After the DNA Wars: Skirmishing with NRC II

Richard O. Lempert
University of Michigan Law School, rlempert@umich.edu

Available at: https://repository.law.umich.edu/articles/2374

Follow this and additional works at: https://repository.law.umich.edu/articles

Part of the Evidence Commons, and the Forensic Science and Technology Commons

Recommended Citation
AFTER THE DNA WARS:
SKIRMISHING WITH NRC II

Richard Lempert*

ABSTRACT: This article traces some of the controversies surrounding DNA evidence and argues that although many have been laid to rest by scientific developments confirmed in the National Research Council’s second DNA report, there remain several problems which are likely to lead to continued questioning of standard ways prosecutors present DNA evidence. Although much about the report is to be commended, it falls short in several ways, the most important of which is in its support for presenting random match probabilities independent of plausible error rates. The article argues that although one can sympathize with the NRC committee’s decision as an effort to say no more than what science reliably tells us, it is not a good forensic science recommendation because following it means that the probative value of DNA evidence is likely to be substantially overstated. Fortunately, it will be the rare case where this matters.


The "DNA Wars," we are told, are over. Two of the key, and at times most effective, participants in the battle, the FBI’s Bruce Budowle and the leading early scientific skeptic, Eric Lander, have declared a private truce and

*Richard Lempert is the Francis A. Allen Professor at the University of Michigan Law School and Professor and Chair in the University of Michigan’s Department of Sociology. He served on the committee that produced the National Research Council’s first DNA report.

Professor Lempert notes that the paper began as a presentation at meetings jointly organized by the Eastern North American Region of the International Biometric Society, the Institute of Mathematical Statistics, and the American Statistical Association at Richmond, Virginia, in March 1996. He is grateful to Shari Diamond and David Kaye, both of whom served on the committee that produced the National Research Council’s second DNA report, for their comments on an earlier draft of this paper and also to Jonathan Koehler and Michael Sobel for their comments. A slightly different version of this paper appears in the Israel Law Review under the title, After the DNA Wars: A Mopping Up Operation.
Lempert suggested that all are included. The controversial report of the first National Research Council Committee on DNA Evidence has in crucial respects been replaced by a second report that has debuted to far better reviews than its predecessor and is likely to erase the earlier report's influence as courts grapple with DNA evidence. Yet NRC II, despite its virtues, does not adequately resolve all questions about the significance of DNA evidence, and some veterans of the DNA wars are not yet content to lay down their arms. We can better understand why the Landers-Budowle truce and NRC II may not resolve all conflict if we first understand why disputes over the population genetics and statistical issues raised by the forensic use of DNA identification evidence became so heated that antagonists and observers came to speak of the "DNA Wars."

One explanation for the heatedness of the scientific dispute over the probative value of DNA evidence is that it was fueled by passions common when civil libertarians are arraigned on one side of an issue and law and order advocates on the other. Although some prosecutors who fought for the admissibility of DNA evidence acted as if those who questioned the scientific case for DNA evidence were opening prison cells for guilty defendants, I do not believe these differences were the prime motivators of the scientific debate. Among the scientists who helped make the case for DNA evidence, I am sure there are many who are devoted civil libertarians, and among the ranks of those who have questioned aspects of the forensic use of DNA evidence, there is no one I know who takes pleasure in seeing the guilty go free. Nor do I think


2. NATIONAL RESEARCH COUNCIL COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) [hereinafter NRC I]. I served on the panel that produced this report. I have also been critical of certain of its aspects. Richard Lempert, DNA, SCIENCE and THE LAW: Two Cheers for the Ceiling Principle, 34 JURIMETRICS J. 41-57 (1993).


4. See B.S. Weir, The Second National Research Council Report on Forensic DNA Evidence, 59 Am. J. Hum. Genetics 497-500 (1996) (invited editorial). Although Weir quarrels with some aspects of the report, he concludes: "The 1996 report will be a valuable resource to the forensic and legal communities. The authors of the 1996 report are to be congratulated on their efforts to make recommendations on the basis of scientific arguments. They have done much to help the proper calculation and presentation of DNA-profile statistics, and the day on which DNA profiles are employed with the same trust as are fingerprints has surely been brought forward by their report." Id. at 500. An early news report in Science notes that NRC II is distinguished from NRC I in that its release seems to have set off no fierce debates. The report notes that "DNA forensics experts like prosecutor Rockne Harmon of Alameda County, California, have embraced these guidelines as 'reasonable'" and concludes "But for the most part forensics experts say, the new NRC rules offer a rationale for practices that the courts are already adopting." Elliot Marshall, Academy's About-Face on Forensic DNA (National Research Council Report on DNA Fingerprinting), 272 Science 803, 803-04 (1996).


440 37 JURIMETRICS
scientists who have opposed each other in court necessarily differ on the value they place on the two types of error the legal system can make: acquitting the guilty or convicting the innocent. If that were the issue, DNA's advocates would win hands down. DNA is an extremely valuable identification technique—far better, for example, than eyewitness testimony, which often is treated as dispositive in determining "who did it." I know no critic of DNA evidence who wants to return to a pre-DNA world. All, I think, agree that DNA evidence leads to more correct convictions of the guilty and fewer mistaken convictions of the innocent than occurred before DNA's arrival on the scene.

What has motivated my past criticisms of how DNA has figured in trials, and, I believe, the criticisms of most of those who have urged caution in ascribing weight to DNA evidence, is a more aesthetic criterion: namely, evidence presented to a jury with the convincing rituals and trappings of scientific evidence, should be good science. This means that the assumptions on which a scientific analysis proceeds should be well-grounded in empirical data, and if statistical inferences are made, the inferences should be fair ones that accurately capture the weight of the evidence as it pertains to questions juries must answer. Critics acknowledge that DNA evidence should be admitted, and most agree that statistics can help juries understand how to weigh DNA evidence, but, the critics hold, the evidence must be presented to juries in such a way that its uncertainties are not obfuscated nor its implications oversold.

DNA's advocates, on the other hand, believe that from a scientific standpoint most doubts about DNA evidence are at best overstated and at worst mere chimera. They can point to studies that questioned the assumptions underlying statistical methods for determining the weight of DNA evidence but that were obviously flawed or that did not hold up to replication. And they can point to theoretical objections to DNA evidence that proved to be empirically unproblematic. In short, I believe that scientists on both sides of the DNA debate sometimes wrote with passion because, as scientists, they are ordinarily passionate about getting things right, and never more so than when core values like lives and justice are at stake.

Scientists in both camps can point to trials where "expert" testimony for or against the admissibility of DNA evidence seemed scientifically unjustified. In some trials, for example, defense experts take potential problems with DNA evidence that are at best hypothetical and quite unlikely in the context of cases and suggest that these hypothetical difficulties pose substantial danger to the reliability of DNA identifications. Prosecution experts, on the other hand, occasionally misstate or let prosecutors misstate the import of their testimony—suggesting, for example, that if a defendant's DNA matches DNA collected

---

6. However, the very fact that scientists came regularly to oppose each other in court or to write articles that might be used against each other in court may explain some of the conflict's intensity.
Lempert

at a crime scene and there is a one in a million chance that a randomly selected person would have matching DNA there is only a one in a million chance that someone other than the defendant committed the crime. There are even cases where prosecution experts have suggested that false positive error is impossible in DNA analysis.

This last claim I once found particularly puzzling, for I do not believe this claim can be made about any technique of forensic science. Then I met a forensic scientist who told me that false positive error never occurred. Questioning her, I learned that she meant that when intragel RFLP tests were run, the DNA in the evidence and suspect lanes would not match unless the DNA was in fact matching. In this forensic scientist’s vocabulary, the test results were not wrong when a match was called because the suspect’s DNA had contaminated the evidence DNA or as a result of pure coincidence. In either case, according to this scientist, the test told the truth—the DNA in the suspect and evidence lanes in fact matched, whatever the reason.

I. THE JURY’S CONCERN

A scientist who testifies that false positive error never happens does not address the question the jury needs answered—namely, how likely is it that a match would be reported if the evidence DNA was not the suspect’s. Moreover, she risks misleading or confusing the jury because she uses language in a specialized way that invites misinterpretation. All I can say in defense of the scientist I talked to is that she seemed honestly to believe what she said, and no doubt would have explained her denial of the possibility of error to a jury as she explained it to me, if only opposing counsel asked the right questions.

The flaw in this scientist’s assertion—that it did not address the question the jury had to answer—is, I think, the most common flaw in statistical testimony on the import of DNA matches. Although DNA statistical evidence is ordinarily presented in terms of the probability of coincidental matches, let me use likelihood ratios to make my point. The likelihood ratio to which scientists

7. Arguments of this type have come to be known as "the prosecutor’s fallacy." See W.C. Thompson & E.L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defense Attorney’s Fallacy, 11 LAW & HUM. BEHAV. 167-87 (1987). To appreciate the fallacy, consider a man who is in jail at the time a rape occurs in the community. His DNA may match the DNA extracted from the rapist’s semen and only one in a million people may have similar DNA, but if we are sure that the man was in jail at the time of the rape, it is certain that someone other than he committed the crime.

The prosecutor’s fallacy need not reflect prosecutorial bias, for there are instances where defense counsel similarly transpose the conditional, suggesting that fallacious reasoning of this type can result from the unbiased misunderstanding of statistics. Thus, NRC II uses to the more general phrase, "the fallacy of the transposed conditional." NRC II, supra note 3, at 153, 198.

(implicitly) testify; that is, the ratio implied by the often vanishingly small probabilities they present, is

\[ \frac{P(E|H)}{P(E|\neg H)} \tag{1} \]

where \( E \) is the event that the DNA in the evidence and suspect lanes of a gel match, and \( H \), the hypothesis in dispute, is that the suspect was the source of the crime scene DNA. Since identification is not the only element of guilt, a suspect may be the source of crime scene DNA and still be innocent. As a practical matter, however, in most cases in which DNA evidence has been used, identification has been tantamount to guilt, for the other elements of the crime have been conceded. So, from the usual jury point of view, we can write the likelihood ratio for DNA evidence as:

\[ \frac{P(\text{DNA Match} | \text{Defendant's Guilt})}{P(\text{DNA Match} | \text{Defendant's Innocence})} \tag{2} \]

or the probability that there would be a DNA match given that the defendant is guilty divided by the probability that there would be a DNA match if the defendant were innocent.

In presenting statistics reflecting this ratio, the state’s witnesses are speaking a specialized scientific language which, like the testimony of the forensic scientist I mentioned, does not address the question the jury is interested in, and, in fact, does only a little better than the scientist I described in allowing for the possibility of false positive error. It does somewhat better because it allows for the possibility of a coincidental match, since the denominator of the ratio is the chance that a randomly selected person would have DNA matching the evidence DNA. The numerator of the ratio is conventionally taken as one, though in practice the fact that a defendant is guilty does not guarantee a DNA match.

From a legal standard, the appropriate likelihood ratio is different. Ratio (2) assumes the jury knows for a fact that the DNA left at the crime scene matches the defendant’s DNA, but the jury does not know this fact. The jury only knows that someone, presumably a reputable scientist, testified that an RFLP or PCR test showed that, by some criterion, DNA that supposedly came from the crime scene matched DNA that supposedly came from the defendant. The jury relies on the witness’s credibility to conclude not just that the DNA matched, but also that the sources of the reportedly matching DNA were the suspect and the crime scene. In this respect, DNA evidence is no different from most evidence a jury receives. The jury seldom knows at first hand facts relevant to a case; it knows only that witnesses have reported certain facts.

For this reason, the likelihood ratio a jury hearing DNA evidence should estimate is not likelihood ratio (2) but rather:
This ratio differs from ratio (2). Consider, for example, O.J. Simpson's case. Had DNA statistics been presented as likelihood ratios, some would have been exceedingly small. Making the conventional assumption that the numerator is one, one prosecution expert testified with respect to one blood sample to a likelihood ratio of one to a trillion, since the expert claimed there was only a one in a trillion chance that a random man's DNA would match the evidence DNA.10 Other implied likelihood ratios were also so low as to seem impossible if Simpson were innocent. But these likelihood ratios assume a denominator that is an established fact rather than a report of a fact. There are, however, at least five ways in which a match might have been reported, assuming that Simpson was innocent. The first is that the real killer, assumed to be a random person, had matching DNA. It is this typically low probability, and only this probability, that the jury is given. The second is that the scientist who reported the match was lying about her observations.11 A third possibility is that the source of the evidence DNA was not a "random person," but a relative of O.J.'s, for example, his child. The fourth is that a laboratory error is responsible for the reported match. The fifth possibility is inadvertent or intentional police contamination of evidence.

To illustrate the difference between the likelihood ratios for actual and reported facts, consider in Simpson the report that DNA extracted from the blood on a sock found in O.J.'s bedroom matched the DNA of Nicole Brown Simpson. Supporting the defense argument that the match was due to police tampering was evidence tending to establish that: (1) two police officers in a position to plant blood on the sock were apparently willing to lie under oath to aid the prosecution's effort.12 (2) The sock was examined three times without

---

9. Even this ratio, which reflects the question the jury confronts, is not quite right in RFLP testing, for it incorporates binning procedures that treat closely matching DNA samples as having the same evidentiary weight as more distantly matching samples so long as in both cases the evidence and suspect samples fall within a prespecified range. I.W. Evett et al., An Efficient Statistical Procedure for Interpreting DNA Single Locus Profiling Data in Crime Cases, 32 J. FORENSIC SCI. SOCY'Y 307-26 (1992); I.W. Evett et al., An Illustration of the Advantages of Efficient Statistical Methods for RFLP Analysis in Forensic Science, 52 AM. J. HUM. GENETICS 498 (1993). On binning, see NRC II, supra note 3, at 142-48.


11. The recent scandal about biased or misleading testimony from FBI scientists makes this possibility appear more plausible than I imagined when I first noted it. See R.L. Jackson & D.G. Savage, FBI Warns of Possible Flaws in Lab Evidence: Courts, Prosecutors, Defense Counsel Nationwide are Told of Potential Problems Due to Alleged Misconduct, LOS ANGELES TIMES, Jan. 31, 1997, A1.

12. Officer Fuhrman's denial at trial that he ever used the word "nigger" was apparently a lie as was Officer Vannatter's statement in a preliminary hearing that O.J. was not a suspect when the police went to his house the night of the killing.
After the DNA Wars: Skirmishing with NRC II

anyone noticing blood, yet the stain when noticed, was larger than a quarter and crusted so that the sock material puckered; (3) the blood had soaked through one side of the sock and left a wet transfer stain on the other side. This was impossible while the sock was on O.J.’s ankle, but might well happen if blood had been dripped on the sock while it was lying flat. It could also have happened had the blood on the sock still been liquid when Simpson took it off, but a defense expert testified that had Simpson acquired the blood at the crime scene, it would have dried by the time he took the sock off. (4) The chemical preservative EDTA, present in the LAPD test tubes in which the victims’ blood was stored, was found in the blood on the sock. The FBI agent who examined the blood admitted there was evidence of EDTA but thought there was too little for the blood to have been in an LAPD test tube. A defense expert testified that the EDTA level found could have characterized blood taken from a LAPD reference tube but was too high for the blood to have been shed by a living human being.

Given the defense’s evidence favoring a police frame up, how likely is it that the blood on the sock would have been reported to match Nicole Brown Simpson’s blood had Nicole not shed it? Even one who trusts the police is likely to estimate the chance of their malfeasance as many orders of magnitude higher than the chance that the DNA of a randomly selected person would match Nicole Brown Simpson’s DNA at four or five loci. O.J.’s defense team claimed that without any need to posit a police conspiracy, a similar story, pointing either to planted evidence or unintentional contamination, could be told about every drop of the prosecution’s blood evidence. Even if the set of defense stories appear unlikely to be true, the collective truth of the possibilities they identify is far more likely than the chance that by sheer coincidence someone with DNA identical at crucial loci to O.J. Simpson’s shed blood at the crime scene while persons with DNA identical to Ronald Goldman’s and Nicole Brown Simpson’s shed blood that found its way to O.J. Simpson’s back

13. Arguments 2, 3, and 4 are taken from William C. Thompson, DNA Evidence in the O.J. Simpson Trial, 67 COLO. L. REV. 827 (1996). Professor Thompson is a Ph.D. psychologist who has acquired considerable expertise in and written a number of fine articles on DNA evidence. He was a member of the Simpson defense team.

14. I am framing the matter this way rather than, in terms of Simpson’s innocence because when police malfeasance rather than laboratory error is the possible cause of a spurious match, the likelihood of a defendant’s innocence may affect the probability of malfeasance. However, a jury may not consider as evidence of guilt, the fact that a police officer thought a person was guilty.

15. Thompson, supra note 13.

16. Reading only the defense theories and supporting evidence, as presented by Thompson, supra note 13, the set of defense explanations for the blood evidence seem to me sufficiently likely that if Simpson could somehow prove his innocence (for example, if proof emerged that the killings occurred while Simpson was on the plane), few would be puzzled about how the mass of blood evidence came to incriminate him.

SUMMER 1997 445
gate, his vehicle, his sock and a glove that may or may not have belonged to
him.

Thus in People v. Simpson, the chance that each match in a series of
incriminating DNA matches would be reported was substantially higher than
the likelihood that each match was coincidental.17 The information that would
have best allowed the jury to appreciate the probative value of the different
blood evidence was not the random match probabilities the state's DNA experts
testified to, but the likelihood of intentional or accidental police malfeasance.
Indeed, both the prosecution and defense agreed that the DNA matches in the
case did not occur because the actual killer, by pure chance, had DNA matching
Simpson's or because the blood of persons other than the victims was in
Simpson's Bronco, on his gate and on his sock. The prosecution and defense
disagreed only on their theories of how the DNA in the various samples
compared came to be identical. The prosecution thought it was because
Simpson was a killer, while the defense argued that O.J. was the victim of both
sloppy police work and a frame-up.

II. THE PROBLEM OF ERROR

Although the likelihood that police contaminated the DNA evidence in
People v. Simpson intentionally or through sloppiness cannot be precisely
determined, the jurors are probably as well equipped to estimate these
likelihoods as anyone, and more likely to estimate them accurately than expert
witnesses in the pay of either the prosecution or defense. Simpson is unusual,
however, in that, on the one hand, considerable evidence supports the police
malfeasance and police sloppiness hypotheses, while, on the other hand, the
prosecution took special efforts to guarantee against laboratory error. Where
possible, evidence samples were split and sent to different laboratories for
analysis. Moreover, there were many different DNA samples and different
methods were used to analyze them. The results of virtually all analyses were
what one would have expected had Simpson alone been the killer. It does not
appear that any laboratory made a mistake in reporting incriminating results,
but even if a reported match resulted from a laboratory error, there are enough
reported matches that the chance a laboratory erred in conducting one test
seems of little moment.

Ordinarily, however, laboratory error rather than malfeasance detracts most
from the probative force of a reported match.18 Although a precise probability

17. I overstate a bit here to avoid awkward writing. Some random match probabilities
presented to the jury were quite high because the analyzed DNA was relatively uninformative.
Koehler, supra note 10, at 861 (Table 1). In these cases the match probabilities given jurors were,
in my view, as large or larger than the police malfeasance probabilities. The textual discussion fits
best the multilocus matches, especially those based on RFLP technology.

18. The likelihood that sloppy evidence handling by the police caused the DNA match is also
probably higher than the probability of intentional police malfeasance, but we know even less
about this probability than we do about the likelihood of laboratory error.
cannot be associated with the chance of laboratory error, error rates in other
types of forensic laboratory tests\textsuperscript{19} and the results in the few blind proficiency
tests done to date suggest that false positive error rates in DNA tests must be
many orders of magnitude higher than the random DNA match probabilities
often given juries. If so, the probative value of a DNA match is always limited
by the chance of false positive error. Jurors may not, however, realize this, even
if they are not misled by experts claiming false positive errors cannot happen.
This is why I once suggested that the honest scientist’s testimony ordinarily
should emphasize the likelihood that she is erroneously reporting that the
evidence and suspect samples match, rather than the probability that a random
person left the evidence DNA.\textsuperscript{20} If jurors are given both the false positive error
and random match probabilities, they may tend to see the probative value of the
evidence as somewhere between the two probabilities rather than treating the
chance of a false positive report as limiting the probative value of the reported
match.\textsuperscript{21}

We do not, however, know what DNA laboratory error rates are, and they
are extremely hard to estimate accurately. The obvious way to estimate
laboratory error rates is through blind proficiency testing. Here we encounter,
at the outset, a problem of nomenclature. Although one might think that if a test
is blind a laboratory does not know it is being tested, in the forensic laboratory
world a blind test can refer to a situation where a laboratory knows it is being
tested and is blind only to what the test results should be. These so-called blind
tests can be expected to understate error rates, for laboratory technicians can
take special care when they know they are dealing with test samples.

Constructing truly blind DNA proficiency tests, which some in the forensic
laboratory world call double blind tests, is difficult and expensive. Not only
must test samples be confounded and degraded as crime scene samples often
are, but agencies that ordinarily submit samples to labs must appear to be the
source of the test samples. This requires contact people who appear to be police
or prosecutors submitting DNA evidence, and the labs must receive the kinds
of crime and suspect information they typically acquire. Almost none of the
proficiency testing done to date or planned for the future is truly blind in these
ways. Moreover, the less susceptible DNA testing is to error, the greater the
number of proficiency tests needed for error to emerge. Thus, absent a legal
mandate and infusion of funds, one cannot expect to generate reliable,

\textsuperscript{19} Joseph Peterson et al., Crime Laboratory Proficiency Testing Research

\textsuperscript{20} Richard Lempert, The Honest Scientist’s Guide to DNA Evidence, 96 Genetica 119

\textsuperscript{21} Lempert, supra note 2. Some support for this hypothesis has emerged from Jonathan
Koehler’s empirical research. Koehler et al., supra note 8. The Koehler study used a highly
simplified stimulus to elicit probability estimates. It is a typical first step in investigating how
jurors behave, but the study should be replicated with a videotaped trial and mock jurors who
deliberate in order to confirm Koehler’s finding.
Lempert

empirically-based estimated error rates through truly blind proficiency testing. Instead the best estimates of error rates we are likely to achieve will come from tests using casework quality samples of unknown character. Because laboratories receiving these samples will know they are test samples, error rates on these tests may substantially underestimate actual casework error rates.22

Assuring adequate proficiency testing is, however, only part of the problem. Even if proficiency tests were frequent, it can be hard to know what to make of test results in subsequent litigation. Indeed, determining error rates can itself be problematic. In 1988, for example, a California crime laboratory returned one false positive error in 50 casework quality DNA samples they were sent to identify.23 This appears to translate into a 2% error rate as one in 50 samples was incorrectly matched. However, as Kathryn Roeder pointed out, each of the 50 samples was compared with every other sample, meaning that there was just one false positive error made in 1225 comparisons, yielding a more respectable error rate of 0.0816%.24 But Roeder's number understates the error rate because some samples matched in pairs or as triads, meaning false positive errors were not possible in these comparisons. Moreover the false positive that was reported was due to cross-contamination when two samples were stored next to each other during the test preparation. With 50 samples none was proximate to most of the others in the proficiency test, but in actual cases evidence and suspect samples might always be proximate if the proficiency test sample handling procedure were followed. If so, the 2% error rate may be a better estimate of the likelihood of casework error than an estimate based on the number of comparisons that were made. The laboratory, however, responded to its mistake by revising its test protocols to emphasize the need to keep stores of evidence and suspect DNA at some distance from each other. So what is one to make of the likelihood that this laboratory would make a false positive error in the next actual case it handled?

22. The extent to which knowledge exists will depend on laboratory practice. A laboratory does not have to tell its DNA analysts that particular samples are test samples, and if a laboratory routinely keeps crime information from its analysts until the analysis is concluded and a report filed, casework quality samples known by the laboratory but not its analysts to be test samples may adequately substitute for truly blind test samples. Since, as discussed below, there are other good reasons for keeping crime-related information from DNA analysts, this should become standard laboratory practice. A laboratory which did this would still have to resist the temptation to assign its most capable analysts to the test samples, as apparently happened in some early California proficiency tests, and to leak information that the samples are tests. Since a laboratory's business may be harmed by a single proficiency test error, the temptation to alert analysts to tests may be hard to resist.

23. The testing was conducted by CACLD, the California Association of Crime Laboratory Directors and involved three laboratories tested over two years. Another laboratory also made a false positive error and several other reports were questionable. See CALIFORNIA ASSOCIATION OF CRIME LABORATORY DIRECTORS, REPORT TO THE DIRECTORS (1988).

If these complications were not enough, there are additional ones. Not all false positive proficiency test results would yield false positive reports at trials if they involved actual casework. In actual cases, tests may involve several samples of evidence DNA, or the defendant may run his own test. Inconsistent results would alert analysts to the possibility of error.

More subtly, the actual likelihood of false positives for a laboratory of a given quality depends on the true positive base rate in samples submitted for testing. Suppose, for example, that a laboratory prone to make one false positive error for every 10 non-matching samples it analyzes participates in a series of blind proficiency tests. If the laboratory over time tested consecutively 100 non-matching test pairs, ten reported matches would be expected. If 10 matches were in fact reported, it might seem we had an accurate measure of the laboratory’s false positive error rate. But what is that error rate? It is true that false positives were reported in only 10% of the cases analyzed, but among the analyzed cases every reported match is false, so the error rate among reported positives is 100%. It is this probability rather than the 10% error rate that directly concerns the jury. That is, the jury wants to know, given that a match has been reported, how likely is it that it is false. A jury which heard each of the ten mistaken cases might, with but a little additional evidence, convict in most of them if told that the laboratory’s error rate was only 10%. The jurors would reach the right answer only if they learned that all of this laboratory’s positive reports were mistaken. Now take a different laboratory with the same propensity to make false positive errors. Suppose it too was tested with 100 paired samples, but 90 of these pairs had matching DNA while 10 did not. This laboratory would presumably make just one false positive error. If juries were presented with the 91 cases in which this laboratory reported a match and were told that its false positive rate was 10%, there would probably be too many acquittals. If on the other hand, the juries knew that the laboratory was mistaken in only one of 91 matches it reported, they would be more conviction-prone and get more cases right.

Now let us return to the real world of proficiency testing. False positive rates are typically measured by the proportion of non-matching comparisons in which matches are reported. But as the examples illustrate, the implications of this error rate for the likelihood that a particular reported match is erroneous depends on the proportion of true matches in the cases the laboratory analyzes. In the real world, unlike the example we created, this likelihood is not only unknown, but it may also vary with time and place. If in one jurisdiction prosecutors, facing substantial budget constraints, screen out most weakly suspected individuals, even a laboratory with a moderately high false positive rate will report mainly true positives. In the extreme case where the police always get their man, even the most sloppy laboratory will report no false positives. The justice problem will be to avoid dismissing cases when the laboratory reports non-matches. Conversely, if a prosecutor uses DNA testing as a brute screening tool to eliminate weakly suspected individuals, even a
careful laboratory, with low false positive rates as measured by proficiency testing, can be mistaken in more than negligible proportion of its reported matches.

Consider, for example, two laboratories, A and B. Proficiency tests reveal that laboratory A reports a match in 10 of every 100 non-matching cases it is given, while laboratory B is similarly mistaken only once in every 100 non-matching cases. Let us locate laboratory A in a state where prosecutors pay for DNA testing from their own budgets and so seek to test only serious suspects, and laboratory B in a state that pays for local DNA testing and so induces prosecutors to use DNA tests for screening weak suspects. To keep things simple, let us also assume that neither laboratory makes false negative errors; that is all true matches are correctly identified. Laboratory A, let us suppose, receives only 109 cases a year of which 99 involve the perpetrator, and 10 do not. Laboratory B gets more cases involving perpetrators because prosecutors spread their net so wide—let us say 180 such cases—but it also receives the cases of 2,000 innocent men. In these circumstances, only one of 100 matches reported by laboratory A will be erroneous, so only 1% of the matches they report will be mistaken. Laboratory B will report 200 matches and be mistaken in 20 of them, or 10% of the matches it reports. Thus, different guilty base rates among those tested yield reporting error rates that reverse the error rates indicated by proficiency testing.

An implication of this example is that the chance that a reported match is a false positive is greatest when the other evidence in the case is weak. (Prior probabilities of guilt could be substituted for base rates in our example.) Of course, these are the cases in which the DNA evidence is most important to securing a conviction. In typical cases where there is substantial non-DNA evidence implicating a defendant, a report of a DNA match, is almost certainly accurate, but it will often be mere window dressing as far as securing a conviction is concerned.

Thus, we see that estimating what, from the jury’s point of view, is the case-relevant error rate is a more difficult enterprise than one might expect, since it turns on more than administering an adequate number of well designed proficiency tests. Nevertheless, the essential point should not be lost. The probability that a positive report is false in those cases where this possibility most matters; that is, in cases where, apart from a DNA match there is little incriminating evidence, is likely to be much greater than the random match probabilities juries are today given and higher than the false positive rates well-conducted proficiency tests would reveal.

25. If DNA evidence were presented in a Bayesian framework, following the suggestions of Professor Kaye in his article, DNA Evidence: Probability, Population Genetics and the Courts, 7 HARV. J. L. & TECH. 101 (1993), the problem would be theoretically, and to some degree in practice, alleviated because juror estimates of prior probabilities of guilt are estimates of base rate guilt probabilities for individuals, who, apart from a reported DNA match, are implicated by a certain amount of other evidence. Realistically, these estimates may be quite inaccurate.
The report of the National Research Council's second DNA panel,26 is a useful volume and in many ways a fine effort, but it falls down in its treatment of error as it ducks fundamental issues. For example, it discusses the question of whether an error rate should be included in random match calculations without ever specifically discussing the implications of an error probability27 for the probative value of a reported DNA match. It concludes that because it is impossible to know from proficiency tests what the probability of an error is in a given current case, there should be no attempt to develop a statistic that deflates a random match probability given a likelihood of error.28 Indeed, unlike the National Research Council's first DNA report,29 NRC II refuses even to recommend that laboratory error rates, as established in blind proficiency testing, be admitted at trials.30 It does this while advocating procedures for estimating random match probabilities that often will yield results certain to be many orders of magnitude smaller than the chance of a false positive error.

We have here a failure of common sense, best explained by the way the NRC, and perhaps the committee that produced NRC II in particular, understood its mission. NRC panels are typically charged with analyzing and summarizing the state of scientific knowledge in a particular policy relevant area and, if appropriate, discussing the policy or action implications of their findings. The committee that produced NRC I was funded by a consortium of government agencies, headed by the FBI. These agencies, it is fair to say, expected the committee's report to reflect law enforcement's great confidence in DNA technology and to ease problems of court admissibility by categorically rejecting certain challenges that criminal defense attorneys had begun routinely to bring.

To say that the law enforcement community was disappointed by NRC I is a serious understatement. Responding to the uncertainty of core assumptions underlying the "product rule" procedures used to calculate random match statistics and believing that the risk of this uncertainty should be borne by the state rather than by criminal defendants, the first DNA committee recommended techniques for calculating random match probabilities that were extremely conservative.31 Although these techniques could be justified by the values they promoted or perhaps as a second best solution to the problem of

26. NRC II, supra note 3.
27. This is not, in a given case, the same as an "error rate." This difference, between the rate at which a laboratory has historically reported false positives and the probability it is reporting a false positive in a specific current case, is offered by NRC II as a justification for avoiding basic issues. NRC II, supra note 3, at 85-86.
28. Id. at 85-87.
29. NRC I, supra note 2.
30. NRC II, supra note 3, at 185.
31. Even so they could often be expected to yield random match probabilities less than one in 100,000 or even one in 1,000,000, figures which are almost certainly smaller than any laboratory's false positive error rate.
laboratory error,\textsuperscript{32} their scientific justification was from the start weak in that the "random match" probabilities they yielded were likely to be further, and perhaps much further, from true random match probabilities than the probabilities yielded by the product rule procedures the committee's recommendation was designed to replace. Moreover, even before NRC I appeared, studies were beginning to appear indicating that the principal scientific justification for the committee's recommendation, the invalidity of crucial assumptions underlying product rule procedures, was empirically of little moment. When NRC I did appear, its recommendation to replace the product rule with a more conservative "ceiling principle" received considerable criticism from statisticians and population geneticists.\textsuperscript{33}

It was this criticism and the fact that a number of courts were following NRC I's recommendation that led the FBI to take the initiative in sponsoring a second DNA report.\textsuperscript{34} This report was supposed to update NRC I and to "specifically rectify those statements regarding statistical and population genetics issues in the previous report that have been seriously misinterpreted or led to unintended procedures,"\textsuperscript{35} a kind way of saying parts of NRC I needed correction without saying that the prior report's recommendations were wrong. Against this backdrop and the fierce criticism of NRC I for allegedly poor science, it is easy to understand why the committee that produced NRC II ducked the statistical implications of possible error. There is no scientifically reliable way to identify the error probability in a given case, and so in any particular case there is no way to combine error probabilities with random match probabilities to give a precise numerical estimate of the likelihood that evidence DNA would be reported as matching the suspect's DNA if the suspect were not the DNA source.

At the same time, common sense and considerable experience tell us that random match probabilities calculated according to NRC II's recommendation will often be far less than any reasonable lower bound estimate of the likelihood

\textsuperscript{32} Lempert, supra note 2.


\textsuperscript{34} Because many courts applied the \textit{Frye} test, interpreted so as to require a scientific consensus before novel scientific evidence could be introduced, NRC I's recommendation for the application of an interim ceiling principle carried great weight. Even though scientists involved in the forensic use of DNA evidence did not believe its application gave a valid random match probability, all could agree that the random match probability was almost certainly no greater than the result that ceiling principle calculations yielded. Thus there was a scientific consensus that ceiling principle figures were a conservative upper bound but, because of the report, an apparent lack of scientific consensus about the validity of the probabilities yielded by product rule calculations.

\textsuperscript{35} NRC II, supra note 3, at 49.
that a reported match falsely incriminates a suspect. However proficient laboratories are, they are still bound to err more than one in some billions of times. False positive errors have occurred in proficiency tests, and we know of one wrongly incriminating report, caused by the switching of the victim’s and defendant’s reference samples, that came to light during a trial only because of the close scrutiny of chain-of-custody documents by a defense expert.\(^{36}\) Moreover, revelations of falsely incriminating reports emanating from the FBI crime lab and of a West Virginia scandal involving scientific evidence that was made up or manipulated to help win convictions in 36 cases\(^{37}\) suggest that corrupt reporting or even planting evidence occurs with substantially greater frequency than the random match probabilities often associated with DNA evidence.

The fact that there is no scientifically valid way of estimating error or corrupt reporting probabilities does not mean we should, as NRC II suggests, present courts and juries with random match probabilities that pack far more of an incriminating wallop than a reported match justifies. We must recognize that we live in a "second best" world. The scientific defensibility of a number that substantially overstates the deserved import of evidence (i.e., the probability that a match will exist if a "random man" left evidence DNA) does not make that number less likely to mislead a jury than a less scientifically defensible number (the probability of error in a case) that addresses the issue that concerns the jury (the probability that a match would be reported if the defendant did not leave the evidence DNA).

Although there is no way to attach an accurate number to the chance of error in a given case, a "second best" solution exists. One can establish a false positive error rate for a given laboratory, if enough proficiency tests have been done, or for laboratories in general, and based on that rate and the number of tests done establish a 95% confidence interval around the rate, even if it is zero.\(^{38}\) One may then use the upper bound of this confidence interval as an estimated error rate. NRC II gives an example of this sort of solution and then rejects it because it would suggest that a laboratory that makes one proficiency test error is worse than one that makes none even where the two laboratories have the same propensity to err. This objection is, however, beside the point. For a jury is not charged with deciding which of two laboratories is superior, but rather with deciding what the probative value of a reported match is. More importantly, the estimation technique is defensible, for procedures like this are often used in situations of inescapable uncertainty, and the number arrived at will in most close cases be far nearer the actual false positive error probability, which sets a limit on the probative value of a reported match, than the random

---

36. Thompson, supra note 13, at n.35.
37. Jackson & Savage, supra note 11.
38. NRC II, supra note 3, at 86.

SUMMER 1997  453
match probability the jury is now given. In clear cases, an error rate estimated at the 95% confidence interval around proficiency test results will probably yield an error probability that far exceeds the likelihood that a match has been mistakenly reported. But with enough proficiency tests and a presumably low error rate, the number will still be small, and together with the other incriminating evidence that makes a case clear, should guarantee a conviction.

NRC II is on track in its emphasis on the need for those handling DNA evidence to act to reduce the chance of error. Its suggestion to split evidence samples at the earliest possible time and to allow defendants the opportunity for retesting is sound and promises to reduce the likelihood of laboratory error so substantially that error rate estimates should suffice to secure convictions even when DNA evidence forms the bulk of the case. Other steps that NRC II advocates, such as accrediting laboratories and technicians, establishing regular proficiency testing and the like, are being taken, often in response to similar suggestions in NRC I.

A possible further step is to routinely use intergel comparisons for RFLP matches. Currently intragel comparisons are routinely used because gel differences introduce variation which requires larger match windows to avoid missing true matches and this in turn leads to larger random match probabilities across the set of common alleles. The loss of probative value is, however, likely to be dwarfed by the elimination of a potential source of false positive error, the accidental contamination of evidence DNA with suspect DNA. Indeed the use of intergel comparisons would allow samples of evidence and suspect DNA to be handled by different people in separate rooms from the time the samples arrive at the lab. Minimizing the chance of laboratory error in this way would make the probative value of intragel matches higher than that of intragel comparisons. Moreover, if enough DNA is available, intergel matches could be confirmed with intragel comparisons, and statistics based on the latter comparison could then be used.

III. WHOSE DNA?

A second source of statistical overstatement in DNA analysis stems from what the jury is told about the probability of a match if the defendant is not the DNA source. When a DNA match is reported, jurors are almost always told the probability that a randomly selected person of the defendant’s race would have

39. In NRC II’s example, the true error rate of each of two laboratories is .10%. Determining the 95% confidence interval estimates an error probability of .30% for the laboratory that makes no errors in 1,000 proficiency test trials and .47% for the laboratory that makes one error in 1,000 proficiency tests. Either of these figures is likely to be far closer to the true probative value of a reported match than the extremely low random match probabilities that the jurors would otherwise hear.

40. NRC II, supra note 3, at 87.

41. This is because the larger the match window, the greater the proportion of the reference population that will have alleles matching those in the tested DNA.
DNA matching the evidence DNA.\textsuperscript{42} Since NRC I, it is also common to provide jurors with similar figures for members of the country’s other gross racial grouping. Thus when the defendant is white, jurors will typically learn not only of the chance that a randomly selected white person would have DNA matching the evidence DNA, but also of the chances that a randomly selected black or Hispanic person would have DNA matching the evidence DNA. These estimated chances rest on the analysis of allele frequencies in different convenience samples, for example, donors to blood banks in a particular city or FBI agents in training. Within each sample, the distribution of allele sizes at the loci analyzed are calculated for different ethnic groups so that for each major ethnic group a frequency is attached to each allele size at each locus. The frequencies attached to the different allele sizes are then multiplied together following the product rule to give an overall probability that a random person would have the same allele combination as that found in both the evidence sample and in DNA taken from the defendant.

This procedure rests on two false assumptions. The first is that the convenience samples from which allele frequencies are calculated are random subsamples of some larger racial population to which the frequencies are attributed. The second is that the Hardy-Weinberg and linkage equilibria that justify the application of the product rule actually exist. Yet, for the alleles commonly used in RFLP analysis, the untenability of these assumptions seems hardly to matter. Empirical studies suggest that conservatism in estimating allele frequencies in the first instance can more than make up for any prejudice an accused suffers from the untenability of the assumptions.

No statistical principle requires juries to be presented with probabilities based on allele frequencies for the country’s three major ethnic groups. Neither is an older procedure, in which only allele frequencies for the defendant’s ethnic group were given, statistically sounder. Rather the allele frequencies of interest are those that characterize the population of possible suspects.\textsuperscript{43} In particular, if the arrest of a suspect is conditioned on or likely to be confined to persons having certain characteristics, allele frequencies should ideally be based on samples of people having that characteristic. Thus if a woman is certain that her assailant had a Polish accent and an immigrant Pole is arrested for the crime, the allele frequencies used to estimate how unusual his DNA pattern is should be based on data generated from a Polish speaking sample. In a case where nothing is known about the perpetrator except that he probably resides somewhere near the crime scene, allele frequencies should be based on a sample that reflects the ethnic mix in the locality in question. Only where the perpetrator’s ethnicity is known before arrest, should the jury be presented with

\textsuperscript{42} "Race" as used by DNA analysts is more of a sociological than a biological concept. Most races DNA analysts recognize are socially constructed gross categorizations, like black and white.

Lempert

DNA results for the defendant's ethnic group. But the practice of presenting juries with information derived from samples of the defendant's coethnics or of presenting a jury with the statistics applicable to the country's major ethnic groups is a harmless one and most likely conservative in its implications. Moreover, even if it does not always make statistical sense to present the jury with statistics drawn from a black sample in the case of a black defendant and from an Hispanic sample in the case of an Hispanic defendant, it makes a kind of common sense and may well avoid juror suspicion of statistics that seem not to take something as salient as the defendant's race into account. Thus, these scientific flaws in estimating relevant allele frequencies matter little.

The serious problems that techniques for calculating allele frequencies raise arise because the question a State's DNA expert answers for the jury—What is the probability that a random man left the evidence DNA?—is not the question of interest to the jury. Rather the jury wants to know the probability that some person other than the defendant left the evidence DNA. The answer to this question differs from the answer to the "random man" question because some people may be far more likely to be the source of evidence DNA than an individual sampled randomly from either the world at large or from the defendant's ethnic group. These people also may have a far greater likelihood than a random man of having DNA that is indistinguishable from the defendant's. Suppose, for example, a man is accused of raping his sister-in-law, and the man's DNA matches DNA extracted from sperm taken from the woman's vagina. It hardly matters that a random person would be unlikely to have matching DNA. When deciding whether someone other than the defendant might have left the evidence DNA, the jury would be most concerned not with the likelihood that the evidence sample came from some random man, but with the chance that it came from the victim's husband, the defendant's brother. This likelihood will be many times greater than the likelihood that a random man would have left the evidence DNA. Consider also the Simpson case. If O.J. did not shed the blood drops found in Nicole Brown Simpson's driveway, they are far more likely to have been shed by O.J.'s younger son, who might have skinned a knee playing there, or by O.J.'s grown son, who might have had his own grievances against Nicole Brown Simpson, than they are to have been shed by any one random person. The jurors should be more interested in the probability that DNA matching O.J.'s DNA came from one of his two sons than they are in the chance that it came from some random person.

These examples, selected to make the point, are not typical, but the analysis applies in ordinary situations. Suppose, for example, that a rape victim believes that a man in a mugshot resembles her assailant. Later, in an identification contaminated by her prior view of the mugshot, she picks the defendant from a six-person line-up and upon hearing him speak says his accent is like that of the rapist. When the suspect's DNA is tested, it matches DNA extracted from the rapist's sperm at four loci. A random man's DNA may have a one in 50 million chance of matching the evidence DNA. A jury should, however, be less
interested in this probability than in the chance that someone in a group consisting of the defendant’s father, five brothers, nine cousins, and two uncles has DNA matching the defendant’s at the loci tested. These people may all live near the crime scene; they may all resemble the defendant and hence the rapist, and they may share a common accent. But for the fact their mug shots were not shown to the victim, they all may have been as likely as the defendant to be arrested for the crime. Moreover, if the defendant had an ironclad alibi, they and not some randomly selected person would be the people on whom suspicion would fall. The probability that at least one of these relatives has DNA matching the evidence DNA is far higher than the likelihood that some random person has matching DNA.44

In determining how likely it is that someone other than the defendant left the evidence DNA, the “others” of concern to the jury are only those who might plausibly be suspected of having left the evidence DNA if the defendant did not. I have elsewhere called this group the “suspect population.” This population is not, however, confined to others who might be suspected of the crime. As in the example of the rape victim’s husband or Simpson’s younger son, it includes people who might have innocently left what is thought to be evidence DNA. In other words, the suspect population includes all people who might plausibly be thought to be the source of what seems to be evidence DNA if the defendant is not the source.

Often there is little reason to believe that anyone with easy access to the crime scene was more or less likely than any other such person to have committed the crime if the defendant did not. Here the likelihood that a random person might have left the evidence DNA is the relevant statistic, except that if a defendant’s close relatives are in the suspect population, the “other” match probability should reflect their presence. Sometimes this requires no special calculations. Consider, for example, a suspect population of one million people, with a one in ten million chance that a random person would have left the evidence DNA. Suppose further that the defendant’s brother is a member of the suspect population, and that there is one chance in a thousand that the brothers have matching DNA. If there is no more reason to suspect the defendant’s brother of the crime, assuming the defendant’s innocence, than there is to suspect any other member of the suspect population, the probability that a random draw from the suspect population would yield an individual with matching DNA is 101 in a billion when the brother’s presence is taken into account, hardly more than the random man estimate of one in 10 million that ignores the brother. In these circumstances, match probabilities may be calculated without giving special attention to the brother’s possible guilt.

Suppose, however, that the criminal is known to be a young man who resides in an apartment building that has 100 young male occupants, including the defendant’s brother. Now the suspect population numbers 100, and the

44. See NRC II, supra note 3, at 113 (equations 4.8, 4.9).
brother’s presence means that there is about a one in 100,000 chance that the defendant’s DNA would match the crime scene DNA if the defendant were innocent. This is quite a bit higher than the random man probability of one in 10 million which would apply if the brother lived elsewhere. Finally suppose that in addition to the DNA match the evidence implicating the defendant consists of the victim’s somewhat hesitant eyewitness identification, the floor she saw her assailant take an elevator to after he left her, and the discovery of the victim’s jewelry in the defendant’s bedroom. The defendant’s brother, who looks like the defendant and shares a bedroom with him may be, we shall assume, about 500 times more likely to have assaulted the victim if the defendant did not than any of the 99 other young males living in the apartment building. Now the probability of finding a DNA match if the defendant is innocent is a little more than eight in 10,000, not much below the one in a thousand probability that the brother’s DNA would match.

We see from this example that the relevant match statistic, if it could be derived, is an average that turns on the number of people in the suspect population and a likelihood that each has DNA matching the defendant’s DNA, weighted by the probability that each committed the crime if the defendant did not. If relatives are no more likely than others to have committed the crime and the suspect population is large, providing a jury with random match statistics is not misleading because the prior probability that the relative did it is very small. But if relatives form a substantial portion of the suspect population, or if evidence apart from the DNA match makes them prime alternative suspects, the random match statistic is a misleading measure of how unlikely it would be to find a DNA match if the defendant were innocent.

The weighted average statistic I propose, unlike the “random man” match statistic, tells the jury how surprising it would be to find a DNA match if the defendant is innocent. The presence of the defendant’s relatives in the suspect population can make this far less surprising than the random man statistic suggests. However, neither the random man statistic nor the statistic I propose addresses the question the jury is most interested in; namely, given a DNA match, how likely is it that the defendant and not some third party committed the crime?

The average likelihood that someone in the suspect population has DNA matching the defendant’s can never exceed the likelihood that the defendant’s closest relative has matching DNA at the loci examined. As people join the defendant’s closest relative in the suspect population, the average probability that someone in the suspect population has matching DNA will diminish, making it appear more likely that the defendant is guilty. But the appearance of the defendant’s increased guilt is misleading. Adding potential suspects can only diminish the likelihood that the defendant is guilty, for adding suspects increases the chance that some other person who might have committed the crime has matching DNA. This increase can be substantial if the additional suspects are related to the defendant. Consider an extreme case. Suppose the
defendant, who has been identified solely on the basis of a DNA match, is an identical twin, but police records show his twin was in jail at the time of the crime and so is not in a large suspect population. The DNA evidence may be powerfully incriminating both in the sense that given the population of suspects it would be very surprising to find DNA matching the crime scene DNA if the defendant were innocent and in the sense that given the distinctiveness of the allele combination, it is very likely that the defendant and not someone else committed the crime. Now suppose it turns out that the defendant’s twin had been inadvertently released before his sentence expired so that his earlier exclusion from the suspect population was mistaken. If the suspect population is large enough, including the twin in it would, if the defendant were innocent, do little to diminish our surprise that the defendant’s DNA matched the evidence DNA, since before the DNA was analyzed there was no particular reason to suspect the twin. Nevertheless, including the defendant’s twin diminishes substantially the probative value of the match, for now we can conclude from the fact of the match a probability of defendant’s guilt that is at best slightly less than 50%. A similar diminution in the probative value of a DNA match occurs when large numbers of a defendant’s close relatives are in the suspect population.

The random match probabilities ordinarily given juries do not attempt to correct for the presence of relatives in the suspect population. This is one reason why they can mislead jurors about the probative implications of a DNA match on the question that most concerns them, which is not how surprising is it to find a match if the defendant is innocent but rather, given other possible perpetrators, what does the fact of a match say about the likelihood of the defendant’s guilt. The danger of this type of confusion exists not just when there are numbers of relatives in the suspect population, but also, in principle, whenever the suspect population is large relative to the reciprocal of the random match probability. Thus, if the random match probability is one in a million and there is a suspect population of five million, it is quite likely that several potential suspects have DNA matching the evidence DNA. Indeed, as surprising as it is that the defendant’s DNA matches the evidence DNA, the match only indicates that the defendant is one of several men who might, if the DNA match is dispositive, be guilty. The small chance of the DNA match does not mean there is little chance the defendant is innocent. The expected chance of the defendant’s innocence based on the DNA evidence alone is, in our example, about eight in ten, which should mandate an acquittal rather than serve as proof beyond a reasonable doubt.

This argument, however, comes perilously close to exemplifying what Thompson and Schumann call “the defense attorney’s fallacy.” It is seldom true that statistical evidence indicating that the defendant is one of several

45. The probability is less than 50% because of the chance of false positive error.
46. Thompson & Schumann, supra note 7.
people who might have committed a crime means that there is only a one in
several chance that the defendant is guilty. Arguments to this effect are
ordinarily fallacious because statistical evidence is seldom the only evidence
pointing to the defendant. Thus, if a man identified by a rape victim has no alibi
and shortly after the rape has scratches on his face and neck, the fact that his
DNA matches the rapist's DNA with a random match probability of one in a
million makes for an overwhelming case of guilt even if the suspect population
is so large that it is likely to contain several other people with matching DNA.
We do not worry about these unknown people because they are unlikely to be
incriminated by the other evidence in the case. But the same argument is not
fallacious if other evidence implicating the defendant is also likely to implicate
those possessing matching DNA. The presence of relatives in the suspect
population poses special problems because not only is the likelihood of
matching DNA far higher for a close relative than for a randomly chosen
person, but also because a close relative is more likely than a randomly chosen
person to share other characteristics, like neighborhood of residence, facial
appearance and accent, that may implicate the defendant in the crime.

As DNA tests become more discriminating, the problems posed by
relatives in the suspect population will become less acute, but today the best
way to deal with the potential problem is to test all fairly suspected relatives
and exclude them as possible contributors of the evidence DNA, or, if a relative
is not excluded, to run further tests until he or the defendant is excluded. This
is the approach recommended by NRC II, which also provides a formula for
estimating the probability that a relative who cannot be tested has DNA
matching the incriminating DNA profile.47

But NRC II does not go far enough. Rather than report the probability of
matching DNA for each untested relative who is a plausible alternative suspect,
the state's expert should report the probability that matching DNA would be
found in at least one suspected relative and, if this number is high, the
probability that more than one relative would have matching DNA. This
recommendation, like NRC II's recommendation to test suspected relatives is,
however, likely to be undermined in practice by failures to perceive that
relatives are plausible alternative suspects. Since investigations typically stop
when police think they have the culprit, it will be easy to overlook relatives
who might have committed the crime because no evidence suggesting they "did
it" will be gathered. If, for example, an eyewitness identifies a criminal in a
lineup and DNA evidence confirms the identification, the police are unlikely
to then present the defendant's brother to the witness in a lineup as a check.
That might reduce the certainty of the witness's earlier identification and make
the case against the defendant harder to prosecute.

Thus the list of fairly suspected relatives for the purpose of triggering the
NRC's recommendation regarding relatives should not be limited to those

47. NRC II, supra note 3, at 113.
whom the police regarded as plausible suspects. Rather the defendant should be allowed to name any close relatives whom he thinks might have committed the crime. This, in turn, should require the state to replace its random match statistic with a statistic showing the likelihood that at least one named relative has DNA like the defendant’s unless the state excludes each named relative through DNA testing or exculpatory non-DNA evidence or by showing that substantial non-DNA evidence of a sort unlikely to implicate the relatives implicates the defendant.48

Arguments that resemble the defendant’s fallacy are also not fallacious when the defendant is picked out by a DNA match from a DNA data bank. If a data bank is large enough and random match probabilities not very small, a match might be expected even if everyone in the data bank is innocent. Moreover, when the DNA of many people is examined, finding a match is less surprising than when the DNA of only one person is examined. NRC I recommended that the resulting statistical problems be dealt with by examining alleles other than those used to select the suspect and, if these alleles confirm the selection, using only their collective frequencies to estimate random match probabilities.49 NRC II rejects NRC I’s proposed solution because of the information it sacrifices and recommends instead simply multiplying the random match probability by the number of people in the data bank.50

NRC II’s solution is in my view a step backward. If, for example, one had a DNA data bank of ten million people, NRC II’s solution would suggest that a match was guaranteed whenever the multiplied random match probability was greater than one in ten million, and this is not true. There is also little reason to reject NRC I’s more cautious approach51 in a situation where a lack of any prior cause for suspicion makes the accidental identification of an innocent person far more likely than when a person already suspected is tested.52

There are two other problems with NRC II’s solution, both of which stem from a failure to consider the suspect population and the fact that the jury’s concern is not with the surprising nature of the match but rather with who else

48. Since the evidence the state is required to produce relates to a preliminary question of admissibility—whether the state is required to limits its statistical evidence to evidence showing the likelihood that at least one named relative has matching DNA—the judge rather than the jury would evaluate the state’s evidence.
49. NRC I, supra note 2, at 124.
50. NRC II, supra note 3, at 161.
51. If DNA data bases grow very large, NRC II’s approach may be more cautious, in the sense of yielding higher random match probabilities, in certain instances. In these circumstances, however, I would argue that the caution is excessive as compared to NRC I’s solution.
52. In certain rare circumstances, the quantity of DNA may counsel against NRC I’s recommendation, but with the development of PCR technology this is unlikely to be a substantial problem. Moreover, the original identification can and should be made using the minimum number of alleles required to select someone uniquely from the data base. Finally, not all information from the selection match is lost since a jury should be able to appreciate the additional probative value that accrues when a person is selected on the basis of a match on some alleles and the selection is confirmed by a match on other alleles.
might have committed the crime. In some situations, a DNA data bank might include virtually all potential suspects, and a unique identification from the data bank can be dispositive of a guilt even if the random match probability is quite high. If, for example, a rape occurred in a remote Arctic air base and the DNA profiles of all men stationed there were available, a unique DNA match with a random match probability of one in 100,000 would be more probative of guilt than a DNA match with a random match probability of one in ten million in a situation where only one soldier was tested and the DNA profiles of the air base’s other men were unknown. Conversely, where the suspect population is large and not confined to the data bank, NRC II’s suggested correction is inadequate because it only reveals how surprising it is to find a match within the data bank but says nothing about how unlikely additional matches are within the suspect population. Since there is no cause, apart from the match, to think the person selected is more likely to be the culprit than any other person in the suspect population, the suspect population’s size must be considered in a way that is not required when other evidence first inculpates a suspect. Suppose, for example, that there are 1,000 people in a data bank and that some person’s DNA matches the evidence DNA with a random match probability of one in a million. Multiplying this figure by 1,000 to yield a match probability of one in 1,000 can be substantially misleading. If the suspect population is ten million, it is far more likely than not that the person is innocent, for one would expect nine other equally plausible suspects to have matching DNA, yet a jury may conclude from the one in 1,000 figure that it is quite likely the person identified is guilty.

IV. THE EITHER-OR OF MATCH-BINNING

The final set of statistical issues I shall address results from the match-binning procedures that are the most common basis for deriving DNA statistics. From a justice standpoint, these are probably the least consequential of the errors that are common in the presentation of DNA statistics, but they nevertheless should not be made and occasionally may prove consequential. In RFLP analyses using match-binning procedures, the DNA analyst first decides if a suspect’s DNA matches evidence DNA. If the analyst concludes that a

53. Two interesting issues I shall ignore are the implications for this analysis of the fact that many DNA data banks consist only of convicted criminals or convicted sex criminals and how specifically incriminating evidence discovered after the DNA identification should qualify my argument. In both situations, the specific details of the crime and the additional information are likely to affect the analysis.

54. The reported match has substantial probative value in the hypothetical, for it makes it far more likely that the person identified is guilty (i.e., was the source of the evidence DNA) than it would be without the evidence. However, the probative value of the match is not nearly enough given the size of the suspect population to justify the conclusion that the person identified is probably guilty. Yet, the small random match probability given the jury may, even after taking account of the data base size, make it appear to the jurors as if the identified person is quite likely to be guilty.
match exists, she then calculates the probability of a coincidental match based on the frequencies with which alleles of the size found appear in a population data base.55 With this technique all matches are treated as if they are of the same quality—perfect matches within specified standards—for purposes of statistical calculations.

Others have pointed out how match-binning procedures lose information. Not all matches are the same. When the allele sizes of the evidence and suspect samples differ by, for example, 2%, the evidence is not as incriminating as it would be if the difference were only half of a percent. Yet, frequency calculations in each case will be based on the same match window and any suspect allele within the window will be declared a match.56 Conversely, if the discrepancy between an allele in the evidence DNA and an allele in the suspect DNA exceeds a match window, a match may not be called, though if the discrepancy is small and other alleles match there may be substantial reason to believe the DNA from the two samples had the same source. As other writers have noted this problem and indicated how Bayesian methods can avoid match calling problems, I shall not deal further with this issue.57

What concerns me more are the statistical implications inherent in the subjectivity of matching procedures.58 While the process of declaring a match is presented as an objective, scientific one, there can be a substantial subjective aspect to it. Although computer-assisted procedures are used to determine whether two alleles fall into the same match window, analysts can, and do, override the computer’s judgment. There are also situations in which evidence and suspect DNA apparently do not match, but analysts nevertheless call matches. For example, evidence and suspect DNA may closely match on seven bands, but there may be an eighth band in one sample that has no counterpart in the other. If the extra band in, let us say, the evidence DNA, indicates an actual allele and if the allele is actually absent from the suspect’s DNA, then the two samples have different sources, and the suspect can be excluded. But an analyst may attribute the extra band, particularly if it is fainter than the others,

55. The frequency count may be based on floating bins which involves searching a data base and counting all alleles whose lengths vary by a certain amount from the size of the evidence allele examined or, more commonly, on fixed bins, which have in advance grouped data base alleles into bins based on size and determined the proportion of the data base population with alleles in each bin. NRC II endorses floating bin based counts where their use is feasible. NRC II, supra note 3, at 161.

56. The match window is the range within which a DNA analyst will call evidence and suspect DNA samples the same. Thus, if a match window is ± 2.5%, any time the length of suspect DNA measured in base pairs is within 2.5% of the length of the evidence DNA, a match will be called.


58. For a detailed treatment of this issue which emphasizes the subjectivity that can exist, see William C. Thompson & Simon Ford, The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints, in FORENSIC DNA TECHNOLOGY (M. Farley & J. Harrington eds., 1991).
to an extraneous cause like bacterial contamination and declare a match regardless.

Analysts who call matches in circumstances like these are not cheating. If they are competent and honest, their decisions are likely to be correct, and we would be throwing away valuable evidence if whenever ideal matching criteria were not met we assumed that evidence and suspect DNA had different sources. Indeed, there is nothing wrong with analysts who, when faced with probative but not perfectly matching patterns, discard their preferences for frequentist procedures and become crude closet Bayesians.

In these circumstances, however, the same match-binning statistics are presented to juries as are presented in the case of ideal matches. A match having been called, the same bin frequencies and calculation algorithms are used to provide random match statistics. But given that the jury is concerned not with the likelihood of a DNA match but with the likelihood that a DNA match would be reported, applying the ordinary match-binning product rule calculations overstates the evidentiary import of the called match. If, for example, a DNA analyst calls a match where seven of an evidence sample’s eight bands match bands in the suspect sample, and the eighth has no counterpart, then the jury should be told not the probability of a seven-band match but the probability that seven bands in a suspect sample would match any seven of eight bands in the evidence sample because, regardless of which band had not matched, a match might have been reported to the jury. Although the random match probabilities presented to juries would diminish somewhat, it is unlikely that much probative force would be lost, and to the extent this happens, the loss is appropriate.

So long as match-binning procedures contain any subjective element, there is a more important further reform to be made in DNA testing. While working on a case, DNA analysts sometimes know of the other evidence against a defendant. Being human, they may be influenced by this information. If they are, jurors considering a match may, without knowing it, be weighing information that reflects not just DNA results, but also either inadmissible evidence or evidence admitted in the case and already considered by the jurors. Moreover, the DNA analyst only hears the untested police side of the story. If hearing this induces an analyst to call a match where she otherwise would have labeled the results indeterminate, the state may secure evidence powerful enough to convict even when the defense proves at trial the unreliability of the evidence that influenced the analyst’s subjective call. Except for a few limited purposes, such as information about the likely number of assailants in a rape

59. This is a conservative suggestion since it might not be all seven band matches that the analyst would call matches but only those in which the eighth band in the evidence sample was faint or otherwise seemed attributable to something other than the evidence DNA. However, in a setting where, as the recent FBI lab scandals indicate, forensic scientists tend to make discretionary calls so as to favor the state, statistical conservatism in interpreting those calls appears justified.
After the DNA Wars: Skirmishing with NRC II

case, there is no scientific reason why a DNA analyst should know anything about the crime or the evidence against the defendant. Until all subjective elements are eliminated from DNA test procedures, laboratory protocols should preclude providing unneeded information about the crime or defendant to DNA analysts. Indeed, even if subjectivity played no part in interpreting DNA test results, analysts should not receive crime specific information, because certainty of the suspect’s guilt can induce carelessness or even corruption. NRC II should have taken a firm position on this.

DNA comparisons are, next to fingerprinting, forensic science’s most powerful identification technique. Techniques of DNA identification and procedures for calculating random match probabilities are, generally speaking, scientifically sound. But this does not necessarily make standard practices regarding the presentation of DNA evidence in court good forensic science. In fact, DNA evidence is routinely evaluated and presented in court in ways that risk substantially overstating its incriminatory power. This is largely because the state’s DNA experts aim their testimony toward the questions which the science of DNA identification can answer: (1) Does the suspect’s DNA match the evidence DNA?, and, if so, (2) if someone were to be selected at random from an appropriate larger population, how likely is it that his DNA would match the evidence DNA? The question that the forensic science of DNA identification should answer—How likely is a report of a DNA match if the defendant did not leave the evidence DNA?—is typically ignored by the state’s experts except to the extent that they are forced to grapple with aspects of it on cross-examination. Even when a prosecution expert does address an issue relevant to the likelihood of a reported match; information about this likelihood does not figure in the statistics the expert uses to convey the probative force of the DNA evidence to the jury.

It is the disjunction between the science questions of DNA identification and the forensic science question that has made this an awkward issue for the

60. For ease of exposition I am assuming here, as I have throughout this paper, that the evidence DNA is DNA containing evidence left by a criminal at a crime. It is, however, also common to finger criminals by showing that blood or other DNA containing evidence from a crime victim is present on the criminal's clothes or on other possessions or in places the criminal has special access to. The points I make apply regardless of whether the evidence DNA is attributable to the criminal or the victim.

61. Usually appropriate populations are considered to be populations of people who share the same racial heritage, in its broadest sense (e.g., black, white, southwest Hispanic, etc.) as the person suspected of leaving the evidence sample. The control made for race is often not scientifically required, but it is usually benign as the defendant is more likely to be helped than hurt by the control. Lempert, supra note 43.

62. While I have identified the core questions in each category, there are other questions that are somewhat differently treated depending on one's perspective. One can, for example, treat questions pertaining to the likelihood that a relative of the defendant will have DNA matching the evidence DNA as a science question and give a scientific answer to it. See I.W. Evett, Evaluating
National Research Council to resolve and has led to the publication of two reports, neither of which is a complete success, although the second is likely to be better received by the science community and the largely prosecution-oriented forensic science community than was its predecessor. When NRC I was being drafted, the science was less clear than it was when NRC II was written, and the committee that produced NRC I, perhaps because it had a broader charge, was more concerned with legal values and the forensic science question than the later committee was. Thus NRC I made recommendations that substantially import a view of legal values into what are supposed to be scientifically derived standards, and it devotes a good deal of attention to ways of improving laboratory practice so that error and analytic subjectivity will be minimized. Even so, NRC I does not fully recognize the statistical implications of laboratory error, perhaps because its recommended technique for calculating random match probabilities is sufficiently conservative that laboratory error—which has more effect on probative value the more its chances exceed random match probabilities—appears as less of a statistical problem. NRC II reflects more current and sounder science than NRC I in areas where it overlaps the earlier report, and in particular contains clear and useful treatments of a variety of statistical issues. But it ignores or postpones for another day important aspects of what I have called the forensic science question and the implications this has for the statistical analyses it recommends. Thus, it argues for giving jurors random match probabilities even if common sense and proficiency test results suggest that the probative force of reported matches cannot be nearly so great as these figures suggest. Worse, it explicitly breaks

63. The first NRC committee considered ethical issues relating to DNA evidence, the construction of DNA data banks, the future monitoring of the technology, and other issues that the second committee saw as beyond the latter committee’s charge. The second committee saw its role as limited to determining how the uncertainty of laboratory findings can be reduced, how the risk of error can be minimized, how to take account of population substructure (including relatives), and what statistical theory and empirical observations allow us to say about the probability of DNA matches. The second committee also discussed some legal issues, focusing particularly on the reception by the courts of the earlier NRC report.

64. The committee made a misjudgment here, for even by its recommended calculation procedures, random match probabilities are often several orders of magnitude less than the likely chance of laboratory error.

65. Professor David Kaye argued, in the course of some extraordinarily helpful comments on an earlier draft of this paper, that given the adversary system it might be appropriate for the state to present random match probabilities, leaving it to the defendant to discuss the implications of possible error. I reject the argument because I believe the state in a criminal prosecution is not just another adversary. As the Supreme Court indicated in *Brady v. Maryland*, 373 U.S. 83 (1963),
After the DNA Wars: Skirmishing with NRC II

with the recommendations of NRC I and encourages those who wish to claim that DNA profiles are uniquely associated with particular individuals, though the chance of error means that the report of an apparently unique DNA match cannot guarantee a suspect's association with a crime. Also, NRC II, unlike NRC I, does not caution against subjective aspects of the DNA matching process and fails to appreciate their possible statistical implications. Finally, NRC II conditions its recommendations for excluding relatives on the existence of evidence that one or more relatives of the suspect are possible perpetrators. This language appears to condition the need to exclude relatives or take account of their likely DNA profiles on the existence of evidence that actually points to relatives as suspects. It misses the way that evidence gathering is an active, focused process. If, for example, an eyewitness thinks a person whose mug shot she sees raped her, the police are likely to seek additional evidence against that person and are unlikely to seek a picture of the defendant's brother to see if he too looks like the rapist. Even hypothetical, unknown relatives, whose possible presence the committee dismisses, may be plausible alternative suspects if the defendant did not commit the crime.

But enough criticism. Let us conclude with the good news about DNA testing. The process is so robust and the chance of error, the greatest of the dangers I have pointed to, is ordinarily sufficiently small that overstating the probative value of a DNA match will seldom lead to unjust convictions. When, and other cases, the state has a special obligation to avoid concealing exonerative evidence and to present its case so as to get at the truth. It is the state's special obligations in these respects that underlie a number of my recommendations and the special burdens I would put on the state and its experts. In my view, the state only does good science when its scientific analyses alone and unrefuted are fair to the defendant. Thus, there are no pure scientific answers to the questions that surround DNA evidence. As NRC I recognized more than NRC II, legal values have something to say about the demands of good science.

66. The committee writes that "[t]he number of loci and the degree of heterozygosity per locus that are needed to meet the criteria illustrated above [for uniqueness] do not seem beyond the reach of forensic science, so unique typing (except for identical twins) may not be far off." NRC II, supra note 3, at 138. This remark is scientifically justified in the sense that people, except for identical twins, have unique DNA profiles and science is now at a point where enough information about an individual's profile can be extracted. Hence, there is good ground for thinking no one else would be the same on even the small fraction of the whole profile we can observe. But chances of error mean one cannot make the claim which seems to follow naturally from the uniqueness claim: namely, that DNA found at a crime scene must be the defendant's because tests show this to match the defendant's DNA, and we can be confident that no one else has DNA that would match. It is, however, possible that uniqueness claims will not be as misleading as low random match probabilities. Juries may be less confused about how to use error probabilities to deflate the likelihood of a true identification when the DNA evidence is said to be unique than when it is claimed that DNA like that of the defendant has a chance of one over some very large number of characterizing a random person. Possibilities like this are one reason why it is important to implement NRC II's wise recommendation that behavioral research be carried out to identify conditions that will cause fact finders to misinterpret DNA evidence (Id. at 42). Also uniqueness claims might be the spur courts need to make DNA experts cease proclaiming, as some have, that false positive errors are impossible.

67. NRC II, supra note 3, at 123.
as is usually the case, evidence other than the DNA match also implicates the defendant, a guilty verdict will be justified even if the probative value of a DNA match is limited to the probability of error.\(^{68}\) For similar reasons NRC I's ceiling principle calculations were unlikely to produce unmerited acquittals, and the committee members knew this when they made the proposal. With respect to error, NRC II has made suggestions of great practical importance: namely, that defendants be allowed, even at state expense, to secure independent laboratory tests of incriminating DNA evidence and that those receiving DNA evidence handle it to enhance the independence of these tests. Although this suggestion will not deal with all sources of error, such as cross-contamination in the field, and although the probabilities associated with independent errors at two laboratories will still dwarf random match probabilities, in most cases the conjoint probability will be so small that we will be able to say with some confidence that laboratory handling or testing errors will not lead to wrongful convictions. Implementing these changes will not, however, mean that echoes of the DNA Wars will no longer remain. The battles, especially as they were joined after NRC I, were never really over whether we would sanction procedures that convicted innocent men or let the guilty go free. Rather they were about the standards that good science and good forensic science demand we achieve. On this issue, despite the National Research Council's second DNA report, there remains room for disagreement.

---

\(^{68}\) Even in the error-filled case of *People v. Castro*, 545 N.Y.S.2d 985 (Sup. Ct. 1989), which revealed serious problems in the way DNA tests were then conducted, it appears that the DNA profiling was in fact accurate and the defendant was guilty.