# Compounding Evidence from Multiple DNA-Tests

SAM C. SAUNDERS
Scientific Consulting Services, Inc.
P.O. Box 458
Kirkland, WA 98083

N. CHRIS MEYER
DANE W. WU
Pacific Lutheran University
Tacoma, WA 98447

**1. Introduction** The sensational trial of a national football hero, O.J. Simpson, for the brutal murder of his former wife, Nicole, and her friend Ron Goldman, has become one of the most controversial trials in American history. It ranks with the 1925 trial of John Scopes, for teaching evolution in biology, as a premier example of a divisive and controversial ruling by the American juridical system (cf. [**5**]).

The American system of jurisprudence requires the jury to hold the defendant innocent until proved guilty beyond a reasonable doubt. If jurors are confused about the degree of incrimination of certain pieces of evidence, then doubt (reasonable or not) will ensue. The outcome of the O.J. trial hinged upon the jury's understanding of the probabilities of DNA blood-matching, a technique to establish guilt which was tested in this trial as never before. Some experts claim that DNA can only be used to exonerate persons accused of a crime, but never to establish guilt (see, e.g., [**7**]). The jury's acquittal was controversial. One columnist ([**2**]) ascribed it to their innumeracy. Others said the prosecution could not explain what they, themselves, did not understand. Justice requires that the probabilities of such events as matching fingerprints or matching DNA from blood samples be understandable. In this note we do not concern ourselves with the chemical problems of DNA matching. Neither do we consider the possibility of blood-sample contamination or of a law enforcement conspiracy to fake evidence. (These complications are discussed in [**4**].) Here we treat only some of the subtleties of conditional probabilities related to DNA matching.

**2. Conditional probabilities and guilt** Let $M$ be the event of a DNA match between the defendant's blood and blood found at the crime scene, let $I$ be the event the defendant is innocent, and let $I'$ be the complement of $I$—the event the defendant is guilty. There are reliable measures for $P(M \mid I)$, the conditional probability of $M$ given $I$. If the defendant is innocent, the blood must be someone else's; hence $P(M \mid I)$ is the probability of a match between blood samples of two different individuals.

Human DNA signatures are so distinctive that the probability of a random match is very low. Depending upon the amount of DNA obtained, it is often found that $P(M \mid I)$ is between $10^{-8}$ and $10^{-10}$. But it is $P(I \mid M)$, the likelihood of innocence given the evidence, that should be considered by the jurors, not $P(M \mid I)$. These two conditional probabilities are related by Bayes' little theorem:

$$P(I \mid M) = \frac{P(I)}{P(M)} P(M \mid I). \tag{1}$$

We see from (1) that $P(I \mid M) \neq P(M \mid I)$ unless $P(I) = P(M)$. The erroneous assumption that $P(I \mid M) = P(M \mid I)$ is sometimes called the *prosecutors' fallacy*. (For a discussion of the confusion this fallacy causes, see **[3]**.)

If in formula (1) we substitute for $P(M)$ from the theorem of total probability, we obtain Bayes' big theorem:

$$P(I \mid M) = \frac{P(I)P(M \mid I)}{P(I)P(M \mid I) + P(I')P(M \mid I')}.$$

It follows, since $(1 + x^{-1})^{-1} < x$ for $x > 0$, that

$$P(I \mid M) < \frac{P(I)P(M \mid I)}{P(I')P(M \mid I')} \approx \frac{P(I)P(M \mid I)}{P(I')}. \tag{2}$$

The approximate equality follows because $P(M \mid I')$, the probability of a match given that the defendant is guilty, is virtually unity. Here $P(I)$ and $P(I')$ are the *a priori* probabilities of innocence and guilt, before the evidence has been introduced. Thus we see from (2) that the prosecutor's fallacy is not so much a fallacy as ignorance of a factor of proportionality (the prior odds, $P(I)/P(I')$) in an upper bound on the true probability of innocence given a DNA match.

**3. The compounding evidence** In many cases, such as the O.J. trial, more than one DNA matching is involved. Our calculations thus far only deal with one piece of evidence, $M$. The jury needs to consider the combined effect of all the evidence (cf. **[6]**). Suppose events $M_1, \dots, M_k$ are all pieces of evidence introduced against the defendant. We want to calculate $P(I \mid \bigcap_{i=1}^{k} M_i)$, the probability of innocence given all the evidence. The theorem below gives an upper bound for this probability. First, we introduce some needed definitions.

For a given event $I$, we denote $P_I(\cdot) = P(\cdot \mid I)$. Two events $M_1$ and $M_2$ are said to be *conditionally independent* with respect to $I$ if $P_I(M_2 \mid M_1) = P_I(M_2)$. Note that the events $M_1$ and $M_2$ can be interchanged in this definition and that the conditional independence becomes mutual independence if $I$ is a sure event.

The ratio $P(M_2 \mid M_1)/P(M_2)$ gives a measure of how strongly associated $M_2$ and $M_1$ are. If the ratio is less than 1 then $P(M_2 \mid M_1) < P(M_2)$, which means that $M_2$ is less likely to occur given $M_1$. Likewise, $M_2$ is more likely to occur given $M_1$ if the ratio is greater than 1. The ratio is 1 if and only if $M_1$ and $M_2$ are mutually independent.

With this in mind, we say that two events $M_1$ and $M_2$ are *more strongly associated conditionally* given $I'$ than given $I$ if

$$\frac{P_{I'}(M_2 \mid M_1)}{P_{I'}(M_2)} \geq \frac{P_I(M_2 \mid M_1)}{P_I(M_2)};$$

this is equivalent with

$$\frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1)P_{I'}(M_2)} \geq \frac{P_I(M_1 M_2)}{P_I(M_1)P_I(M_2)}. \tag{3}$$

$M_1$ and $M_2$ are said to be *equally associated conditionally* under $I$ and $I'$ if equality holds in (3) (see **[1]**). We leave to the reader the proof of the following fact: If $M_1$ and $M_2$ are conditionally independent with respect to both events $I$ and $I'$, then they are equally associated conditionally under $I$ and $I'$.

The following simple example illustrates equal association. Consider throwing two fair dice, say $(X, Y)$, with events $M_1 = \{(x, y): y = 3, 4, \text{ or } 5\}$, $M_2 = \{(x, y): x = 1 \text{ or } 2\}$, and $I = \{(x, y): x + y = 7\}$. In this case $M_1$ and $M_2$ are equally associated conditionally under $I'$ and $I$, since

$$\frac{P_I(M_1 M_2)}{P_I(M_1) P_I(M_2)} = \frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1) P_{I'}(M_2)} = 1.$$

If we modify event $M_2$ as $M_2 = \{(x, y): x = 1 \text{ or } 2\} \cup \{(3, 3)\}$, then $M_1$ and $M_2$ are more strongly associated conditionally given $I'$ than given $I$, since, in this case,

$$\frac{P_I(M_1 M_2)}{P_I(M_1) P_I(M_2)} = 1 \quad \text{and} \quad \frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1) P_{I'}(M_2)} = \frac{12}{11}.$$

In general, events $M_1, \dots, M_k$ are said to be *more strongly associated conditionally* given $I'$ than given $I$ if

$$\frac{P_{I'}\left(\bigcap_{i=1}^{k} M_i\right)}{\prod_{i=1}^{k} P_{I'}(M_i)} \geq \frac{P_I\left(\bigcap_{i=1}^{k} M_i\right)}{\prod_{i=1}^{k} P_I(M_i)}. \tag{4}$$

We have equal conditional association when the quality in (4) holds. (This definition is related to the concept of association; see **[1]** for more details.)

THEOREM. *If events $M_1, \dots, M_k$ are more strongly associated conditionally given $I'$ than given $I$, then*

$$P\left(I \mid \bigcap_{i=1}^{k} M_i\right) \leq \frac{P(I)}{P(I')} \prod_{i=1}^{k} \frac{P(M_i \mid I)}{P(M_i \mid I')}. \tag{5}$$

*Proof.* From Bayes' big theorem we obtain

$$P\left(I \mid \bigcap_{i=1}^{k} M_i\right) = \frac{P(I) P\left(\bigcap_{i=1}^{k} M_i \mid I\right)}{P(I) P\left(\bigcap_{i=1}^{k} M_i \mid I\right) + P(I') P\left(\bigcap_{i=1}^{k} M_i \mid I'\right)}$$

$$\leq \frac{P(I) P\left(\bigcap_{i=1}^{k} M_i \mid I\right)}{P(I') P\left(\bigcap_{i=1}^{k} M_i \mid I'\right)}. \tag{6}$$

From the fact that $M_1, \dots, M_k$ are more strongly associated conditionally given $I'$ than given $I$, we have

$$\frac{P\left(\bigcap_{i=1}^{k} M_i \mid I\right)}{P\left(\bigcap_{i=1}^{k} M_i \mid I'\right)} \le \frac{\prod_{i=1}^{k} P(M_i \mid I)}{\prod_{i=1}^{k} P(M_i \mid I')}. \tag{7}$$

Thus (5) follows by substituting (7) into (6).

**4. An application of the theorem** We are concerned in the O.J. trial with evaluating the probability of guilt given the totality of DNA evidence. Let $M_1$ be the event that a blood drop found near the victims' bodies is consistent with the defendant's blood. Cellmark Diagnostics, the DNA laboratory of record in the trial, said that only one person in 170 million could be expected to match the genetic markers identified in the blood drop. Thus $P(M_1 \mid I) = (1.7 \times 10^8)^{-1} = 5.88 \times 10^{-9}$. Let $M_2$ be the event that blood on a sock belonging to the defendant and found in his bedroom is consistent with Nicole's blood. Cellmark said the probability was one in 6.8 million that another person would match the genetic markers they found in the victim's blood (see [3]). Thus $P(M_2 \mid I) = (6.8 \times 10^9)^{-1} = 1.47 \times 10^{-10}$. Moreover, as mentioned above, $P(M_1 \mid I') \approx 1$ and $P(M_2 \mid I') \approx 1$.

In order to apply the theorem, one must argue that $M_1$ and $M_2$ are more strongly associated conditionally given guilt than given innocence. If $I$ is true, then $M_1$ and $M_2$ are independent. But assuming guilt, one can imagine many scenarios in which the occurrence of one would increase the probability of the other. The two blood-match events would reasonably be judged to be more strongly associated conditionally given $I'$ than given $I$. Substituting these numbers into (5) gives

$$P(I \mid M_1 M_2) \le \frac{P(I)}{P(I')} \times P(M_1 \mid I) \times P(M_2 \mid I) \approx \frac{P(I)}{P(I')} \times 8.65 \times 10^{-19}.$$

Now we give an upper bound for the ratio $P(I)/P(I')$. If we go to the extreme of saying that O.J. Simpson was no more likely to be the killer than anyone else in the world, then $P(I') \approx 10^{-10}$ (realistically, $P(I')$ is much larger than this), and, of course, $P(I) \approx 1$. Hence $P(I)/P(I') \approx 10^{10}$ and $P(I \mid M_1 M_2) \le 8.65 \times 10^{-9}$. The conditional probability of innocence given both DNA matches is so small as to place the defendant's guilt beyond any reasonable doubt.

Many other events could be used to condition the events in the trial of the defendant. For example, the glove found at the defendant's house, matching the one found at the crime scene, revealed genetic markers not only matching the victims' DNA but that of the defendant as well; label this match $M_3$. Let $M_4$ be the match of the defendant's blood to drops leading from the murder scene to the gate. According to the California state DNA analyst, this could have been left by one person in 240,000, including O.J. Let $M_5$ be the match of the DNA extracted from a drop of blood smeared in the white Bronco, which was consistent with that of Ron Goldman. Earlier we claimed that $M_1$ and $M_2$ were independent given $I$. Now $M_1, \dots, M_5$ are not independent given $I$, because some involve the same persons. However, we need not assert such independence in order to apply the theorem; we need only assert that $M_1, \dots, M_5$ are more strongly associated conditionally given $I'$ than given $I$.

**5. An instructive analogy** Comments of some jurors after the trial revealed the opinion that DNA blood-matching is no more certain than fingerprint matching. A simple analogy can illustrate the reliability of DNA evidence. Every human carries DNA in each blood cell containing information which is analogous to the information from one permutation of a deck of cards. There are $52! \approx 8.0658 \times 10^{67}$ possible card permutations. Suppose that, from a drop of blood at the crime scene, there are only a few DNA sites (corresponding to card positions in the deck) from which information can be extracted.

Sometimes, moreover, depending upon the amount of DNA or the degree of contamination, not all the information from a site can be determined. The analogous situation for cards would be that only the number, the suit, or the color (say B or R) of the card can be determined. How incriminating can this be? Assume that partial information is available at only seven card positions, as follows:

$$(\spadesuit, \diamondsuit, 2-\heartsuit, 3-\clubsuit, 10-\heartsuit, 4-\diamondsuit, B).$$

Correspondingly, suppose that at the same DNA site locations in the defendant's blood sample we find

$$(2-\spadesuit, 10-\diamondsuit, 2-\heartsuit, 3-\clubsuit, 10-\heartsuit, 4-\diamondsuit, K-\clubsuit).$$

This corresponds to a DNA match comparable to the suspect's blood being consistent with that found at the crime scene. Assuming that all permutations are equally likely, the probability of obtaining this match is

$$\frac{13}{52} \times \frac{13}{51} \times \frac{1}{50} \times \frac{1}{49} \times \frac{1}{48} \times \frac{1}{47} \times \frac{24}{46} \approx 6.0154 \times 10^{-9}.$$

Only about six persons in a billion would yield a match.

**Acknowledgment.** We thank the referees for their constructive suggestions.

## REFERENCES

1. R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, To Begin With, Silver Spring, MD, 1981, 31.
2. Dave Barry, Happy Newt Year, column in the December 31, 1995, Knight-Ridder newspapers.
3. Robin Clark, Will DNA turn the tide?, *Philadelphia Inquirer*, May 21, 1995.
4. R. Gomulkiewicz and N. A. Slade, Legal standards and the significance of DNA evidence, *Human Biology* 69, 1997.
5. Edward J. Larsen, *Summer for the Gods: The Scopes Trial and America's Continuing Debate over Science and Religion*, Basic Books, New York, NY, 1997.
6. S. C. Saunders, An application of conditional probability, or, O.J. + DNA = QED, Mathematics Notes from Washington State University, Vol. 38, No. 2, 1995.
7. David Pringle, Who's the DNA fingerprinting pointing at?, *New Scientist*, January 29, 1994.