DRAFT – NO QUOTES

Misconceptions About Statistics and Statistical Evidence

Jonathan J. Koehler[1]

The University of Texas at Austin

January 25, 2007

[1] McCombs School of Business, The University of Texas at Austin, 1 University Station B6500, Austin, TX 78712-0212. 512-471-7856 (w), 512-471-0587 (fax), koehler@mail.utexas.edu.

Thanks in large part to advances in computing and information technology, statistics are everywhere. Whether the information concerns business, health, politics, sports, or nearly anything else, it is likely to appear in statistical form. The front page of the country's highest circulation daily newspaper (*USA Today*) is littered with descriptive statistics and graphical depictions of those statistics. The star of a popular prime time television show called *Numbers* solves fictional legal cases each week through the innovative use of statistics and statistical reasoning. And so it is no surprise that statistics and statistical arguments find their way into the American courtroom at an unprecedented rate. Feinberg (1989) reported "dramatic growth" in the use of statistical evidence from the 1960s through the 1980s. He noted that the terms "statistic" or "statistical" appeared in thousands of reported district court opinions (p. 7). I performed a Westlaw search on these terms and found a 56% increase in the use of these terms in the Federal Cases database from 1990 to 2004. I also found that the phrase "statistical analysis" appeared 94% more often in 2004 than in 1990; "regression analysis" appeared 95% more often.

In most cases, expert witnesses are responsible for introducing statistical evidence in the courtroom. Many of these experts have graduate level training in quantitative fields, and speak easily about correlation coefficients, p values, regression coefficients, and random match probabilities. However, there is scant reason to believe that factfinders can interpret this information properly. Jurors are not screened for their quantitative reasoning skills prior to appearing in cases in which key evidence is statistical or probabilistic. Nor are they taught how to think about beta coefficients and likelihood ratios once this evidence is presented. When John Allen Paulos (1988) speaks of the crippling "innumeracy" (or mathematical illiteracy) that permeates American culture, the legal profession should take special notice. Paulos argues persuasively that otherwise intelligent laymen are confused by numbers, probabilities, and elementary statistical principles. If true – and the evidence suggests that it is – then jurors may misweight the statistical evidence they hear and render unjust verdicts.

Trial consultants need to be aware of the documented problems that jurors and others have working with statistical evidence and arguments (for reviews, see Arkes, Connolly, & Hammond, 2000; Kahneman, Slovic and Tversky, 1982; Saks & Kidd, 1980-81; Thompson & Schumann, 1987?). They also should be prepared to provide the conceptual background information to attorneys, judges, and jurors that will help them understand and reason with case-specific quantitative evidence.

This chapter identifies nine misconceptions surrounding statistical evidence and arguments at trial. These are not the only misconceptions, but they are important for two reasons. First, each misconception arises with some regularity in trials, particularly criminal trials. Second, many talented judges, attorneys and experts will insist that some of these misconceptions are not misconceptions at all. Such insistence should not be regarded as mere difference of opinion with the perspective offered here. Instead, it is a testament to the difficulty of reasoning with and about statistics.

Misconception #1: Jurors Overweight Statistics

A famous law review paper by Harvard Law Professor Lawrence Tribe argued that statistical evidence has no place in the courtroom because jurors will overweight this type of evidence. He wrote:

> "The problem of the overpowering number, that one hard piece of information, is that it may dwarf all efforts to put it into perspective with more impressionistic sorts of evidence. . . . The syndrome is a familiar one: If you can't count it, it doesn't exist. . . . [E]ven the most sophisticated user is subject to an overwhelming temptation to feed his pet the food it can most comfortably digest. Readily quantifiable facts are easier to process--and hence more likely to be recognized and then reflected in the outcome--than are factors that resist ready quantification" (Tribe, 1971, p. 1360, 1361-2).

At first blush, the argument seems persuasive. When confronted with, say, a 0.01% chance that semen found in a rape victim would match an innocent suspect by sheer chance, it would seem that jurors might not give much weight less quantifiable evidence that the defendant might offer in support of his innocence. However, a large body of literature suggests that the opposite is true (Kaye & Koehler, 1991; Schklar & Diamond, 1999; Thompson, 1989). Far from being blinded and hoodwinked by statistical evidence, studies indicate that jurors and other decision makers attach *too little* weight to this evidence. In my own mock jury studies, I found that about half of jurors are unwilling to vote for conviction in the face of a DNA match, even when they are told that the chances that the match is merely coincidental is less than one in a billion (see e.g., Koehler, Chia, & Lindsey, 1995). Explanations for the relative underweighting of statistics include their complexity, their abstractness (Nisbett & Ross, 1980), their causal insufficiency (Bar-Hillel, 1980), and their tendency to speak more to generalities than to specific instances. Regardless of the sufficiency of these explanations, there is no evidence that people overweight statistical evidence when confronted with evidence of a more qualitative sort such as eyewitness testimony or alibi evidence.

Misconception #2: Small Samples Are Not Informative

Suppose that the defendant in a medical malpractice case introduces a study in support of his decision to treat the plaintiff with drug B rather than drug A. In this study, drugs A and B were assigned at random to each of eight patients. The dependent measure of interest was the amount of time it takes for each patient to be discharged from the hospital. Hypothetical data in Table 1 show that the four patients who received drug A were discharged in an average of 3 days and the four patients who received drug B were discharged in an average of 6 ½ days. A statistical test, called an independent samples t-test, indicates that this difference is statistically significant ($t=4.58$, $df=6$, $p < .01$). On cross-examination, the plaintiff argues that the study is worthless because of the small sample size. Does the plaintiff have a point? Not really. If the study is well-designed (e.g., double-blind, random assignment to treatment conditions, etc.), the fact that the treatment groups differed significantly on the dependent variable in a small study speaks to the probable magnitude of the difference between the two treatments. On the other hand, if a small, well-designed study did *not* reveal significant differences between the two treatment groups, the study wouldn't deserve much attention. The reason is that a small study may not have sufficient power to detect differences between treatment groups that actually are different.

Insert Table 1 About Here

A related point of confusion is whether the results of a sample of, say, 1,000 respondents in a survey can adequately represent the views of a broader population of millions of people. Counterintuitive though it may be, one need not survey a large proportion of a population in order to represent it. The reliability of a survey (often described as a margin of error) has nothing to do with the size of the population that the survey seeks to describe. A more critical consideration is the extent to which members of the sample were selected at random from the population of interest. Trial consultants should be prepared to discuss this issue when presenting survey data to guard against the false intuition that a sample cannot predict the views of a much larger population.

Misconception #3: $p < .05$ = Important; $p > .05$ = Unimportant

Suppose that a defendant is indicted by a grand jury that is comprised of a list of registered voters in a community. However, the defendant argues that there was discrimination in the selection of the grand jurors. As proof, he offers uncontroverted evidence that the proportion of registered minority voters in the community is 10%, but that there were no minority grand jurors on the broader panel from which his grand jury was chosen. Holding aside issues of intent, is this compelling proof of discrimination? The answer may well turn on whether the difference between 10% and 0% is "statistically significant."

The phrase "statistically significant" is frequently uttered, but the word "significant" has a specialized meaning that does not correspond with what jurors and other laymen may associate with it. Many and perhaps most laymen would think that the exclusion of minority voters from a grand jury panel is significant indicator of discrimination – provided that the grand jury panel was reasonably large. But how large? If the panel included just 10 people, then it would only be mildly surprising if there were no minorities. If the panel included 1000 people, then the lack of minorities would be extremely surprising. This is where statistical analyses – and determinations of statistical significance – come in. The difference between the observed (0%) and expected (10%) proportion of minorities on the grand jury panel is statistically significant if the probability of observing so few minorities by chance alone – the so-called "p value" – is less than 5%. If the panel included just 10 people, the p value would be much greater than 5%, a result that falls short of statistical significance. This means that the lack of minorities on this panel is not strong evidence of discrimination.

Statistical novices often assume that a non-significant p value proves that the observed disparity arose by chance alone. Some will think that the p value identifies the probability that there was no discrimination, and that 1-p is the probability of discrimination. These interpretations are not correct. In our example, the p value identifies the probability that a disparity as large (or larger) than that observed would occur if, in fact, the selection of grand jurors was made at random (i.e., without regard for minority status). This probability, P(No Minorities | No Discrimination), is helpful for determining whether discrimination occurred, but it does not identify the probability of interest, which is P(No Discrimination | No Minorities). This point receives more consideration in Misconception #7.

Statistical novices also mistakenly assume that statistically significant differences are *practically* important and that statistically insignificant differences are practically unimportant. However, it is important to distinguish between these two types of significance. Sample size plays a central role in determining whether a difference is statistically significant. Large sample sizes are more likely to yield statistically significant differences. But sample size plays no role at all in determining whether a difference is practically significant. A salary disparity of $1,000 between two groups of employees may not be statistically significant, but this doesn't mean that the $1,000 difference has no practical significance. Similarly, a statistical analysis of the chip speed of two personal computers may show that chip A is significantly faster than chip B, but this difference may not have any noticeable or practical significance to the user.

In sum, trial consultants should have a thorough understanding of what p values are and what they are not. They should also be prepared to educate the courts on the role that sample size places in the production of p values. Finally, consultants should consider educating the courts about subtler measures of the importance of statistical results such as effect sizes.[2]

---

[2] An effect size is the standardized difference between two groups. It is computed as the difference between the mean values of two groups divided by the standard deviation. As a rough rule of thumb, effects sizes of .2 are small, .5 are medium and .8 are large (Cohen, 1988).

Misconception #4: Correlation Implies Causation

One of the take-away principles in Statistics 101 is "correlation does not imply causation." This means that if two variables are correlated or associated with one another, one may not necessarily conclude that one of the variables caused the other. A lurking third variable may well be the causal agent. There is a strong correlation between amount of ice cream consumed at the beach and the number of drownings. But this does not mean that the ice cream consumption *caused* the drawings. The weather is a more likely causal agent for changes in both variables.

Although the idea that correlation does not imply causation is well-known, people rely on correlation as a primary indicator of causality (Einhorn & Hogarth, 1986). When the causal connection is intuitively plausible, people may equate degree of correlation with degree of causal strength. Public health information is a prime example. Every week seems to bring the results of yet another correlational study that finds an association between a plausible-sounding cause (e.g., diet, stress level, temperament, etc.) and an important effect (e.g., illness, longevity, etc.). But because most of these studies are observational rather than experimental (in which treatment levels are assigned at random), the correlation provides scant grounds for the causal claim.

One way to make jurors think twice before inferring causality from causal claims is to give them a reason to believe that a 3rd (or 4th or 5th) variable may be the lurking causal force. This is particularly important in cases involving multiple regression analyses. Multiple regression analyses are useful for measuring the amount of variability in an outcome variable (e.g., cancer) that is explained or predicted by variability in two or more predictor variables (e.g., family history, exposure to toxins, etc.). Although such analyses are routinely offered in legal cases to establish causal relationships, regression analysis only identifies correlations among variables (controlling for the influence of other variables in the regression model) that may or may not have causal significance. Moreover, some correlations that are statistically significant in one regression model, may become insignificant as predictor variables are added to or subtracted from the model. For example, a strong association between the predictor variable "age" and the outcome variable "salary" might disappear once the predictor variable "years of job experience" is introduced into the model. Consequently, it would be hasty to conclude that significant associations identified by a regression model demonstrate causal relationships.

This is not to say that regression analyses cannot provide any relevant information about casual claims. To the extent that (a) the many assumptions that underlie multiple regression models are met, (b) the analyses control for all plausible lurking variables that could explain a significant association, and (c) the data are replicated by other high quality studies, a causal inference may be more reasonable. But such an inference is just that: an inference that reaches beyond the correlational data.

Misconception #5: Base Rates Don't Matter

One of the most important and pervasive misconceptions about probabilistic evidence is that background probabilities – or base rates – are irrelevant for case-specific judgments and decisions. Consider, for example, a person who wants to estimate the chance that a randomly selected couple will get divorced. On the one hand, the fact that 40-50% of U.S. marriages end in divorce indicates that there is a substantial chance that the selected couple will divorce. On the other hand, those who know the couple are likely to assign little, if any, weight to the base rate. Instead, they are likely to focus on such individuating features of the couple such as whether they share common interests, appear to be in love, are committed to religious values, etc. Although such individuating features might be informative, the base rate *also* provides relevant information. After all, if the divorce rate in some sub-population were 99%, then 99 out of 100 randomly selected couples will, on average, get divorced. A random couple in this sub-population that shares many interests, appears to be in love, and practices a religion that does not tolerate divorce is *still* quite likely to be divorced, though this probability may be less than 99%. The same principle applies to probability estimates in the legal domain. If most heroin in the U.S. is illegally imported, then this fact is relevant for estimating the chance that the drug dealer's heroin has been illegally imported.[3]

Though base rate evidence sometimes "feels" different from other types of evidence, it is no more or less inherently probative than individuating or direct evidence. An 80 percent probability of guilt based entirely on a base rate carries with it the same 20 percent chance of a false conviction as an 80 percent probability of guilt based on, say, the somewhat unreliable testimony of an eyewitness.

Nevertheless, higher courts frequently reject general base rate evidence in cases that should include more individuating sorts of evidence. For example, the Wyoming Supreme Court ruled that testimony that "eighty to eighty-five percent of child sexual abuse is committed by a close relative of the child" was irrelevant in a sexual abuse case. The court wrote, "It is difficult, however, to understand how statistical information would assist a trier of fact in reaching a

[3] Relevant evidence may be excluded in court if the trial judge determines that it's potential for unfairly prejudicing factfinders exceeds its probative value. Because base rate evidence is often perceived to be prejudicial, judges must understand its probative value before ruling on its admissibility.

determination as to guilt in an individual case."[4] However, when base rates are closely linked to the particulars of a case, courts may be more willing to admit them. For example, an Ohio appellate court saw nothing untoward about admitting the base rate for multiple gunshot wound suicides (0.4%) when it was offered for the purpose of rebutting the defense's theory that the decedent's gunshot wounds resulted from a suicide (State v. Sage, 1983).

Base rate evidence proponents are likely to face resistance in the courts. However, parties that wish to introduce such evidence may be able to do so if they can convince courts that the evidence is needed to rebut a claim or is otherwise probative in the instant case.[5]

Misconception #6: A Small Match Probability Implies Source Identification

In many television courtroom dramas, there comes a time when an expert or attorney reveals that the blood found on the murder weapon belongs to Jones, or the fiber recovered from the victim's clothing came from the carpet in Smith's bedroom. Such evidence is presented as irrefutable and often seals a suspect's fate. In reality, forensic science evidence is probabilistic at best. Courtroom claims notwithstanding, forensic scientists cannot establish source identity. They cannot establish that a particular object created a particular marking. Instead, forensic scientists who have access to large, well-constructed databases can estimate the frequency with which people or objects that have various characteristics exist in the population.

---

[4] Id. at 64.

[5] Koehler (2002) identifies a host of other factors that predict when courts are more and less likely to find base rates relevant. These include cases that have a statistical case structure, and those in which the base rates (a) rebut a "chance" hypothesis, (c) are derived from sufficiently refined reference classes, or (c) are offered when individuating proof is in short supply due to the nature of the dispute.

When forensic scientists find a "match" between, say, genetic material left at a crime scene and a set of genetic markers from a suspect, they can help factfinders understand the significance of that match by estimating the frequency with which such matches would occur in the population. If the match is relatively common (e.g., if 50% of people would match), then the evidence has little probative value. If the match is rare, then the evidence has much more probative value. If the match is extremely rare – e.g., if a set of genetic markers has an estimated frequency less than one in every 6.5 billion people (the approximate number of people on the planet) – the forensic scientist will often assert that the matchee is the source of the genetic material. On first consideration, such an assertion may seem reasonable. After all, if the match is rarer than the number of people on earth (e.g., 1 in 10 billion), there *couldn't* be a second person who matches as well, right? Wrong. The probability that at least one other person in a population of 6.5 billion will match a set of genetic markers that occurs just one time in 10 billion is about 48%.

The point is that even very small match probabilities cannot establish that this bullet came from that gun, that this fiber came from that carpet, or even that this fingerprint came from that suspect.

Misconception #7: The Match Probability Identifies the Chance of Innocence

In many criminal cases, the police identify a genetic match between a suspect and trace evidence recovered from the crime scene. The reported match is commonly accompanied by an estimate of the frequency with which the genetic profile occurs in one or more populations. In the language of conditional probability, this frequency is approximately equal to the probability of finding a match with a person who is not the source of the trace evidence, i.e.,

$P(Match|\overline{Source})$. This probability is easily confused with its inverse, $P(|\overline{Source}|Match)$, which is of much greater interest to jurors. That is, jurors do not really want to know how likely a match would be if someone weren't the source of the blood stains. The more central question for them is "how likely is it that the person who matched those blood stains is (or is not) the source?" However, research shows that experts, attorneys, judges, and jurors alike confuse $P(|\overline{Source}|Match)$ with $P(Match|\overline{Source})$ and thereby commit the "source probability error" (Koehler, 1993).

A more extreme form of this "inverse error" (Kaye & Koehler, 1991) is the "prosecutor's fallacy" (Thompson & Schumann, 1987). Those who commit this error treat the match probability as a statement about the probability that the suspect is innocent. Although the match probability provides information about the strength of the forensic science match (up to a point), it does not translate into a probability that describes the chance that a defendant is (or is not) the source of the trace evidence, or the chance that he is (or is not) guilty of the crime. More information than a forensic scientist can supply is needed to make these assessment.

An example clarifies the point. Suppose that a defendant matches genetic material recovered from a rape. Suppose further that the match probability is 1 in 10,000. This means that the approximate probability that a man who is not the rapist (assuming that the material was left by the rapist) will match by sheer coincidence is about 1 in 10,000. In a large city, many men will match a 1 in 10,000 genetic profile. It would therefore be ludicrous to suggest that each of these matching men has only a 1 in 10,000 chance of not being the rapist because this implies that each has a 9,999 in 10,000 (i.e., 99.99%) chance to be the rapist. It would be more accurate to say that, based on the genetic evidence alone, each matchee has only a 1/n chance to be the rapist where n is the estimated number of matchees.

This error detailed above is so subtle and tempting that the entire field of paternity testing appears to have been victimized by it (Kaye, 1989). In paternity testing, forensic scientists typically identify the probability that a man who is not the father of a particular child would have a genetic profile that is consistent with paternity. In other words, the test identifies $P(Match|\overline{Father})$. In court, however, the forensic scientists commonly claim that the test identifies $P(\overline{Father} \mid Match)$. This is the inverse error: the conditional probability $P(\overline{Father} \mid Match)$ cannot be identified from a paternity test, and it is certainly not the same as $P(Match|\overline{Father})$. However, once this error is made, many experts then subtract this value from one to obtain $P(Father|Match)$. This latter probability describes the chance that a person whose genetic profile is consistent with paternity is, in fact, the father of the child. In other words, it is the central issue that jurors must decide in a paternity case, and forensic scientists who testify about it have committed a blunder that effectively usurps the function of the jury.

Consultants should familiarize themselves with these inverse errors and, at a general level, be prepared to explain why the conditional probability P(A|B) is different from P(B|A).

<u>*Misconception #8*: Error Rates: Nice to Have But Not Essential</u>

When DNA evidence burst into U.S. courtrooms in the late 1980s and early 1990s, jurors were exposed to match probability estimates on the order of one in millions, billions, and trillions. These numbers had defense attorneys scrambling for cover. Maybe the computations were wrong. Maybe the statistical databases were incomplete. Maybe the experts were unqualified. Or maybe the analyst made a mistake and called a match where there wasn't one.

This last possibility is critical: if the chance that the analyst made a mistake is several orders of magnitude larger than the match probability (e.g., .001 vs. .00001), then the diagnostic value of the match probability is controlled by the chance that the analyst made a mistake. That is, if the chance that the analyst called match on a nonmatch is .001, and the chance that the match is merely coincidental is .000001, then the .001 value rather than the .000001 value is the key to understanding the value of the match report.

In many cases involving matching evidence, jurors are provided with very small match probabilities but they are not provided with probabilities that will help them identify key error rates, such as the rate at which nonmatches are identified as matches. As a result, they are in a poor position to assess the probative strength of the reported match evidence. When appellate courts have considered this problem, they generally conclude that failure to provide error rate data is insufficient grounds for overturning a conviction, even if the match evidence played an important role in securing that conviction. As one California court reasoned:

> "We believe the issue of whether laboratory error rates should have been presented to the jury in addition to the profile frequency is not one that goes to the very integrity or reliability of the DNA results.…Whether calculation and presentation of laboratory error rates could have improved the FBI's scientific procedure went to the weight of the evidence, not its admissibility" (People v. Pizarro, 2002).

Such reasoning poses a challenge to trial consultants and others who wish to help courts make informed judgments about scientific information. Courts need help understanding that the absence of error rate guidance in cases that include very small match probabilities leaves jurors with a misleading perspective on the importance of the match evidence. In such cases, error rates provide an upper boundary on the probative value of a reported match. Even among judges who understand this point, there is still the problem of persuading them that general error rate data are sufficiently relevant to the instant case.

Misconception #9: Non-unique Matches Are Worthless

Thompson and Schumann (1987) coined the phrase "the Defense Attorney Fallacy" to describe the erroneous belief that non-unique match evidence is not probative. Of course, since match evidence is nearly always non-unique, the Defense Attorney Fallacy amounts to a claim that match evidence is nearly always worthless. Obviously this is not true. Still, there is something seductive about this fallacy. Suppose, for example, that a partial DNA profile from an evidentiary sample occurs with a frequency of 1 in 1,000. Suppose further that the suspect population includes 30,000 adult males in a city. In this scenario, the Defense Attorney Fallacy occurs if one argues that a DNA match on a particular suspect is worthless because approximately 30 adult males in the city would match. This argument is fallacious because it fails to distinguish between probative evidence and conclusive evidence. Non-unique evidence may be highly probative if the size of the inclusion group is small relative to that of the exclusion group. In the example above, the partial DNA match is highly probative because 99.9% of people who were not the source would have been excluded, yet the defendant happens to be a member of the very small group (n=30) of non-excluded people. Holding aside error rate considerations, the discovery of this match means that a juror should believe that the defendant is now about 1000 times more likely to be the source of the DNA evidence than he was prior to the discovery of the match. Certainly this evidence is probative.

Studies show that significant minorities of jurors are susceptible to the Defense Attorney Fallacy (Thompson & Schumann, 1987; Nance & Morris, 2005), particularly when it is explicitly offered as an argument by a defense attorney. Hans, Kaye, Dann, Farley, and Albertson (2006) found that fully 40% of jurors agreed with the statement "The mtDNA evidence in this case is completely irrelevant because a substantial number of other people could also be the source of the hairs" (p. 17). At the same time, many jurors in this study regarded the DNA evidence to be completely persuasive, including a smaller number who may have fallen victim to the Prosecutor's Fallacy. Perhaps the best explanation for why people commit Defense Attorney and Prosecutor's Fallacies is the simplest one: people are easily confused and misled by statistical evidence, particularly when the relevant probability values are very small.

Conclusion

As the law increasingly leans on scientific evidence to solve tough cases, the ability to reason with quantitative data becomes increasingly important to the administration of justice. Yet the unfortunate reality is that attorneys, judges, jurors and other legal actors have limited knowledge of how to reason with quantitative evidence. Trial consultants can do a great service to our legal system by studying the problems that jurors and others are likely to have with scientific and statistical evidence, and by sharing their quantitative expertise with various legal actors. For example, they can help jurors understand the significance of very low match probabilities, and they can help attorneys zero in on the weaknesses in the statistical arguments of opposing parties. They may also need to find ways to get through to those who dismiss statistics and statistical arguments in their entirety because they simply don't trust statistics. These skeptics pose a serious challenge to the trial consultant because they are unlikely to be moved by the subtleties of the statistical misconceptions and corrective points discussed above. Instead, trial consultants may need to persuade the skeptics that truth is harmed when any particular type of evidence or argument is dismissed. Toward this end, I recommend that the trial consultant quote the late, great statistician Fred Mosteller: "Sure, it's easy to lie with statistics – but it's easier to lie without them."

References

Arkes, H. R., Connolly, T., & Hammond, K. R. (Eds.) (2000). Judgment and Decision Making: An Interdisciplinary Reader. Cambridge: Cambridge University Press.

Bar-Hillel, M. (1980) The base-rate fallacy in probability judgments. Acta Psychologica, 44, 211 -233.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd Ed.). Hillsdale, NJ: Erlbaum.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. Psychological Bulletin, 99, 3-19.

Hans, V. P., Kaye, D. H., Dann, B. M., Farley, E., & Albertson, S. (2006). Science in the jury box: Do jurors understand mtDNA evidence? (Unpublished manuscript).

Kahneman, D., Slovic, P. & Tversky, A. (Eds.) (1982). Judgment Under Uncertainty: Heuristic and Biases. Cambridge: Cambridge University Press.

Kaye, D. H. (1989). The probability of an ultimate issue: The strange cases of paternity testing. Iowa Law Review, 75, 75-109.

Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? Journal of the Royal Statistical Society, Series A, 154, 75-81.

Koehler, J. J. (1993). Error and exaggeration in the presentation of DNA evidence. Jurimetrics Journal, 34, 21-39.

Koehler, J. J. (2002). When do courts think base rate statistics are relevant? Jurimetrics Journal, 42, 373-402.

Koehler, J. J., Chia, A., & Lindsey, J. S. (1995). The random match probability (RMP) in DNA evidence: Irrelevant and prejudicial? Jurimetrics Journal, 35, 201-219.

Nisbett, R. & Ross. L. (1980). Human Inference: Strategies and Shortcomings of Social Judgment. Prentice Hall.

People v. Pizarro, 123 Cal.Rptr.2d 782, *921 to 123 Cal.Rptr.2d 782, *922

Saks, M. J. & Kidd, R. F. (1980-1) Human information processing and adjudication: Trial by heuristics. Law and Society Review, 15, 123-160.

Schklar, J. & Diamond, S. (1999). Juror reactions to DNA evidence: Errors and expectancies. Law and Human Behavior, 23, 159-184.

State v. Sage, No. 82AP-983 (Ct. App. Ohio 1983).

Thompson, W. C. (1989). Are juries competent to evaluate statistical evidence? Law and Contemporary Problems, 52, 9-41.

Thompson, W. C. & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutors' fallacy and the defense attorney's fallacy. Law and Human Behavior, 11, 167-187.

Tribe, L. H. (1971) Trial by mathematics: Precision and ritual in the legal process. Harvard Law Review, 84, 1329-1393.

Table 1:

Hypothetical Data: Amount of Time Until Patient is Discharged

Patient	Drug Received	Days to Discharge
1	A	2
2	A	3
3	A	3
4	A	4
5	A	5
6	B	6
7	B	7
8	B	8