Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003).
Communicating statistical DNA evidence. Jurimetrics, 43, 147-163.

ARTICLES

COMMUNICATING STATISTICAL DNA EVIDENCE

Samuel Lindsey*
Ralph Hertwig
Gerd Gigerenzer**

ABSTRACT: There is a growing need to present statistical scientific evidence in a form that judges and jurors can understand and evaluate. After examining statistical issues surrounding forensic DNA evidence, this article presents research that demonstrates a way to improve judges' and jurors' understanding of evidence involving probabilities and statistics.

CITATION: Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer, Communicating Statistical DNA Evidence, 43 Jurimetrics J. 147–163 (2003).

Advances in forensic science over the last decade are revolutionizing the possibilities of criminal investigation. In particular, the once controversial use of DNA analyses to link previously identified suspects to crimes[1] is being supplemented by large databases of the DNA types of convicted offenders or arrestees.[2]

---

*Samuel Lindsey is Schlossman Research Fellow, Max Planck Institute for Foreign and International Criminal Law, Günterstalstrasse 73, D-79100 Freiburg im Br., Germany.

**Ralph Hertwig is Research Scientist, and Gerd Gigerenzer is Director, Max Planck Institute for Human Development, Lentzeallee 94, D - 14195 Berlin, Germany.

1. Christopher Anderson, *Academy Approves, Critics Still Cry Foul*, 356 NATURE 552 (1992); Eric S. Lander, *DNA Fingerprinting on Trial*, 339 NATURE 501 (1989); R.C. Lewontin & Daniel L. Hartl, *Population Genetics in Forensic DNA Typing*, 254 SCIENCE 1745 (1991); William C. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the "DNA War,"* 84 J. CRIM. L. & CRIMINOLOGY 22 (1993).

2. *See, e.g.*, The DNA Identification Act of 1994, 42 U.S.C. §§ 3751, 3753, 3793, 3797 (1994); Police and Criminal Evidence Act, 1984, c. 60, §§ 62, 63, 63A, 64 (Eng.); Fred Barbash, *British*

These databases make it possible to identify possible perpetrators of crimes from the DNA that is on file.[3]

Unfortunately, the proper interpretation of a DNA analysis is tangled in a web of statistical complexities that DNA databanks only exacerbate.[4] There is thus a growing need to present such statistical scientific evidence in a form that judges and jurors can understand. This article examines some of the statistical issues surrounding forensic DNA evidence and the inferences they require. It then reviews psychological findings concerning how people respond to different forms of statistical information and offers a way, based on psychological research,[5] to spontaneously improve judges' and jurors' understanding of evidence involving probabilities and statistics.[6]

Part I discusses what DNA analyses can reveal, what they cannot, and how the two are easily confused. Part II examines the statistical inferences and Bayesian reasoning associated with forensic DNA analyses. Two experiments designed to address these issues are then presented, followed by a discussion of their legal implications. The results suggest that the choice of how to present statistics associated with DNA evidence may have significant legal consequences.

I. FORENSIC DNA ANALYSIS

A. Clarifying the Result

A DNA analysis, by itself, can establish only that someone could be the source of a genetic evidentiary sample. Whether that person is in fact the source depends on the integrity of the analysis, the rarity of the DNA profile in question, and any other evidence implying that the suspect is or is not the source of the evidence.

Even if the suspect is in fact the source, however, that does not mean he is guilty of any crime. There may be ample reason for genetic traces of someone other than the perpetrator to be at a crime scene. Alternatively, those genetic traces could have come from someone who has never been there, if, for example, the evidence was planted.

That said, it is not possible to identify the source of any genetic trace based on a DNA analysis alone, even if every individual, with the exception of monozygotic twins, ultimately has a unique genotype. Many forensic DNA

Authorities Launch 1st National DNA Database, WASH. POST, Apr. 11, 1995, at A15; Nicholas Wade, *F.B.I. Set to Open Its DNA Database for Fighting Crime*, N.Y. TIMES, Oct. 12, 1998, at A1.

3. *E.g.*, Michael Fleeman, *First DNA Database "Hit" Leads to Arrest*, ORANGE COUNTY REG., Nov. 9, 1995, at A22.

4. COMMITTEE ON DNA FORENSIC SCIENCE, NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE (1997) [hereinafter NRC II]; Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 MICH. L. REV. 931 (1999).

5. Ulrich Hoffrage et al., *Communicating Statistical Information*, 290 SCIENCE 2261 (2000).

6. *Cf.* D.H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J. ROYAL STAT. SOC'Y (A) 75 (1991), *available at* http://www.jstor.org/view/09641998/di993074/99p0146b/0.

profiles reflect only a limited number of the genotypic features that cannot reasonably be said to be unique. More than one person could have the same such DNA profile, just as more than one person could pick the same combination of numbers in a lottery, even though the probability of that particular combination winning is extremely small.

Although it is unlikely that a criminal suspect would share a DNA profile with a piece of incriminating evidence by coincidence, quantifying this probability requires estimating the population frequency of the varying genetic features (alleles) in a specified reference class, such as a racial group.[7] In the simplest computation, the population frequencies of each of the varying genetic features are deemed to be inherited independently of one another. They are then multiplied together, along with various coefficients, according to the product rule.[8] The result is commonly known as the "random match probability," and it is tantamount to the chance of a DNA match by sheer coincidence. It is also the statistic usually reported to a judge or jury.[9] This statistic indicates the estimated rarity of the DNA profile, although that specific figure may be misleading if it induces the factfinder to ignore the much more likely chance of a laboratory error or errors in subjective judgments made in interpreting the results.[10] Despite past expert testimony that false positive laboratory errors are *impossible*,[11] they do occur, and usually with a frequency several orders of magnitude higher than the chance of a coincidental match.[12]

All this applies equally to DNA matches found by searching through DNA databanks. However, as the number of DNA profiles analyzed increases, so too does the likelihood of finding one that matches by chance alone.[13] DNA matches

---

7. On the estimation of these frequencies, see NRC II, *supra* note 4, at 21. As for which population is an appropriate reference class, see Richard Lempert, *The Suspect Population and DNA Identification*, 34 JURIMETRICS J. 1 (1993); R.C. Lewontin, *Which Population?*, 52 AM. J. HUM. GENETICS 205 (1993).

8. This can produce infinitesimal numbers, the reciprocals of which are well beyond the size of the human population. For example, the chance that the semen on Monica Lewinsky's dress was not former President Bill Clinton's was reported to be 1 out of 7.87 trillion. Final Report of the Office of the Independent Counsel to the United States House of Representatives, "Grounds," at § I(A)(4) (Sept. 9, 1998), *available at* http://icreport.access.gpo.gov/report/2toc.htm. But even this infinitesimal number is consistent with a chance of 1 in 1,430 that someone else in a population of 5.5 billion unrelated people shares this DNA profile.

9. NRC II, *supra* note 4, at 192–204.

10. *Id.* at 203; Jonathan J. Koehler et al., *The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?*, 35 JURIMETRICS J. 201 (1995); William C. Thompson, *Subjective Interpretation, Laboratory Error, and the Value of Forensic DNA Evidence: Three Case Studies*, 96 GENETICA 153 (1995).

11. *See, e.g.*, Commonwealth v. Teixeira, 662 N.E.2d 726, 728 (Mass. App. Ct. 1996).

12. Armstead v. State, 673 A.2d 221, 245 (Md. 1996) (testimony that the rate of false positive errors could be as high as 0.7%); NRC II, *supra* note 4, at 24–25, 85–87; Koehler et al., *supra* note 10. The rate of laboratory errors will vary with individual laboratories, their techniques, and the types of proficiency tests. Experts can testify to errors even if there have not been any detected in proficiency tests. *See* Williams v. State, 679 A.2d 1106, 1118–20 (Md. 1996).

13. The same is true of DNA analyses of samples taken in mass police screenings of men residing in certain geographical regions, which amount to *ad hoc* DNA databanks.

found by searching a large databank at a small number of loci could identify people who have nothing to do with the crime in question. Again, as in a lottery, the chance that any one ticket matches the winning combination may be extremely small, but the chance that some combination on at least one of a large number of tickets matches may not be.

Going through all the other tickets is sure to find the winner eventually, and the chance of finding it before going through all the tickets improves with each ticket examined.[14] The more samples analyzed, the more likely it becomes to find a match purely by chance. How much more likely is a direct function of the number of samples examined, so the chance of a random match must take into consideration the size of the databank.[15] But that does not change the fundamental question of whether the person whose DNA profile matches the evidentiary sample is in fact its source.[16]

B. Confusing the Result

Whatever statistics are reported with the results of a DNA analysis, the interpretation of its probative value is ultimately left to the courts. But these results are widely confused and misinterpreted—by judges, jurors, and sometimes even DNA experts themselves. The estimated population frequency of a DNA profile has often been misinterpreted, for example, as the chance that an accused person is innocent, for example, "a one in 5 billion chance that anybody else could have committed the crime."[17] Even the president of the *Deutschen Gesellschaft für Rechtsmedizin* (German Society for Forensic Medicine) stated that a DNA match identifies a perpetrator with 100% certainty.[18]

The rarity of the DNA profile (or its complement) is also misinterpreted as the likelihood that some person is not (or is) the source of DNA evidence. A leading European expert assumes in his calculations that it is certain (i.e., $p = 1$) that a matching defendant is the source of the evidence, and it is FBI policy to infer the defendant is the source if the profile frequency is smaller than 1 in 260 billion.[19] Other DNA experts have misinterpreted the profile frequency as the probability that the DNA evidence came from anyone other than the defendant.

---

14. Assuming there is at least one ticket with a winning combination in the lottery, or at least one DNA sample with a matching profile in the databank.

15. The situation is slightly more complicated depending on the specifics of the case. For a detailed analysis, see Donnelly & Friedman, *supra* note 4.

16. Nor does it diminish the diagnostic value of the DNA evidence. Contrary to the conclusions of the National Research Council, testing and eliminating others as potential sources of the evidence slightly increases its diagnostic value. *Id.* The case against a suspect based solely on a DNA-databank match arguably may be weaker than a case based on other evidence in addition to DNA evidence, but not because the databank match itself is less probative.

17. Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS J. 21 (1993) (quoting from the transcript of Bethune v. State, 821 S.W.2d 222 (Tex. Crim. App. 1991)).

18. *Eine Hundertprozentige Sicherheit, Ihn Zu Identifizieren* (*With 100% Certainty, It Identifies Him*), FRANKFURTER ALLGEMEINE ZEITUNG, Apr. 14, 1998, at 13.

19. *DNA Fingerprinting Comes of Age*, 278 SCIENCE 1407 (Constance Holden ed.,1998).

This may lead judges to misunderstand it similarly in their opinions as, for example, "the probability of someone else leaving" the genetic trace.[20]

II. STATISTICAL THINKING

All this confusion is consistent with three decades of research finding that people have difficulty understanding probabilities, and not just because expressing probabilities in natural language can be confusing.[21] Countless studies show people struggling in the face of probabilities and falling short of normative standards,[22] routinely confusing, for example, the probability that a hypothesis H is true, given a certain piece of evidence E, or $p(H|E)$, with the probability that a certain piece of evidence occurs given the truth of a particular hypothesis, or $p(E|H)$, and conflating conditional probabilities with conjunctive ones.[23]

These findings are particularly germane to DNA evidence because the probative value of this evidence hinges on conditional probabilities. A judge or juror faced with a DNA match must determine how likely it is that the person whose DNA profile matches the incriminating profile is actually the source of the incriminating evidence. But how likely that is depends, among other things, on the frequency of the DNA profile in the population. For this reason, likelihood ratios incorporating that frequency have been recommended to indicate the significance of DNA evidence.[24]

A likelihood ratio compares the probability of finding the evidence under one hypothesis with the probability of finding that same evidence under its opposite.[25] In formal terms, the ratio is

$$\frac{p(E \mid H)}{p(E \mid \overline{H})},$$

the ratio of the probabilities of the evidence E under the alternative hypotheses of H and $\overline{H}$. But with DNA evidence, the ultimate concern is the ratio of the transposed conditional probabilities, given the DNA match, what is the relative likelihood that the defendant is the source:

$$\frac{p(H \mid E)}{p(\overline{H} \mid E)}.$$

---

20. Wilson v. State, 962 S.W.2d 905, 807 (Ark. 1998).

21. *See, e.g.*, Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859 (1996); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167 (1987).

22. Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman et al. eds., 1985).

23. Hoffrage et al., *supra* note 5; Kaye & Koehler, *supra* note 6.

24. NRC II, *supra* note 4, at 127–29.

25. *See* Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977) (suggesting that under Federal Rule of Evidence 401, evidence would be relevant as long as the likelihood ratio $\neq 1.0$).

Bayes' Theorem bridges the inference from the one ratio to the other. The relative likelihood of the hypotheses, given the evidence, is the product of the first likelihood ratio and some prior estimate of the relative probability of the two hypotheses:

$$\frac{p(H \mid E)}{p(\overline{H} \mid E)} = \frac{p(H)}{p(\overline{H})} \times \frac{p(E \mid H)}{p(E \mid \overline{H})}.$$

Scientific disagreement over calculating these probability estimates and likelihood ratios has largely abated,[26] but some statistical issues remain. The National Academy of Sciences has twice evaluated the use of forensic DNA analysis,[27] but the two evaluations disagree on various points and leave some questions unresolved. Two areas of particular importance are the significance of DNA matches identified through DNA databanks[28] and the ability of judges and jurors to understand the significance of DNA matches when different methods of presentation are used.[29] The following sections explore and present findings from psychological research that address these issues.

III. PRESENTATION

A. Focus and Frame

One of the earliest studies of the effects of presenting statistical evidence in different ways documented a common confusion mentioned earlier[30] by giving mock jurors expert testimony in one of two written forms. The first form stated that there is only a two percent chance the defendant's hair would be indistinguishable from that of the perpetrator if he were innocent.[31] The other form stated that only 2% of the people have hair that would be indistinguishable from that of the defendant and in a city of 1,000,000 people there would be 20,000 such individuals. Jurors were much less likely to convict when given the latter version.

These forms of presenting a statistic exhibit two characteristics that have a systematic effect on how people understand and use statistical evidence.[32] The

---

26. *See* Eric S. Lander & Bruce Budowle, *DNA Fingerprinting Dispute Laid to Rest*, 371 NATURE 735 (1994).

27. NRC II, *supra* note 4, at 1; Richard Lempert, *After the DNA Wars: Skirmishing with NRC II*, 37 JURIMETRICS J. 439 (1997).

28. NRC II, *supra* note 4, at 39–40; Donnelly & Friedman, *supra* note 4.

29. NRC II, *supra* note 4, at 174, 192–99, 203–04.

30. *See* Thompson & Schumann, *supra* note 21, at 177 (stating that a deputy district attorney believed that, if a defendant has a trait found in only a small sample of a population, e.g., 1%, then the probability that the defendant is guilty is 99%).

31. This form states a conditional probability of a match assuming the defendant is innocent. As discussed earlier, this sort of statement may be misleading because, even if the defendant is indeed the source of the indistinguishable hair, the question of guilt or innocence requires more than physical evidence.

32. There are, of course, many ways to present a particular piece of evidence, and DNA statistics will be interpreted in light of the circumstances surrounding their presentation. Choices of

first form focuses on the individual defendant and states the probability of a single event, whereas the second formulation focuses on other people and states how many of them have a particular trait.[33] In a series of studies,[34] Koehler found that when the DNA match statistic is "framed in the language of probability (e.g., 0.1%) in a way that highlights a particular suspect's chance of matching by coincidence, it tends to be persuasive" evidence that the suspect is the source; "[b]ut when the statistic is framed in the language of frequencies (e.g., one in one thousand) in a way that highlights the chance that others will match by coincidence, it is much less persuasive."[35] Similarly, match statistics that target an individual suspect are more persuasive than the equivalent statistic that targets a broader population.[36]

To cite another example, compare the two different but mathematically comparable expressions of a 1 in 1,000 DNA-match statistic: "The probability that the suspect would match the blood specimen if he were not the source is 0.1%" versus "One in one thousand people in Houston who are not the source would also match the blood specimen."[37] Once again, the former focuses on an individual suspect and states the probability of a match, while the latter focuses on other people and states the frequency of the matching trait. As before, mock jurors were less likely to convict based on the latter wording than they were based on the former.[38]

The latter statement directly cues people to think about the other people besides the individual suspect who might also match, and the findings suggest that people judge the probative value of a DNA match based on, among other things,

---

presentation organize the facts, focus attention, force associations and discriminations, and bring out relationships that may be illuminating, intriguing, subtle, or banal. Expert testimony is one example, and much could be made of the incidental influences lurking in just these two sentences. For example, the first statement invokes the subjunctive mood with an implication about the guilt or innocence of the defendant. *See generally* HOWARD SCHUMAN & STANLEY PRESSER, QUESTIONS AND ANSWERS IN ATTITUDE SURVEYS: EXPERIMENTS ON QUESTION FORM, WORDING AND CONTEXT. (1981); Denis. J. Hilton, *The Social Context of Reasoning: Conversational Inference and Rational Judgment*, 118 PSYCHOL. BULL. 248 (1995); Norbert Schwarz et al., *Base Rates, Representativeness, and the Logic of Conversation: The Contextual Relevance of "Irrelevant" Information*, 9 SOC. COGNITION 67 (1991).

33. *Cf.* LEONARD J. SAVAGE, THE FOUNDATIONS OF STATISTICS (2d ed. 1972); RICHARD VON MISES, PROBABILITY, STATISTICS AND TRUTH (2d ed. 1957); Hoffrage et al., *supra* note 5; Daniel Kahneman & Dan Lovallo, *Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking*, 39 MGMT. SCI. 17 (1993).

34. Jonathan J. Koehler, *The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient*, 74 S. CAL. L. REV. 1275, 1278 (2001). Others have contributed as well. *See, e.g.*, David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process*, 12 LAW & HUM. BEHAV. 1 (1988); Jane Goodman, *Jurors' Comprehension and Assessment of Probabilistic Evidence*, 16 AM. J. TRIAL ADVOCACY 361 (1992); Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 LAW & HUM. BEHAV. 159 (1999).

35. Koehler, *supra* note 34, at 1278.

36. *Id.*

37. *Id.*

38. Koehler, *supra* note 34.

how readily others who might also match come to mind. The more people are aware of others who might also match, the less compelling the evidence will seem, and vice versa.[39] Strikingly, the majority of sample jurors provided with the first statement above were "99% certain" that the suspect was the source of the evidentiary sample, while more than a third who were provided with the second statement were equally convinced that the suspect was not the source.

Unless the subjects who were "99% certain" that the defendant was the source thought that the prior probability was very low (e.g., less than 10%), its combination with the DNA-match statistic generally should have produced higher posterior probability estimates than those reported.[40] This is in line with previous studies showing that people do not usually conform to Bayesian rules when reasoning with probabilities.[41] Research over the past decade, however, has identified a way to improve statistical reasoning.[42]

B. Natural Frequencies

Expressions of probabilities that are mathematically equivalent are not necessarily psychologically equivalent. Probabilities expressed as non-normalized frequencies of conjunctive events sampled from a single population, which are more like simple counts, are more easily understood than are fractions bounded by zero and one. We may call such frequencies with their original numerators and denominators "natural frequencies."[43]

In the case of DNA evidence, the probability that someone randomly selected would have a particular DNA profile must be combined with the probabilities of finding a match for both those who share that profile and those who do not. For individuals who are not skilled in or aware of the use of Bayes' Theorem, combining these probabilities is simpler and easier if they are expressed as natural frequencies rather than as standardized or relative frequencies. The reason why is that converting natural frequencies into probabilities eliminates the base rate of the DNA profile in the relevant population.[44]

---

39. To disentangle the effect of focusing on the individual versus some group of others from the effect of stating a probability versus a frequency, Koehler, *id.*, employed two other expressions yielding less extreme effects: "The frequency with which the suspect would match the blood specimen if he were not the source is one in one thousand." and "One tenth of one percent of the people in Houston who are not the source would also match the blood specimen." *Id.* at 1278. The results indicate that targeting an individual and stating a probability is the most compelling form of the evidence and that targeting others and stating a frequency is the least compelling. Conviction rates for those who were given both expressions of the evidence were in between these extremes. These effects, however, appear to vary with the value of the DNA match statistic, with lower frequencies leading to higher conviction rates.

40. Jonathan J. Koehler, *When Are People Persuaded by DNA Match Statistics?*, 25 LAW & HUM. BEHAV. 493, 503 (2001).

41. Kaye & Koehler, *supra* note 6; Tversky & Kahneman, *supra* note 22.

42. Hoffrage et al., *supra* note 5.

43. *See* Gerd Gigerenzer & Ulrich Hoffrage, *Overcoming Difficulties in Bayesian Reasoning: A Reply to Lewis & Keren and Mellers & McGraw*, 106 PSYCHOL. REV. 425 (1999).

44. The conversion neutralizes the base rates by setting them to the same value (e.g., 100).

Consider the probability that a patient who tests positive on a diagnostic test for a disease actually has the disease. Ascertaining $p(\text{disease}|\text{positive})$ requires determining how many people have the disease out of all those who test positive:

$$p(\text{disease}|\text{positive})=\frac{N(\text{positive}\cap\text{disease})}{N(\text{positive})}.^{45}$$

That proportion hinges on how common the disease is, as well as the chance of a positive result for someone who has the disease and the chance of a positive result for someone who does not have the disease.

With DNA evidence, the comparable inference involves the proportion of people who have a certain DNA profile out of all those who appear to match in a DNA analysis:

$$p(\text{profile}|\text{match})=\frac{N(\text{profile}\cap\text{match})}{N(\text{match})}.^{46}$$

If a particular DNA profile occurs in a population of 10 million with a frequency of one in a million, for example, one might expect approximately 10 people to share that DNA profile. But if the chances of false-positive laboratory errors are as high as sometimes estimated,[47] then there could be some 30,000 people in the population without that DNA profile who would nonetheless match in DNA analyses. Thus, without any other incriminating evidence, if someone is said to share a DNA profile with a forensic sample, the conditional probability of that suspect sharing the profile given the match, that is, $p(\text{profile}|\text{match})$ is equal to 10 out of 30,010.

Probabilities and frequencies name the same rational numbers, but natural frequencies lay bare the numerical dependencies required for Bayesian inference. Because these frequencies are mathematically equivalent to probabilities, the Bayesian bridge of statistical inference spans both. But the equivalent computation with probabilities requires that the base rates[48] be computed and multiplied together with conditional probabilities to get a correct answer:

$$p(\text{profile}|\text{match})=\frac{p(\text{profile})\,p(\text{match}|\text{profile})}{p(\text{profile})\,p(\text{match}|\text{profile})+p(\overline{\text{profile}})\,p(\text{match}|\overline{\text{profile}})}.$$

However, frequencies require only a count of the people who would actually have the profile out of all the people who match:

---

45. In this expression, N (positive ∩ disease) indicates the number of individuals with a positive test result *and* the disease. N(positive) is the number with a positive test result.

46. As in the previous equation, N refers to the number of individuals with the characteristics noted in parentheses.

47. *See supra* notes 10 & 12.

48. The base rates are the frequency of the profile and its complement, i.e., the number of those who do and those who do not share the matching profile.

$$p(\text{profile}|\text{match}) = \frac{N(\text{match} \cap \text{profile})}{N(\text{match})}.$$

In other words, natural frequencies facilitate Bayesian reasoning because part of the calculation is already "done" by the form of presentation itself.

As noted earlier,[49] however, not just any variety of frequency expression will facilitate Bayesian inference. Normalized frequencies do not facilitate Bayesian inference.[50] Nevertheless, some authors have not been sensitive to the distinction,[51] and others have made new distinctions, such as "information type" versus "information structure"[52] or "partitive" versus "nonpartitive" formats,[53] which redescribe the distinctive properties of natural frequencies. Natural frequencies should not be confused with other kinds of frequencies. Natural frequencies are non-normalized counts of conjunctive events sampled from a single population, and they therefore retain the leverage of the base-rates and simplify Bayesian calculations.

Because Bayesian calculations are so much simpler with natural frequencies, using these expressions could offset confusion surrounding statistical analyses and help judges and jurors understand the uncertainties associated with DNA evidence. Perhaps a more transparent way of presenting the statistics would afford a better understanding of the evidence itself.

---

49. *See* Gerd Gigerenzer & Ulrich Hoffrage, *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, 102 PSYCHOL. REV. 684 (1995).

50. *See id.*

51. *E.g.*, Jonathan St. B.T. Evans et al., *Frequency Versus Probability Formats in Statistical Word Problems*, 77 COGNITION 197 (2000). Here, normalized frequencies (called "frequency formats—hard") do not improve Bayesian reasoning.

52. Vittorio Girotto & Michel Gonzalez, *Solving Probabilistic and Statistical Problems: A Matter of Information Structure and Question Form*, 78 COGNITION 247 (2001). The information type is either frequencies or probabilities, with no distinction between natural frequencies and other types. Information structure depends on whether relevant conjunctive events are given. The authors suggest that information structure, rather than type, facilitates Bayesian inference. Yet this information structure is exactly what natural frequencies provide.

53. Laura Macchi, *Partitive Formulation of Information in Probabilistic Problems: Beyond Heuristics and Frequency Format Explanations*, 82 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 217 (2000). Here, problems with a frequentist formulation in partitive format produced a high percentage of Bayesian responses, whereas problems in nonpartitive format produced a high percentage of non-Bayesian responses, implying that "a frequentist formulation is not the crucial element for eliciting correct reasoning." *Id.* at 225. True enough. Any kind of frequency formulation will not work, and the result is to be expected since nonpartitive frequencies are not sampled from a single population, whereas partitive frequencies are.

IV. EXPERIMENT: SAME EVIDENCE— DIFFERENT EXPRESSIONS

A. Design and Method

To investigate this possibility, we asked a sample of 127 advanced law students and 27 professional jurists in Germany to evaluate two realistic criminal court rape case files.[54] In one case the victim had first identified the defendant, whose DNA profile then matched that of samples recovered from the crime scene. In the other case, however, a DNA match was found through a DNA databank that contained a sample from the defendant. Aside from the DNA match, there was little reason to suspect the defendant was the perpetrator, and in neither case was the victim absolutely sure that the defendant was the rapist.

In each case, an expert who performed a DNA analysis testified that the particular DNA profile recovered from the crime scene occurred in one in a million (0.0001%) Caucasians. The expert also testified that it was practically certain that the DNA analysis would find a match if a person actually has a matching DNA profile (i.e., true-positive result). Prompted by the defense, he also reported the rates of technical and human mishaps leading to false-positive results in laboratory proficiency tests.[55]

This numerical information, however, was presented as expert testimony in two different formats. One format stated all the information as probabilities (e.g., 0.0001%), and the other format stated it as natural frequencies (e.g., 1 out of 1,000,000). That was the only difference.[56]

After reading the expert testimony in one or the other format, each participant was asked to provide two probabilities. The first question asked for the probability that someone said to match in the DNA analysis would actually have the DNA profile in question: p(profile|match). The second question asked for the probability that someone said to match in the DNA analysis is the source of the trace recovered from the crime scene: p(source|match). Immediately after answering these two questions, the participants were asked to render a verdict of guilty or not guilty.

After examining one case file in either a frequency or probability format, each participant was given a second case file in which the expert testimony was in the other format. They then answered the same questions as before. All participants were randomly assigned to conditions where the different formats of expert testimony (frequency or probability), their order (first or second), and the type of case (databank hit or witness identification) were all varied systematically

---

54. In Germany, a sealed criminal court file (*Strafakten*) documents the course of the legal proceeding from the initiation of charges to the final disposition of the court. Our case files matched actual case files in all respects except those requiring experimental manipulation and anonymity of the parties.

55. The rates were an average of error rates once found in actual proficiency tests of DNA laboratories. *See* Koehler et al., *supra* note 10, at 205–11.

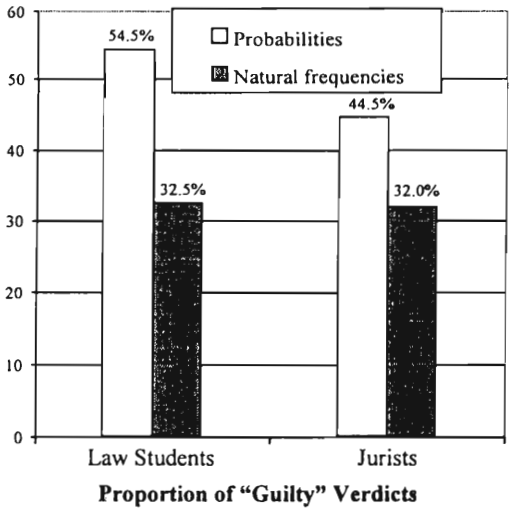
56. An example of the expert testimony in each format is in the Appendix, as the rest of the material in the case files was extensive and detailed but otherwise identical in every respect.

and counterbalanced. To boost motivation, we offered a monetary incentive for better performance. Two-thirds of the student sample received a flat fee, and one third received a slightly smaller flat fee but with bonuses for each correct numerical response.

B. Results and Discussion

We found that the same DNA evidence had dramatically different effects in the different formats on both statistical reasoning and decision-making. Consistent with previous findings,[57] the same evidence always led to higher conviction rates in the probability format than in the frequency format. The proportion of *guilty* verdicts in the probability format was 50 to 100 percent greater than the proportion of *guilty* verdicts in the frequency format in every condition. Figure 1 shows the overall pattern of results combined across all conditions. Although the case involving eyewitness identification always garnered more *guilty* verdicts than the case based solely on a databank match, these differences were not statistically significant.[58] It appears, then, that the differing expressions of the statistical evidence had a stronger impact than the kind of evidence.

Figure 1



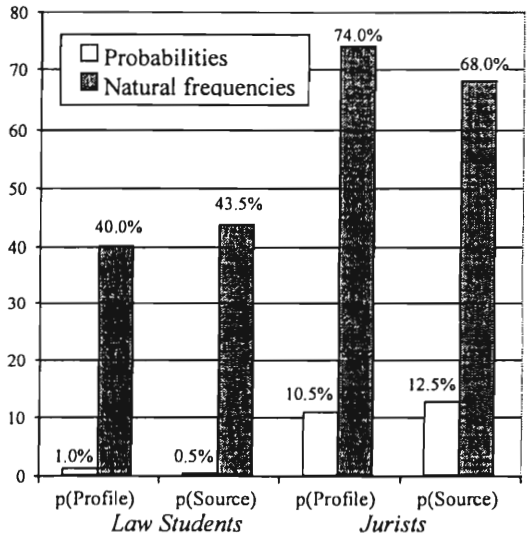
The disparate outcomes in the different numerical formats were likely due to a disparate understanding of the statistical evidence. All participants had to infer

57. Koehler, *supra* notes 21, 34 & 40.

58. All of the findings presented, except this one, are statistically significant using a χ^2, $p < 05$ standard. Overall, 53% returned *guilty* verdicts in the witness-identification case whereas 48% did so in the databank-hit case.

the conditional probabilities of (1) having a DNA profile given a DNA match and (2) being the source of the evidence given the match. In general, across all conditions there were far more correct answers in the frequency format than in the probability format, irrespective of the type of case or monetary incentive.[59] Only one of the students and a few of the professionals could derive the correct probabilities in the probability format, whereas, by contrast, about 40 to 50 percent of the law student sample and 70 to 75 percent of the jurist sample spontaneously derived the correct answers using natural frequencies, as shown in Figure 2.

Figure 2



These response patterns were no different for those who were paid more for deriving the correct answers. Indeed, most participants, regardless of payment condition, spent about two hours evaluating the cases, so the incorrect answers in the probability format were not for lack of trying.[60]

---

59. The offer of additional payment did make the notes written on scratch paper more numerically oriented, typically with those paid for performance attempting various misguided calculations in the probability format and then giving up.

60. The time spent ranged from 0:55 to 2:45, with an average of 1 hour and 35 minutes. Most of this time was used on the materials in the probability format, although the time needed to complete the tasks in the separate formats was not measured.

In contrast, the correct Bayesian answers in the frequency format often came easily, without need for pencil and paper calculations. When the evidence was expressed as frequencies, those who answered correctly seemed to grasp the structure of the Bayesian problem, albeit without fully understanding the mathematics, since these same participants could not derive the correct answers in a probability format that was mathematically equivalent.

While the correct answers in the frequency format required no written calculations, the calculations attempted in the probability format were uniformly misguided. The most common numerical operation in the probability format was subtracting the base rate of the DNA profile from 100% or subtracting the false positive error rate from 100%. The most common alternative was simply to restate one of the statistics mentioned in the expert testimony. These two types of responses account for nearly all the numerical responses in the probability format. Subtracting the base rate of the DNA profile from 100% accounts for 50 to 60 percent of the probability format responses (e.g., 99.999%) and restating a statistic mentioned in the testimony accounts for another 35 to 40 percent (e.g., .0001%).[61]

These results are all the more surprising because half of the participants saw the statistical evidence expressed as natural frequencies almost immediately before having to evaluate that same evidence expressed as probabilities—and, again, the statistics in the two formats were mathematically identical. They were not psychologically identical, however, and not without potential legal consequences.

V. IMPLICATIONS

If judges and jurors are to evaluate DNA evidence, they should understand that a forensic DNA analysis can identify a potential source of incriminating evidence, but not with certainty. The inherent uncertainty depends on the different population frequencies of different DNA profiles and the possibility of laboratory error, although anomalies such as police misconduct or planted evidence are also possible. Judges and jurors should understand the statistics surrounding both of these inherent uncertainties and their significance to a DNA match. Their understanding may be helped or hindered by different ways of expressing those statistics.

Experts and nonexperts alike are easily confused by statistics expressed as probabilities. But this confusion diminishes when the same statistics are expressed as natural frequencies. We found that presenting the same statistical evidence as natural frequencies rather than conditional probabilities dramatically increased the proportion of correct statistical inferences by trained legal decisionmakers and influenced the verdicts in each case.

---

61. These kinds of erroneous responses are comparable to those found in other studies concerning probabilistic reasoning. *Cf.* Gigerenzer & Hoffrage, *supra* note 49; Kaye & Koehler, *supra* note 6.

Different expressions of the statistical evidence, which was the only difference in the evidence in each case, significantly changed the likelihood of conviction.[62] That is most likely because confusion generated by the probability expressions leads people to infer that the posterior probability that the defendant is the source is much higher than it actually is, while the clarity maintained in the frequency expressions leads them to infer that the posterior probability is lower, and therefore less incriminating, since it is appropriately weighted with the population frequency of the DNA profile. Any inference that neglects the profile frequency will produce erroneously high probability estimates.[63]

When statistics were expressed as natural frequencies rather than probabilities, more people could, without instruction, successfully combine the population frequency of the DNA profile with the rates of laboratory error to derive the correct conditional probability of someone being the source of the incriminating DNA evidence—the ultimate meaning of a DNA match. Understanding that meaning and the statistical uncertainties will not guarantee that the evidence is evaluated appropriately, but it would seem to be a crucial first step. Our experiments suggest that the natural frequency expressions reveal the Bayesian structure that links these uncertainties. Clarifying this structure for legal decisionmakers might alleviate considerable confusion.

Beyond the statistics, though, other evidence besides a DNA match is also important in deciding a case. Deciding whether a defendant is the true source of some matching DNA sample, not to mention rendering a verdict, depends on other evidence, of which there may be plenty or precious little. The latter case is especially pertinent to matches found solely through searching DNA databanks when there is little other evidence. Our results suggest that people may be as likely to convict based solely on a DNA match taken from a databank as they are based on a confirmatory match coupled with eyewitness identification.[64] Indeed, the different ways of presenting the same statistical evidence had a much greater impact on the tendency to convict than even eyewitness identification did. Probability expressions sharply increased conviction rates in comparison with

---

62. Both the probability format and frequency format used here cued people directly to think of others who could also match in the analysis. *Cf.* Koehler, *supra* notes 34& 40.

63. Koehler, *supra* note 40, assessed sample jurors' understanding of a 1 in 1000 (0.1%) match statistic by asking them how many others would also match in a city of 500,000 and found that 60.7% versus 42.1 % of subjects gave the correct answer in the frequency versus the probability formats.

64. In *Regina v. Adams*, [1998] 1 Cr. App. R. 377 (Eng. C.A. 1997), the only evidence linking the defendant to a rape was a DNA match (and his residing in the general area where the rape occurred). Even though he did not, according to the victim, look very similar to the actual rapist, and, even though he presented an alibi that could be corroborated, he was convicted on the basis of the DNA evidence. Thus, if someone who does not resemble the assailant can be convicted on the basis of a DNA match alone, then it would seem even more likely that someone who does resemble the assailant could be wrongfully convicted on that basis together with eyewitness testimony, especially considering the notorious prevalence of eyewitness identification error. *See* Gary L. Wells & Amy L. Bradford, *Distortions in Eyewitnesses' Recollections: Can the Postidentification-Feedback Effect Be Moderated*, 10 PSYCHOL. SCI. 138 (1999); Gary L. Wells & Elizabeth A. Olson, *The Other-Race Effect in Eyewitness Identification: What Do We Do About It?*, 7 PSYCHOL. PUB. POL'Y & LAW 230 (2001).

frequency expressions regardless of whether the DNA match was found through a databank or was coupled with eyewitness testimony. Beginning with a DNA match may be unfairly prejudicial.[65] There is also the danger of constructing a case to augment the match with the bias of hindsight[66] and a shifting focus toward confirmatory evidence.[67]

Since the chance of a coincidental DNA match increases with the number of samples analyzed, the population frequency of the DNA profile should be multiplied by the size of the databank to reflect the new chance of a random match. But that does not affect the chance of laboratory error or its relation to the new chance of a coincidental match.[68] The chance of a laboratory error is usually greater than the chance of a coincidental match, but it could be the other way around given a databank large enough.

Since the U.S. Supreme Court has said courts should consider the "known or potential rate of error"[69] before admitting scientific evidence, arguments involving these statistics may become more common.[70] Representing those statistics one way or another does not matter from a mathematical point of view, but it should matter from psychological and judicial point of view. When verdicts hinge on DNA evidence, understanding that evidence is crucial, and how it is presented may make an important difference.

---

65. *Cf.* FED. R. EVID. 403.

66. Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgement Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL.: HUMAN PERCEPTION & PERFORMANCE 288 (1975).

67. RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT 167–92 (1980).

68. In a between-subjects sample of 20 university students given only the case materials involving a databank hit from the experiments reported earlier, 19 successfully multiplied a profile frequency of one in a million by a databank size of 500,000, with 60% then going on in the frequency format to infer the correct conditional probability of the defendant's being the source of the evidentiary sample. As always, the difficulty of any calculation will vary with the specific numbers in question—which is a moot point if a calculator is allowed.

69. Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 594 (1993).

70. In *Williams v. State*, 679 A.2d. 1106 (Md. 1996), the court of appeals reversed the trial court for not allowing more extensive questioning about laboratory proficiency tests and the possibility of contamination in a laboratory. *Id.* at 1120.

APPENDIX

A. Probability Format

In a country the size of Germany there are as many as 10 million men who fit the description of the perpetrator. The probability of a randomly selected person having a DNA profile that matches the trace recovered from the crime scene is .0001%. If someone has this DNA profile it is practically certain that this kind of DNA analysis would show a match. The probability that someone who does not have this DNA profile would match in this type of DNA analysis is .001%. In this case, the DNA profile of the sample from the defendant matches the DNA profile of the trace recovered from the crime scene.

B. Frequency Format

In a country the size of Germany there are as many as 10 million men who fit the description of the perpetrator. Approximately 10 of these men would have a DNA profile that matches the trace recovered from the crime scene. If someone has this DNA profile it is practically certain that this kind of DNA analysis would show a match. Of the some 9,999,990 people who do not have this DNA profile, approximately 100 would be shown to match in this type of DNA analysis. In this case, the DNA profile of the sample from the defendant matches the DNA profile of the trace recovered from the crime scene.

EDITOR'S PREFACE

This issue of *Jurimetrics, The Journal of Law, Science, and Technology* includes articles on the psychology of evidence,[1] medical records and privacy,[2] genetic research and bioethics,[3] and the computation of lost profits in patent infringement cases.[4] Inasmuch as these papers confront difficult issues and stake out new positions, we expect them to prompt further debate and a better understanding of the questions they seek to answer.

The many legally trained readers of *Jurimetrics* should appreciate the merits of the legal and ethical arguments in several of these articles with no further editorial gloss. However, the psychological experiment, previously reported in substantially less detail in *Science*,[5] may be less accessible to the modal reader. For this reason, a few comments on the analysis seem appropriate. In *Communicating Statistical DNA Evidence*, Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer discuss the presentation of probabilities associated with DNA and other trace evidence. They propose using "natural frequencies" rather than decimals or percentages,[6] especially when a defendant has been identified by searching through a database of DNA profiles from convicted offenders. The research adds to our understanding of how the manner in which trace evidence is presented affects assessments of that evidence.[7] The authors report that students and judges given a written description of certain evidence return fewer guilty

---

1. Samuel Lindsey et al., *Communicating Statistical DNA Evidence*, 43 JURIMETRICS J. 147 (2003).

2. Wendy J. Netter, *Curing the Unique Health Identifier: A Reconciliation of New Technology and Privacy Rights*, 43 JURIMETRICS J. 165 (2003).

3. David E. Winickoff, *Governing Population Genomics: Law, Bioethics, and Biopolitics in Three Case Studies*, 43 JURIMETRICS J. 187 (2003).

4. Clement G. Krouse, *But-For Markets and Reasonable Royalties: The* Rite-Hite v. Kelley *Misdirection*, 43 JURIMETRICS J. 229 (2003).

5. Ulrich Hoffrage et al., *Communicating Statistical Information*, 290 SCIENCE 2261 (2000).

6. The authors define "natural frequencies" as "[p]robabilities expressed as non-normalized frequencies of conjunctive events sampled from a single population . . . ." Lindsey et al., *supra* note 1, at 154. An example would be "ten out of the ten million people in the city" rather than "0.000001."

7. For another recent and more comprehensive study on this general issue, see Dale A. Nance & Scott B. Morris, *An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability*, 42 JURIMETRICS J. 403 (2002).

verdicts and arrive at the correct figure for particular probabilities much more often when relevant probabilities are presented as "natural frequencies."[8]

In interpreting this finding, the precise meaning of "natural frequencies" must be appreciated. The experiment involves more than the translation of decimal numbers such as 0.0001% into fractions such as 1 in a million. The subjects in the "frequency condition" were told the number of men in the population who actually possess the incriminating DNA type—that is, the *number* of true positives. They were told how many men who do not possess the type would be erroneously reported to have it—the *number* of false positives. Because they were provided such numbers rather than mere frequencies, they did not have to understand the tricky concept of conditional probability to find the posterior probability.[9] All that was necessary was to divide the specified number of false positives by the total number of positives. In contrast, the subjects in the "probability condition" were given no clues as to how to transpose a conditional probability.[10] Whether merely presenting the statistical information as frequencies expressed as natural numbers would have the same effect as giving the numbers of true and false positives in a particular population therefore remains unknown.

A second facet of the paper also warrants a brief comment. The authors suggest that the probative value of a DNA match depends greatly on (1) whether the defendant is identified by searching a database of DNA types of convicted offenders, as opposed to (2) whether the defendant is identified on the basis of non-DNA evidence, and then the DNA match is made. They write that in the initial database-search case, "the population frequency of the DNA profile should be multiplied by the size of the databank to reflect the new chance of a . . . coincidental match."[11] An accompanying footnote states that "the correct conditional probability of the defendant's being the source of the evidentiary sample" requires multiplication by the "databank size."[12] As discussed in a symposium in this journal, however, this supposed correction is controversial at best.[13] Thoughtful statisticians have argued strongly that when the hypothesis in

---

8. *See* Lindsey et al., *supra* note 1, at 158–59. This result is consistent with other findings reported by Gigerenzer and his colleagues at the Max Planck Institute. *See, e.g.*, Gerd Gigerenzer & Ulrich Hoffrage, *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, 102 PSYCHOL. REV. 684 (1995); *cf.* Craig R. Callen, *Informational Indigestion and Irrationality*, 41 JURIMETRICS J. 513 (2001) (reviewing SIMPLE HEURISTICS THAT MAKE US SMART (Gerd Gigerenzer et al. eds., 1999)).

9. *Cf.* Lindsey et al., *supra* note 1, at 156 ("natural frequencies facilitate Bayesian reasoning because part of the calculation is already 'done' . . . .").

10. The procedure required to arrive at the answers counted as correct apparently involves Bayes' theorem, an initial assumption that all men are equally likely to be the source of the DNA found at the crime scene, and a further assumption involving the independence of having a particular DNA type and the commission of the crime.

11. Lindsey et al., *supra* note 1, at 162.

12. *Id.* at 162 n.68.

13. Symposium, *The Evaluation of Forensic DNA Evidence*, 37 JURIMETRICS J. 395 (1997).

question is whether the defendant is the source, such multiplication is grossly incorrect.[14]

These statistical and psychological issues, like the bioethical and economic questions in the other articles in this issue, are far from trivial. They require careful analysis and explication for the legal system intelligently to regulate science and technology and to apply scientific methods to its own needs. As such, we look forward to publishing further dialogue on these topics.

March 2003

—D.H. Kaye
Editor

14. *See, e.g.*, David H. Kaye & George F. Sensabaugh, Jr., *DNA Typing*, *in* 3 MODERN SCIENTIFIC EVIDENCE § 25–2.6.2[3][d], at 277–78 (David L. Faigman et al. eds., 2d ed. 2002) (describing and citing this literature).