Fundamentals of statistical evidence—a primer for legal professionals

Citation for published version:
Aitken, C & Taroni, F 2008, 'Fundamentals of statistical evidence—a primer for legal professionals' International Journal of Evidence & Proof, vol. 12, no. 3, pp. 181-207. DOI: 10.1350/ijep.2008.12.3.296

Digital Object Identifier (DOI):
10.1350/ijep.2008.12.3.296

Published In:
International Journal of Evidence & Proof

Aitken, C & Taroni, F 2008, 'Fundamentals of statistical evidence - a primer for legal professionals' International Journal of Evidence and Proof, vol 12, no.3, pp. 181-207.

Fundamentals of statistical evidence—a primer for legal professionals

By C. G. G. Aitken and*
School of Mathematics, University of Edinburgh

F. Taroni†
School of Criminal Sciences, University of Lausanne, Switzerland

Abstract Criminal courts are today frequently confronted with statistical evidence, notably in relation to DNA profiling. Recent experience tends to confirm both widespread perceptions and more systematic research indicating that probability and statistics are not handled confidently, or always competently, by lawyers, judges, jurors or even by forensic scientists. Conceived as a primer for legal professionals, this article reviews basic statistical terminology and its forensic applications, and explores the options for presenting statistical information to fact-finders effectively. In raising awareness of the issues and by encouraging improved comprehension of probability and statistics amongst legal and forensic science professionals, we aim to contribute directly to the administration of justice by promoting more successful applications of forensic statistics in legal adjudication.

Keywords Evaluation of evidence; Likelihood ratio; Presentation in court; Probability; Statistics.

* Email: c.g.g.aitken@ed.ac.uk.
† Email: f.taroni@unil.ch. The authors acknowledge support from research grant ESRC RES-000-23-0729 and very helpful comments from the Editor.

doi:1350/ijep.2008.12.3.296

Criminal courts have frequently been confronted with statistical evidence in recent years, not least in relation to proof of identification through DNA profiling.[1] One of the best illustrations is *R* v *Denis Adams*,[2] in which the accused was charged with rape. The evidence linking Adams to the crime was, first, a match between his DNA and that of semen obtained from the victim and, second, the fact that he lived locally. A 'match probability' of 1 in 200 million for the DNA was reported. The defence challenged this, countering that a figure of 1 in 20 million or even 1 in 2 million could be more appropriate.

More ambitiously, the defence attempted to invite the jury to employ formal statistical methods in their deliberations on the evidence presented in the case. In particular, the jury were instructed in the statistical calculus known as Bayes' Theorem. In order to apply Bayes' Theorem to the factual scenario presented in *Adams* it was necessary to assign numerical values, not only to the explicitly probabilistic DNA evidence, but to all pertinent information before the jury. For example, the victim gave a description of her attacker which was hard to reconcile with the defendant. She also failed to pick out Adams at an identity parade. For the purposes of illustration, a probability of 0.1 was assigned to this (palpably weak) identification evidence on the assumption that Adams was guilty, as against a probability of 0.9 assuming his innocence. This gives a 'likelihood ratio' of 9 in support of innocence. In addition, a former girlfriend of Adams gave an alibi which was not challenged at trial. This evidence was assigned a probability of 0.25 if the defendant were guilty and a probability of 0.50 if he were innocent. The jury were then instructed in how to combine these probabilities with 'prior odds' of guilt of 200,000 to 1 against and DNA evidence of 1 in 20 million (as calculated by the defence).

The jury returned a verdict of guilty. In allowing Adams' first appeal against conviction, the Court of Appeal was dismissive of the attempt to introduce probabilistic reasoning into court, objecting that 'it trespasses on an area peculiarly and exclusively within the province of the jury'. The court continued:

> [W]hatever the merits or demerits of the Bayes Theorem in mathematical or statistical assessments of probability, it seems to us that it is not appropriate for use in jury trials, or as a means to assist the jury in their task ... [T]he attempt to determine guilt or innocence on the basis of a mathematical formula, applied to each separate piece of evidence,

1 On the broader significance of probability, as the basis of all juridical evidence, see e.g. R. Allen and P. Roberts (eds), Special Issue on the Reference Class Problem (2007) 11 E & P 243--317.

2 *R* v *Adams* [1996] 2 Cr App R 467, CA; *R* v *Adams (No. 2)* [1998] 1 Cr App R 377, CA.

> is simply inappropriate to the jury's task. Jurors evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them ... [T]o introduce Bayes Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task.[3]

A retrial was ordered, at which further attempts were made to describe the Bayesian approach to the integration of all the evidence. Once again the jury convicted, once again the case went to appeal and once again the Bayesian approach was rejected as inappropriate to the courtroom: but this time Adams' appeal was dismissed. In *Adams (No. 2)* the Court of Appeal was even more emphatic in its condemnation of probabilistic approaches to non-statistical evidence:

> [W]e regard the reliance on evidence of this kind ... as a recipe for confusion, misunderstanding and misjudgment, possibly even among counsel, but very probably among judges and, as we conclude, almost certainly among jurors. It would seem to us that this was a case properly approached by the jury along conventional lines ... We do not consider that [juries] will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion. We are very clearly of opinion that in cases such as this, lacking special features absent here, expert evidence should not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial.[4]

Whether or not jurors would in practice be assisted by explicit reference to it, the probabilistic foundation of evidence is a fact of life, with apparently increasing salience for criminal courts reflected in cases such as *Adams* and *Sally Clark*[5] (to which we return in conclusion). An alternative to the Court of Appeal's reflex defensiveness and suspicion of statistics is for lawyers and courts to try to get to grips with the fundamentals of probabilistic reasoning and its associated terms of art, including Bayes' Theorem, 'prior odds', 'likelihood ratio' and 'match probability'.

3 [1996] 2 Cr App R 467 at 481–2.
4 [1998] 1 Cr App R 377 at 384–5.
5 *R* v *Clark* [2003] EWCA Crim 1020.

This article offers a primer for legal scholars and practitioners with limited background in statistical method. Section 1 introduces basic conceptions of evidential value from a probabilistic perspective. Section 2 then proceeds to explore these ideas through a glossary of statistical terminology, culminating in a proof of Bayes' Theorem. In Section 3 we consider the competing merits of five alternative ways of presenting statistical information to jurors at trial. Although we endorse the likelihood ratio as our preferred approach, continuing difficulties of juror (and lawyer) comprehension must be acknowledged. Finally, the threads of the discussion are drawn together in Section 4, where we emphasise the crucial importance of renewed efforts to provide lawyers, judges, jurors and scientific experts with a basic knowledge and understanding of probability and statistics.

1. Probabilistic conceptions of evidential value

The admission of evidence describing the relative frequency of a DNA profile has encouraged the courts to be somewhat more open to the admission of probabilistic evidence in general than perhaps they were before the advent of DNA profiling. However, there is still much confusion surrounding the interpretation of evidence to which a measure of uncertainty is attached in explicitly probabilistic terms. It is notable that uncertainty in relation to such evidence can be measured quantitatively, in contrast to other types of evidence such as eyewitness identification, for which there is no mathematical metric for gauging reliability.

A small probability for finding incriminating evidence on an innocent person does not imply a large probability of guilt for a person on whom the evidence is found. This seemingly innocuous statement has been the source of much confusion in the interpretation of evidence in which probabilities have been mentioned. John Darroch, an Australian statistician, used the following example in his evidence to an Australian Royal Commission.[6] Consider a town in which a woman has been raped. Trace evidence has been found at the crime scene which indicates that there was one, and only one, perpetrator. It is also known that the assailant was a miner. There are 200 miners in the town and 19,800 other men capable of having committed the crime. Now, there is obviously room here for a debate as to the number of people in the population to which the criminal may be presumed to belong, i.e. the *relevant population* for our purposes. It is a matter of contention whether such a population may be defined and, if so, whether its size may be determined.[7]

6 J. Darroch, 'Probability and Criminal Trials: Some Comments Prompted by the Splatt Trial and the Royal Commission' (1987) 6 *The Professional Statistician* 3; J. Darroch, 'Probability and Criminal Trials' (1985) 30 *Newsletter of the Statistical Society of Australia* 1.

7 Generally, see C. G. G. Aitken and F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists* (Wiley: Chichester, 2004) 274–81.

In the example, suspects may be identified by means other than consideration of the trace evidence. Any individual suspect must work as a miner, otherwise he would have been excluded from the investigation earlier. Trace evidence is found in his environment (on his clothes, in his car, in his home, etc.) which in some sense matches the trace evidence at the crime scene but which is trace evidence that is also associated with all other miners in the town. This numerical information can be represented in a table with two rows and two columns, known as a two-by-two (or '2 × 2') table, as shown in Table 1:

Table 1. *Data on the frequency of trace evidence amongst miners and other males*

Trace evidence	Guilt	Innocence	Total
Present	1	199	200
Absent	0	19,800	19,800
Total	1	19,999	20,000

Table 1 illustrates why a small probability for finding incriminating evidence on an innocent person does not imply a large probability of guilt for a person on whom the evidence is found. Consider the 'Innocence' column, containing in total 19,999 males. Of these, 199 have trace evidence present in their environment. The ratio of these two numbers is 199/19,999, which approximately equals 1/100 or 0.01. This small number is an estimate of the probability of finding the trace evidence on an innocent person. To reiterate: *there is this small probability of finding the evidence on an innocent person.*

Now consider the probability of *guilt* of a person for whom trace evidence is found in his environment. There are 200 such people, as shown in the row labelled 'Present'. Yet *ex hypothesi* one, and only one, of these individuals is guilty. The ratio of these two numbers (1/200, or 0.005) estimates the probability of guilt of a person on whom the evidence is found. Thus, we can now clearly see that the small probability (0.01) of finding the evidence on an innocent person is not equivalent to a large probability that a person on whom the evidence is found is actually guilty. In other words, the probability 199/19,999 is not the probability of innocence, but the probability that trace evidence will be detected on an innocent person. Conversely, the complement of this fraction, 1 − 199/19,999 or 19,800/19,999, is obviously not the probability of guilt. To the contrary, given that there are 199 'false positives' in the total sample of 20,000 males and only one truly guilty individual, it is very likely (199/200) that any person chosen at random with trace evidence in their environment *is innocent.*

The crucial consideration is to distinguish correctly between (1) the event about whose outcome one is uncertain, and (2) the known—or presumed known—event on which one is conditioning an assessment of probability. Suppose that the event on which one is conditioning is the innocence of the person of interest. We might then say: assuming that this individual is innocent, what is the probability of trace evidence being found in his environment? There are 19,999 innocent people in the relevant sample, as depicted in Table 1's 'Innocence' column. The event, the outcome of which one is uncertain, is the presence or absence of the evidence. There are 199 people, known by assumption to be innocent, on whom the evidence is present. The probability of the event of interest (presence of the evidence on an innocent person) is therefore 199/19,999.

Now consider, again, the probability of guilt of a person on whom the evidence is found. The event on which one is conditioning is the discovery of the evidence, that is, we are now assuming that trace evidence has been found on a particular individual. We can see from Table 1's 'Present' row that there are 200 people on whom the evidence will be discovered. The event which is uncertain is the guilt or innocence of the suspect. There is one, and only one, person, known by assumption to have the evidence on him, who is guilty. The probability of the event of interest (guilt of a suspect with trace evidence in his environment) is then 1/200. The complementary probability, the probability of innocence, given discovery of the evidence, is 1 – 1/200, or 199/200.

The fallacy of equating (1) the probability of the presence of the evidence, assuming innocence, to (2) the probability of innocence, assuming presence of the evidence, is known to mathematicians as the fallacy of the transposed conditional. In legal circles, it is more popularly known as 'the prosecutor's fallacy'. The prosecutor's fallacy was denounced by the Court of Appeal in *Adams*, but the lesson has apparently not always been properly understood or taken to heart.[8]

The crucial distinction between different forms of conditioning is further illustrated by another memorable example. The statement 'If I am a monkey, then I have two arms and two legs' is true. However, the statement 'If I have two arms

8 A recent report produced by the Nuffield Council on Bioethics states that '[t]he prosecutor's fallacy has bedevilled the use of DNA evidence in courts': *The Forensic Use of Bioinformation: Ethical Issues* (Nuffield Council on Bioethics: 2007) para. 5.20. During the course of its investigations, the Nuffield Council made the startling discovery that one accredited forensic laboratory was routinely committing the prosecutor's fallacy in its written reports adduced in evidence in criminal trials. As the Council remarks (ibid. at paras. 5.30–5.32), if forensic scientists can make such errors, it is hardly surprising that lawyers, judges and—presumably—jurors are susceptible to making them as well.

and two legs, then I am a monkey' is not always true! There is a direct analogy to the interpretation of DNA profiles in a forensic context. The statement 'If I am guilty, then my DNA profile will match that of the DNA profile at the crime scene' (let us stipulate) is true. The statement 'If my DNA profile matches that of the crime scene DNA profile, then I am guilty' is not always true. This is not a question of innocent contamination or deliberate evidence tampering, etc., possibilities which can be set aside for present purposes. The point is that DNA evidence is inherently probabilistic, such that the simple fact of a match does not exclude even a strong likelihood of innocence—as Table 1 demonstrated.

2. The language of probability

The examples presented in the last section made use of terminology which statisticians routinely employ when discussing probability. Since this terminology is probably unfamiliar to most lawyers, but comes naturally to statisticians—not least when they are testifying in court as expert witnesses—it may be helpful for readers to have statistical jargon explained more systematically. In providing such an explanation, this section also simultaneously further explores basic concepts of probability and their forensic applications.

As a starting point we suppose, simplistically, that before any evidence is heard 'innocent until proven guilty' means that every person in the relevant population is equally likely to be guilty. If the relevant population were taken to be the population of the whole world it is fairly straightforward to think of evidence that will eliminate most of the people in the world from serious consideration as potential suspects.

(a) Proposition

A proposition is taken to be the hypothesis put forward by one side in adversarial trial proceedings. In criminal cases there are prosecution propositions and defence propositions. Examples of the prosecution proposition include 'the defendant is guilty', 'the defendant was at the scene of the crime', 'the DNA at the scene of the crime is that of the defendant'. The defence does not necessarily need to have a proposition, since a blanket denial of the charges puts the prosecution to proof. Where the defence chooses to argue an affirmative case, however, examples of the defence proposition could include 'the defendant is innocent', 'the defendant was not at the scene of the crime', 'the DNA at the scene of the crime is not that of the defendant'. These are all complements of the corresponding prosecution proposition, but the defence proposition is not the complement of the prosecution proposition in every case. Examples include 'my brother committed the crime' and 'I acted in self-defence'. If the defence does not put forward an explicit proposition, evaluation of the evidence may proceed on the assumption that the defence proposition is simply the complement of the prosecution proposition, whatever that happens to be.

(b) Relevant population

The relevant population is determined from the circumstances of the crime, and refers to the class of individuals to which the criminal, as yet unknown, can be said to belong. This population may be used to help determine—or, more strictly speaking, *estimate*—the probability of particular evidence, for example DNA frequencies.

The relevant population is partly defined by the defence proposition, as an example proposed by Robertson and Vignaux demonstrates. An English tourist was murdered in Hamilton, New Zealand, in 1992.[9] A man of Maori appearance was seen running away from the scene and subsequently washing himself in a nearby river. Blood which did not belong to the victim was found at the scene and analysed to produce a DNA profile. The frequency of DNA profiles is known to vary from race to race. A Maori male was subsequently arrested and identified by the eyewitness as the person seen running away from the scene. In this example, the prosecution proposition is, obviously, that this individual was the perpetrator of the murder.

The defence might have argued that the accused was indeed the person seen running away from the crime scene, but that he was not the murderer. If this proposition were accepted, and the murderer was not, in fact, the man seen running away from the scene, then it turns out that we have no information about the murderer. In particular, we have no information about the ethnicity of the murderer. The conditioning event for the assessment of the DNA profile in this case should therefore be that the murderer is a person of unknown ethnic origin. The probability of the DNA profile should be determined from consideration of the population of New Zealand as a whole (assuming for the sake of argument that it was a New Zealander who committed the crime).

Alternatively, the defence might argue that the accused was *not* the person seen running away from the scene, and that the eyewitness's identification of the defendant as that man was mistaken. Were this proposition accepted, then the accused is innocent but there is still reason to believe that the murderer was a man of Maori origin. The conditioning event for the assessment of the DNA profile in this case is that the murderer is a person of Maori origin. The probability of the DNA profile should be determined from consideration of the Maori population of

9 B. W. Robertson and G. A. Vignaux, *Interpreting Evidence* (Wiley: Chichester, 1995) 36–7.

New Zealand, not from the entire population containing Maoris and non-Maoris. The 'relevant population' is thus determined, not only by the nature of the charge and the evidence adduced to prove it—as reflected in the prosecution proposition—but also by the arguments and evidence advanced by the defence.

(c) Prior odds

The probability that a person chosen at random from the relevant population is guilty can be calculated, in the absence of any other information, by dividing 1 by the number of people in the relevant population. Thus, if all we know is that there are 1,000 individuals in the relevant population, of whom one and only one is guilty, the probability of any individual chosen at random being the guilty individual is 1/1000. This may be taken as a numerical representation of the belief that the person chosen at random is just as likely, and no more likely, to be guilty as anyone else similarly chosen at random from the relevant population. This is an obvious simplification of reality, but it nonetheless supplies a useful working assumption, sometimes designated the 'prior probability' (that is, prior to considering the impact of any other evidence).

The complement of the (prior) probability of guilt is the (prior) probability that a person chosen at random from the relevant population is innocent. This probability may be taken to be the number of people in the population minus 1, divided by the total number of people in the relevant population. Where there are 1,000 individuals in the relevant population, the prior probability of innocence is (1,000 − 1)/1,000 = 999/1,000. Referring back to Table 1, above, the prior probability of guilt for a person chosen at random from the relevant population would be 1/20,000. The prior probability of innocence for a person chosen at random from the relevant population would be (20,000 − 1)/20,000 = 19,999/20,000.

Now, the ratio of these two prior probabilities is 1 to 19,999 (i.e. 1:19,999), which can also be written as 1/19,999 (because [1/20,000] / [19,999/20,000] = 1/19,999). This is known as the 'prior odds' in favour of guilt. Notice that the prior odds are very small, much less than 1 but fractionally larger than 1/20,000. The reciprocal of the prior odds is 19,999 (because 19,999/1 = 19,999), which can be read as 19,999 to 1 against guilt. Odds of 1 (sometimes expressed as 50–50) are equivalent to a prior probability of 0.5 for each of the two relevant events (here, guilt and innocence) since 0.5/0.5 = 1. Odds greater than 1 arise when the prior probability of guilt is greater than the prior probability of innocence. Whenever, conversely, the prior probability of innocence is greater than the prior probability of guilt (which is the assumption at the start of all criminal trials), the prior odds will be some fraction (much) less than 1. As odds are ratios of two probabilities, they are never negative and only equal to zero when the numerator is equal to zero, i.e. when there are no individuals of interest in the relevant population. This is equivalent to saying, in the forensic context, that the perpetrator is not within the relevant population, such that the probability of guilt for any individual selected at random is zero. When the denominator equals zero (no chance of innocence) the corresponding odds are infinite (guilty by definition).

(d) Posterior odds

The probability of guilt for a person chosen at random from the population on whom the evidence has been found is taken to be 1 divided by the number of people in that population. The population of individuals on whom the evidence is found is always a subset of the relevant population used in the determination of the prior odds. In other words, the members of the population on whom the evidence has been found constitute a subset of the initial relevant population. This subset may be denoted the *posterior population*. By extension, a probability determined from this population is known as a *posterior probability*. It is determined after the evidence in question has been heard (that is, posterior to the consideration of that evidence).

It is expected that the size of the posterior population will be very much smaller than that of the original population used in the determination of prior odds. The complement of the posterior probability of guilt (for a person on whom the evidence has been found) is the probability that a person chosen at random from the population on whom the evidence has been found is nonetheless innocent. This complement is taken to be the number of people in the posterior population minus 1, divided by the number of people in the posterior population. In Table 1, above, the probability of guilt for a person chosen at random from the population on whom the evidence has been found would be 1/200. The probability that a person chosen at random from this population is innocent would then be (200 − 1)/200 = 199/200. The ratio of these two probabilities is 1:199, or 1/199. This number represents the *posterior odds* in favour of guilt. Conversely, its reciprocal is 199, which can be expressed in words as 'the posterior odds are 199 to 1 against guilt'. Notice that this analysis assumes that all 200 people on whom the evidence is found are equally likely to be guilty, without (yet) having taken into account any other evidence that might bear on the issues in the case. This is a very big assumption, which may or may not be warranted in the instant case. Reliance should only be placed upon it with appropriate circumspection.

(e) False negatives and false positives

Theoretical probabilities are axiomatic, true by definition. In the empirical world, however, allowance has to be made for errors of various kinds. Two particularly significant kinds of error, which we must mention in order to set aside, are known as 'false negatives' and 'false positives'. These two types of error can be illustrated by considering the results of comparisons between DNA crime-stain samples and comparison samples taken from known suspects.

In some circumstances a negative (non-match) result may be reported where the two samples have a common source, and hence should provide a positive (match) result. This is a 'false negative': the result reported is negative, but that result is false. A 'false positive' report, conversely, occurs when a positive (match) result is reported when the two samples in reality have different sources, and therefore should have been reported as a non-match. False negative and false positive reports can arise for a host of reasons (the details of which need not concern us here) including contamination of samples, laboratory testing error and misinterpretation of test results.

Analogous terms are employed in medical diagnosis. A false negative outcome of a test for the presence of a particular disease would be one in which the patient does, in fact, have the disease but the test appears to rule it out. A false positive outcome of a test for the presence of a particular disease would be one in which the patient is actually disease-free but the test indicates they do have the disease. The consequences of such errors are patently potentially very serious. Errors in medical diagnosis can lead to inappropriate treatment, or to vital treatment being withheld until it is too late to intervene successfully. In a forensic context, a false negative report may result in a guilty suspect being excluded from an ongoing investigation, whilst a false positive report could potentially implicate an innocent suspect and precipitate a serious miscarriage of justice.

(f) Probability of the evidence if the person is guilty (or is innocent)

For the purposes of undertaking probabilistic analyses for DNA profiles with a stain from a single source, it is conventional to assume that false negatives cannot occur, since the profile is a discrete entity rather than a measurement with which random error is associated. On this assumption, if an individual is guilty, then the evidence found in that individual's environment will be certain to match that found at the crime scene; for example, it is certain that the suspect's DNA will match the crime-stain sample. The probability of an event which is certain is 1. Thus, assuming no false negatives, the probability of the evidence (for example, a DNA match) if the person is guilty is 1. All our calculations here are premised on this assumption.

Referring back to the mining example summarised in Table 1, above, the assumption of no false negatives is reflected in the fact that trace evidence will definitely be found on the guilty suspect (depicted by '1' in the first cell of the 'present' row). There are also 19,999 innocent people in the relevant population of males in the town. Of these, 199 have evidence on them which is linked to the crime scene (as shown in the second cell of the 'present' row). Thus, the probability of finding the evidence in the environment of a person who is (nonetheless) innocent—the probability of the evidence assuming innocence—equals 199/19,999.

(g) Likelihood ratio

The ratio of the probability of the evidence if the person is guilty, divided by the probability of the evidence if the person is innocent, is known as the *likelihood ratio*. In the mining example, the likelihood ratio is calculated by dividing 1 by (199/19,999), which is 19,999/199. This in turn is approximately equal to 20,000/200, or—simply—100. A verbal interpretation of this result is that 'the evidence is 100 times more likely if the person is guilty than if he is innocent'.

We can once again invoke the mining example to illustrate this important general result, which in fact underpins the entire probabilistic approach to the evaluation of evidence:

Prior odds (of guilt, for a person chosen at random) = 1/19,999
Likelihood ratio (of the evidence) = 19,999/199
Posterior odds (of guilt, for a person chosen at random) = 1/199.

Note that:

$$1/199 = (19{,}999/199) \times (1/19{,}999)$$

or in words:

Posterior odds = Likelihood ratio × Prior odds.

This is a numerical verification of the general result known as Bayes' Theorem, after the 18th century Reverend Bayes who is credited with discovering it.[10] The

10 See e.g. Richard Lempert, 'Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes' (1991) 13 *Cardozo Law Review* 303. The now-celebrated 'Essay Toward Solving a Problem in the Doctrine of Chances' was discovered amongst the papers of the Reverend Thomas Bayes (1702–61), and published posthumously by the Royal Society in 1763: see David A. Schum, *The Evidential Foundations of Probabilistic Reasoning* (Wiley: New York, 1994) 47–52.

likelihood ratio is the factor which converts prior odds into posterior odds. More fully, the posterior odds in favour of guilt are equal to the product of (i) a ratio of the probability of the evidence if the suspect is guilty, to the probability of the evidence if the suspect is innocent; and (ii) the prior odds in favour of guilt.

Consider two propositions, the proposition put forward by the prosecution and the proposition put forward by the defence. In their most simple forms, these two propositions may be, respectively, that the suspect is guilty and that the suspect is innocent. There could just as easily be other pairs of propositions: the suspect was at the scene of the crime and the suspect was not at the scene of the crime; or the DNA of the crime-stain sample came from the suspect and the DNA of the crime-stain sample came from some other person (unrelated to the suspect), for example. The general result given by Bayes' Theorem may then be written as:

> the posterior odds in favour of the prosecution's proposition are equal to the product of (i) a ratio of the probability of the evidence if the prosecution's proposition is true, to the probability of the evidence if the defence proposition is true; and (ii) the prior odds in favour of the prosecution's proposition.

This result has several interesting implications, some of which have important forensic applications:

(i) A likelihood ratio greater than one means that the posterior odds are greater than the prior odds. Evidence for which the likelihood ratio is greater than one may be said to support the prosecution's proposition.

(ii) A likelihood ratio less than one means that the posterior odds are less than the prior odds. Evidence for which the likelihood ratio is less than one may be said to support the defence proposition.

(iii) A likelihood ratio equal to one means that the posterior odds are equal to the prior odds. Evidence for which the likelihood ratio equals one may be said to be *irrelevant*, both logically and legally, in that the evidence leaves the probability of the truth of either proposition exactly the same as it was before the evidence was taken into account. Evidence which is incapable of affecting the prior odds, either in favour of the prosecution (likelihood ratio greater than 1) or in favour of the defence (likelihood ratio smaller than 1), has no utility in adjudication. It cannot logically assist the fact-finder to arrive at a decision, because the probability of the accused's guilt or innocence is wholly unaffected by that evidence.

(iv) As we have already seen, a likelihood ratio is a ratio of two probabilities which takes a value between zero (when the probability of the evidence if the prosecution proposition is true equals zero, implying that its complement, the defence proposition, is certainly true) and infinity (when the probability of the evidence if the defence proposition is true equals zero, implying that the prosecution proposition is definitely true). Note that, whereas probabilities take values between 0 and 1, likelihood ratios take values between 0 and infinity. The likelihood ratio is sometimes taken to be a measure of support for the relevant proposition.

(v) There is symmetry about 1 in the values of the likelihood ratio. A value of, say, 1000 means that the posterior odds are greater than the prior odds by a factor of 1000. As a numerical illustration, prior odds of 1/10, for example, would be converted to posterior odds of 1000/10, that is, 100. A value of 1/1000, by contrast, means that the posterior odds are smaller than the prior odds by a factor of 1000: thus, prior odds of 10 would be converted to posterior odds of 10/1000, that is, 1/100.

(vi) Evaluation of a relative frequency does not require a study of the entire population, of which the relevant population is a subset. A sample from the entire population is sufficient. Thus, frequencies derived from forensic databases of fingerprints, shoe-prints, glass, and DNA samples, etc., can be used for probabilistic inference. It is not necessary to collect samples from every conceivable source or donor.

(vii) In criminal adjudication, the values of the prior odds and the posterior odds are matters for the judge and jury, in accordance with the normal division of labour in forensic fact-finding. The value of the likelihood ratio, however, is a matter for the forensic scientist or other expert witness, as it is an assessment of the objective probative value of their evidence. Assessments of prior and posterior odds require subjective opinions which are the responsibility of the fact-finders. The scientist does not need to know values for either the prior or the posterior odds. The likelihood ratio, or a range of such ratios, can be calculated on the basis of the assumed truth of the propositions put forward by the prosecution and defence.

(viii) If a value of zero is assigned to the prior probability either of the truth of the prosecution's proposition or the truth of the defence proposition, the corresponding posterior probability will also necessarily be zero, regardless of the value of the likelihood ratio. This is a logical consequence of the arithmetic result that the product of zero and any number is zero. It follows that any potential juror who believed that 'innocent until proven guilty' equates to a probability of zero for the truth of the prosecution's proposition should be barred from jury service, because such a person is also logically committed to finding a posterior probability of guilt to be zero, *regardless of the strength of the evidence against the defendant*. An alternative interpretation of the dictum 'innocent until proven guilty' would be to assume that the accused is 'just as likely to be guilty as anyone else'.[11] This interpretation has been challenged on the basis that it is not normally realistic to assume that the accused is just as likely to be guilty, no more and no less, than any other person in the relevant population, which could conceivably be the population of the entire world. It must be understood, however, that what is being advocated is a default interpretation to be adopted prior to the consideration of any detailed information (i.e. evidence) actually bearing on the case. Once evidence is led, for example in relation to the location of the crime, the vast majority of theoretical suspects will be eliminated from any further consideration, and most of those still remaining will have probabilities of guilt considerably lower than the accused (whom the prosecution must, presumably, be able to place at the scene, otherwise the case against him would never have been brought to trial).

3. Presenting statistics in court

It is implicit in the foregoing discussion that uncertainty can be described in different ways, for example as relative frequencies, likelihood ratios or posterior odds. This is one way of characterising a central bone of contention in *Adams*,[12] where the Court of Appeal seemed drawn to whatever mode of presentation would best steer clear of mathematical formulae and jargon likely to bemuse jurors. Moreover, researchers have shown that experimental subjects tend to assign different degrees of weight to pieces of evidence with identical probabilities depending merely upon the form of its presentation.[13] Some verbal formulations are prone to make evidence *appear* stronger, whilst other formulations make that same evidence *appear* weaker in the eyes of laypeople—those same laypeople who

11 For the rejection of a third alternative, see Richard D. Friedman, 'A Presumption of Innocence, Not of Even Odds' (2000) 52 *Stanford Law Review* 873.

12 Also see Mike Redmayne, 'Presenting Probabilities in Court: The DNA Experience' (1997) 1 E & P 187; and Ian W. Evett, Lindsey A. Foreman, Graham Jackson and James A. Lambert, 'DNA Profiling: A Discussion of Issues Relating to the Reporting of Very Small Match Probabilities' [2000] Crim LR 341.

13 F. Taroni and C. G. G. Aitken, 'Probabilistic Reasoning in the Law' (1998) 38 *Science & Justice* 165–77 and 179–88; Jonathan L. Koehler, 'The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient' (2001) 74 *Southern California Law Review* 1275.

sit on juries and return verdicts in criminal trials. Choice of mode of presentation for statistical evidence is therefore neither a purely theoretical matter nor a trivial issue that can be casually brushed aside.

This section examines the relative merits of five competing approaches to presenting statistical evidence to jurors in criminal trials: (a) profile frequency; (b) percentage exclusion; (c) numerical conversion; (d) likelihood ratio; and (e) posterior odds. The analysis is informed by our own empirical research, in which Scottish law students, forensic science students, practising lawyers and forensic scientists were exposed to various methods of presenting DNA profile evidence.[14]

(a) Profile frequency

The profile frequency is the frequency of the characteristic of interest in the relevant population. For example, the profile frequency of blood evidence found at a crime scene and associated with the perpetrator might be 10%, meaning that 10% of the relevant population shares that particular blood group. These calculations assume that individuals in the relevant population are not genetically related; an assumption which could of course be rebutted in practice, i.e. where the relevant population includes blood relatives.

(b) Percentage exclusion

The percentage exclusion is the proportion of the population excluded from the relevant population by particular evidence, for example because their blood group does not match that of the crime-scene sample. The percentage exclusion is the complement of the profile frequency, which can also be described as the proportion of the population *not* excluded from the relevant population by the evidence in question. If the profile frequency of blood analysis evidence is 10%, the percentage exclusion is 90%. If the percentage exclusion is known to be 60%, then the profile frequency must be 40%.

The percentage exclusion formulation is related to the probability that two people share the same profile. It is sometimes known as the random man not excluded (RMNE) mode of analysis.[15] In *Harrison* v *Indiana*,[16] for example, the frequency of the genetic characteristic in the relevant population was 7.4%. The prosecution expert testified that although her test had been able to exclude all but 7.4% of the

14 Taroni and Aitken, above n. 13.

15 See P. Gill, C. Brenner, J. S. Buckleton, A. Carracedo, M. Krawczake, W. R. Mayr, N. Morling, M. Prinz, P. M. Schneider and B. S. Weir, 'DNA Commission of the International Society of Forensic Genetics (ISFG): Recommendations on the Interpretation of Mixtures' (2006) 160 *Forensic Science International* 90–101.

16 *Harrison* v *State*, 644 NE 2d 1243 (1995), Supreme Court of Indiana.

relevant population (13,000 white males living locally) as the source of the evidential specimen, the defendant had not been excluded. In other words, the percentage exclusion for the genetic test in *Harrison* v *Indiana* was (100 − 7.4)% = 92.6%.

For a suspect who has not been excluded, the weight of the evidence against him grows in proportion to the size of the percentage exclusion. Evidence with a percentage exclusion of 99% is plainly more probative than evidence with a percentage exclusion of 75% for the same relevant population, since in the first case the accused is amongst only 1% of individuals with a matching profile, as opposed to 25%—1 in 4—in the second scenario. In *Harrison*, where the relevant population numbered 13,000 white males and the evidential specimen could only have come from one of the 7.4% of that population not excluded by the expert's test, the accused was one of 962 people (7.4% of 13,000) who may have committed the crime (or, more correctly, who might have been the source of the specimen). If the percentage exclusion had been 99%, the field of eligible suspects would have been further narrowed down to only 130 individuals. Conversely, a profile frequency of 25%—corresponding to a percentage exclusion of 75%—would have left 3,250 eligible suspects.

As these examples imply, RMNE gives a broad indication of the probative value of particular evidence, but limited assistance with drawing inferences and arriving at conclusions in the specific case under examination. On the facts of *Harrison*, for example, the statistical evidence still left the fact-finder with 961 other potential candidates as the perpetrator in addition to the accused. But the fact-finder needs to determine whether or not *this* particular accused currently standing trial is guilty, not only whether he is a member of a bigger or smaller class of potentially guilty individuals. RMNE does not offer a balanced or comprehensive assessment of the evidence, because it is focused on the denominator of the equation, specifying the classes of individuals excluded or not excluded by a particular test. It does not pay sufficient attention to the numerator, relating to the individual accused. A more balanced assessment can only be achieved using the likelihood ratio, as we explain below.

Harrison foreshadowed guidelines for presenting DNA evidence at trial laid down by the English Court of Appeal in *Doheny and Adams*.[17] The court suggested that 'provided that the expert has the necessary data, it may ... be appropriate for him to indicate how many people with the matching characteristics are likely to be found in the United Kingdom or a more limited relevant sub-group, for instance,

17 *R* v *Doheny and Adams* [1997] 1 Cr App R 369 at 375.

the Caucasian, sexually active males in the Manchester area'. Dubbing this profile frequency the 'random occurrence ratio', the Court of Appeal endorsed the following direction as a model for trial judges to adopt, appropriately tailored to the instant facts, when directing juries on the interpretation of DNA evidence:

> Members of the jury, if you accept the scientific evidence called by the Crown this indicates that there are probably only four or five white males in the United Kingdom from whom that semen stain could have come. The defendant is one of them. If that is the position, the decision you have to reach, on all the evidence, is whether you are sure that it was the defendant who left that stain or whether it was possible that it was one of that other small group of men who share the same DNA characteristics.

Compared to the expert's conclusion in *Harrison* that the accused was one of 962 men from whom the crime sample might have come, a direction informing the jury that the accused is one of 'only four or five white males in the United Kingdom' who could be the crime-stain donor certainly appears more illuminating. However, DNA technology has developed apace since the Court of Appeal's pronouncement in *Doheny*, such that the figures associated with a matching DNA profile are now so small that the *Doheny* direction makes little sense. For example, with a 'random occurrence ratio' (probability of a random match) calculated at 1 in a billion, the jury would have to be told that the accused is one of no more than six people *in the world* with a matching profile.

(c) Numerical conversion

A third approach to organising and presenting statistical information is to calculate the expected number of people that would have to be examined before another person with a profile matching the crime stain and the accused is found. In *Ross* v *State*,[18] for example, the frequency of the genetic characteristic in the relevant population was 1 in 209,100,000. As an explanation of this number, the expert said he had a database of blood samples from all over the country and he posed the rhetorical question: 'How many people would we have to look at before we saw another person like this?' The answer he gave to the jury was 209,100,000. However, this testimony was in fact an example of 'numerical conversion error', which we have discussed more fully elsewhere.[19]

18 *Ross* v *State*, 1992 WL 23575 (Tex. App.-Hous. (14 Dist.) 13 February 1992), also discussed in Taroni and Aitken, above n. 13.

19 Aitken and Taroni, above n. 7 at 83–5.

The question may be rephrased in terms which are more explicitly probabilistic. Given a profile frequency of 1 in 209,100,000, how many people would need to be tested such that the probability of finding a matching profile is at least 0.5? The answer is 145 million. The expert's answer of 209,100,000 overstates the value of the evidence (by increasing the size of population required to find another match) in favour of the prosecution by a factor of over 25%, and to that extent was prejudicial to the defence. If a total of 209,100,000 people were tested, the probability of finding at least one other match is actually 0.632, easily better than a 50–50 bet. Match profiles greater than 0.5 might be specified, to model more closely the criminal standard of proof beyond reasonable doubt. If a probability of 0.9 were specified instead, for example, one would expect to have to test around 480 million individuals, more than double the number specified by the expert in *Ross*, before encountering a match. The expert's 209 million figure thus understates the value of a match.

In summary, the expert's conclusion is necessarily problematic, owing to the inherent uncertainty in the calculations and expressions of probability. With uncertain events, and a finite set of possible outcomes, it is impossible to state with certainty how many occurrences would have to be observed before the outcome of interest, for example a DNA match, would be seen. An example more familiar to UK readers makes this point explicit. The probability of winning the jackpot in the UK lottery is approximately 1 in 14 million. However, someone who plays the lottery regularly is not certain to hit the jackpot only on the 14 millionth play, and not before. A punter might never win the lottery in many more than 14 million attempts. Alternatively, he could win on the very next draw—which is, presumably, precisely the allure of lotteries for punters. A win on the next draw is an outcome that has a probability of 1 in 14 million. However, the probability that *at least one person* will win on the next draw if there are 14 million players is roughly 0.632. For populations larger than the reciprocal of the profile frequency, the probability that at least one other person has the same profile is greater than 0.632 and the probability increases as the population size increases. Thus for a frequency of 1 in 1,000 and a population size of 2,000 the probability is 0.865; for a population of size 5,000, the probability is 0.993. Beyond such discrete probabilistic statements, nothing more can be said in terms of predicting concrete (empirical) outcomes.

This type of analysis may inform further criticisms of the direction endorsed by the Court of Appeal in *Doheny*. Consider a hypothetical case in which there is a DNA match probability of 1 in 1,000 and a relevant population of 1,000. A *Doheny* ruling would be to the effect that 'there is probably only one person in the population from which the stain could have come', which appears to be pretty damning evidence against the accused.[20] However, the posterior odds of guilt in this case are actually approximately 1:1 (evens, or 50:50),[21] equating to a posterior probability of only 0.5 which nobody would regard as proof 'beyond reasonable doubt'.

It is difficult to interpret the importance of the value attached to the probability that there is at least one other individual in the relevant population with a matching profile. This probability does not contribute directly to an evaluation of the evidence. What is required, but not supplied by numerical conversions of this probability, is an explicit statement about the relationship between the odds in favour of guilt before the presentation of the evidence and the odds in favour of guilt after the presentation of that evidence. This is the role of the likelihood ratio in Bayes' Theorem.

(d) Likelihood ratio

The likelihood ratio was defined and explained in Section 2(g). It is the ratio of the probability of the evidence assuming guilt divided by the probability of the evidence assuming innocence. In many cases the numerator of the likelihood ratio equals 1, assuming no false negatives. However, in some important cases this is not so. Examples include bodily fluid stains consisting of a mixture of contributors or cases involving measurements such as the elemental compositions of glass fragments or the chemical compositions of drugs. The likelihood ratio allows prior odds to be updated into posterior odds in accordance with Bayes' Theorem. Trial judges have occasionally emphasised the likelihood ratio in directing juries. In the New Zealand High Court case of *Pengelly*, for example, the forensic scientist remarked:

> In the analysis of the results I carried out I considered two alternatives, either that the blood samples originated from Pengelly or that the ... blood was from another individual. I find that the results I obtained were at least 12,450 times more likely to have occurred if the blood had originated from Pengelly than if it had originated from someone else.[22]

20 Also see Mike Redmayne, *Expert Evidence and Criminal Justice* (Oxford University Press: Oxford, 2001).

21 A match probability of 1 in 1,000 gives a likelihood ratio of [1/(1/1000)] = 1000. When multiplied by the prior odds of 1 in 999 ((1/1000)/(999/1000) = 1/999) we obtain posterior odds of 1/999 × 1000 which is approximately 1. This is because the calculation uses a match probability of 1 in 1,000 with a population of 999 innocent people among whom there is an expectation that one has the incriminating profile, together with one guilty person who also has the profile.

22 *R* v *Pengelly* [1992] 1 NZLR 545, NZCA, quoted in Robertson and Vignaux, above n. 9 at 23–4.

Notice the conditioning on particular propositions. In the *Pengelly* direction, the crucial phrases are 'the blood samples originated from Pengelly' and 'from another individual'. The relevant likelihoods are the likelihoods of obtaining such results under the assumptions of these propositions. In this instance, the forensic scientist accurately characterised the strength of the evidence by reference to its likelihood ratio. Unfortunately, courts and even experts themselves are not always so commendably careful. In the English case of *Gordon*, for example, the scientist summarised their evidence as:

> there was a visual match between the critical samples and the appellant's sample which showed a likelihood that the appellant was the rapist in each case.[23]

What the expert in *Gordon* should have referred to is the likelihood of *the evidence* if the accused were the donor of the sample, compared to the likelihood of *the evidence* if somebody else donated it, their ratio being the likelihood ratio. By (apparently) commenting on the likelihood of a conditioning *proposition*—the prosecution's proposition of guilt—the expert in *Gordon* trespassed on the terrain of the fact-finder and consequently perpetrated the prosecutor's fallacy. If the Court of Appeal's summary of the evidence is accurate, it would appear that this error was committed by the *defence* expert as well as by experts instructed by the prosecution.

Even if the likelihood ratio is properly presented by expert witnesses and faithfully summarised by the trial judge, it does not necessarily follow that juries will interpret it correctly. Indeed, in our survey research respondents were notably nonplussed by the use of the likelihood ratio in a summary similar to that in *Pengelly*, commenting that the likelihood ratio was 'too difficult to understand' and 'very confusing', or even that they had 'no idea what it means'.[24] It is worth recalling that our respondents were all people associated with legal process, as students or practitioners of forensic science or law. If likelihood ratios defeat this relatively well-informed audience, it may well be unrealistic to imagine that lay jurors could grapple with them successfully (though the Nuffield Council's report still advocates the education of jurors).

(e) Posterior odds

We saw in the last section that posterior odds can be calculated by multiplying prior odds by the likelihood ratio. Rather than presenting jurors with likelihood

23 R v *Gordon* [1995] 1 Cr App R 290 at 295, CA.
24 Taroni and Aitken, above n. 13.

ratios, which they might not be able to interpret properly, a fifth presentational alternative is for experts to cite posterior odds. This could be done on an explicitly conditional basis, to avoid any impression that the expert is attempting to usurp the role of the fact-finder in calculating prior odds for the other evidence in the case. For example, an expert might testify:

> If your prior odds in favour of guilt were *x* before you heard this evidence, then your posterior odds in favour of guilt, after hearing the evidence, would be the product of *V* and *x* (where *V* is the value of the evidence).

For a value *V* of 1 million, for example, the expert might explain to the jury:

> If your prior odds were 1000:1 against guilt, then your posterior odds, having heard my evidence, would be 1000:1 in favour of guilt.

It would be possible to express posterior odds in terms of a verbal scale rather than in raw numbers.[25] In *Bilal*[26] a handwriting expert, testifying to the similarity between the appellant's script and the writing on stolen cheques, proposed a 10-point scale, with level-1 representing a 'conclusive' match and level-10 'did not write' (conclusive non-match). In the instant case, the degree of similarity between the appellant's handwriting and the questioned documents (cheques) was assessed at level-4, defined as 'distinct possibility/could well have been written by'. It is plausible to suggest that verbal formulations are more easily grasped by jurors (and doubtless by lawyers and judges, too) than numerical statistics or mathematical probabilities. Whether such intuitive appeal translates into an accurate grasp of the probative value of evidence is another question, however. In *Bilal*, the expert further characterised her evidence as demonstrating that it was 'more likely than not' that the appellant was the author of the writing on the stolen cheques. One is left to ponder whether this level of confidence is better expressed as 'distinct possibility' or 'level-4 on a 10-point scale', or in some other way, such as a likelihood ratio or profile frequency.

Some may suggest that it is better for the jury to hear multiple alternative formulations of statistical evidence, with their accompanying explanations and divergent meanings. However, we believe that such multiple formulations will serve only to confuse lay fact-finders. A single statement of a likelihood ratio,

25 As proposed by I. W. Evett, 'Towards a Uniform Framework for Reporting Opinions in Forensic Science Casework' (1998) 38 *Science and Justice* 198–202.
26 *R* v *Bilal* [2005] EWCA Crim 1555.

accompanied by general explanatory information as recommended by the Nuffield report, will communicate the true value of the evidence more effectively.

4. Conclusions

The manner in which statistical evidence is presented can influence an audience's assessment of its probative value. In our survey research, students of forensic medicine and forensic science who received evidence in the form of percentage exclusions and relative frequencies gave higher posterior assessments of the proposition that the suspect committed the crime, and were more willing to find the suspect guilty, than those who received the evidence in the form of a likelihood ratio or as posterior odds. However, all survey respondents were conservative in their assessments of probabilities. In all samples analysed, the respondents' posterior probabilities were lower than those calculated using Bayes' Theorem. This suggests that even forensic professionals consistently *under*-estimate the true probative value of probabilistic evidence.[27]

The Bayesian framework set out in this article is a powerful tool for interpreting evidence, not least for forensic scientists and other experts who write reports submitted to court and sometimes testify in person at trial. Bayes' Theorem forces one to think carefully about the events to which probabilities need to be assigned, and to be explicit about the propositions on which any statement of probability is conditioned. It also requires the forensic practitioner to view physical evidence from opposing viewpoints, since, in order to calculate the likelihood ratio, it is necessary to consider both (i) the probability of the evidence, *assuming that the accused is guilty*; and (ii) the probability of the evidence, *assuming that the accused is innocent*. It is always salutary to recall, and to articulate expressly, that this probability calculus necessarily anticipates adventitious matches—false positives, where the crime-sample matches the profile of an innocent suspect—albeit that the probability of a random match may be very small for certain kinds of forensic technology, the most prominent being DNA profiling. The difficulty of assigning numerical values to the uncertainties surrounding the outcomes of certain events may be perceived as a practical disadvantage of this approach. However, even the somewhat arbitrary assignment of quantitative determinations to certain subjective beliefs (for example the guilt or innocence of a defendant) may facilitate illuminating comparative assessments of probative value.

The recent report by the Nuffield Council on Bioethics concluded that:

27 A finding replicated by D. A. Nance and S. B. Morris, 'Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability' (2005) 34 *Journal of Legal Studies* 395–444.

In view of the difficulties with the presentation of complex statistical information in the courtroom, we recommend:

- that professionals (including judges) working within the criminal justice system should acquire a minimum standard of understanding of statistics, particularly with regard to DNA evidence;
- that trial judges ensure statistical evidence is accurately presented during trials, and that the decision in the *R.* v. *Doheny and (Gary) Adams* (1997) 1 Cr App R 369 judgment regarding the correct presentation of DNA evidence is adhered to; and
- that in all cases where bioinformation evidence is adduced, introductory information should be made available to jury members, to ensure some basic understanding of the capabilities, and also the limitations, of such evidence.[28]

Although we have argued in this article that the *Doheny* direction is no longer viable for DNA evidence with very small match probabilities, we welcome the general thrust of the Nuffield Council's recommendations. Basic training in statistical method is essential for criminal justice professionals, who increasingly encounter statistics in their daily working lives. Fact-finders must also grasp the rudiments of statistical interpretation if their decisions are to be securely rooted in logic and evidence. Contrary to some of the pronouncements of the Court of Appeal in *Adams*, common sense is not always a reliable guide to probative value; many probabilistic results are counter-intuitive. Bayes' Theorem is superior to common sense in this regard.[29]

Debate about the role of statistics and probabilistic reasoning in the courts rumbles on. The extent of the confusion still routinely engendered by forensic statistics is encapsulated in the well-known case of Sally Clark. At Clark's trial for murdering her two infant sons, an expert paediatrician testified that the probability of two cases of sudden infant death syndrome (SIDS) occurring in the same family was 1 in 73 million. This figure was arrived at by squaring the relative frequency of a single SIDS case, calculated at 1 in 8,543 live births for a middle-class family like the Clarks. With around 650,000 live births per annum in

28 *The Forensic Use of Bioinformation: Ethical Issues* (Nuffield Council on Bioethics: 2007) para. 534.

29 In the recent case of Keran Henderson, a childminder was convicted of shaking a baby to death after the jury heard evidence from a dozen medical and forensic experts. One juror reportedly commented: 'Ultimately, the case was decided by laymen and laywomen using that despicable enemy of correct and logical thinking, that wonderfully persuasive device, common sense': *The Times*, 19 December 2007.

England and Wales, an instance of double-SIDS could be expected to occur once in about 100 years (100 × 650,000 = 65 million births in 100 years).

At Clark's first appeal against conviction,[30] written evidence was considered by the court on the relative frequency of infant homicides. This evidence established that there are between 20 and 30 infant homicides a year in England and Wales. Put together with a figure of 650,000 live births annually, these figures suggest upper and lower estimates of the probability that an infant will be murdered of roughly 1 in 21,700 and 1 in 32,500. The overall rate of SIDS (for all socio-economic groups) was roughly 1 in 1,300 in the same period. Dividing that SIDS rate by the infant murder rates produces ratios of approximately 17 (based on the lower boundary-estimate of 20 infant homicides per year) and 25 (based on the upper estimate of 30 infant homicides). In other words, it can be said that an infant is somewhere between 17 and 25 times more likely to be a SIDS victim than a homicide victim. An exact analogue with the figure of 1/8,543 for families like the Clarks is difficult to determine since infant homicide rates for families like the Clarks are difficult to determine. However, it is a reasonable assumption that the ratio in favour of homicides will still be greater than 1.

On appeal, it was argued for Clark that the jury might have taken a different view of the evidence if they had been presented with this statistical comparison directly. Instead:

> the prosecution invited the jury to adopt the figure of 73 million as having a significance in itself when, without reference to the likelihood of a competing possibility, the figure has no significance or relevance.[31]

Dismissing Clark's appeal on all grounds, the Court of Appeal specifically found no merit in the statistical argument being advanced. Observing that the 'competing possibility' in question was 'double infant murder by a mother', the court continued:

> That may be capable of being expressed in terms of a statistical probability, but legally speaking the exercise is not realistic ... [I]t is not an exercise the courts would perform.[32]

30 *R* v *Clark*, 2 October 2000, CA (LEXIS Transcript).
31 Quoted from the appellant's skeleton argument, ibid. at [168].
32 Ibid. at [176].

No explanation was given as to why the exercise was not 'legally' realistic. The court referred with approval to *Adams (No. 2)*, where it was said, emphatically, that 'expert evidence should not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial'.[33] The Court of Appeal's final conclusion in *Clark (No. 1)* was unequivocal:

> If there had been no error in relation to statistics at the trial, we are satisfied that the jury would have convicted on each count. In the context of the trial as a whole, the point on statistics was of minimal significance and there is no possibility of the jury having been misled so as to reach verdicts they might not otherwise have reached.[34]

Despite the conclusory appearance of this determination, this was not the end of the legal drama. At the second attempt, Sally Clark's appeal was allowed, her convictions were quashed and she was released from prison.[35] The Court of Appeal's change of heart was mainly inspired by fresh medical evidence. This fresh evidence cast doubt on the original post-mortem reports which had been presented to the trial jury and which had indicated foul play. The court nonetheless went out of its way to comment on the statistical evidence which the jury had heard, and was particularly critical of the figure of 1 in 73 million invoked by Professor Sir Roy Meadow:

> If there had been a challenge to the admissibility of the [statistical] evidence we would have thought that the wisest course would have been to exclude it altogether. ... Quite what impact all this evidence will have had on the jury will never be known but we rather suspect that with the graphic reference by Professor Meadow to the chances of backing long odds winners of the Grand National year after year it may have had a major effect on their thinking notwithstanding the efforts of the trial judge to down play it. ... Thus it seems likely that if this matter had been fully argued before us we would, in all probability, have considered that the statistical evidence provided quite a distinct basis upon which the appeal had to be allowed.[36]

So, in the space of two appeals in the same case, statistical evidence which was of such 'minimal significance' that there was 'no possibility of the jury having been misled so as to reach verdicts they might not otherwise have reached' was

33 *R* v *Adams (No. 2)* [1998] 1 Cr App R 377 at 385, CA.
34 *R* v *Clark*, 2 October 2000, CA, at [272].
35 *R* v *Clark* [2003] EWCA Crim 1020.
36 Ibid. paras. 177–178, 180.

transformed into 'a distinct basis upon which the appeal had to be allowed'! It is a notable feature of this litigation history that the expert witnesses involved in the Sally Clark case proved no more adept in their handling of statistics than the lawyers (defence as well as prosecution), judges or—one must presume—the jurors who grappled, evidently with limited success, with the evidence they were called upon to interpret.

A broad education in forensic statistical method cannot be acquired overnight, or imparted through a single journal article. But it is necessary to start somewhere, and the broad dissemination of such learning is urgently required. Conceived as a primer for legal professionals, this article has reviewed basic statistical terminology and its forensic applications and explored the options for presenting statistical information to fact-finders effectively. Some problems remain unresolved. Although Bayes' Theorem and likelihood ratios are in principle our preferred methods for presenting probabilities in court, it remains to be seen whether information presented in this form can be interpreted correctly by lay fact-finders, or by the judges who direct them on the evidence. Meanwhile, in raising awareness of the issues and by encouraging improved comprehension of probability and statistics amongst legal and forensic science professionals, we hope that this article may contribute directly to the administration of justice by promoting more successful applications of forensic statistics in legal adjudication.[37]

37 The Royal Statistical Society has established a Working Group on Statistics and the Law, under the chairmanship of the first author, to improve the understanding and use of statistics in the administration of justice. The group provides an interface for the Society with the legal, forensic scientific and justice communities. It is working with forensic scientists, barristers, advocates and members of the judiciary to develop educational programmes on the role of statistics and probabilistic reasoning in the law and forensic science. It welcomes comments on any of these matters at any time.