Biometrics 59, 727–732
September 2003

READER REACTION

The Evidential Value in the DNA Database Search Controversy and the Two-Stain Problem

Ronald Meester[1,*] **and Marjan Sjerps**[2,**]

[1]Divisie Wiskunde, Faculteit der Exacte Wetenschappen, Vrije Universiteit,
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
[2]Nederlands Forensisch Instituut, P.O. Box 3110, 2280 GC Rijswijk, The Netherlands
[*]*email:* rmeester@cs.vu.nl
[**]*email:* m.sjerps@nfi.minjus.nl

Summary. Does the evidential strength of a DNA match depend on whether the suspect was identified through database search or through other evidence ("probable cause")? In Balding and Donnelly (1995, *Journal of the Royal Statistical Society, Series A* **158,** 21–53) and elsewhere, it has been argued that the evidential strength is slightly larger in a database search case than in a probable cause case, while Stockmarr (1999, *Biometrics* **55,** 671–677) reached the opposite conclusion. Both these approaches use likelihood ratios. By making an excursion to a similar problem, the two-stain problem, we argue in this article that there are certain fundamental difficulties with the use of a likelihood ratio, which can be avoided by concentrating on the posterior odds. This approach helps resolving the above-mentioned conflict.

Key words: Database search; DNA; Likelihood ratio; Posterior odds; Two-stain problem.

1. The Database Controversy

The statistical evaluation of the strength of DNA evidence at the scene of a crime has been the subject of debate among statisticians over the last decades. One of the issues that is still under debate is the interpretation of the evidential value of a single so-called *database match.*

Suppose that at the scene of a crime, a blood stain is found, which we assume comes from the criminal who committed the crime. The stain has a DNA profile which we denote by A. This profile is compared to DNA profiles in a database, and a *single* match results: one and only one person in the database has this profile A; suppose this person is named John Smith. How strong is the case against John Smith, compared to the situation where Smith is identified through other evidence, the so-called probable cause case? In particular, do we need to adjust the evidential value with the number of people that are compared to the stain?

Balding and Donnelly (1995), and Dawid and Morterra (1996) independently approached the problem using a likelihood ratio, and they reached the same conclusion: the evidential value of a DNA database match is slightly larger than the same DNA match in a probable cause case. Balding (2002) claims that this is intuitively obvious: indeed, in a database search, all nonmatching individuals are excluded as being the criminal, and therefore it should become more likely that Smith is our man. In an attempt to quantify this heuristic argument, both above-mentioned articles used the following approach.

Assume that the profile A occurs at a fraction p_A of all people. That is, if we randomly select a person from the population, the probability that this person has profile A is p_A. We let N be the population size, and $n \leq N$ is the size of the database. Furthermore, we ignore, for the sake of the argument, all relatedness between members of the population. The procedure to evaluate the weight of the DNA evidence against suspect John Smith now runs as follows. Consider the two competing hypotheses:

H_p: John Smith is the donor of the crime stain;
H_d: Someone else is the donor of the crime stain.

We now compute the probability of the evidence, that is, the single match of John Smith, under both competing hypotheses and look at the *ratio* of these two numbers. If this number, the *likelihood ratio,* is very high, then this is taken as strong evidence against the suspect. In the computation, we need (prior to considering the DNA evidence), the probabilities

$$\delta = P(\text{John Smith left the crime stain})$$

and

$$\pi = P(\text{someone in the database left the crime stain}).$$

These prior probabilities are based on the non-DNA evidence in the case against Smith, such as witness statements.

We then compute (Stockmarr, 1999; Dawid, 2001)

$$\begin{aligned}\text{LR} &= \frac{P(\text{Smith is the only match} \mid H_p)}{P(\text{Smith is the only match} \mid H_d)} \\ &= \frac{(1-p_A)^{n-1}}{p_A(1-p_A)^{n-1}\frac{1-\pi}{1-\delta}} \\ &= \frac{1}{p_A}\frac{1-\delta}{1-\pi}.\end{aligned}$$

In the case of uniform priors, that is, $\delta = 1/N$ and $\pi = n/N$, this leads to

$$\text{LR} = \frac{1}{p_A}\frac{N-1}{N-n}.$$

This last ratio is slightly larger than $1/p_A$, the likelihood ratio in a probable cause case, thereby giving support to the heuristic argument in Balding (2002) that the evidence is slightly stronger in a database search case. LR is usually taken as a measure of the weight of the evidence against suspect Smith.

Stockmarr (1999) criticized the above hypotheses H_p and H_d as being "data-dependent"; the fact that John Smith was the only match can only be known *after* the database search and therefore depends on the data. According to Stockmarr, the above hypotheses H_p and H_d of Balding and Donnelly are not appropriate, since they ignore the way that suspect Smith was found. Instead, Stockmarr suggested the following hypotheses:

$\tilde{H}_p$: The donor of the crime stain is in the database;
$\tilde{H}_d$: The donor of the crime stain is not in the database.

The corresponding likelihood ratio turns out to be $\delta/\pi p_A$ (Stockmarr, 1999). In case of uniform priors, this reduces to $1/np_A$, which differs from the likelihood ratio corresponding to the hypotheses H_p and H_d by a factor of about n. Formulated this way, the evidence is (possibly much) stronger in a probable cause case than in a database search case. The conclusion of Stockmarr supported Recommendation 5.1 in the second report of the National Research Council (NRCII, 1996), which suggested (in our terminology) dividing the likelihood ratio by the number of people in the database.

Stockmarr's arguments were dismissed by various people for various reasons; see Donnelly and Friedman (1999), Evett, Foreman, and Weir (2000), Dawid (2001), and Balding (2002).

2. The Two-Stain Problem

We shall return to the database search soon, but first we investigate what the likelihood procedure yields in another forensic statistical problem, namely, the *two-stain problem.*

Consider a crime that was committed by two people, each of whom left a blood stain at the scene of the crime; one of the stains is on a pillow, the other on a sheet. The DNA profiles of the stains are investigated: the stain on the pillow has profile A and the stain on the sheet has profile B. The population frequencies of these profiles are p_A and p_B, respectively.

John Smith is arrested for completely different reasons, unrelated to the above crime, but it turns out that Smith's DNA profile matches with A. As a result, poor John Smith becomes a suspect, and the question arises as to how strong the DNA evidence against Smith actually is. In this case, various hypotheses can be set up, without a clear *a priori* preference for either of them.

The following set of hypotheses was advocated by Evett (1987) and has been followed by many people since then; see, e.g., Aitken (1995).

H_p: John Smith was one of the crime stain donors;
H_d: John Smith was not one of the crime stain donors.

However, the following set of hypotheses is just as (or perhaps even more) natural:

H'_p: John Smith left the stain on the pillow;
H'_d: Two unknowns left the two stains.

Note the difference between the two sets of hypotheses: in the first set, there is no distinction between the stain on the pillow and the stain on the sheet, whereas in the second set, this distinction is consciously made.

Finally, there is a third natural set of hypotheses which completely ignores the stain on the sheet:

H''_p: John Smith left the stain on the pillow;
H''_d: Someone else left the stain on the pillow.

Our problem becomes clear when we compute the likelihood ratios for the various sets.

The likelihood ratio for H_p against H_d is computed in Evett (1987) as follows: Let F_1 be the event that the profiles found at the scene of the crime are A and B, *without specifying the stain that has profile A,* and let F_2 be the event that John Smith's profile is A. We then compute the likelihood ratio as follows:

$$\begin{aligned}\text{LR} &= \frac{P(F_1 \text{ and } F_2 \mid H_p)}{P(F_1 \text{ and } F_2 \mid H_d)} \\ &= \frac{P(F_1 \mid F_2 \text{ and } H_p)P(F_2 \mid H_p)}{P(F_1 \mid F_2 \text{ and } H_d)P(F_2 \mid H_d)}.\end{aligned}$$

Clearly, the probability that F_2 occurs does not depend on the assumption that Smith was or was not involved in the crime, and hence

$$P(F_2 \mid H_p) = P(F_2 \mid H_d),$$

and LR reduces to

$$\text{LR} = \frac{P(F_1 \mid F_2 \text{ and } H_p)}{P(F_1 \mid F_2 \text{ and } H_d)}.$$

It is easy to see that the numerator is equal to p_B, and that the denominator is equal to $2p_Ap_B$. Hence,

$$\text{LR} = \frac{p_B}{2p_Ap_B} = \frac{1}{2p_A}.$$

Aitken (1995) comments on this as follows:

> This is intuitively reasonable. If there are two criminals and one suspect, one would not expect the evidence of a matching blood stain to be as valuable as in the case in which there is one criminal and one suspect.

Next we compute the likelihood ratio for H'_p against H'_d, denoting this ratio by LR′. Since the hypotheses this time do specify which stain has which profile, the evidence is treated similarly: F'_1 is the event that the stain on the pillow has profile A and the stain on the sheet has profile B. The reduction of LR′ is the same as above, so we again find that

$$\text{LR}' = \frac{P(F'_1 \mid F_2 \text{ and } H'_p)}{P(F'_1 \mid F_2 \text{ and } H'_d)}.$$

The numerator is again p_B. The denominator, however, is different now. Indeed, since we specify the types of the two stains, the probability that we do find A on the pillow and B on the sheet is not $2p_Ap_B$, but just p_Ap_B. Hence, we find that

$$\text{LR}' = \frac{1}{p_A},$$

which is quite different from LR. The quote of Aitken loses its significance, since H'_p is just as natural as H_p.

The likelihood ratio for the third set of hypotheses, denoted by LR″, is different from both LR and LR′. To compute LR″, we need the prior probability δ that John Smith was involved in the crime (and therefore left one of the stains). Indeed, we still have

$$\text{LR}'' = \frac{P(F'_1 \mid F_2 \text{ and } H''_p)}{P(F'_1 \mid F_2 \text{ and } H''_d)},$$

and the numerator is p_B as before. However, in the denominator, we compute the probability that the stains on the pillow and sheet have profile A and B, respectively, if we know that John Smith has profile A and someone else left the stain on the pillow.

The probability that Smith left neither of the two stains, *given* that he did not leave the pillow stain, is the ratio of the probability that he left neither of the stains and the probability that he did not leave the pillow stain. The first probability is $1-\delta$, and the second is $\delta \times 1/2 + (1-\delta) = 1-\delta/2$, where the first term comes from the possibility that Smith was involved, but left the other stain, and the second comes from the possibility that he was not involved at all. Hence, under the assumption that Smith did not leave the pillow stain, the probability that he left no stain at all is $(1-\delta)/(1-\delta/2)$. So, given that Smith did not leave the pillow stain, the probability that he left no stain at all is $(1-\delta)/(1-\delta/2)$ and then the probability of finding the two profiles A and B is p_Ap_B. If Smith did leave the sheet stain; then the probability of finding B on the sheet is zero, since Smith has profile A himself. We conclude that the denominator of the likelihood ratio is equal to $p_Ap_B(1-\delta)/(1-\delta/2)$, leading to

$$\text{LR}'' = \frac{(2-\delta)}{2p_A(1-\delta)}.$$

So here is the problem: we have a number of natural sets of hypotheses, and each of these gives a *different* likelihood ratio. Yet it is precisely such likelihood ratios that are used as a measure of the strength of the evidence against suspect John Smith. Our problem is abundantly clear: which set of hypotheses is the "correct" one? If we do not know whether this question has a meaningful answer, how do we know which likelihood ratio we have to use? (See also Koehler, 1996, for a related discussion.) It should be clear from our discussion that likelihood ratios are not appropriate for the two-stain problem, and they are not appropriate for the database search problem either, for the very same reason.

3. Posterior Odds instead of Likelihood Ratios

The fact that the three sets of hypotheses lead to different likelihood ratios is at first blush alarming, but, as already discussed in Dawid (2001), not really surprising when you look at the role played by likelihood ratios in Bayes' rule. To apply Bayes' rule, we need *prior odds* for one hypothesis against another; the *posterior odds* are then found via

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds}. \tag{1}$$

Let us now see what the posterior odds for the three sets of hypotheses are.

For the first case, we have prior odds $\delta/(1-\delta)$. The posterior odds are then found with (1) to be

$$\begin{aligned}\text{LR} \times \text{prior odds} &= \frac{1}{2p_A} \times \frac{\delta}{1-\delta}\\ &= \frac{\delta}{2(1-\delta)p_A}.\end{aligned}$$

For the second case, the prior probability that John Smith left the stain *on the pillow* is simply $\delta/2$, so $P(H_p) = \delta/2$. The probability that two unknowns left the stains is $1-\delta$; hence, the prior odds are $(\delta/2)/(1-\delta)$. The posterior odds are then

$$\text{LR}' \times \text{prior odds} = \frac{1}{p_A} \times \frac{\delta/2}{1-\delta} = \frac{\delta}{2(1-\delta)p_A},$$

which is the same as for the first case!

Finally, the priors for H''_p and H''_d are $\delta/2$ and $1-\delta/2$, respectively, giving prior odds of $(\delta/2)/(-\delta/2)$. The posterior odds are

$$\text{LR}'' \times \text{prior odds} = \frac{(2-\delta)}{2p_A(1-\delta)} \times \frac{\delta/2}{1-\delta/2} = \frac{\delta}{2(1-\delta)p_A}.$$

Again we get the same answer. The reason that all posterior odds are the same is that the hypotheses pairs are conditionally equivalent, that is, they are logically equivalent given the evidence (Dawid 2001).

What do we learn from this? The above computations confirm the claim of Dawid (2001) that

> ...one should avoid talk of "the likelihood ratio" as if this term designated a well-defined objective measure of evidence: at best, it can only be regarded as such relative to a chosen specification of the hypotheses.

The above analysis shows that there are situations where there is no obvious single "relevant" pair of hypotheses. We consider it misleading to choose one of the possible hypotheses pairs, and report merely its likelihood ratio as a measure of evidential strength for these hypotheses. Jurors may tend to focus on this likelihood ratio, and not know about counterbalancing prior odds, since these are not mentioned. Fortunately, the posterior odds are invariant for pairs of conditionally equivalent hypotheses, and they incorporate the prior odds. We therefore agree with Dawid (2001) that the posterior odds are more meaningful than a likelihood ratio.

At this point, we should elaborate on the distinction between strong *evidence* and a strong *case*. We have been very careful in using of these two words, since they express something different: the likelihood ratio is a measure of the strength of the evidence, while the posterior odds are an indication for the strength of the case. Hence, when priors are exceedingly small, strong evidence need not lead to a strong case. The discussion so far suggests that we need to look at the strength of the case, not of the evidence.

4. Consequences for the Database Controversy

We now return to the database search problem, where John Smith was the only match with the DNA profile at the scene of the crime.

It is clear from our discussion above that as far as the strength of the *case* is concerned, it doesn't matter which set of hypotheses is used. We can confirm this by computing the posterior odds for both sets of hypotheses (see Stockmarr, 1999; Dawid, 2001). The prior odds for H_p against H_d are easily seen to be $\delta/(1-\delta)$, and the prior odds for $\tilde{H}_p$ against $\tilde{H}_d$ are $\pi/(1-\pi)$, both leading to the *same*

$$\text{posterior odds} = \frac{1}{p_A}\frac{\delta}{1-\pi}. \qquad (2)$$

As mentioned before, the reason for this is that the two sets of hypotheses are conditionally equivalent.

How can focusing on the posterior odds now settle the database controversy? The original basis for the database search controversy was the difference between the case of a suspect already identified on other grounds ("probable cause"), and a suspect found through a database search. In our opinion, the main difference between these cases is in the prior odds (as explained by, e.g., Donnelly and Friedman, 1999). In a database search case, there need not be other evidence, and consequently the prior odds can be extremely low. By contrast, in a probable cause case the prior odds should be relatively high. Low *posterior* odds despite a powerful DNA match are therefore expected to occur mainly in database search cases (or similar cases, such as huge mass screens). Hence, the same DNA match that leads to a strong case in a probable cause setting, may lead to a weak case in a database search setting. If the forensic expert reports merely Balding and Donnelly's (1995) likelihood ratio, the juror may not know about the counterbalancing prior odds, and thus overestimate the strength of the DNA case against Smith. On the other hand, if the expert reports merely NRCII/Stockmarr's likelihood ratio, the juror may not know how to use evidence from a database ($\tilde{H}_p$ versus $\tilde{H}_d$) in a case against Smith, and might as a consequence underestimate the strength of the DNA case against Smith. Fortunately, we can avoid these misinterpretations by reporting the posterior odds, which are of direct interest to the court, and this should be acceptable to all discussants.

From our point of view, it is obvious that the database controversy is a *false* controversy. In the controversy, people argue about which pair of hypotheses is the "correct" one. We have argued that there is no single "correct" or "relevant" hypotheses pair. The only interesting quantity is given by the posterior odds, which should then make it clear that there is no dilemma at all. We end this section with three remarks.

1. Note that the posterior odds *increase* as we add more people to the database. Hence, a larger database means a stronger case against suspect Smith, but the reason for this is not that the likelihood ratio increases (as claimed in Donnelly and Friedman, 1999), but that the posterior odds increase. We emphasize, however, that the effect will be negligible in most situations, in particular when the database is not so large and π is very small.
2. Evett, Foreman, and Weir (2000) claim that Stockmarr's (1999) hypotheses give an answer to the wrong question. They claim that it is not so relevant to the case against John Smith whether or not the criminal is in the database, but so far as John Smith is concerned, it is relevant. But the above discussion makes it clear that this is not an issue. Since the likelihood ratio is not the quantity we have to look at, Stockmarr's (1999) approach, if used with posterior odds, is just as good as the hypotheses suggested by Balding and Donnelly (1995).
3. Dawid (2001) explains that the difference between Stockmarr and Balding/Donnelly is superficial and, furthermore, warns that the likelihood ratio is not a well-defined objective measure of evidence. However, it escapes us as to why Dawid nevertheless does recommend using a mere likelihood ratio in the courtroom.

5. What to Report? A Suggestion

Our approach rests on the combination of prior odds and likelihood ratios. The latter can be determined by the forensic expert or the statistician, but the former is typically outside the province of the expert. How do we treat these prior odds, then?

Well, since the expert has nothing to say about the priors in an actual case, we think that the best he or she can do is to provide the juror with a table relating a range of priors to the corresponding posteriors. (This was suggested earlier; see, e.g., NRCII 1996.) The juror can then see to what posterior odds his or her prior odds lead. We illustrate this with two examples.

Example 1. Consider the two-stain problem, where the profile A of the suspect and the pillow stain has match probability 1 in a billion (10^9). The forensic expert can now report the information in Table 1.

A juror can now see that the posterior odds are much larger than the prior odds; the match of the profile does indeed strengthen the case against John Smith. However, it is also clear that if other evidence is completely lacking, very weak, or in favor of the suspect, which results in small prior odds,

Table 1

The relationship between prior and posterior odds of "the suspect left the pillow stain" versus "two unknowns left the stains" in the two-stain case, with p_A equal to one in a billion, assuming uniform priors and no relatedness

Prior odds	Posterior odds
10^{-9}	1
10^{-6}	1000
10^{-3}	10^6
10^{-2}	10^7
1	10^9
10	10^{10}
100	10^{11}

then the posterior odds are larger, but perhaps still too small to convict John Smith.

Example 2. As a second example, we consider the case of the murder of a young girl in the Netherlands. A partial profile of a male individual was recovered from dirt under her fingernails, and the match probability of that profile A was about 1/4000. A database of 600 persons was searched, and a single match, Mr. C., resulted. Note here the problems with the use of likelihood ratios: according to, for instance, Balding and Donnelly (1995), the likelihood ratio is about 4000, while the Stockmarr (1999) approach yields a likelihood ratio of about 7, a dramatic difference. The expert, who was aware of the controversy and unable to decide which figure to report, decided to report the match probability of 1 in 4000, as well as the size of the database.

Clearly, to compute the odds in this case, one needs the prior probability δ that John Smith left the stain, and the prior probability π that the donor is in the database. In the Netherlands, the juror is allowed to know how the suspect was found; he (the juror) therefore knows that in the above-mentioned case, the suspect was found as the only hit in a database search. In principle, it is therefore possible to ask the juror his or her prior probabilities δ and π and create a table as above, for various combinations of δ and π. This, however, might be confusing for a juror. It might be effectively impossible to make a prior simultaneous statement about δ and π. Moreover, there are also legal systems where the juror is not allowed to know that the suspect is in a database.

Therefore, we suggest the following: the expert uses δ, the prior probability that John Smith left the stain, and then *computes* π, the probability that the donor is in the database, by using δ, plus the conservative assumption that all other people in the population are equally likely to be the donor. This leads to

$$\pi = \delta + (1-\delta)\frac{n-1}{N-1}.$$

For this, the size N of the relevant population is needed, and this value may be discussed with the court or, alternatively, multiple tables for various values of N can be reported.

Table 2 is an example of a table, where $n = 600$, $N = 10{,}000{,}000$, and the match probability of the stain is 1 in 4000.

A number of remarks on Tables 1 and 2 are appropriate:

Table 2

The relation between prior and posterior odds of "John Smith left the stain" versus "someone else left the stain" in the database case, with $n = 600$, $N = 10{,}000{,}000$, and $p_A = 1/4000$, assuming uniform priors and no relatedness

Prior odds	Posterior odds
0.00001	0.04
0.0001	0.4
0.001	4
0.01	40
0.1	400
1	4000
10	40,002
100	400,024
1000	4,000,240

1. This format is also suitable for combining the DNA evidence with other kinds of evidence, for example, witness statements. The juror could use the other evidence, either to choose prior (first column) odds, or to update the posterior odds (second column). The expert should carefully make clear in the report how the table should be used and interpreted, and also explain the underlying assumptions.
2. The numbers in the second column are independent of which set of conditionally equivalent hypotheses has been used.

6. A Comparison between Two-Stain and Database

Finally, it is interesting to compare the hypotheses chosen in the database search problem and in the two-stain problem. A school of thought in forensic statistics, dubbed the "likelihood ratio approach to interpreting evidence," advocates choosing the hypotheses summarized in Table 3 (see, e.g., Evett, 1987; Evett et al., 2000).

There are two interesting inconsistencies in this approach when comparing the two problems:

1. In the two-stain problem, only *summarized* evidence is used: indeed, the only information about the stains that is used is the fact that there is a stain with profile A; it is not specified that this is the stain on the pillow. On the other hand, in the database search problem, *fully described* evidence is used, namely, the profiles of all people in the database, instead of summarizing the evidence as "only one profile in the database matches profile A."
2. In the two-stain problem, the hypotheses are data independent, since we can formulate them before it is known that Smith matches the stain on the pillow. By contrast, in the database search problem, the hypothesis that Smith is the donor is data dependent.

As a result of these differences, in the two-stain case the likelihood ratio is divided by a factor 2, while in the database case, the likelihood ratio is *not* divided by n (in the case of uniform priors). This clearly is potentially confusing.

It is clear from the discussion that concentrating on the *case* rather than on the evidence, that is, on the posterior odds rather than on a likelihood ratio, avoids this confusion.

Table 3
Choice of hypotheses advocated by the likelihood ratio approach to interpreting evidence, and resulting odds and likelihood ratio, assuming uniform priors

	Two-stain	Database search
H_p	Smith is donor of one of the stains	Smith is donor
H_d	Smith is not donor of either stain	Unknown is donor
Prior odds	$2/(N-2)$	$1/(N-1)$
Evidence	• Smith has profile A • One stain has profile A • One stain has profile B	• Smith has profile A • Profiles of all people in database
Likelihood ratio	$1/2p_A$	$(N-1)/(N-n)p_A$
Posterior odds	$1/(N-2)p_A$	$1/(N-n)p_A$

Acknowledgements

We are grateful to Ton Broeders, Andre Hoogstrate, Aart Spek, and Hans de Moel from the Netherlands Forensic Institute for many helpful discussions.

Résumé

La force probante d'un DNA compatible dépend elle de la façon dont le suspect a été identifié: par recherche dans une base de données ou par d'autres preuves ("cause probable")? Dans les articles de Balding et Donnelly (1995) et d'autres auteurs, il est argumenté que celle-ci est légèrement plus élevée dans les identifications par recherche dans une base de données que par cause probable, alors que Stockmarr (1999) aboutit à la conclusion inverse. Ces deux approches utilisent le rapport de vraisemblance. En examinant un problème similaire, le problème des deux taches, nous argumentons dans cet article qu'il y a certaines difficultés fondamentales avec le rapport de vraisemblance qui peuvent être évitées en utilisant les odds a posteriori. Cette approche permet de résoudre la contradiction énoncée ci-dessus.

References

Aitken, C. G. G. (1995). *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester, U.K.: Wiley.

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

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

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

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

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

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

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

Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review* **67**(1), 859–886.

NRCII (1996). *The evaluation of forensic DNA evidence*. Report, National Research Council. Washington D.C.: National Academy Press.

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

Received October 2002. *Revised March* 2003.
Accepted March 2003.