J Forensic Sci, Jan. 2003, Vol. 48, No. 1
Paper ID JFS2001171_481
Available online at: www.astm.org

William C. Thompson,[1] *J.D., Ph.D.; Franco Taroni,*[2,3] *Ph.D.; and Colin G. G. Aitken,*[4] *Ph.D.*

How the Probability of a False Positive Affects the Value of DNA Evidence

ABSTRACT: Errors in sample handling or test interpretation may cause false positives in forensic DNA testing. This article uses a Bayesian model to show how the potential for a false positive affects the evidentiary value of DNA evidence and the sufficiency of DNA evidence to meet traditional legal standards for conviction. The Bayesian analysis is contrasted with the "false positive fallacy," an intuitively appealing but erroneous alternative interpretation. The findings show the importance of having accurate information about both the random match probability and the false positive probability when evaluating DNA evidence. It is argued that ignoring or underestimating the potential for a false positive can lead to serious errors of interpretation, particularly when the suspect is identified through a "DNA dragnet" or database search, and that ignorance of the true rate of error creates an important element of uncertainty about the value of DNA evidence.

KEYWORDS: forensic science, DNA typing, statistics, Bayes theorem, likelihood ratio, error rate, false positive, proficiency testing, prosecutor's fallacy, database, DNA dragnet

When evaluating the strength of DNA evidence for proving that two samples have a common source, one must consider two factors. One factor is the probability of a coincidental match (sometimes called the random match probability). A coincidental match occurs when two different people have the same DNA profile. The second factor is the probability of a false positive. A false positive (as we use that term here) occurs when a laboratory erroneously reports a DNA match between two samples that actually have different profiles. A false positive might occur due to error in the collection or handling of samples, misinterpretation of test results, or incorrect reporting of test results (1–3). Either a coincidental match or a false positive could cause a laboratory to report a DNA match between samples from different people. Consequently, one must consider both the random match probability and the false positive probability in order to make a fair evaluation of DNA evidence.

Although both factors affect the value of a reported match, forensic scientists and courts have been far more concerned about having a solid scientific basis for determining random match probabilities than for determining false positive probabilities. Efforts to

[1]Department of Criminology, Law & Society, University of California, Irvine, CA.
[2]Institut de Police Scientifique et de Criminologie, Law Faculty, University of Lausanne, B.C.H., 1015 Lausanne-Dorigny, Switzerland.
[3]Institut de Médecine Légale, Faculty of Medicine, University of Lausanne, 21 Rue du Bugnon, 1005 Lausanne, Switzerland.
[4]Department of Mathematics and Statistics, The University of Edinburgh, King's Buildings, Mayfield Road, EH93JZ Edinburgh, Scotland.
Received 22 May 2001; and in revised form 14 Nov. 2001; accepted 19 July 2002; published XXXX.

establish rates of laboratory error through empirical study have, to date, received relatively little attention compared to efforts to establish the frequency (and hence the random match probability) of DNA profiles (4). When DNA evidence is presented in court, juries typically receive statistical data on the probability of a coincidental match (5,6). For example, a jury might be told "that the probability of selecting an unrelated individual at random from the population having a DNA profile matching [the defendant's] [is] approximately 1 in 351,200 blacks and approximately 1 in 572,000 Caucasians" (7). But juries rarely hear statistics on the frequency or probability of false positives (5,6).

Courts in many jurisdictions refuse even to admit evidence of a DNA match unless it is accompanied by statistical estimates of the random match probability, and they require that these statistics be computed in a manner that is valid and generally accepted by the scientific community (6). By contrast, no court has rejected DNA evidence for lack of valid, scientifically accepted data on the probability of a false positive (5,6). It is considered essential to know, with a high degree of scientific certainty, whether the frequency of random matches is 1 in 1,000, 1 in 10,000, or one in one million, but unnecessary to have comparable estimates on the frequency of false positives.

Why are the two possible sources of error in DNA testing treated so differently? In particular, why is it considered essential to have valid, scientifically accepted estimates of the random match probability but not essential to have valid, scientifically accepted estimates of the false positive probability?

In this article we will consider several possible explanations for the difference. We will argue that it arises, in part, from failure to appreciate the importance of the false positive probability for determining the value of DNA evidence. We will present a framework for considering the role that error may play in determining the probative value of forensic DNA evidence. We will show that even a small false positive probability can, in some circumstances, be highly significant, and therefore that having accurate estimates the false positive probabilities can be crucial for assessing the value of DNA evidence.

Errors Happen

When DNA evidence was first introduced, a number of experts testified that false positives are impossible in DNA testing (6,8). This claim is now broadly recognized as wrong in principle (1,9–12), and it has repeatedly proven wrong in practice (3,13,14). But it has been repeated frequently, without skepticism, in appellate court opinions (6,8).

Why did experts offer this questionable testimony? One commentator has suggested that avid proponents of DNA evidence sought to allay judicial concerns about the potential for error by engaging in "a sinister semantic game" (8). They were able to deny that a DNA test could produce an error by excluding consideration of human error in administering or interpreting the test. Sinister or not, it is misleading to exclude considerations of human error in DNA testing when humans are necessarily involved in the administration and interpretation of DNA tests. For those who must evaluate DNA evidence, it makes little difference what causes a false match; what matters is how often false matches might be expected (9,15).

False positives have occurred in proficiency tests (2,3,11,13,16) and in actual cases (14,17). For example, the Philadelphia City Crime Laboratory recently admitted that it had accidentally switched the reference samples of the defendant and victim in a

rape case. The error led the laboratory to issue a report that mistakenly stated that the defendant was a potential contributor of what the analysts took to be "seminal stains" on the victim's clothing (18). The report also stated that the defendant's profile was "included" in a mixed sample taken from vaginal swabs. After the sample switch came to light, the laboratory reassessed the evidence and concluded that the "seminal stains" were actually bloodstains that matched the victim's DNA profile and that the defendant was excluded as a potential contributor to the vaginal sample (19).

In 1995, Cellmark Diagnostics admitted that a similar sample-switch error had caused it to report, incorrectly, that a rape defendant's DNA profile matched DNA found in vaginal aspirate from a rape victim. After the error came to light during the defendant's trial, Cellmark issued a revised report that stated that the vaginal sample matched the victim's own DNA profile and that the defendant was excluded as a potential donor (20).

False positives can also arise due to misinterpretation of test results. One such error led to the false conviction of Timothy Durham (14,17). In 1993, a Tulsa, Oklahoma jury convicted Durham of the rape of an eleven-year-old girl. He was sentenced to 3000 years in prison. The prosecution presented three pieces of evidence against him: the young victim's eyewitness identification, testimony that Durham's hair was similar (in microscopic examination) to hair found at the crime scene, and a DNA test (DQ-alpha) that reportedly showed that Durham's genotype matched that of the semen donor. Durham presented eleven witnesses who placed him in another state at the time of the crime, but the jury rejected his alibi defense. Fortunately for Durham, post-conviction DNA testing showed that he did not share the DQ-alpha genotype found in the semen. He was also excluded at several other genetic loci in multiple tests. The initial DNA test result that helped convict Durham was proven to have been a false positive. The error arose from misinterpretation. The laboratory had failed to completely separate male from female DNA during differential extraction of the semen stain. The victim's alleles, when combined with those of the true rapist, produced an apparent genotype that matched Durham's. The laboratory mistook this mixed profile for a single source result, and thereby falsely incriminated an innocent man. Durham was released from prison in 1997 (14).

Although experience has shown that false positives can occur, the rate at which they occur is difficult to estimate on the basis of existing data. Most laboratories participate in periodic proficiency tests, which can cast some light on the potential for error. European forensic laboratories have carried out collaborative exercises involving analysis of stains from known sources (21–26). However, this work is designed more to test the uniformity of DNA test results among laboratories using the same protocol than to determine the rate of errors. In the United States, TWGDAM guidelines call for each analyst to take two proficiency tests each year (27), and proficiency testing is a requirement for laboratory certification under the program administered by ASCLAD-LAB (28). However, these tests generally are not well designed for estimating the rate of false positives. The tests typically are not blind (i.e., the analysts know they are being tested), they involve limited numbers of samples, and the samples may be easier to analyze than those encountered in routine casework.

In 1992, a report of the National Research Council called for more extensive proficiency testing, declaring that "laboratory error rates must be continually estimated in blind proficiency testing and must be disclosed to juries" (1). The NRC called for

external, blind proficiency tests "that are truly representative of case materials (with respect to sample quality, accompanying description, etc.)." Thereafter, the Federal DNA Identification Act of 1994 required the director of the National Institute of Justice (NIJ) to report to Congress on the feasibility of establishing an external blind proficiency testing program for DNA laboratories. But the move toward external blind proficiency testing lost momentum when the NIJ director raised a number of practical concerns. It was dealt another blow by the 1996 report of the National Research Council, which downplayed the need for proficiency testing. The 1996 NRC report suggested that the problem of laboratory error be addressed through a variety of means and concluded that the best way to safeguard against error is to allow re-testing of samples (28).

Do We Need Scientifically Valid Estimates of Laboratory Error Rates?

Although re-testing is undoubtedly helpful, it does not eliminate the need to consider error when evaluating DNA evidence. Re-testing cannot catch every error. A critical error, such as cross-contamination of samples, may occur before samples can be split for duplicate testing (29,30). Some errors, such as the error of interpretation that falsely incriminated Timothy Durham, may simply be repeated on re-test. And re-testing cannot be done in every case because critical samples are sometimes exhausted by the first test. Re-testing may reduce the likelihood of a false positive, but no one claims that it can eliminate false positives. Hence, the availability of re-testing does not by itself explain why less importance is placed on having accurate estimates of false positive probabilities than random match probabilities.

Another explanation, suggested by some court opinions, is that jurors have less need of statistical estimates when evaluating the probability of a false positive because they can appreciate the potential for false positives based on common sense and experience. "Shortcomings such as mislabeling, mixing the wrong ingredients, or failing to follow routine precautions against contamination may well be amenable to evaluation by jurors without the assistance of expert testimony" (31). By contrast, there is nothing in jurors' everyday experience that would allow them to estimate the probability of a coincidental match between two DNA profiles; hence, experts must present statistical estimates of the random match probability.

The problem with this argument is that it equates the ability to appreciate the potential for a laboratory error with the ability to accurately estimate the probability of an error. It is not clear that the latter will necessarily follow from the former. Even if jurors understand the various ways in which a false positive might occur, it requires a leap of faith to conclude that they will therefore be able to determine accurately, based on common sense, whether, for example, the probability of such an error in a particular case is 1 in 100 or 1 in 10,000. In the absence of solid empirical data there is considerable disagreement among experts about what the rate of laboratory error might be (3,8,13,15,16). To rely on jurors' common sense to produce accurate estimates when experts cannot agree seems unduly optimistic.

It might be argued, however, that jurors do not need precise estimates of the false positive probability—they need only know that the probability of error is low enough to make a false positive unlikely in the case at hand. If, as commentators have suggested, the rate of false positives is between 1 in 100 and 1 in 1000, or even less (3,8,12,13,16), then one might argue that the jury can safely rule out the prospect that the reported match in their case is due to error and can proceed to consider the probability of a coincidental

match. For reasons we will explain more fully below, this argument is fallacious and profoundly misleading. The core of the fallacy is the erroneous assumption that the false positive probability, which is the probability that a match would be reported *between two samples that do not match,* is equal to the probability that a false match was reported in a particular case. As we will explain below, the probability that a reported match occurred due to error in a particular case can be much higher, or lower, than the false positive probability.

How the Potential for Error Affects the Value of DNA Evidence

We now present a framework for considering the role that error may play in determining the value of DNA evidence. Our approach relies on Bayes' theorem, a basic principle of logic. Bayes' theorem indicates how a rational evaluator should adjust a probability assessment in light of new evidence (32–34). Our analysis shows how the probability of a false positive should influence a rational evaluator's belief in the proposition that a particular individual is the source of a biological specimen. We use Bayes' theorem here solely to illustrate the logical connection between the false positive rate and the value of DNA evidence. We do not address the separate issue of whether Bayes' theorem should be used to explain the value of DNA evidence to juries.

Suppose that a rational evaluator is considering whether a biological specimen could have come from a particular suspect. The evaluator must assess the probability of two alternative propositions:

S: the specimen came from a suspect;
$\overline{S}$: the specimen did not come from a suspect.

The evidence to be evaluated is a forensic scientist's report of a DNA match between the suspect's profile and the profile of the sample. We will call the report of a match R. Under the conventional expression of Bayes' theorem:

$$\frac{P(S \mid R)}{P(\overline{S} \mid R)} = \frac{P(S)}{P(\overline{S})} \cdot \frac{P(R \mid S)}{P(R \mid \overline{S})} \qquad (1)$$

Bayes' theorem describes the relationship between three components: the prior odds, the posterior odds, and the likelihood ratio. The term to the immediate right of the equal sign is the prior odds, which reflect the evaluator's assessment of the odds that a proposition is true before the receipt of new evidence. The term to the left of the equal sign is the posterior odds, which reflect the evaluator's belief in the odds that the proposition is true after receipt of new evidence. The remaining term, to the right of the multiplication sign, is the likelihood ratio. It specifies the evaluator's belief in the relative probability that the new evidence would arise if the proposition is true and if it is not true. Bayes' theorem specifies that the posterior odds of a proposition equal the prior odds multiplied by the likelihood ratio.

Bayes' theorem can be used to show the effect that DNA evidence should have on belief in the propositions S and $\overline{S}$. Suppose, for example, that the evaluator initially

(before considering the DNA evidence) thinks there is a 20% chance that the suspect is the source of a specimen. In terms of Eq 1, $P(S) = 0.20$ and $P(\overline{S}) = 0.80$. Therefore, the prior odds would be 0.25 (often expressed as 1:4 odds). Suppose further that the evaluator thinks the match is certain to be reported if the suspect was the source of the specimen, hence $P(R \mid S) = 1.00$, and the evaluator thinks that there is only one chance in 1000 that a match would be reported if the suspect was not the source of the specimen, hence $P(R \mid \overline{S}) = 0.001$. Accordingly, the likelihood ratio is $1.00/0.001 = 1000$. To determine the posterior odds, one simply multiplies the prior odds by the likelihood ratio; hence the posterior odds should be $0.25 \cdot 1000 = 250$. In other words, the evaluator should now believe that proposition S is 250 times more likely than proposition $\overline{S}$.

The conclusion can be restated as a probability by simply converting the posterior odds to a probability using the formula: Probability = Odds/(Odds + 1). Thus, one can say that the evaluator should now believe the probability that the suspect is the source of the specimen is $250/251 = 0.996$. In other words, if the evaluator believes that the DNA evidence is 1000 times more likely to arise under S than under $\overline{S}$, then the evaluator should revise his estimated probability that the suspect is the source from 0.20 to 0.996 after receipt of the DNA evidence.

In the conventional expression of Bayes' theorem, the likelihood ratio takes into account all variables that affect the value of the evidence. The likelihood ratio for a reported DNA match is affected by both the probability of a random match and the probability of a false positive, because both factors contribute to the denominator of the likelihood ratio, $P(R \mid \overline{S})$. In order to assess the relative impact of the random match probability (*RMP*) and the false positive probability (*FPP*) on the value of DNA evidence, we must expand the likelihood ratio in order to show the separate effect of these two variables. As explained in the Appendix, the likelihood ratio can be expanded as follows:

$$\frac{P(R \mid S)}{P(R \mid \overline{S})} = \frac{1}{RMP + [FPP \cdot (1 - RMP)]} \qquad (2)$$

Using this version of the likelihood ratio, it is easy to show how the potential for a false positive affects the value of DNA evidence. Table 1 shows how variations in the prior odds, random match probability, and false positive probability should affect a rational evaluator's assessment of the posterior odds that the suspect was the source of a biological specimen. The posterior odds presented in the table were calculated by multiplying the prior odds by the likelihood ratio as stated in Eq 2.

The prior odds presented in Table 1 are designed to correspond to four distinct case types that vary in how strongly the suspect is implicated as the source of the specimen by evidence other than the DNA match. Prior odds of 2:1 describe a case in which the other evidence is fairly strong but not sufficient, by itself, for conviction. It has been reported that DNA testing leads to the exclusion of approximately one third of suspects in sexual assault cases. Hence, prior odds of 2:1 might describe a typical sexual assault case submitted for DNA testing.

Prior odds of 1:10 and 1:100 describe cases in which the other evidence indicates a relatively low initial probability that the suspect is the source, as might occur if the match were found during a "DNA dragnet," in which the police tested many possible contributors in a particular locality with little reason to suspect any of them in particular other than their proximity to the crime. Prior odds of 1:1000 describe a case in which there is almost no evidence apart from the DNA match, as might occur in a "cold hit" case in which the suspect is selected by scanning a databank of thousands of people for matching DNA profiles.

The random match probabilities presented in Table 1 are chosen to represent a range of values that might plausibly arise in actual cases. Random match probabilities on the order of one in one billion (10^{-9}) are often reported when laboratories are able to match two single source samples over ten or more STR loci. Random match probabilities closer to one in one million (10^{-6}) are common when fewer loci are examined, when the laboratory can obtain only a partial profile of one of the samples, or when one of the samples contains a mixture of DNA from more than one person. Random match probabilities near 1 in 1000 (10^{-3}) often result from the use of less discriminating tests, such as DQ-alpha/polymarker, particularly when the comparison involves a mixed sample.

The false positive probabilities presented in Table 1 are also chosen to represent a plausible range that might arise in actual cases. Although the probability of a false positive in any particular case will depend on a variety of factors, commentators generally have estimated the overall rate of false positives to be between 1 in 100 (0.01) and 1 in 1000 (0.001) (8,13,16). Of course, these estimates may overstate the probability for cases in which special steps, such as repeat testing, have been taken to reduce the chance of error. So for purposes of illustration we also present a false positive probability of 1 in 10,000 (0.0001). If two independent tests comparing the same samples each had a false positive probability of 1 in 100, then the probability of a false positive on both tests would be 1 in 10,000. A false positive probability of zero is also included for purposes of comparison with the other values (although zero is not a plausible value for this variable).

As Table 1 shows, the posterior odds are strongly influenced by the prior odds, the random match probability, and the false positive probability. This result indicates that a rational evaluator should consider all three factors when assessing the likelihood that the suspect is the source of a particular sample.

One aspect of these results that may be counter-intuitive is that the importance of the false positive probability for determining the posterior odds varies dramatically depending on the value of the random match probability. As Table 1 shows, changes in the false positive probability have a much greater effect on the posterior odds when the random match probability is low than when it is higher. For example, when the random match probability is one in one billion (10^{-9}), the posterior odds diminish by five orders of magnitude when the false positive probability increases from 0 to 1 in 10,000. In contrast, when the random match probability is 1 in 1000 (10^{-3}) the same increase in the false positive probability produces only a small change (much less than one order of magnitude) in the posterior odds.

These results may seem counter-intuitive given that the false positive probability and the random match probability are combined in a manner that is approximately additive in Eq 2. However, the effect of changing one of these variables on the value of

the likelihood ratio depends on the size of the change *relative to* the other variable. Receiving $100 changes my net assets more dramatically if I started with $1 than if I started with $200. Similarly, an increase of given size in the false positive probability will affect the likelihood ratio more dramatically when the random match probability is very small than when it is larger. Hence, it may be far more important to have an accurate estimate of the probability of a false positive when evaluating a reported match on a rare DNA profile than when evaluating a reported match on a more common profile.

Another important lesson to be learned from Table 1 is that the posterior odds can be rather low notwithstanding an impressive random match probability. When the random match probability is one in one billion, for example, one might assume that the odds the suspect is the source of the sample will necessarily and always be very high. Not so. If the prior odds are 1:1000 because the suspect was selected by trawling through a large data bank to find a matching profile, and there is little other evidence of his guilt, then the posterior odds will be only 10 if the false positive probability is 1 in 10,000, only 1.00 if the false positive probability is 1 in 1000, and only 0.10 if the false positive probability is 1 in 100. Hence, a rational evaluator who thought the false positive probability was between 1 in 100 and 1 in 1000 should conclude that the suspect probably is not the source of the sample, notwithstanding the reported match on a profile found in one person in a billion.

Posterior Odds and the Standard of Proof

One way to understand the posterior odds presented in Table 1 is to relate them to the traditional standard of persuasion in criminal trials. How high should the posterior odds be to convince a rational juror "beyond a reasonable doubt" that the suspect is the source of the sample?

A number of legal commentators have linked the criminal standard of persuasion to posterior odds (35). For example, Professor Richard Friedman (36) has argued that a rational adjudicator should treat an accused as guilty if and only if

$$Oy > \frac{Ep}{En} \tag{3}$$

where *Oy* is the odds of guilt, *Ep* is the social cost (disutility) of a false conviction, and *En* is the social cost (disutility) of a false acquittal. If one accepts Blackstone's famous statement that "it is better that ten guilty persons escape, than that one innocent suffer" then, according to Friedman's analysis, one should convict only if the posterior odds of guilt are at least 10:1 (37).

The United States Supreme Court has quoted with apparent approval Thomas Starkie's statement that "it is better that ninety-nine … offenders should escape than that one innocent man should be condemned" (38). If one accepts Starkie's statement, then the posterior odds of guilt should exceed 99:1 to justify conviction. Although there is no apparent consensus among experts on this issue, Ceci and Friedman (37) have recently argued that Blackstone's ratio "understates" the correct legal standard for conviction and that Starkie's ratio "appears closer to the mark."

This analysis casts additional light on the data presented in Table 1. To appreciate what the data tell us about the strength of DNA evidence, we can consider the circumstances under which DNA evidence would meet the Blackstone and Starkie standard of proof. We are not proposing that these quantitative standards be employed in actual trials. We invoke these standards merely as a framework for understanding what the data in Table 1 tell us about the value of DNA evidence. In the discussion that follows, we will assume a hypothetical criminal case in which a laboratory reports a DNA match between a sample known to have come from the perpetrator and a reference sample from the defendant. We will assume that identity is the only issue in the case, and hence that the jurors should convict if they are convinced beyond a reasonable doubt that the defendant is the source of the sample. Under what circumstances should a rational jury convict the defendant?

When the prior odds are 2:1, the posterior odds are well above both the Blackstone and Starkie threshold for all levels of random match probability and false positive probability presented in Table 1. Because the case against the defendant is relatively strong even without the DNA evidence, the reported DNA match is sufficient to push a rational evaluator over the threshold of conviction even under the worst-case scenario in which both the random match probability (10^{-3}) and the false positive probability (0.01) are high.

When the prior odds are 1:10, the situation becomes more complicated. Here the other evidence against the defendant is weaker and the DNA evidence must therefore be a bit stronger to push a rational evaluator across the threshold of conviction. For this type of case, the posterior odds are well above the Starkie threshold only when the random match probability is one in one million (10^{-6}) or less and the false positive probability is 1 in 10,000 or less. When the false positive probability is 1 in 100, the posterior odds are at or below the Blackstone threshold for all random match probabilities. Thus, for cases of this type, it appears very important to know whether the false positive probability might be as high as 1 in 100. If so, there is "reasonable doubt" about the defendant's guilt.

When the prior odds are 1:100, the DNA evidence must be very powerful to justify conviction. The posterior odds barely meet the Starkie threshold when the random match probability is one in one million or less and the false positive probability is 1 in 10,000. The posterior odds exceed the Blackstone threshold only when the random match probability is one in one million or less *and* the false positive probability is 1 in 10,000 or less. For this type of case, it is again crucial to know the exact value of the false positive probability in order to determine whether the DNA is strong enough to justify conviction. If the false positive rate is as high as 1 in 1000, there is "reasonable doubt" about the defendant's guilt.

In the weakest case, when the prior odds are 1:1000, DNA evidence is insufficient to meet the Starkie standard under any of the values listed in Table 1, except when the false positive probability is (unrealistically) assumed to be zero. Even when the random match probability is one in one billion and the false positive probability is 1 in 10,000, the posterior odds barely reach the Blackstone threshold. For a case of this type, a false positive probability of even 1 in 1000 should render the DNA evidence insufficient to justify conviction. Indeed, when the random match probability is 1 in 1000, a DNA match is insufficient even to prove that the suspect is more likely than not to be the source of the sample.

The False Positive Fallacy

The key conclusion to emerge from this analysis is the importance of having accurate information about *both* the random match probability and the false positive probability when evaluating DNA evidence. Ignoring or underestimating the potential for a false positive can lead to serious errors of interpretation, particularly when the other evidence against the suspect (apart from the DNA evidence) is weak.

We return therefore to the question raised at the beginning of this article. Why is it considered essential to have valid scientific data on the random match probability but unnecessary to have valid data on the false positive probability?

We believe the explanation lies partly in a common logical fallacy that we shall call the false positive fallacy. We suspect that people mistakenly assume that *if* the false positive probability is low *then* the probability of a false match must also be low in every case. For example, a forensic scientist who thinks that there is only a 1% chance (1 chance in 100) of falsely declaring a match between the samples in a case *if they really do not match*, might assume that there is, necessarily, a 99% chance (99 chances in 100) that the reported match is a true match. This assumption is fallacious, although the mistake is not easy to see.

The fallacy arises from mistakenly equating the conditional probability of a match being reported *when the samples do not match* (the false positive probability) with the probability that the samples do not match *when a match has been reported.* These two probabilities are not the same. The *false positive probability* is the probability of a match being reported under a specified condition (no match). It does not depend on the probability of that condition occurring. By contrast, the probability that the samples do not match *when a match has been reported* depends on both the probability of a match being reported under the specified condition (no match) and on the prior probability that that condition will occur. Consequently, the probability that a reported match is a true match or a false match cannot be determined from the false positive probability alone.

In formal terms, the fallacious assumption is that $P(M \mid R) = 1 - P(R \mid \overline{M})$, where M is the event that the suspect and the perpetrator have matching DNA profiles, $\overline{M}$ is the event that they does not have matching profiles, and $P(R \mid \overline{M})$ is the false positive probability, i.e., the probability of a match being reported given that the samples do not have matching profiles. This assumption is fallacious because it ignores the prior odds that the suspect's profile matches the sample profile. Let the prior odds, $P(M)/P(\overline{M})$, equal $1/k$ where k is large. Then:

$$\frac{P(M \mid R)}{P(\overline{M} \mid R)} = \frac{P(R \mid M)}{P(R \mid \overline{M})} \cdot \frac{1}{k} \tag{4}$$

Assume $P(R \mid M) = 1$. Then $P(M \mid R) = 1/\left[1 + k \cdot P(R \mid \overline{M})\right]$ which can be much lower than $1 - P(R \mid M)$ when k is large.

For example, suppose that the prior odds the suspect will match are 1:1000 because the suspect is selected through a large DNA dragnet and appears, initially, to be an unlikely perpetrator. Suppose further that a DNA match is reported and that the false

positive probability is 0.01 (1 in 100). The probability that this reported match is a true match is, therefore, $1/[1+1000 \cdot 0.01] = 0.0999$. In other words, the probability that this reported match is a true match is not 0.99 (99 chances in 100), as the false positive fallacy would suggest; it is less than 0.1 (one chance in ten).

Thus, when the prior odds that a particular suspect will match are very low, as might be the case if the suspect is identified during a "DNA dragnet" or database search, the probability that the samples do not match when a match has been reported can be far higher than the false positive probability. For cases of this type, true matches are expected to be rare. Therefore, the probability in a particular case that a non-match will mistakenly be reported as a match, even if low, may approach or even surpass the probability that the suspect truly matches.

The false positive fallacy is similar in form to the well known "prosecutor's fallacy" (39), but differs somewhat in content. Victims of the false positive fallacy mistakenly assume that $P(M/R) = 1 - P(R/\overline{M})$. Victims of the prosecutor's fallacy mistakenly assume that $P(S/M) = 1 - P(M/\overline{S})$ (39). Both fallacies arise from failure to take account of prior probabilities (or odds) when evaluating new evidence; both can lead to significant overestimation of the posterior probability when the prior probability is low. The prosecutor's fallacy is an erroneous way of estimating the probability that the suspect is the source of a sample based on evidence of a matching characteristic; the false positive fallacy is an erroneous way of estimating the probability of a true match based on a reported match. It is important that forensic scientists, and others who evaluate DNA evidence, understand and appreciate both fallacies.

False Positives and Cold Hits

When first introduced, DNA testing was used primarily for "confirmation cases," that is, cases where other evidence pointed to a likely suspect (40). In recent years, the growing use of offender databanks and "DNA dragnets" has created a new class of cases, sometimes called "cold hit" or "trawl cases," in which the DNA match itself makes the defendant a suspect (40,41). In such cases there may be little evidence against a suspect other than a DNA match.

The evidentiary value of "cold hit" DNA matches has been debated. The National Research Council, in reports on forensic DNA evidence issued in 1992 (1) and 1996 (28), argued that DNA matches obtained in database searches are less probative than those obtained when testing a previously identified suspect because the probability of finding a match by chance increases when one trawls though a database comparing large numbers of profiles.

However, statisticians David Balding and Peter Donnelly have argued persuasively from a Bayesian perspective that the likelihood ratio describing the value of a DNA match does not depend on the nature of the search that produced the match and hence that a cold hit is just as powerful as any other DNA match (assuming the same random match probability) (41). By their account, the strength of the overall case may sometimes be weak when the suspect was identified in a database search because the prior probability of guilt in such cases can be very low, but the trawl through the database does not diminish the probative value of the DNA match. In fact, they argue that a database DNA match may provide slightly stronger evidence of identity than a

confirmation case match if, as typically happens, the search of the database rules out (excludes) a large number of other individuals while finding a match to only one (40,41).

The Balding and Donnelly analysis seems correct, as far as it goes. However, Balding and Donnelly acknowledge that they "ignore the possibility of handling or laboratory error leading to a 'false positive' match, although this possibility must be addressed in practice" (41). The analysis reported in the present article goes beyond that of Balding and Donnelly to demonstrate the implications of false positives for both confirmation and trawl cases and thereby casts important new light on the question of the evidentiary value of database matches.

The potential for false positives may be a particularly important consideration when evaluating DNA evidence in trawl cases where the prior probability that any particular suspect is the source of an evidentiary sample is very low. In such cases, a key issue is whether the DNA match is sufficiently probative to create a high posterior probability that the suspect is the source despite the low prior probability. The results reported in Table 1 suggest that the probability of a false positive may be a critical factor in determining whether the DNA evidence is indeed strong enough.

Consider, for example, the hypothetical cases illustrated in Table 1 in which the prior odds that the suspect is the source of an evidentiary sample are 1:1000 and the random match probability is one in one billion (10^{-9}). If the probability of a false positive is zero, then the posterior odds are a million to one in favor of the suspect being the source, which certainly seems high enough to justify confidence in that conclusion. In other words, the DNA evidence has more than enough probative value to make up for the low prior probability. However, if the false positive probability is even 1 in 10,000, the posterior odds in favor of the suspect being the source are reduced drastically to only 10:1. It is very important for those evaluating DNA evidence to understand that a false positive probability on the order of 1 in 10,000, which may seem low enough to be "safe," may nevertheless undermine the value of a one-in-a-billion DNA match sufficiently that, when combined with a low prior probability, there is still room for doubt about whether the suspect is the source of the matching sample.

Of course, the assessment of hypothetical cases cannot tell us whether, as a practical matter, the false positive probability could be as high as 1 in 10,000 in a given case. As Donnelly and Friedman have noted, "what matters is not the probability of any laboratory error, but rather only the probability of those errors that would lead to the false declaration of a match in the given case—a probability that will vary widely with the circumstances of the DNA testing" (40). The false positive probability is undoubtedly affected by such factors as the quality of laboratory work and the clarity of the results. Dangerous laboratory practices, such as handling and processing evidentiary and reference samples in close physical and temporal proximity, might increase the false positive probability. Loose interpretive standards that allowed a match to be called based on incomplete or problematic data might also increase the false positive probability. Fortunately, the particular circumstances of database searches would seem to rule out, or at least greatly reduce, the likelihood of some types of errors, such as those arising from switching or cross-contaminating samples, because samples are tested at different times and, often, in different laboratories. However, other types of errors, such as those arising from misinterpretation of test results, might still produce false matches. Whether the chance of a false match is high enough to be of concern is a question that should be

considered carefully in each case by those who evaluate DNA evidence. The practical value of this article is in showing circumstances under which even low false positive probabilities should be of concern.

Conclusion

The present article does not address the difficult question of how to estimate the false positive probability, but it shows the importance of knowing how high that probability might be. Whether a suspect should be judged guilty or not guilty depends, in some cases, on whether the false positive probability is closer to 1 in 100, 1 in 1000, or 1 in 10,000. Particularly in cases in which there is little other evidence against the suspect, ignorance of the true probability of error creates a disturbing element of uncertainty about the value of DNA evidence. Commentators have noted the difficulty of generating accurate estimates of the probability of a false positive in a particular case (14,16,28). However, the task is no less important for being difficult.

External blind proficiency testing is said to be the best source of information about laboratory error rates (1,13,42). Of course, the rate of error in a proficiency testing does not necessarily equate to the false positive probability in a particular case because the unique circumstances of each case may make various types of errors more or less likely than average. Nevertheless, data on the rate of various types of errors in proficiency testing provides insight into the likely range of values for a particular case (14,42). When considering the probability of a false positive due to a sample switch error, it would clearly be helpful to know, for example, whether the rate of such errors in forensic laboratories in general is 1 in 50 or 1 in 20,000. Similarly, when considering the probability of a false positive due to inadvertent cross-contamination of samples, or misinterpretation of test results, it would be helpful to know how often cross-contamination, or misinterpretation, occurred in proficiency tests.

There has been continuing debate over the feasibility of external blind proficiency testing of forensic DNA laboratories. The National Institute of Justice funded a major study of this issue in which small-scale blind proficiency tests were conducted to assess their practicality and costs (43). The study found that blind proficiency testing is possible, although costly and "fraught with problems." The estimated annual cost of administering two blind proficiency tests (involving simulated cases) to each of the 150 DNA testing laboratories in the United States was $450,000 to $3,020,000. The directors of the study recommended to NIJ that a program of blind proficiency testing be deferred in order to allow assessment of less costly alternative programs, such as external laboratory audits, that might achieve many of the same goals. It remains to be seen whether an audit program will be implemented and whether such a program will produce useful data on laboratory error rates.

In the absence of such data, the problem of error will not go away. It will only become more acute as DNA testing is used in a widening range of cases. If DNA evidence is to achieve its full promise and potential, forensic scientists and legal professionals must give more attention to this issue.

APPENDIX

Here we describe how the traditional Bayesian likelihood ratio may be expanded to show the separate effect of the random match probability (RMP) and the false positive probability (FPP) on the value of a reported DNA match. Our analysis follows a method

first described by David Schum and his colleagues for distinguishing reliability and diagnosticity of evidence in "cascaded inference" (33,44).

We begin by distinguishing *R*, a reported match, from *M*, a true match. We assume there are two possible underlying states of reality:

M : The suspect and the specimen have matching DNA profiles;

$\overline{M}$: the suspect and the specimen do not have matching DNA profiles.

However, it is impossible to know with certainty whether M or $\overline{M}$ is true because the only information available about M, $\overline{M}$ is the laboratory report, which might be mistaken.

The numerator of the conventional likelihood ratio, $P(R \mid S)$, is equivalent to the expression $P(R \cap S)/P(S)$, where $P(R \cap S)$ means the probability that *both* *R* and *S* occur. Furthermore, $P(R \cap S)$ can be written as the disjoint union of two compound events, $P(R \cap M \cap S)$ and $P(R \cap \overline{M} \cap S)$. Therefore, $P(R \cap S) = P(R \cap M \cap S) + P(R \cap \overline{M} \cap S)$.

Because

$$P(R \cap M \cap S) = P(R \mid M \cap S) \cdot P(M \mid S) \cdot P(S)$$

and

$$P(R \cap \overline{M} \cap S) = P(R \mid \overline{M} \cap S) \cdot P(\overline{M} \mid S) \cdot P(S),$$

we can eliminate $P(S)$ and write:

$$P(R \mid S) = P(R \mid M \cap S) \cdot P(M \mid S) + P(R \mid \overline{M} \cap S) \cdot P(\overline{M} \mid S)$$

The denominator of the likelihood ratio can be expanded in similar fashion. Hence, the likelihood ratio, in expanded form, can be written as:

$$\frac{P(R \mid S)}{P(R \mid \overline{S})} = \frac{P(R \mid M \cap S) \cdot P(M \mid S) + P(R \mid \overline{M} \cap S) \cdot P(\overline{M} \mid S)}{P(R \mid M \cap \overline{S}) \cdot P(M \mid \overline{S}) + P(R \mid \overline{M} \cap \overline{S}) \cdot P(\overline{M} \mid \overline{S})} \quad (5)$$

In order to simplify this rather cumbersome statement of the likelihood ratio, we will assume that $P(R \mid M)$ is independent of S, $\overline{S}$. In other words, we assume the probability that a match will be reported if there really is a match is not affected by whether the match is coincidental. Consequently, $P(R \mid M \cap S) = P(R \mid M \cap \overline{S}) = P(R \mid M)$. Because the suspect and specimen will

necessarily have matching DNA profiles if the suspect is the source of the specimen, $P(M \mid S) = 1.00$ and $P(\overline{M} \mid S) = 0.00$. Finally, because $\overline{M}$ can only arise under $\overline{S}$, $P(R \mid \overline{M} \cap \overline{S})$ can be simplified to $P(R \mid \overline{M})$. Accordingly, Eq 5 can be re-stated as:

$$\frac{P(R \mid S)}{P(R \mid \overline{S})} = \frac{P(R \mid M)}{P(R \mid M) \cdot P(M \mid \overline{S}) + P(R \mid \overline{M}) \cdot P(\overline{M} \mid \overline{S})} \qquad (6)$$

In this expanded version of the likelihood ratio, the term $P(R \mid M)$ is the probability that the laboratory will report a match if the suspect and the specimen have matching DNA profiles. If the samples are adequate in quantity and quality, and the laboratory is competent, we would expect $P(R \mid M)$ to be close to 1.00. Estimates of less than 1 imply that the laboratory may fail to detect a true match due, for example, to error (a "false negative") or inadequately sensitive procedures. For present purposes, we will simply assume that $P(R \mid M) = 1.00$.

The term $P(M \mid \overline{S})$ is the probability of a coincidental match. For a comparison between single-source samples, $P(M \mid \overline{S})$ is the random match probability, *RMP*, or the frequency of the matching profile in a relevant reference population. Because *M* and $\overline{M}$ are mutually exclusive and exhaustive, $P(\overline{M} \mid \overline{S})$ is the complement of the *RMP*. Finally, the term $P(R \mid \overline{M})$ is the false positive probability, *FPP*. Substituting terms, the expanded likelihood ratio can be restated as in the form presented in the text as Eq 2:

$$\frac{P(R \mid S)}{P(R \mid \overline{S})} = \frac{1}{RMP + [FPP \cdot (1 - RMP)]} \qquad (2)$$

References

1. National Research Council. DNA technology in forensic science. Washington, D.C.: National Academy Press, 1992.
2. Thompson WC, Ford S. The meaning of a match: sources of ambiguity in the interpretation of DNA prints. In: Farley M, Harrington J, editors. Forensic DNA technology. New York: CRC Press, Inc, 1991;93–152.
3. Thompson WC. Subjective interpretation, laboratory error and the value of forensic DNA evidence: three case studies. Genetica 1995;96:153–68.
4. Taroni F, Aitken CGG. Forensic science at trial. Jurimetrics 1997;37:327–37.
5. Kaye DH, Sensabaugh GF. Reference guide on DNA evidence. In: Cecil J, editor. Reference manual on scientific evidence. Washington, DC: Federal Judicial Center, 2000;2:485–576.
6. Thompson WC. Forensic DNA evidence. In: Black B, Lee P, editors. Expert evidence: a practitioner's guide to law, science and the FJC manual. St. Paul, MN: West Group, 1997;195–266.

7. Jerome Smith v. State. Southern Reporter, Alabama Court of Criminal Appeals, 1995;677:1240–48.
8. Koehler JJ. Error and exaggeration in the presentation of DNA evidence. Jurimetrics 1993;34:21–39.
9. Kaye D. DNA evidence: probability, population genetics, and the courts. Harv J Law Technol 1993;7:101–72.
10. Jonakait RN. Stories, forensic science and improved verdicts. Cardozo L Rev 1991;13:343–52.
11. Koehler JJ. DNA matches and statistics: important questions, surprising answers. Judicature 1993;76:222–9.
12. Thompson WC. Comment. In Roeder K, DNA fingerprinting: a review of the controversy. Stat Sci 1994;9:263–6.
13. Koehler JJ. The random match probability in DNA evidence: irrelevant and prejudicial? Jurimetrics 1995;35:201–19.
14. Thompson WC. Accepting lower standards: The National Research Council's second report on forensic DNA evidence. Jurimetrics 1997;37(4):405–24.
15. Mueller L. The use of DNA typing in forensic science. Acct in Res 1993;3:1–13.
16. Roeder K. DNA fingerprinting: a review of the controversy. Stat Sci 1994;9:222–47.
17. Scheck B, Neufeld P, Dwyer F. Actual innocence. New York: Doubleday, 2000.
18. Brenner L, Pfleeger B. Investigation of the sexual assault of Danah H. Philadelphia, PA: Philadelphia Police Department DNA Identification Laboratory; 1999 Sept. 24; Lab No.: 97-70826.
19. Brenner L, Pfleeger B. Amended report: investigation of the sexual assault of Danah H. Philadelphia, PA: Philadelphia Police Department DNA Identification Laboratory; 2000 Feb. 7; Lab No.: 97-70826.
20. Cotton RW, Word C. Amended report of laboratory examination. Germantown, MD: Cellmark Diagnostics; 1995 Nov 20; Case No.: F951078.
21. Schneider PM, Fimmers R, Woodroffe S, Werrett DJ, Bär W, Brinkmann B, et al. Report of a European collaborative exercise comparing DNA typing results using a single locus VNTR probe. Forensic Sci Intl 1991;49:1–15.
22. Gill P, Woodroffe S, Bar W, Brinkmann B, Carracedo A, Eriksen B, et al. A report of an international collaborative experiment to demonstrate the uniformity obtainable using DNA profiling techniques. Forensic Sci Intl 1992;53:29–43.
23. Gill P, Kimpton C, D'Aloja E, Anderson JF, Bar W, Brinkmann B, et al. Report of the European profiling group (EDNAP): Towards standardisation of short tandem repeat (STR) loci. Forensic Sci Intl 1994;65:51–9.
24. Kimpton C, Gill P, D'Aloja E, Anderson JF, Bar W, Holgersson S, et al. Report on the second collaborative STR exercise. Forensic Sci Intl 1995;71:137–52.
25. Wiegand P, Amgach E, Augustin C, Bratzke H, Cremer U, Edelman J, et al. GEDNAP IV and V. The 4th and 5th stain blind trials using DNA technology. Intl J Legal Med 1995;108:79–84.
26. Anderson JF, Martin P, Carracedo A, Dobosz M, Eriksen B, Johnsson V, et al. Report on the third EDNAP collaborative STR exercise. Forensic Sci Int 1996;78:83–93.
27. Technical Working Group on DNA Analysis Methods (TWGDAM). Established guidelines for a quality assurance program for DNA testing laboratories; including RFLP and PCR technologies. Crime Lab Dig 1995;18:44–75.
28. National Research Council. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press, 1996.
29. Balding DJ. Errors and misunderstandings in the second NRC report. Jurimetrics 1997;37:469–76.
30. Thompson WC. DNA evidence in the O.J. Simpson trial. U Colorado L Rev

1996;67(4):827–57.
31. People v. Venegas: California Reporter, California Supreme Court, 1998;18:47–88.
32. Aitken CGG. Statistics and the evaluation of evidence for forensic scientists. Chichester: J. Wiley & Sons, 1995.
33. Schum DA. Evidential foundations of probabilistic reasoning. New York: John Wiley & Sons, 1994.
34. Robertson B, Vignaux GA. Interpreting evidence. Evaluating forensic science in the courtroom. Chichester: J. Wiley & Sons, 1995.
35. Lempert RO. Modeling relevance. Michigan L Rev 1975;75:1021–101.
36. Friedman RD. Answering the bayesioskeptical challenge. Intl J Evid Proof 1997;1:276–8.
37. Ceci SJ, Friedman RD. The suggestibility of children: scientific research and legal implications. Cornell L Rev 2000;86(1):33–108.
38. Schlup v. Delo: United States Reports, U.S. Supreme Court, 1995;513:298–322.
39. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy. Law Hum Behav 1987;11:167–87.
40. Donnelly P, Friedman RD. DNA database searches and the legal consumption of scientific evidence. Michigan L Rev 1999;97:931–84.
41. Balding DJ, Donnelly P. Evaluating DNA profile evidence when the suspect is identified through a database search. J Forensic Sci 1996;41(4):603–7.
42. Koehler JJ. Why DNA likelihood ratios should account for error (even when a national research council report says they should not). Jurimetrics 1997;37:425–37.
43. Peterson JL, Gaensslen RE. Developing criteria for model external DNA proficiency testing: final report. Chicago, IL: University of Illinois at Chicago; 2001 May.
44. Schum DA, DuCharme WM. Comments on the relationship between the impact and the reliability of evidence. Org Behav Human Perf 1971;6:111–31.

Additional information and reprint requests:
William C. Thompson, J.D., Ph.D.
Department of Criminology, Law & Society
University of California,
Irvine, CA 92697-7080
E-mail: wcthomps@uci.edu

TABLE 1—*Posterior odds that a suspect is the source of a sample that reportedly has a matching DNA profile, as a function of prior odds, random match probability, and false positive probability.*

Prior Odds	Random Match Probability	Probability of a False Positive	Posterior Odds
2:1	10^{-9}	0	2 000 000 000
2:1	10^{-9}	0.0001	20 000
2:1	10^{-9}	0.001	2000
2:1	10^{-9}	0.01	200
2:1	10^{-6}	0	2 000 000
2:1	10^{-6}	0.0001	19 802
2:1	10^{-6}	0.001	1998
2:1	10^{-6}	0.01	200
2:1	10^{-3}	0	2000
2:1	10^{-3}	0.0001	1818
2:1	10^{-3}	0.001	1001
2:1	10^{-3}	0.01	182
1:10	10^{-9}	0	100 000 000
1:10	10^{-9}	0.0001	1000
1:10	10^{-9}	0.001	100
1:10	10^{-9}	0.01	10
1:10	10^{-6}	0	100 000
1:10	10^{-6}	0.0001	990
1:10	10^{-6}	0.001	100
1:10	10^{-6}	0.01	10
1:10	10^{-3}	0	100
1:10	10^{-3}	0.0001	91
1:10	10^{-3}	0.001	50
1:10	10^{-3}	0.01	9
1:100	10^{-9}	0	100 00 000

1:100	10^{-9}	0.0001	100
1:100	10^{-9}	0.001	10
1:100	10^{-9}	0.01	1
1:100	10^{-6}	0	10 000
1:100	10^{-6}	0.0001	99
1:100	10^{-6}	0.001	10
1:100	10^{-6}	0.01	1
1:100	10^{-3}	0	10
1:100	10^{-3}	0.0001	9
1:100	10^{-3}	0.001	5
1:100	10^{-3}	0.01	1
1:1000	10^{-9}	0	1 000 000
1:1000	10^{-9}	0.0001	10.0
1:1000	10^{-9}	0.001	1.0
1:1000	10^{-9}	0.01	0.1
1:1000	10^{-6}	0	1000
1:1000	10^{-6}	0.0001	9.9
1:1000	10^{-6}	0.001	1.0
1:1000	10^{-6}	0.01	0.1
1:1000	10^{-3}	0	1.00
1:1000	10^{-3}	0.0001	0.91
1:1000	10^{-3}	0.001	0.50
1:1000	10^{-3}	0.01	0.09