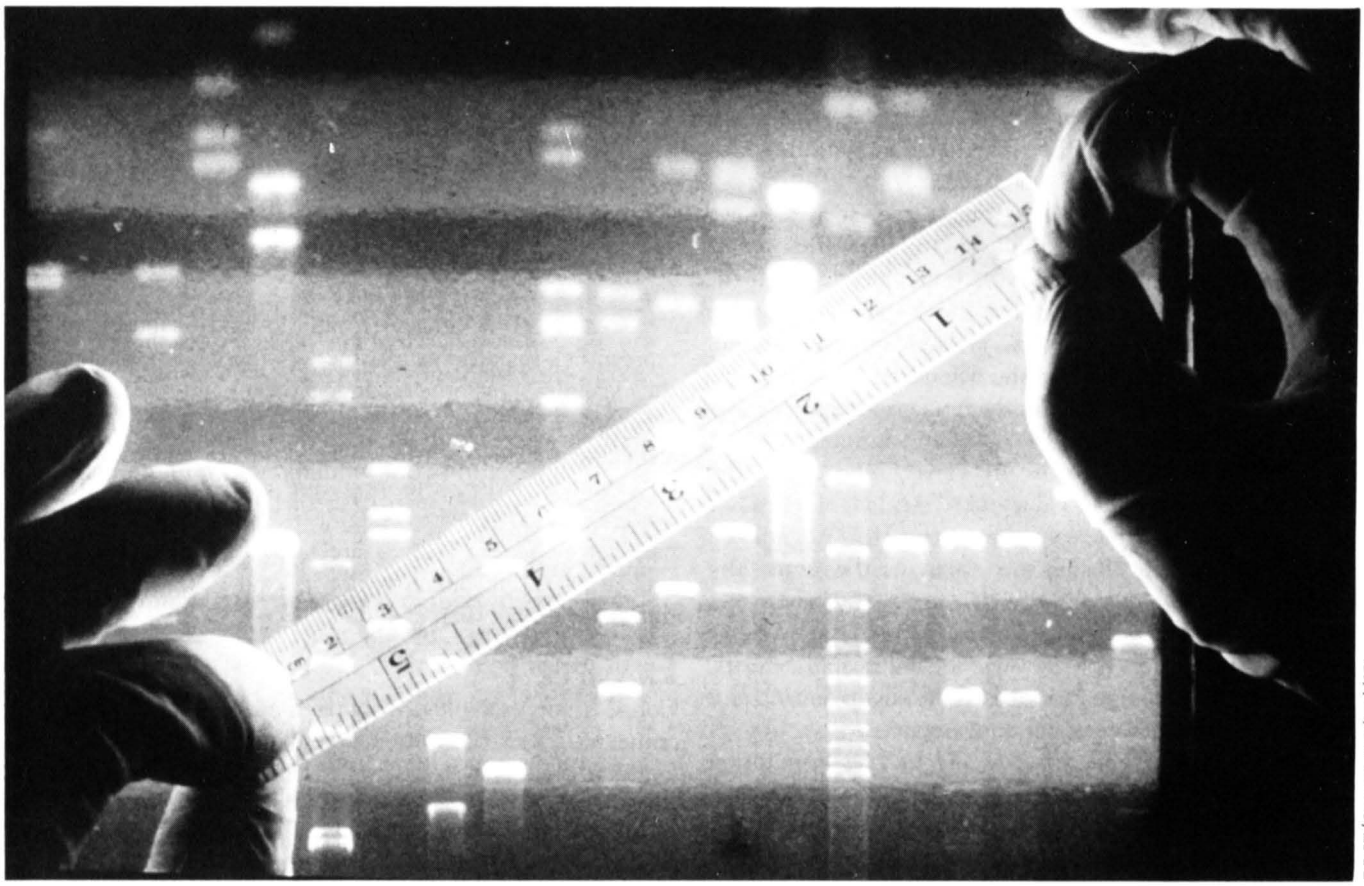
> "the odds in favour of the presence of the defendant at the crime scene are increased by a factor of 32 because of the evidence of the blood stain."

A qualitative scale

A qualitative scale has been suggested[3,4] for reporting the value of the support given by V for H_0, and is given in Table 1.

DNA

DNA profiling is a very powerful tool in forensic science. It may be thought that it would provide a good example of the use of statistics in the evaluation of scientific evidence. However, there is still much controversy surrounding its use and a definitive approach based on likelihood ratios does not exist. However, it is possible to determine values for the



DRCT/Custom Medical/SPL

Table 1

A qualitative scale for reporting the value of the support of the evidence for H_0 against H_1

$1 < V \leq 10^{1/2}$	Slight increase in support
$10^{1/2} < V \leq 10^{3/2}$	Increase in support
$10^{3/2} < V \leq 10^{5/2}$	Great increase in support
$10^{5/2} < V$	Very great increase in support

evidence, based on a simpler approach known as match-binning described later. Even this simpler approach has its critics, however, and some criticisms are given below. First, however, there is a brief description of the technique.

DNA (Deoxyribonucleic acid) is composed of two weakly connected strands of molecules that spiral to form a double helix. Each strand of DNA consists of an almost endless chain of four different chemical building blocks or *bases*, often referred to by their initials: A for adenine, C for cytosine, G for guanine and T for thymine. The C on one strand always pairs with the G on its complementary strand; the A on one strand always pairs with the T on its complementary strand. There is a very large number of possible orderings of these so-called *base-pairs* in a lengthy stretch of DNA. A sample of DNA may be cut by a restriction enzyme and fragments are separated by an electric field, a process known as electrophoresis. Fragments of particular interest are identified by a labelled probe which fuses to the fragments since it has a DNA sequence complementary to them.

The final product from a single-locus probe (a probe which attaches itself to a single locus) is a maximum of two bands which are represented on an autoradiograph, each band representing a separate fragment of DNA. The bands correspond to the two alleles, one from each parent, at the particular locus specific to the probe. The distance the band has migrated on the autoradiograph is measured in numbers of basepairs (*bp*) or kilobasepairs (*kb*). A comparison of a sample of DNA assumed to be from the criminal (known here as the crime sample) and a sample of DNA from a suspect, who may or may not be the criminal, is made by comparing the positions of the bands on two autoradiographs (one for the crime sample, one for the sample from the suspect).

DNA profiles may be evaluated by using one of two methods, match-binning or likelihood ratios, though both are a long way from universal acceptance.

Match-binning

Consider one fragment with an estimated length x. Construct an interval $x \pm \delta$ centred on x. Two fragments, each with an appropriately constructed interval centred on their estimated length, are said to match if these intervals overlap, otherwise they are said not to match. There is measurement error associated with the determination of the lengths of these fragments and the standard deviation of any measurement is taken to be proportional to the fragment length. For example, the FBI take $\delta \simeq 2.5\%$ of x, Lifecodes take $\delta \simeq 1.8\%$ of x and Cellmark take δ to correspond to a cell migration distance of 1 mm.[5]

The relative frequency of fragments which will be found to match the sample (with estimated length x say) under consideration is determined by reference to a database from a relevant population and consideration of the interval which will contain such matching fragments. This interval is known as a bin and corresponds to a cell in a histogram. This is the binning procedure of match-binning.

As an example of match-binning in practice, consider not one but k loci, fully heterozygose, with k probes used for a crime sample of DNA (ie a sample known to be associated with a criminal). There are then $2k$ alleles. Let p_i $(i = 1, \ldots, 2k)$ be the population relative frequency of matching alleles for the ith allele observed in the crime sample profile, as determined from the binning procedure. The overall profile relative frequency P for these probes is then

$$P = 2^k \prod_{i=1}^{2k} p_i. \tag{3}$$

This result assumes Hardy–Weinberg equilibrium and linkage equilibrium, assumptions which have been shown to be reasonable. The value P, given in (3), may be taken to be the probability a person chosen at random from the relevant population matches the crime sample.

Suppose, for one probe, the relative frequency $2p_1p_2$ is taken to be approximately 1/50. Then, for four independent probes, the frequency is approximately $(1/50)^4$ which equals 1 in 6.25 million. It is not difficult to see how very small probabilities of a match by chance can arise.

Note that the p_i are not well-defined allelic frequencies. They are population relative frequencies of bands which fall sufficiently close to the crime sample DNA to match, under whichever definition of a match is being used. Also, different frequencies may be derived from databases of different racial types present in the population. Thus, several different values for P may be quoted, one for each of several different ethnic groups such as Caucasians, Blacks and Hispanics. It is a matter for judge and jury to decide which group is relevant for the case in hand. Note that it is the racial structure of the population from which the criminal is thought to have come that is relevant. The racial structure of the suspect cannot be used on its own to determine the relevant population.

The method is very easy to use but it is open to criticism on two counts. One is the problem of defining a match; three different definitions have been given above. The other problem is the dichotomy between the presence and absence of a match. A match occurs if two bands have overlapping intervals. Suppose two bands just fail to match under this criterion. Then the suspect, who has provided one of the bands, will be removed from suspicion. However, two bands which just do match will result in the suspect remaining under suspicion. A very small difference in measurements can lead to a very large difference in consequences for the suspect.

Likelihood ratio

The match-binning approach can provide a crude value for the likelihood ratio (1). This is done by taking the numerator to be 1 and the denominator to be P from (3). However, another approach has been developed[6,7] in which the numerator uses known variation in measurement errors and the denominator uses the variability over a database from a relevant population. The likelihood ratio in this case is then a comparison of probability density functions. It is a continuous measure and is not subject to the dichotomy of the match-binning approach. Note, however, that the probability density functions are not necessarily Normal. The denominator is a smoothed version of an appropriate histogram of relative frequencies.

Finally, consider correlation of measurement errors. The

match-binning approach has considered individual bands and no account is taken of any correlation in measurement error that may exist when combining results from the two bands of a single-locus probe. Note that the positions of the two bands from a single-locus probe are independent but it is the measurement error which is correlated between the two bands from a single-locus probe. A method which accounts for this correlation is given in [7].

Note: The National Research Council of the USA set up a committee on DNA Technology in Forensic Science in order to resolve the issues surrounding its use. This committee reported in 1992.[8] However, the report raised more new issues than it resolved. A second committee was set up and at the current time its report is still awaited.

EXPLANATION OF THE PROSECUTOR'S AND DEFENDER'S FALLACIES

Consider the crime described at the beginning of the section entitled "Fallacies" where the fallacies are expounded and consider (1), where

$$V = \frac{Pr(E \mid H_0)}{Pr(E \mid H_1)}.$$

Let H_0 be taken to mean guilty and H_1 be taken to mean truly innocent (not just "found not guilty"), so that $Pr(H_0) + Pr(H_1) = 1$.

The evidence E is such that $Pr(E \mid H_1) = 0.01$ and $Pr(E \mid H_0) = 1$. Thus $V = Pr(E \mid H_0)/Pr(E \mid H_1) = 100$.

The prosecutor claims that there is a 99% chance that the suspect is guilty. This is to say that $Pr(H_0 \mid E) = 0.99$ and hence that $Pr(H_1 \mid E) = 0.01$. The posterior odds in favour of H_0 are then 99, or approximately 100, which is equal to V. From (1), therefore, the prior odds also equal 1, that is

$$\frac{Pr(H_0)}{Pr(H_1)} = 1.$$

In other words, $Pr(H_0) = 1/2$, since $Pr(H_0) + Pr(H_1) = 1$, and the prior probability of guilt is $1/2$.

The prosecutor's statement is only true therefore if the suspect has a probability of $1/2$ of being guilty before the evidence is heard. This is hardly in accord with the dictum that a person is "innocent until proven guilty". The alternative terminology, the fallacy of the transposed conditional, arises from the equation of $Pr(E \mid H_1)$ with $Pr(H_1 \mid E)$.

Unfortunately, the fallacy is still being made. In one recent trial, involving DNA evidence, a match in some sense was found between the DNA profile of the sample associated with the crime and the DNA profile of the defendant, such that if the defendant were not the source of the crime sample then the probability of a match was 1 in 3 million. A statement, however, was made by a forensic scientist that

> "the likelihood of this [the source of the crime sample] being any other man but [the defendant] is 1 in 3 million".

In other words, the probability the defendant is innocent is 1 in 3 million. But this is not a valid inference from the DNA evidence. The prior odds and other information, I, has to be taken into account before a value may be assigned to the probability of innocence.

Consider now the defender's fallacy. The defender in the crime discussed in Section 2 claims that the probability of guilt is 1 in 8,000, i.e., $Pr(H_0 \mid E) = 1/8{,}000$, and thus that the evidence is of no relevance for determining the guilt of his client. Note that $Pr(H_1 \mid E) = 7{,}999/8{,}000$ and the posterior odds are therefore

$$\frac{Pr(H_0 \mid E)}{Pr(H_1 \mid E)} = \frac{1}{7{,}999} \approx \frac{1}{8{,}000}.$$

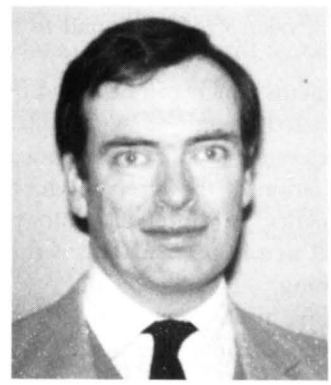
The value of the evidence $V = 100$, as before. Thus, the prior odds are

$$\frac{Pr(H_0)}{Pr(H_1)} = \frac{1}{800{,}000}.$$

In a population of 800,000 the prior odds of guilt of 1 in 800,000 imply that *a priori* the suspect is just as likely to be guilty as anyone else. This seems a perfectly reasonable definition of "innocent until proven guilty". The fallacy lies in the statement that the evidence has no relevance, for the evidence does in fact have considerable relevance. Before it was presented, the odds in favour of the defendant's guilt were 1 in 800,000; after it was presented, the odds in favour of the defendant's guilt were 1 in 8,000, a change by a factor of 100. Such a large change is surely of relevance. Of course, it may not be enough to find the suspect guilty but that is another matter. It would be odd indeed if a person were found guilty solely on the evidence of a match in a blood type which was present in 1% of the population.

References

1. Aitken, C G G. (1995). *Statistics and the evaluation of evidence.* John Wiley, Chichester.
2. Good, I J. (1991). Weight of evidence and the Bayesian likelihood ratio. *The use of statistics in forensic science* (Eds. C G G Aitken and D A Stoney), Ellis Horwood, Chichester, 85–106.
3. Jeffreys, H. (1983). *Theory of probability,* 3rd edition. Clarendon Press, Oxford.
4. Evett, I W. (1987). Bayesian inference and forensic science: problems and perspectives. *The Statistician,* **36,** 99–105.
5. Weir, B S. (1992). Population genetics in the forensic DNA debate. *Proceedings of the National Academy of Sciences,* **89,** 11654–11659.
6. Berry, D A. (1991). Inferences using DNA profiling in forensic identification and paternity cases. *Statistical Science,* **6,** 175–205.
7. Berry, D A, Evett, I W and Pinchin, R. (1992). Statistical inference in crime investigations using deoxyribonucleic acid profiling (with discussion). *Applied Statistics,* **41,** 499–531.
8. National Research Council. (1992). *DNA Technology in Forensic Science.* National Academy Press, Washington, DC, USA.



Colin Aitken is a senior lecturer in statistics at the University of Edinburgh with theoretical research interests in multivariate Bayesian analysis and non-parametric density estimation and applied interests in forensic science and the evaluation of evidence. He was one of four instructors at a course for forensic scientists in Florida in January in which they tried to persuade the participants that a Bayesian approach is the only way to evaluate evidence.