

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

**On Conveying the Probative Value of DNA Evidence:  
Frequencies, Likelihood Ratios, and Error Rates**

Jonathan J. Koehler<sup>1</sup>

The University of Texas at Austin

March 20, 1996

---

<sup>1</sup> **Associate Professor of Behavioral Decision Making, Graduate School of Business, University of Texas at Austin. Ph.D. 1989, University of Chicago. A version of this paper was presented at Symposium: A Criminal Procedure? O.J. Simpson and the Criminal Justice System on Trial, University of Colorado School of Law, January, 1996.**

The first numerical probability that was offered in the O. J. Simpson case came from the defendant himself when he pleaded "Absolutely, 100% not guilty."<sup>2</sup> There isn't much question about what Mr. Simpson meant by that probability statement. But hundreds of other probability statements were offered in this trial that left considerable room for ambiguity, interpretation, and misinterpretation.

Most of these probability statements were associated with DNA evidence that purportedly linked Mr. Simpson to the crime. What do the numbers mean? How should such evidence be presented to jurors? What exactly is meant by "the evidence"? What role, if any, should the possibility of error play in the interpretation of DNA frequencies?

Section I of this paper discusses the issue of what the frequencies associated with DNA evidence do and do not mean. Section II describes an alternate way of presenting DNA statistics in court based on Bayesian likelihood ratios. Issues associated with identifying the hypothesis of interest and characterizing the evidence are addressed. Section III offers a likelihood ratio and estimates its magnitude based on proficiency test data that were analyzed by Koehler, Chia and Lindsey.<sup>3</sup> Section IV identifies the arguments that have been offered for ignoring the possibility of error when computing DNA likelihood ratios. These arguments are considered, then rejected. Section V asks whether likelihood ratios belong in court, and concludes that they may not if experts and jurors don't understand them. Section VI reports an empirical study that compares mock jurors' reactions to various numerical representations of DNA evidence. It is concluded that there are good reasons to suspect that likelihood ratio presentations will confuse jurors. Section VII is a conclusion.

---

<sup>2</sup> B. Drummond Ayers Jr., "'Absolutely' not guilty, a confident Simpson says," *New York Times*, July 23, 1994, at A6.

<sup>3</sup> Jonathan J. Koehler, Audrey Chia, & J. Sam Lindsey (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal*, 35, 201-219, at p. 206-211.

## I. What Does One in One Trillion Mean?

In the Simpson case, numerous blood stains that matched either the defendant or one of the two victims were introduced into evidence.<sup>4</sup> The prosecutorial highlight of testimony about the matching blood stains was the introduction of frequency statistics associated with each stain. Table 1 shows the frequencies that were offered by one witness during cross-examination. On this afternoon, at least 76 different estimates were offered for various blood drops. These estimates ranged from 1 in 1 (i.e., a characteristic that everyone shares) to 1 in 1 trillion. Numbers like 1 in 1 trillion boggle the mind. There is little, if anything, in ordinary experience that prepares us to comprehend a 1 in 1 trillion frequency. Yet jurors in the Simpson case--like jurors in thousands of other cases throughout the country--were asked to make sense of and use such numbers with little or no guidance.

-----  
Insert Table 1 About Here  
-----

What does a frequency like one in one trillion mean? If we assume that the frequency was derived correctly, and if we assume that there is no possibility of error, lying, or misinterpretation of the data, then it means that there is 1 chance in 1 trillion that a single randomly selected person would share the observed characteristics. In other words, assuming that the data and its interpretation are infallible, we would expect to see this DNA profile in approximately one out of every trillion people.

Notice that this is not identical to the probability that there exists someone else who shares the observed profile. Although it may be extremely unlikely that a single randomly selected person would share a DNA profile with another person, it may be quite likely that

---

<sup>4</sup> The victims were Ms. Nicole Brown Simpson and Mr. Ronald Goldman. Ms. Brown Simpson was Mr. Simpson's ex-wife, and Mr. Goldman was a friend of Ms. Brown Simpson.

others share this profile.<sup>5</sup> Thus, there is only one chance in 1 billion that a random person shares a DNA profile that is common to one in every billion people, but there is a 99.6% chance that there are others on earth who share the profile.<sup>6</sup> Even for a frequency as small as 1 in 1 trillion, the chance that there are others who share the profile is greater than one in two hundred.<sup>7</sup>

One should also bear in mind that these tiny frequencies do not themselves tell us (a) the probability that the matching suspect committed the crime, (b) the probability that someone other than the suspect committed the crime, or even (c) the probability that the suspect is or is not the source of the genetic evidence. That DNA frequencies do not identify the probabilities of guilt or innocence should be apparent because otherwise all other evidence (e.g, eyewitness testimony, motive, opportunity, etc.) would be irrelevant. That DNA frequencies do not identify source probabilities is a more subtle point that is best understood by way of example.

Consider the blood evidence that was introduced in a pre-trial hearing in the Simpson case by Los Angeles Police Department blood expert Gregory Matheson. Mr. Matheson testified that approximately 1 in every 200 people (.43%) have the blood characteristics that were found in a trail of blood drops that led away from the scene of the murders.<sup>8</sup> Mr. Matheson also testified that the blood drops matched Mr. Simpson, but did not match either of the victims.

---

<sup>5</sup> Koehler, J. J. (1993). Error and exaggeration in the presentation of DNA evidence. *Jurimetrics Journal*, 34, 21-39.

<sup>6</sup> P(At least one other person on earth has the DNA profile) = 1 - P(No other person on earth has the DNA profile). Assuming there are 5.5 billion people on earth (other than the matching individual), the computation is as follows:  
 $1 - (1 - 1/1,000,000,000)^{5,500,000,000} = .996 = 99.6\%$ .

<sup>7</sup>  $1 - (1 - 1/1,000,000,000,000)^{5,500,000,000} = .0055$ . Thus, the probability that at least one other person on earth has a DNA profile that is common to 1 in 1 trillion people is about 5.5 in 1000 or about 1 in 182.

<sup>8</sup> B Drummond Ayers Jr., *The Simpson Case: The overview; Simpson ordered to stand trial in slaying of ex-wife and friend*. *New York Times*, Pg. A1, 7/9/94.

Two features of this evidence must be understood.<sup>9</sup> First, it is valuable evidence because it excludes a large proportion of people as possible contributors of the blood, but it fails to exclude Mr. Simpson. On the other hand, the evidence is limited because it fails to exclude many people who are not the source of the blood. Indeed, the non-excluded group to which Mr. Simpson belongs might well include thousands of people in a large city like Los Angeles, a few of whom might reasonably be considered potential suspects in the case.<sup>10</sup>

For this reason, one cannot hear the 1 in 200 frequency and conclude either that (a) there is therefore 1 chance in 200 (i.e., 0.5% chance) that the blood drops are not Simpson's, or (b) there are therefore 199 chances in 200 (i.e., 99.5% chance) that the blood drops are Simpson's. How can we conclude that there is a 99.5% chance that Simpson is the source of the matching blood when we know that there are hundreds or thousands of others who share this blood profile? Surely, it cannot be the case that all members of the non-excluded group have a 99.5% chance of being the source of the matching blood. Though surprising, it is true: forensic science analyses alone cannot identify the probability that O. J. Simpson--or any other criminal defendant--is or is not the source of recovered genetic evidence.

Here, the skeptical reader may say, "Yes, but how many of those others who might share Mr. Simpson's blood profile had cuts on their hands, were married to Nicole, or had a history of violence against Nicole?" The point is well-taken, and reinforces my own point that one cannot identify the probability that someone is or is not the source of genetic evidence through genetic testing alone. Nongenetic considerations, such as those listed above, must be factored into any equation that purports to identify the chance that someone is the source of

---

<sup>9</sup> **This statistical evidence presented by Mr. Matheson was based on non-DNA blood tests. However, the statistical principles are identical to those that accompany DNA blood matches.**

<sup>10</sup> **This point was made by the Simpson defense team. On cross-examination, Mr. Matheson agreed that approximately 40,000 people in Los Angeles County share these blood characteristics (Ayers, New York Times, 7/9/94).**

genetic samples.

## II. Likelihood Ratios: An Appropriate Way to Describe DNA Evidence?

For those accustomed to thinking in terms of Bayes's Theorem, this point reflects the fact that one's beliefs about a proposition in the face of new evidence should be a function of both prior probabilities and likelihood ratios.<sup>11</sup> According to Bayesian logic, one's prior beliefs (i.e., the beliefs one holds prior to the introduction of new evidence) are combined with a quantitative measure of the probative value of the new evidence to form posterior beliefs.

The probative value of evidence--such as evidence of a DNA match between a suspect and a crime scene sample--is described by the likelihood ratio,  $\frac{P(\text{Evidence} | \text{Hypothesis})}{P(\text{Evidence} | \text{Hypothesis})}$ .<sup>12</sup>

### What Is The Hypothesis?

But what is the "hypothesis" in cases involving DNA evidence? And what exactly is the "evidence"? Subtle differences in the answers scientists provide for these questions can have a substantial impact on the magnitude of the resultant likelihood ratios. For example, when genetic samples are examined that appear to contain the DNA from more than one source, assumptions about the number of donors affect the size of the corresponding likelihood

---

<sup>11</sup> The odds form of Bayes's theorem provides a normatively justifiable mechanism for updating one's beliefs in the face of new evidence. For some hypothesis H and Evidence E,  $\frac{P(H | E)}{P(\bar{H} | E)} = \frac{P(H)}{P(\bar{H})} \times \frac{P(E | H)}{P(E | \bar{H})}$ . **Error! Main Document Only.** That is, the posterior odds ratio  $\frac{P(H | E)}{P(\bar{H} | E)}$  **Error! Main Document Only.** is the product of the prior odds ratio  $\frac{P(H)}{P(\bar{H})}$  **Error! Main Document Only.** and the likelihood ratio  $\frac{P(E | H)}{P(E | \bar{H})}$ . **Error! Main Document Only.**

<sup>12</sup> Richard O. Lempert, *Modeling Relevance*, 75 Mich. L. Rev. 1021, 1026 (1977) ["evidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue"]. See also Richard D. Friedman, *A close look at probative value*, 66 Boston U. L. Rev. 733 (1986), and Friedman, *Conditional probative value: Neoclassicism without myth*, 93 Mich. L. Rev. 439, 456 (1994).

ratios.<sup>13</sup>

But even where number of donors is not at issue, a more fundamental concern about what the hypothesis of interest is remains. One hypothesis is "The defendant is the source of the genetic evidence." A second hypothesis is "The defendant had contact with the crime scene." A third hypothesis is "The defendant is guilty of the crime." Each of these hypotheses may be offered but, as the discussion below indicates, the values of the corresponding

---

<sup>13</sup> Letter from William C. Thompson to Professor James Crow, Chair, National Research Council Committee on DNA Forensic Science, re: Statistical Characterization of Mixed DNA Samples July 5, 1995. The issue was also highlighted in a contentious and sometimes bizarre exchange between defense attorney William Thompson and prosecution DNA expert Bruce Weir in People v. Simpson:

**Q:** Now, but in this case, we cannot tell by looking at most of the mixtures how many contributors there are, can we?

**A:** Oh, I think the evidence is overwhelming that there were two contributors when we have seven RFLP loci and seven PCR loci and I have at most four alleles. I can't imagine any other possible interpretation.

...

**Q:** But in order to draw conclusions about the relative likelihood [of there being two versus three contributors], you have to make some assumptions drawing on other evidence, don't you?

**A:** The assumption I choose to do these frequencies is that there are either two or three contributors. Based on the evidence, there are two contributors. I have chosen to do it for three. I could have done it for four. I see no point.

...

**Q:** And you say -- but you can't determine, based on those three alleles, whether there are two people or three people there, can you?

**A:** That is -- that is a foolish question in the sense that the evidence --

**Q:** Will you answer it, please, sir?

**The Court:** Wait, wait, wait, wait. Counsel, he gets to answer the question.

**Mr. Thompson:** All right?

**The Witness:** There are three alleles. We know that each person has two. People may share an allele, so two people may have two alleles, one of which is shared, which would end up with three. I can't imagine why you would invoke three people when there are only three alleles. There may be three people. There may be six people. There is nothing there which would require me to assume more than two.

**Q:** All Right. And so based on the genetic evidence alone, there is no basis for determining whether there is two or three or four?

**A:** The evidence is that there are only at most four with here only three alleles. The polymarker had four -- I think -- I would have to look. I think only three alleles. So the evidence says there are three alleles. That means more than one person. It does not say anything about there being three people.

**Q:** Uh-huh. And so you can't tell whether there are two versus three based on this evidence alone?

**A:** I can't tell for certainty.

People v. Simpson, vol. 173, June 22, 1995 (cross-exam)

likelihood ratios can vary greatly.<sup>14</sup>

The likelihood ratio for DNA evidence that corresponds with a "source" hypothesis is  $\frac{P(\text{Evidence} | \text{Suspect is Source})}{P(\text{Evidence} | \text{Suspect is Not Source})}$ .<sup>2</sup> The denominator of the likelihood ratio requires consideration of

the possibility that evidence of a DNA match would emerge when, in fact, the suspect is not the source of the genetic sample. Two scenarios must be considered. First, it is possible that the laboratory erred, and that the defendant's DNA profile actually does not match the genetic sample. Second, it is possible that the match is correct, but that it was purely coincidental. That is, one must consider the possibility that the defendant is not the source of the genetic evidence, but that he just happens to share a DNA profile with the person who actually is the source.

In most cases, the possibility of laboratory error is substantially larger than the possibility of a coincidental match. This is not because DNA laboratory work is particularly sloppy or unreliable. Instead, it is because the chance of a coincidental match is usually small.

In the Simpson case, both the possibility of laboratory error and coincidental matching appeared to be remote. Laboratory error was unlikely because many blood samples were tested at different laboratories using two different DNA typing methods.<sup>15</sup> Coincidental match was unlikely for many of the blood stains, particularly those that were associated with very small frequencies.<sup>16</sup> Because the denominator of the likelihood ratio associated with the

---

<sup>14</sup> Occasionally, these hypotheses are not treated as distinct. For example, one excellent statistician failed to distinguish between the source and guilt hypotheses ("The question of interest is whether the suspect is guilty (G); that is, whether the suspects' and crime sample's DNA fragments are from the same person" Donald A. Berry (1991), *Inferences using DNA profiling in forensic identification and paternity cases*, *Statistical Science*, vol. 6, 175-205. However, because the alternative hypotheses under the source and guilt hypotheses require different considerations, these hypotheses should not be treated synonymously.

<sup>15</sup> Both RFLP and PCR tests were conducted on many samples.

<sup>16</sup> For example, the frequency statistics that were provided for a blood stain that was recovered from Nicole Brown Simpson's back gate and that matched Mr. Simpson ranged from 1 in 57 billion (for African Americans) to 1 in 150 billion (for Caucasians). *People v. Simpson*, vol. 171, Tuesday, June 20, 1995.

"source" hypothesis reflects these two unlikely possibilities, the likelihood ratio will probably be quite large. Thus, under the source hypothesis, the DNA evidence appeared to be extremely probative.

But when we shift to a "contact with crime scene" hypothesis, the evidence is less impressive. The reason is that the denominator of the likelihood ratio,  $P(\text{Evidence} | \text{Suspect Had Contact With Crime Scene})$ , must now reflect a third possibility in addition to laboratory error and coincidental matching. Now one must consider the possibility that the defendant did not have contact with the crime scene, yet is indeed the source of the evidence.

In the Simpson case, a prosecution expert offered the "contact with crime scene" hypothesis, but failed to take this third possibility into account.<sup>17</sup> His omission was significant because the defense alleged that some of the blood evidence was planted.<sup>18</sup> Thus, the defense suggested that the possibility that Mr. Simpson was the source of the blood but did not lose blood at the crime scene was large. To the extent that the defense made a convincing case for this proposition, the denominator of the likelihood ratio should have increased. For example, if a juror believed that there was a 20% chance that the blood evidence was planted by police, then the probability that a DNA blood match on Mr. Simpson would have been found if, in fact, he did not have contact with the crime scene would be at least 20%. Here, the likelihood ratio associated with the "had contact with crime scene hypothesis" could not have been larger than 5:1.<sup>19</sup>

---

<sup>17</sup> Bruce S. Weir, *People of the State of California v. Orenthal James Simpson, Statistical Interpretation of the DNA Evidence, June 20, 1995* ("These notes are concerned with assessing the strength of that DNA evidence E by determining the probabilities of the evidence under the alternative hypotheses that either C: OS had contact with the Bundy scene or -C: OS did not have contact with the Bundy scene," p. 2).

<sup>18</sup> David Margolick, *Blood used in frame-up, Simpson lawyers assert. New York Times, A16, May 5, 1995.*

<sup>19</sup> The maximum likelihood ratio corresponding to a denominator of .20 is  $1:.20 = 5:1$ .

DNA evidence becomes still less probative under a "guilt" hypothesis. This is because the denominator of the likelihood ratio must reflect a fourth possibility, in addition to laboratory error, coincidence, and frame-up: the possibility that a defendant is not guilty, but that he did leave his blood at the crime scene. This scenario was also raised by the Simpson defense team.<sup>20</sup> They argued that because Mr. Simpson was a frequent visitor to the crime scene in the weeks and months prior to the murders, he may have innocently deposited his blood in and around this area during one these visits. If this probability is judged to be nonzero, then the denominator of the likelihood ratio should be even larger than it was under the "had contact with crime scene" hypothesis. This, in turn, creates a smaller (i.e., less probative) likelihood ratio.

In sum, the apparent probative value of DNA evidence is driven, in part, by the hypothesis that one chooses to examine. Hypotheses that require incorporation of additional considerations in the denominator of the likelihood ratio lead to smaller likelihood ratios. The Simpson case is instructive in this respect. Under a "source" hypothesis, the DNA evidence appears to be impressive. But under a "contact with crime scene" or "guilt" hypothesis, the DNA evidence yields less impressive likelihood ratios.

From a policy perspective, DNA experts should avoid using such broad hypotheses as "contact with crime scene" and "guilt" because they require speculation about probabilities that have little to do with DNA genetics. Nothing in a DNA expert's background or knowledge of the evidence qualifies him to assess the likelihood of a frame-up or the chance of an innocent deposit of genetic material at an earlier time. Yet these possibilities must be assessed and quantified when likelihood ratios are constructed against the broader hypotheses.

---

<sup>20</sup> "From a statistical standpoint, Professor, is it important to consider the chance that it may really be Mr. Simpson's blood, but that the drops were left by Mr. Simpson at some earlier date?" Question asked by defense attorney Peter Neufeld to defense expert Professor Terence Speed, *People v. Simpson*, vol. 201, August 7, 1995.

### What is the Evidence?

Different construals of the "evidence" can also affect the size of DNA likelihood ratios. Many researchers who have discussed the use of likelihood ratios in the context of DNA evidence implicitly assume that the evidence is "a match between the suspect and crime sample."<sup>21</sup> But a more accurate description of the evidence is "a report of a match between the suspect and crime sample." Like all evidence, there is a risk of error associated with DNA evidence. When an eyewitness to a robbery testifies, in essence, that a suspect "matches" the individual he saw holding a gun to the bank teller's head, we do not and should not treat this testimony as certain proof that the suspect is the culprit. Instead, we treat this evidence as an eyewitness "report," the reliability of which depends on a host of considerations (e.g., visibility conditions, duration, incentive to lie, etc.). Reports of DNA matches should be treated similarly. The evidence in question is not a DNA match, but a report of a DNA match.

### III. The Role of Error Rates in DNA Likelihood Ratios

When the hypothesis and the evidence are identified in the manner suggested here, the likelihood ratio of interest becomes  $\frac{P(\text{Match Report} | \text{Suspect is Source})}{P(\text{Match Report} | \text{Suspect is Source})}$ .<sup>4</sup> Estimates of this ratio appear in Koehler, Chia and Lindsey (1995).<sup>22</sup>

These estimates were derived from an examination of several different sets of DNA proficiency test data that included "match"/"no match" conclusions. Data from dozens of DNA laboratories between 1988 and 1994 were included. In the typical proficiency test, DNA

---

<sup>21</sup> Bernard Devlin, Neil Risch, Kathryn Roeder (Feb. 5, 1993), *Statistical evaluation of DNA fingerprinting: A critique of the NRC's report*, *Science*, 259, 748; David Jarjoura, James Jamison, & Stavroula Androulakakis, *Likelihood ratios for deoxyribonucleic acid (DNA) typing in criminal cases*, 39 *Journal of Forensic Sciences* 64 (1994); Kathryn Roeder, *DNA fingerprinting: A review of the controversy* 9 *Stat. Sci.* 222 (1994).

<sup>22</sup> Jonathan J. Koehler, Audrey Chia & J. Sam Lindsey (1995). *The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial?* *Jurimetrics Journal*, 35, 201-219, at p. 210.

analysts are provided with sets of genetic samples from known and unknown sources and are asked to determine which match and which do not. When estimating the denominator of the likelihood ratio, Koehler et al. (1995) examined the number of times that a laboratory reported matches between samples that were not from common sources.<sup>23</sup> They also made the maximally generous assumption that the numerator the likelihood ratio is one.<sup>24</sup>

Due to occasional match reports on non-matching samples,<sup>25</sup> Koehler et al. (1995) obtained denominators of one out of several hundred or several thousand. This yielded likelihood ratios on the order of several hundred to several thousand. Although likelihood ratios in the hundreds or thousands constitute strong evidence that the suspect is the source of the sample, they are less persuasive than the likelihood ratios of millions, billions and trillions that are sometimes claimed by those who treat the evidence as a "certain match."

#### IV. "Error Rates are Irrelevant" and Other Counter-Arguments

The major difference between treating DNA evidence as a certain match and treating DNA evidence as a reported match is that only the latter requires consideration and explicit incorporation of the possibility of error. Because likelihood ratios vary dramatically depending on whether error rates are considered or not, the recommendation to treat DNA evidence as a

---

<sup>23</sup> For simplicity, our computations assumed that proficiency tests do not include samples from different sources that share common DNA profiles.

<sup>24</sup> The assumption is "generous" in terms of diagnosticity. It assumes that false negative errors do not occur. In fact, such errors are made. See William C. Thompson and Simon Ford (1991). The meaning of a match: Sources of ambiguity in the interpretation of DNA prints. In M. A. Farley & J. J. Harrington (Eds.). *Forensic DNA Technology*, Chelsea, MI: Lewis; Jonathan J. Koehler (1993). DNA matches and statistics: Important questions, surprising answers. *Judicature*, 76, 222-229. But because the impact of small deviations from one in the numerator of the likelihood ratio have a minimal impact on the likelihood ratio, we produced an upperbound likelihood ratio value by setting the numerator equal to one.

<sup>25</sup> Jonathan J. Koehler, Audrey Chia & J. Sam Lindsey (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? *Jurimetrics Journal*, 35, 201, at p. 206-210.

match report rather than a certain match is sometimes challenged. This section identifies and evaluates some of the counter-arguments that have been raised against the suggestion that errors and error rates should play a role in identifying the probative value of DNA evidence.

**"DNA typing errors are impossible--in theory and in practice."** This is the oldest--and most untenable--counter-argument. In courtrooms throughout the country, DNA experts have testified that false positive matches in DNA profiling are "impossible."<sup>26</sup> However, a review of proficiency test data indicates not only that errors are possible, but that they have occurred in a variety of laboratories.<sup>27</sup>

In a 1993 Jurimetrics article, I noted that some people object to the inclusion of human errors into analyses of DNA false positive match errors.<sup>28</sup> They prefer to confine discussion of error to those errors that can or have resulted from technical failures in the DNA typing process. This may explain--though not excuse--testimony about the impossibility of false DNA match reports. Most of the errors that arise in proficiency tests appear to be due to human errors (e.g., mislabeling, contamination, interpretive error, etc.) rather than to technical errors in the DNA typing process itself. But it does not justify unmitigated enthusiasm about the accuracy of DNA evidence. Factfinders need to be aware of the chances that some error will occur rather than the chance that a particular error, or a particular type of error, will occur. Because human error is responsible for many of the errors observed to date, it must not be

---

<sup>26</sup> *People (CA) v. Marlow*, 39 Cal. App. 4th 343, 363, 1995 ("You simply don't get a result if you haven't done it right . . . You can't get a false positive"); *Commonwealth v. Teixeira*, Mass. 1993 ("There have never been any false positives made with DNA testing at this point"); *Hicks v. State (TX)*, 1993 ("A false positive finding [is] impossible because if the procedures were not correctly followed, no match could be obtained."); *State (OH) v. Pierce*, 1992 ("You can't get a false positive."); *People (CO) v. Fishback*, 1991 ("DNA analysis is failsafe"); *State (AL) v. Yelder*, 1991 ("[T]here's no way to make or create a false positive with this test"). For other examples, See Jonathan J. Koehler (1993). Error and exaggeration in the presentation of DNA evidence. Jurimetrics Journal, 34, 21, 23.

<sup>27</sup> For reviews Koehler (1995) at 206-210; William C. Thompson (1995), Subjective interpretation, laboratory error and the value of forensic DNA evidence: Three case studies, Genetica, 96, 153-268.

<sup>28</sup> Koehler (1993), at 24.

ignored.

**"Even if errors do occur in proficiency test situations, errors have never been documented in DNA case work"**

The claim that errors have never been committed in DNA case work is not only untrue,<sup>29</sup> but it is one that should be accorded little weight even if it were true. A failure to uncover errors in case work gives no more reason to believe that all case work results are accurate than a failure to produce documented proof of human rights abuses in Iran gives reason to believe that all Iranian citizens are treated with dignity. The fact is that there is usually no independent way to determine the veracity of reported DNA matches in case work.<sup>30</sup>

Proficiency tests, on the other hand, do not have this problem. In test situations, it is known whether two genetic samples share a common source prior to DNA testing. Thus, when a laboratory reports that two samples do or do not match in a test situation, the veracity of the report can be determined. However, because it is not ordinarily known whether two casework samples match prior to DNA testing, documented instances of true and false match reports will not be abundant.<sup>31</sup> For this reason, one must exercise caution when drawing

---

<sup>29</sup> Thompson (1995); Eric Lander (1989), DNA fingerprinting on trial, *Nature*, 339:501-505; National Research Council, 1992. *DNA Technology in Forensic Science*, National Academy Press, Washington D.C. A false positive error was recently committed by Cellmark in a San Diego rape case after it mistakenly switched the reference samples from the defendant and victim.

<sup>30</sup> When multiple samples are tested at multiple laboratories and identical similar results are found, our confidence in the accuracy of the reported results should increase. But even in such cases, the possibility that an error occurred in the collection and preparation stages remains. For example, an early mislabeling or contamination of a sample could lead each of many laboratories to falsely report a match between recovered genetic material and a suspect even where no match exists.

<sup>31</sup> Some prosecutors and DNA experts argue that the testing of control samples in casework constitutes a proficiency test. But match/no match declarations are not ordinarily made on control samples. Moreover, even if such declarations were made, analyses of control samples are not "tests" in the usual sense of the word because the person who performs the analysis knows the answer that he should get. Obviously, the chance that one will report the correct answer under such conditions is increased. And even if analyses of control samples are regarded as "tests" in some loose sense of the word, laboratories do not track their control sample errors and/or count these as errors on their scorecards.

inferences about the dearth of documented case work errors. Instead, one should attend to the frequency of documented errors on well-constructed proficiency tests.<sup>32</sup>

**"Even if errors do occur in DNA case work, proficiency test data do not provide a measure of the chance of error in any particular case."**

The claim that proficiency test data cannot be used to estimate case-specific error rates is also unpersuasive. The argument runs roughly as follows: At best, proficiency test data provide an indication of a general, industry-wide error rate in DNA typing. But the specific testing features associated with each laboratory and each test render the industry-wide error rate irrelevant to any given case.<sup>33</sup>

This argument is an extreme form of the base rate fallacy.<sup>34</sup> No one would argue that the unique testing features associated with a particular laboratory in a particular case should be ignored when these features are demonstrably related to a reduced error rate. But it is a fallacy to believe that the fact that such individuating features may exist denies relevance to industry-wide error rate estimates.

By this reasoning, one should predict that nearly all newly married couples will stay married, that most major-college football players will play professional football, and that most

---

<sup>32</sup> **Blind external proficiency tests yield more reliable information than internal open tests. ("Most important, there is no substitute for rigorous external proficiency testing via blind trials. Such proficiency testing constitutes scientific confirmation that a laboratory's implementation of a method is valid not only in theory, but in practice. No laboratory should let its results with a new DNA typing method be used in court, unless it has undergone such proficiency testing via blind trials." National Research Council, 1992. *DNA Technology in Forensic Science*, National Academy Press, Washington D.C., p. 55).**

<sup>33</sup> **Kathryn Roeder (1994), *DNA Fingerprinting: A review of the controversy (Rejoinder)*, *Statistical Science*, vol. 9, 222, 274 ("If a probability of error is presented, it should be the posterior probability of an error for the case at hand").**

<sup>34</sup> **Maya Bar-Hillel (1980) *The base-rate fallacy in probability judgments*. *Acta Psychologica* vol. 44, 211-233; Daniel Kahneman & Amos Tversky (1973) *On the psychology of prediction*. *Psychological Review* vol. 80, 237-251. For a discussion of the base rate fallacy in the law, see Jonathan J. Koehler (1993) *The normative status of base rates at trial. Individual and group decision making* (pp. 137-149), ed. N. J. Castellan, Erlbaum. Also, see Jonathan J. Koehler (in press). *The base rate fallacy reconsidered: Normative, descriptive and methodological challenges*. *Behavioral and Brain Sciences* for an argument that the base rate fallacy may not be as widespread as others have suggested.**

law professors who submit articles to the Harvard Law Review can expect an acceptance letter. Why? Because, in each case, it is easy to identify a host of individuating features that support the favorable outcome. The Jones newlyweds love each other deeply, Smith the college football player has an influential agent, and Taylor the law professor is working in a hot area of the law.

But depressing base rate frequency statistics tell us that 50% of marriages end in divorce, 99% of major college football players do not make it into professional football,<sup>35</sup> and more than 99% of articles submitted to the Harvard Law Review are rejected. Surely a person who takes these base rate statistics into account will make more accurate judgments than a person who relies solely on a select sample of individuating information that is consistent with the favorable outcome.

Likewise, ignoring the industry-wide DNA error rate statistics because there exist favorable individuating features in this laboratory or in this test will produce overly optimistic estimates of the chance that the laboratory made an error in the instant case. **Because favorable individuating features always exist (e.g., the analyst is experienced, the analyst was observed, the samples were clean, the results were double-checked, etc.) extreme caution must be exercised before concluding that the chance of error in the instant case is lower than the chance of error in the industry as a whole.**<sup>36</sup> It cannot be that every laboratory in the industry has an error rate that is less than the industry-wide average. Yet this is exactly the claim that is made every time laboratory personnel reject the industry-wide

---

<sup>35</sup> R. Harvey (3/17/91). Win or lose on field, athletes lose out big-time in education. Los Angeles Times.

<sup>36</sup> **Scientific standards should be employed when evaluating the impact that individuating information should have on judgment. For example, a claim that the chance of error is reduced by special training or by the use of a particular protocol should be bolstered by data that show that specially trained analysts and analysts who follow a particular protocol commit fewer mistakes than other DNA analysts. The burden of proof rests with those who argue that industry-wide figures are inapplicable to make a scientifically compelling case in favor of their position.**

figures, and argue that their own error rate is substantially less or zero.

### Laboratory-Specific Error Rates

Even among those who understand that base rate frequencies are relevant to case-specific prediction, many are uncomfortable using industry-wide error rates to reflect the quality of a specific laboratory. Error base rates are important and useful, but shouldn't we use the error rates for this laboratory, rather than those for the entire industry?

If a sufficiently reliable set of such data were available, it would indeed be reasonable to use these data rather than the industry-wide figures. The problem, however, is that few laboratories have participated in enough high quality proficiency tests to permit computation of reliable error rate estimates. Even error-free performance on a small number of proficiency tests will not ordinarily provide a sufficiently reliable basis for replacing the industry-wide error rate estimates. This elementary principle is poorly understood by many attorneys and judges.<sup>37</sup>

Suppose a laboratory makes 10 correct match calls out of 10 tries on proficiency tests. Based on these data, one may be 95% "confident" that the laboratory's error rate is somewhere

---

<sup>37</sup> **Commonwealth of Virginia v. Antonio Hodges, June 16, 1995, cross exam of Professor Jonathan Koehler by Mr. Arthur Karp:**

**Q: "Well, tell me what you know about how errors are made in processing DNA?"**

**A: I'm not here to comment at all about how errors are made in the processing of DNA. I'm here to comment on the inferential significance of errors that have been committed.**

**Q: What errors have been committed?**

**A: Well, there have been some errors committed in proficiency tests.**

**Q: By the Virginia lab?**

**A: No, not that I'm aware of.**

**Q: Well, tell me what you can say about the implications of error rates for a laboratory that doesn't make errors?**

**A: Well, there is an assumption implicit in your question, [that] the lab doesn't make errors... If there was such a thing as a zero error rate, which I categorically reject and which any scientist should categorically reject, then I really wouldn't have much to say about errors..." (transcript p. 20).**

**Comment by The Court: "We are not going into proficiency tests where there were errors when there is no evidence that the Virginia lab had any errors. Okay." (transcript p. 28).**

between 0 and 26%.<sup>38</sup> In other words, error rates of 0%, 5%, 10%, 20% and 25% are all consistent with these data (i.e., zero errors in ten tries). If a laboratory participated in 100 tests without error then we could be 95% confident that its error rate is between 0% and 3%.<sup>39</sup>

Even when errors are not made, proficiency testing at the individual laboratory level would have to be quite extensive to produce error rate estimates that are substantially below current estimates.<sup>40</sup> Unless and until such extensive testing has been completed, industry-wide error rates should be used when estimating a DNA laboratory's match report error rate.<sup>41</sup>

In sum, the arguments that have been offered to rebut the proposition that error rate estimates should not be used to assess the probative value of reported DNA matches are unconvincing. When scientifically defensible estimates of laboratory error rates are not incorporated into likelihood ratios or related quantitative indexes of probative value, the numbers that jurors receive may exaggerate the strength of the DNA evidence. The possibility of human error imposes practical boundaries on the probative value of DNA evidence. To be sure, estimating the frequency of errors in case work is tricky. But the difficulty of doing so does not justify the production of likelihood ratios that are based on the false assumption that errors do not occur.

---

<sup>38</sup> Let  $e$  = upperbound error,  $n$  = number of successful reports, and  $\alpha$  = Error! Main Document Only. 1-confidence level, then  $(1 - e)^n \leq \alpha = (1 - e)^{10} \leq .05 \Rightarrow e = .2589$ . Error! Main Document Only.

<sup>39</sup>  $(1 - e)^n \leq \alpha = (1 - e)^{100} \leq .05 \Rightarrow e = .0295$ . Error! Main Document Only.

<sup>40</sup> Jonathan J. Koehler, 1993, p. 229 [01-.04]; Richard Lempert, 1991, Some caveats concerning DNA as criminal identification evidence: With thanks to the Reverend Bayes, 13 *Cardozo Law Review* 303, 325 [01-.02]; K. Roeder, 1994 [0008]; William Thompson, 1994, Commentary on Kathryn Roeder, DNA fingerprinting: A review of the controversy, 9 *Stat. Sci.* 263 [00666].

<sup>41</sup> Proficiency data from individual laboratories may be used to supplement these data provided that an explanation of their relative diagnostic value is provided.

## V. Do Likelihood Ratios Belong in Court?

An important question not reached by disagreements about error rates and likelihood ratio values is whether likelihood ratios belong in court. Even if scientists agreed on the proper form and approximate value of DNA likelihood ratios, it is not clear that those who may be charged with describing and using those likelihood ratios know what they mean.

### Experts Do Not Understand Likelihood Ratios

In State of Texas v. Griffith (1996), the prosecution called the director of a Texas DNA laboratory to testify about the significance of a paternity DNA match. The expert used likelihood ratios to describe the significance of the match. He said that the likelihood ratio was 14,961:1, and that this meant the following: "Given this evidence, it is 14,961 times more likely that the defendant is the father than a random man."<sup>42</sup>

This statement is problematic for several reasons. First, it is not clear that the defendant's chances of paternity should be compared to those of a "random man." Instead, the defendant's chances of paternity should be compared to other men who are among those who could be the father of the child.<sup>43</sup> If ten other men could be the father, then the comparison should be between the hypothesis that the defendant is the father, and the alternative hypothesis that at least one of the other ten men is the father.

Second, the expert's statement is not a likelihood ratio. It is a posterior odds ratio. The distinction between posterior odds ratios and likelihood ratios is discussed below.

### Jurors Will Not Understand Likelihood Ratios

Even when likelihood ratios are properly conveyed, there is little reason to believe that jurors will understand what they mean and how they should be used. Although they have

---

<sup>42</sup> State of Texas v. Griffith, Lubbock, TX, January 24, 1996.

<sup>43</sup> The appropriate reference population should be either the suspect population or the potential source population. See Koehler, 1993 and Lempert, 1991.

scientific merit, likelihood ratios--which are the ratios of conditional probabilities--are not easy to understand.

In the Simpson case, prosecution DNA expert Professor Bruce Weir prepared a report in which he presented a series of likelihood ratios to describe the probative value of some blood matches between crime scene samples and Mr. Simpson.<sup>44</sup>

The defense opposed admission of Professor Weir's likelihood ratios on several grounds. First, it claimed that they failed to quantify and take account of all relevant variables.<sup>45</sup> For example, Professor Weir's likelihood ratios did not incorporate the possibility of laboratory error.<sup>46</sup> The defense also objected to the introduction of likelihood ratios on grounds that jurors might confuse them with statements about the probability that Mr. Simpson was, in fact, the source of the samples.<sup>47</sup> It argued that the significance of the DNA matches should be described using statistics that identify the frequency with which the matching profiles exist in various populations.<sup>48</sup>

Judge Ito agreed to a hearing on the issue outside the presence of the jury. Two days before the hearing, the prosecution abandoned its plans to introduce likelihood ratios and agreed to use frequency statistics instead.<sup>49</sup>

---

<sup>44</sup> Weir, 1995.

<sup>45</sup> See Defense motion: Notice of objections to testimony concerning DNA evidence and memorandum in support thereof, at page 39 citing *People v. Cella* (1983) 139 Cal.App.3d 391, 405, 188 Cal. Rptr. 675, 684 ("If mathematical probabilities are to be of any use in the courtroom setting, all crucial variables must be quantified exactly" (emphasis in original)).

<sup>46</sup> When asked why he didn't incorporate error rates into his computations, Professor Weir said "I reject the notion of error rates in case work. I don't think that those are relevant." *People v. Simpson*, vol. 175, June 26 1995.

<sup>47</sup> Defense Motion, p. 38.

<sup>48</sup> Defense Motion, p. 36-7 ("The correct way to characterize statistics on genotype frequency is to call them what they are--statistics on the frequency of certain characteristics in a particular reference population").

<sup>49</sup> Although this likelihood ratio issue was settled in advance, a hearing was held to determine which of two methods would be used to compute the frequencies.

Holding aside the issue of whether Professor Weir's likelihood ratios adequately accounted for all relevant factors, suppose the jury heard the following likelihood ratio: "It is 270 million times more likely that we would see this evidence if Mr. Simpson were the source of the blood than if Mr. Simpson were not the source." Compare this to the following statement: "Given this evidence, it is 270 million times more likely that Mr. Simpson is the source of the blood than that Mr. Simpson is not the source." Is it reasonable to expect that jurors will understand that these two statements are different? The first statement is a likelihood ratio because it describes the value of the evidence in light of the competing hypotheses. The second statement is a posterior odds ratio because it describes the probability of each of the two competing hypotheses in light of the evidence. It seems unlikely that jurors will appreciate this distinction any more than did the expert in *Griffith*. Indeed, one recent study showed that 77% of people verbally confused likelihood ratios with posterior odds!<sup>50</sup>

A review of the psychological literature on how people reason with probabilities provides good reason to believe that jurors will have trouble understanding and using likelihood ratios. Studies have shown that people (a) confuse the conditional probability  $P(H|D)$  with the conjunctive probability  $P(H\&D)$ ,<sup>51</sup> (b) commit the inversion fallacy (i.e., erroneously equate  $P(D|H)$  with  $P(H|D)$ ),<sup>52</sup> and (c) are less likely to engage in sound

---

<sup>50</sup> Christopher R. Wolfe, 1995, *Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account*, *Journal of Behavioral Decision Making*, vol. 8, 85-108, 97.

<sup>51</sup> Gerd Gigerenzer & Ulrich Hoffrage, 1995, *How to improve Bayesian reasoning without instruction: Frequency formats*, *Psychological Review*, vol. 102, 684-704; Stephen E. Edgell, Robert M. Roe, and Clayton H. Dodd, *in press*, *Base rates, experience and the big picture*, *Behavioral and Brain Sciences*, vol. 19.

<sup>52</sup> David H. Kaye, & Jonathan J. Koehler, 1991, *Can jurors understand probabilistic evidence?* *Journal of the Royal Statistical Society, Series A*, vol. 154, 75-81, 77 (use the term "inversion fallacy"); William C. Thompson, 1989, *Are juries competent to evaluate statistical evidence*, *Law and Contemporary Problems*, vol. 52, 9-41; Casscells, W., Schoenberger, A., & Graboys, T. B. (1978) *Interpretation by physicians of clinical laboratory results*. *New England Journal of Medicine* 299:99-1001 (45% inverse errors among Harvard physicians); Leda Cosmides & John Tooby (in press) *Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty*. *Cognition* (56% inverse errors among Stanford students).

probabilistic reasoning when provided with information in conditional probability form than when provided with information in frequency form.<sup>53</sup> Thus, not only is there reason to suspect that jurors will reason poorly with likelihood ratios (which are the ratio of conditional probabilities), but there is at least some reason to suspect that the use of likelihood ratios invites more problems than the use of frequencies. Whereas the meaning of a frequency statement such as "One out of every 270 million African-Americans share this DNA profile" seems clear, the meaning of a likelihood ratio statement such as "It is 270 million times more likely that we would see this evidence if Mr. Simpson were the source of the blood than if Mr. Simpson were not the source" is much less clear.

With this in mind, Professor Weir's contention that it is "immaterial" whether DNA statistics are presented as likelihood ratios or as frequencies is unpersuasive.<sup>54</sup> Although Professor Weir may believe that the mathematics of likelihood ratios and frequencies are the same,<sup>55</sup> the psychology of these statistical representations is not. Studies show that people treat

---

<sup>53</sup> Cosmides & Toobey (in press) (inversion error among Stanford students reduced to 5% (from 56%) when frequencies used; William C. Thompson & Edward L. Schumann, 1987, Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, vol. 11, 167-187, exp 1 (22% committed inverse fallacy on blood matching evidence in the context of a hypothetical robbery case when the evidence was presented in P(E|G) form. However, a frequency presentation of the blood evidence produced inverse fallacies only 4% of time).

<sup>54</sup> "Whether or not we give likelihood ratios is, in a sense, immaterial" Hearing: cross-exam of Professor Bruce Weir by Mr. William Thompson, *People v. Simpson*, 6/22/95.

<sup>55</sup> At the hearing on June 22, 1995, the following exchange took place between Mr. Thompson and Professor Weir on cross-exam in *People v. Simpson*:

**Q: And this report also used likelihood ratios to characterize the value of mixtures?**

**A: Oh, as does my current report. The phrasing is still there. But my current report is focusing on this component of the likelihood ratio.**

**Q: Uh-huh. And in fact the report says likelihood ratios are essential (emphasis in original) in interpreting the evidentiary value of mixed stains, doesn't it?**

**A: Yes, that is true, yes. Whether that is single or mixed—and, of course, we have them throughout this case—every number that is on the board is in fact giving us a likelihood ratio. The numbers I have discussed this morning for the mixture stains are giving us likelihood ratios by inverting them.**

**Q: Likelihood ratios for distinguishing between what hypotheses, Dr. Weir?**

conditional probabilities and frequencies differently, and reasoning appears to be superior with the latter. Because likelihood ratios are ratios of conditional probabilities, courts should exercise caution before admitting likelihood ratios.

## VI. An Experiment: Likelihood Ratios vs. Frequencies

In order to test the impact of likelihood ratio vs. frequency statistics in DNA cases more directly, I conducted an experiment at UT Austin in February, 1996. In this study, 256 jury-eligible subjects<sup>56</sup> were provided with a written summary of a rape case (see Appendix). The suspect in the case was identified when the victim selected his photograph from a larger set of police photographs. At trial, the victim testified that she was "pretty sure" (though not positive) that the defendant was the one.

The case included evidence of a DNA match between the defendant and a semen trace that was recovered from a rape kit examination of the victim. Subjects were presented with additional information about this DNA match in one of three forms: a frequency statistic, a likelihood ratio, or a posterior odds ratio statement. The statistical information was based on a profile frequency of either 1 in 100 or 1 in 1000. Holding aside error rate considerations,<sup>57</sup> the profile frequencies of 1 in 100 and 1 in 1000 are roughly equivalent to likelihood ratios of 100:1 and 1000:1 respectively.

---

**A: The hypothesis of the two contributors being known versus them both being unknown.**

**Q: Okay. Are you planning to tell the jury that those frequency numbers are in fact likelihood ratios?**

**A: If I am asked under oath I would have no choice but to say 'Of course they are.'**

**(Transcript, *People v. Simpson*, June 22, 1995).**

<sup>56</sup> **Subjects were felony-free U.S. citizens, eighteen years of age or older.**

<sup>57</sup> **As noted earlier, it is not scientifically acceptable to treat fallible evidence as if it were not. Nor is it scientifically acceptable to rely on population frequency estimates in cases where the suspect population or potential source population include close genetic relatives of the defendant. See Koehler, 1993, p. 26-7, Lempert 1991, p. 310; Lempert, The suspect population and DNA identification, 34 *Jurimetrics J.* 1 (1993). But because likelihood ratios that are merely the inverse of population frequencies may soon be offered in court, this study examined mock jurors' reactions to likelihood ratios that were created this way.**

Subjects in the frequency condition were told: "We would expect to see the DNA profiles observed in the semen and the defendant in approximately 1 person out of every 100 (1000)." Subjects in the likelihood ratio condition were told: "It is approximately 100 (1000) times more likely we would see this match if the defendant is the source of the semen than if the defendant is not the source of the semen." Subjects in the posterior odds ratio condition were told: "Given that we see this match, it is approximately 100 (1000) times more likely that the defendant is the source of the semen than that he is not the source of the semen." Subjects were assigned at random to one of these conditions. Error rates were not divulged, although subjects heard the expert testify that "there's always room for error and coincidental matches." Subjects also heard the defense argue that the match was probably coincidental (see Appendix).

After studying the case, subjects estimated the probability that the defendant was the source of the sample, and rendered a verdict (guilty, not guilty). The probability data were analyzed using analysis of variance, and the verdict data were analyzed using chi-square ( $\chi^2$ ).<sup>5</sup> Because there were no main effects or interactions involving either (a) the numerical size of the profile frequency (100 vs.1000), or (b) gender (male vs. female), the data were collapsed across these variables. The results are provided in Figures 1 and 2 below.

-----  
Insert Figure 1 About Here  
-----

Figure 1 gives the mean probability that subjects in each of the three conditions assigned to the proposition that the suspect was the source of the recovered semen sample. Significant differences among the conditions emerged ( $F_{2,248}=5.41$ ,  $p=.005$ ). Specifically, subjects who received profile frequencies assigned lower probabilities to the proposition that the suspect was the source than did subjects who received likelihood ratios or posterior odds ratios (65% vs. 75% and 80% respectively). Post-hoc contrasts indicated that the difference

between the frequency and likelihood ratio conditions were statistically significant ( $t_{166}=1.95$ ,  $p=.05$ ), as was the difference between the frequency and posterior odds conditions ( $t_{164}=3.14$ ,  $p=.002$ ). The difference between the likelihood ratio and posterior odds conditions was not significant ( $t_{166}=1.31$ ,  $p=.19$ ).<sup>58</sup>

-----  
Insert Figure 2 About Here  
-----

Figure 2 gives the proportion of subjects in each of the three conditions who rendered guilty verdicts. Significant differences emerged ( $\chi^2(df=2, n=XXX[INSERT])=7.78$ ,  $p=.02$ ).<sup>6</sup> The pattern is similar to that which was found for the source probabilities. Subjects who received profile frequencies were less likely to render a guilty verdict than were subjects who received likelihood ratios or posterior odds ratios (21% vs. 35% and 41% respectively).<sup>59</sup> Post-hoc comparisons indicated that the difference between the frequency and posterior odds conditions was significant ( $\chi^2(df=1, n=169)=7.65$ ,  $p=.006$ ),<sup>7</sup> and the difference between the frequency and likelihood ratio conditions was marginally statistically

---

<sup>58</sup> By convention, p values less than or equal to .05 are widely regarded as "statistically significant." This means that, if the null hypothesis that differences between conditions do not exist is true, then differences smaller than that observed would occur at least 95% of the time in repeated sampling. See Stephen E. Fienberg, *The evolving role of statistical assessments as evidence in the courts*, Springer Verlag, NY, 1989, p. 196.

<sup>59</sup> As an aside these data suggest that jurors may not be as impressed with DNA matching evidence as many believe. Fewer than half of the subjects voted to convict in any condition. Even in the posterior odds ratio condition, where subjects were implicitly told that the probability that the defendant was the source of the semen was at least 99% (because a posterior odds ratio of 100:1 is approximately equal to a 99% source probability), the mean source probability estimate was only 80%. This is all the more intriguing when we consider that subjects were also told that the victim made a partial identification of the suspect. As suggested elsewhere, it may be that jurors use an averaging heuristic to assess the strength of the evidence rather than the rules of probability theory. See Koehler, Chia and Lindsey, 1995, p. 212; R. Lempert DNA, science and the law: Two cheers for the ceiling principle, 34 *Jurimetrics* 41, 54 (1993). Thus, DNA evidence—in combination with a less than certain eyewitness identification—may actually create more uncertainty in the minds of jurors about whether the defendant is the perpetrator relative to situations in which there is less (but unequivocal) evidence. These data do not speak directly to this hypothesis, but it is one that deserves empirical scrutiny.

significant ( $\chi^2(df = 1, n = 171) = 3.61, p = .058$ ).<sup>8</sup> The difference between the likelihood ratio and posterior odds conditions was not significant ( $\chi^2(df = 1, n = 172) = 0.82, p = .36$ ).<sup>9</sup>

These data suggest that people do not treat frequencies and likelihood ratios similarly. Subjects who received likelihood ratios were more sure that the suspect was the source, and were more willing to convict than were subjects who received the frequency statistics. Differences between the likelihood ratio and posterior odds groups were not found on either source probability estimates or verdicts.

Although these data should be replicated by a more comprehensive jury study, they appear to rebut the contention that it is "immaterial" whether jurors are provided with frequencies or likelihood ratios. When viewed in the context of the inversion fallacy literature, the data suggest that people treat likelihood ratios more like posterior odds ratios than like frequencies. This is problematic because posterior odds ratios speak directly to the issue of the probability that the defendant is the source of the genetic material. But because DNA evidence cannot answer this source probability question directly without making unfounded assumptions about the strength of the nongenetic evidence,<sup>60</sup> the introduction of likelihood ratios may give some jurors a false sense that DNA evidence does answer this question directly.

## VII. Conclusion

The DNA Wars<sup>61</sup> have moved to a different level during the past few years. At one time, debate centered around issues biological and genetic. Today, the most controversial issues are statistical, as the new debate focuses on ways to convey the probative value of DNA evidence to jurors.

---

<sup>60</sup> Koehler, 1993 (Judicature), p. 223.

<sup>61</sup> William C. Thompson, 1993, Evaluating the admissibility of new genetic identification tests: Lessons from the "DNA war," J. of Crim. Law and Criminology, vol. 84, 22-104, 22.

Resolution of the new debate requires scientists to address such questions as "what is the hypothesis of interest?" "what is meant by 'the evidence'?" and "how often do errors occur?" Is the evidence "the suspect matches a crime scene sample," or is it "the suspect reportedly matches a crime scene sample?" I have argued that it is the latter. Is the hypothesis of interest it "the suspect is the source of the genetic evidence," "the suspect had contact with the crime scene," or "the suspect is guilty of this crime"? I have argued that the first statement should be used. Can we estimate the frequency with which error occurs? I have argued that we can and we must, even if our estimates rely more on industry-wide data than case-specific data.

In light of the politically charged atmosphere that swirls about DNA evidence,<sup>62</sup> disagreement will not doubt persist. But even if agreement is reached on these thorny issues, thornier psychological issues must also be addressed. How do people process quantitative evidence in general, and likelihood ratio and frequency information in particular? Can jurors be taught to aggregate different types of quantitative evidence and to give it the weight that it deserves?<sup>63</sup> These psychological questions may yet turn out to be the most important of all. If people do not process quantitative information effectively, and if we continue to inundate jurors with statistical data without bothering to train them in the use of such information, then much of the debate over what the right numbers are will have been a waste of time.

---

<sup>62</sup> Leslie Roberts, *Science in court: A culture clash*, 257 *Science* 732 (1991).

<sup>63</sup> For an early review, see Kaye and Koehler, 1991; Thompson, 1989.

## **Appendix: DNA Experiment**

### **Likelihood Ratio Condition**

On March 8th at approximately 2:00 p.m., a woman in a small Louisiana town was raped. After receiving treatment at a local hospital, the victim was sent to the police station. There she was presented with photographs of possible assailants from police files. After a few hours, the victim selected the photograph of George Edwardson as one that "looks a lot like" the man who attacked her.

After being questioned by police, Mr. Edwardson agreed to provide blood and hair samples for DNA analysis. The DNA laboratory reported that Mr. Edwardson's DNA profile matched the DNA profile associated with sperm that was recovered from the victim. Mr. Edwardson was arrested shortly thereafter.

At trial, the prosecution called two witnesses. First, he called the victim. She testified that, although she could not positively say that Mr. Edwardson was the perpetrator, she was "pretty sure" that he was the one.

The prosecution then called the scientist who conducted the DNA analysis. After explaining the logic and science of DNA analysis, the scientist testified that the DNA profile from Mr. Edwardson matched the DNA profile from the sperm recovered from the victim. He also testified that this meant that Mr. Edwardson could not be ruled out as a possible source of the recovered semen. In addition, the following exchange took place:

Q: "And Dr., can you positively identify Mr. Edwardson as the source of the semen recovered from the victim?"

A: "No, a DNA analysis is not capable of positively identifying anyone as the source of genetic material that gets analyzed in the laboratory. There's always room for error and coincidental matches. However, in this case, I can tell you that Mr. Edwardson did match the semen sample that we analyzed. Furthermore, it is approximately 100 times

more likely we would see this match if Mr. Edwardson were the source of the semen than if Mr. Edwardson were not the source of the semen."

The defense did not call any witnesses, but argued that the prosecution had failed to prove its case. Moreover, the defense argued that the DNA match was probably coincidental since Mr. Edwardson was not the only person who could not be ruled out as a possible source of the recovered semen.

### Questions

What would you say is the numerical probability that Mr. Edwardson is the source of the semen? \_\_\_\_\_ %

What verdict would you return: Not Guilty or Guilty? \_\_\_\_\_