Statistics and the Law: Will the Twain Ever Meet?

Mary Gray
Mathematics and Statistics, American University, 4400 Massachusetts Avenue, NW, Washington, DC, 200016-8050

Abstract

Statisticians and lawyers often must work together, but conflicts in their approaches and objectives can lead to difficulties. Some examples of such instances and cautions that, if observed, might avoid conflict are discussed.

Keywords: law, forensics, methodology

Statisticians think of their work as reaching objective conclusions from the analysis of data. While these conclusions may have policy or legal implications, drawing these implications or acting as an advocate is not the role of the statistician as a statistician. In particular, expert testimony cannot reach legal conclusions; for example, the statistician can say that there is a statistically significant difference between the salaries of men and women unexplained by legitimate factors in the model she has constructed, but cannot say that there is discrimination

Unlike research (at least in theory), legal proceedings are *adversarial:* plaintiff versus defendant or the government versus defendant in a criminal case. The two competing sides are represented by legal counsel who must think and act in accordance with their responsibilities of zealous representation of their clients. Each side will present evidence and interpretations of the evidence to bolster its theory of the case. This often leads to two quite different interpretations of the some of the evidence, both presented by statisticians.

Lawyers are bound by certain ethical considerations in a more formal way than are statisticians, but statisticians bear responsibility for the methods they use and the conclusions they reach. Obviously, doctoring the data or misstating results is unacceptable, but expert witnesses can feel the pressure of the adversarial system, compounded by the fact that the attorneys with whom they work rarely have any in-depth knowledge or intuitive understanding of the statistics involved.

1. Historical background

The use of statistics has been compelling in resolving matters of public policy for many years. In *Yick Wo v. Hopkins* (1886), the Supreme Court noted that statistics showed that not a single permit to operate a laundry had been issued to a San Francisco resident of Chinese ethnic origin although many had been issued to whites, an early instance of the power of what has been termed "the inexorable zero." In *Baker v. Carr* (1962) statistical evidence was used to establish the principle of "one man, one vote."

Under Title VII of the anti-discrimination act of 1964 two types of discrimination are prohibited: disparate treatment (for example, men are treated differently from women by job listings stating that only men may apply) and disparate impact, the use of a facially neutral employment policy that affects men and women differently (for example, imposing a minimum height requirement not necessary for the job). Obviously, the concept of "disparate impact" requires the use of statistics, as first recognized by the U.S. Supreme Court in *Griggs v. Duke Power Company* (1971). However, the techniques in *Griggs* and a line of cases that followed were descriptive, difficult in detail perhaps for mathphobic lawyers, judges and juries, especially when Disraeli's categorization of "lies, damned lies and statistics," is presented with a flourish, but more was to come. It has come to be believed that statistical significance was introduced into Supreme Court deliberations by a former physicist who was serving as a clerk, first in *Castenada v. Partida* (1977), a death penalty case, and then in *Hazelwood School District v. United States* (1977), an employment case.

Before *Castenada,* descriptive statistics could lead to anomalies in deciding whether racially exclusionary practices in jury selection deprived defendants of their constitutional right to be tried by a jury of their "peers." The Court was inclined simply to look at the raw difference in the percentage of minorities in the eligibility pool and that on the jury panels. Thus in *Swain v. Alabama* (1965) the Court found a discrepancy of 26% in the pool versus 16% on the panels

unprobative of discrimination; in fact the probability of such a result's occurring by chance under the circumstances is 1 in 10^8. On the other hand, a comparison of 5% versus 0% was found to be evidence of discrimination in *Avery v. Georgia* (1953) even though the relevant probability is .046. However, the inexorable zero here combined with the fact that the names of potential white jurors were written on white slips of paper and those of black jurors on yellow slips before being drawn from a transparent fish bowl. In the precedent-setting *Castaneda* the comparison of 79% versus 39% resulted in a *p*-value of 1 in 10^{140}.

Although courts have come to rely on statistical inference, there is a residual inclination to believe that "statistics don't prove anything." Although in a legal sense that may be true, there is a prevalent failure to understand the role that statistics can play, namely to make a result sufficiently improbable that it should strain the credulity of the legal system. For example, in *McCleskey v. Kemp* (1987), a regression analysis including 39 explanatory variables showed the death penalty to be 4.3 times more likely when the victim was white. The evidence was rejected by the Supreme Court because, among other reasons, death penalty sentencing is at "the heart of the … criminal justice system," and thus should not be effectively challenged by statistical evidence alone.

An unfortunate outcome of the courts' reliance in general on statistical evidence has been some tendency, shared unfortunately by much research in the social sciences, to seek a bright line *p*-value cutoff for reliance on the admission, not just the strength, of statistical evidence. Even more unfortunately there has not been agreement on where the cutoff should come.

Although DNA evidence has come to be decisive, particularly in freeing long-serving prisoners through efforts of such groups as the "Innocence Project," when it first received widespread public attention in the O.J. Simpson trial it tended to inspire widespread defection from the TV audience—with resultant peaks in water usage. This is not to say that there have not been careless or inappropriate uses of DNA evidence. Just as finger-printing has come under fire for relying too much on subjective "matching" standards, so too with DNA.

2. More recent developments

Admissibility of "scientific" evidence has evolved in recent years; in federal courts the controlling standard is U.S. Federal Evidence Rule 702, which states:

If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge skill, experience, training or education, may testify thereto in the form of an opinion or otherwise, if

(1) the testimony is based upon sufficient factors or data

(2) the testimony is the product of reliable principles and methods, and

(3) the witness has applied the principles and methods reliably to the facts of the case.

In the early 1900's the "Commercial Marketplace Test," that is, that the testimony would be accepted in the marketplace, determined whether testimony was sufficiently reliable to be admitted. Subsequently, the *Frye v. United States* (1923) decision required that expert testimony be "generally accepted," which usually was interpreted to mean peer-reviewed. The situation was made more cloudy by the decision in *Daubert v. Merrell Dow Pharmaceuticals* (1993), where the Supreme Court declared that judges must evaluate the methodology of expert testimony according to the following:

- Testing and validation
- Peer review
- Existence and maintenance of standards
- Controlling the use of the technique
- Rate of error
- "General acceptance"

Subsequently *Joiner v. General Electric* (1997) and *Kuhmo v. Carmichael* (1999) further clarified the role of the judge by extending *Daubert* to evaluating the way methodology is applied and expanded the definition of who is an expert.

Unfortunately the establishment of what would appear to be a more rigorous review of evidence appears not to have kept bad statistics out of court. Although of course not controlled by U.S. rules, English courts have had similar problems. In a "cot death" (SIDS, Sudden Infant Death Syndrome) case in England a physician

who had testified several times in similar cases asserted that the probability of a "cot death" in the population from which the defendant came was one in 8500. In the case at the bar a mother was being tried for the death of a second of her children with a diagnosis of "cot death." The "expert" computed:

$$1/8500 \times 1/8500 = 1/72{,}250{,}000,$$

where 72,250,000 is more than the total population of England. Statisticians on behalf of Royal Statistical Society learned of the testimony and objected that the events were not independent since evidence existed of genetic and behavioral factors indicating otherwise. In overturning the conviction, the Court noted that the physician had no evidence for the testimony he presented; he was "struck off" the registry of physicians. Unfortunately, the defendant whose career had been destroyed committed suicide shortly after her release from prison. Several other convictions involving the same physician have been voided and over 250 murder convictions in "cot death" cases are being reviewed.

In *Maryland v. Wilson* (2002), the father of two children who died of SIDS was convicted of murder of the second partly on the basis of similar testimony, although the faulty statistics of the expert testifying for the prosecution produced only a probability of 1 in 4 million. The prosecution's putative probability was further reduced by the presence of another factor observed in the autopsy. Finally the prosecutor told the jury "If you multiply his numbers, instead of 1 in 4 million, you get 1 in 10 million that the man sitting here is innocent. That was what a doctor, their expert, told you." Clearly this goes beyond the problem with multiplying non-independent probabilities. The defense counsel's motion for a mistrial was denied and, instead, the court merely gave a curative instruction. Wilson was convicted and spent two years in prison before the conviction was overturned explicitly because of the misuse of statistics.

Two cases recently decided by the U.S. Supreme Court were on appeal from decisions in which the courts below refused to be misled by the faulty statistics admitted into evidence. *Gonzales v. Carhart* and *Gonzales v. Planned Parenthood* involved state regulation of "partial birth abortions" with the states relying in part on the *Chasen* study (2004). In the study the null hypothesis was that two different procedures led to the same rate of subsequent premature births, with the evidence resulting in a probability of $p = 0.30$. However, the government's expert testified that 30% is just "stretching it a little bit" from 5% and "There is a 30 percent chance this occurred by chance and a 70 percent chance that it in fact is a true, meaningful, increased risk." An *amicus* brief by a group of statisticians attempted to convince the Supreme Court of the error of this interpretation, but the statistical issue did not affect its decision to reverse the lower courts.

3. Is there a better way?

To reconcile the competing roles and responsibilities of a statistician and a lawyer in a legal setting requires an effective partnership of the two. The role of the statistician is

- To present the evidence clearly and ethically
- To prepare the litigator to deal with statistical evidence.

The American Statistical Association has made some effort at outreach by devising a short course for lawyers and others in law-related professions: *Statistics in a Legal Context: A Gentle Introduction.* In the other direction, efforts have been made by some to convince their colleagues that the statistical profession ought to adapt the principle of the legal profession that there is an obligation for *pro bono* service, much of which might be in a legal context.

It has often been proposed, but seldom implemented, that judges take the initiative to appoint a statistician to advise the court. In the Australian system parties must agree on the selection of an expert and are bound by her conclusions.

Absent moving to such a system, what can be done to reduce the potential for clashes from the differing mindsets and ways of working of statisticians and lawyers? Certainly better training of lawyers and judges through efforts such as the ASA course would help. However, one must beware of the hazards of the lawyer who thinks she is a statistician. A little knowledge can be a dangerous thing, even if it is usually better than no knowledge at all. In a situation of cross-examination the expert witness can be at a disadvantage if the questioner's approach is statistical nonsense. For similar reasons, it is usually better to have an opposing expert who is

knowledgeable so long as she is ethical. It should be noted that expert witnesses, unlike "fact" witnesses, are permitted, and indeed expected, to testify on the basis of hypotheticals. Dealing with bizarrely formulated hypotheses can be tricky.

What sorts of things are useful for lawyers to know?

- First and foremost, it is important to understand the necessity of early detailed consultation, particularly with respect to obtaining data.
- Then it is necessary that the lawyer understand the strength and limitations of the evidence.

For example, in one case involving allegations of discrimination against Latinos in hiring at the National Security Agency, actual employment figures were classified. The lawyers need to realize that percentages alone are not going to be very useful. And on the other hand, it is important that statisticians not allow unrealistic expectations of what they can accomplish, both from the point of view of what data are available and what statistical analysis might actually show.

Sometimes even statisticians forget that what data need to be gathered depends on what questions are to be answered; this needs to be understood by everyone before the "discovery" period, when evidence that is held by the opposition can be obtained, is over.

4. Specific topics

Having established that early consultation and continued communication is important, what specifically would we want the lawyer with whom we are working to understand (in an ideal world, of course)?

- The concept of sampling.
- The role played by sample size—often evidence will be challenged on the grounds that the sample was small when that has already been taken into account.
- Probability. This may be too much to hope for.
- Statistical methodology—at least to the extent of understanding what is appropriate for quantitative and qualitative data.
- *t*-tests—uses and limitations
- Regression—linear and other. Lawyers often are enamored of the concept, with little comprehension of the purpose and value.
- Non-parametric tests
- Standard deviations, *p*-values and the connection between the two
- Statistical significance

Thinking of the lawyer with whom one is working, the judge, and the jury as an undergraduate statistics class for whom one needs to provide a very quick and very clear introduction to statistics is probably the best approach to take.

On the other hand, what is most important for the statistician to know?

- The limitations of her role—she is an expert, not an advocate, and cannot come to legal conclusions.
- Statisticians bear responsibility for the methods they use and the conclusions they reach.
- Statisticians can and should be held accountable for their work.
- Because of the nature of their role in adversarial legal settings, it is essential that statisticians carefully guard their reputations.
- What is expected and is it acceptable from an ethical and scientific point of view
- What questions are to be answered in the case at hand and what data are needed to answer them (and what may in fact be available).
- What evidence will contribute to an understanding of the case.
- How to make an effective presentation of evidence and how to respond to cross examination.
- What graphics might aid in making the points that need to be made without confusing the trier of fact (judge or jury).

In general there needs to be a clear understanding of what the statistician can and is expected to do. This begins with being certain that the statistician understands exactly what questions must be answered. Only then is it possible to know what data are needed; the importance of the relevant data needs to be

conveyed to the lawyers on the case. Getting the data can be a major problem, often requiring extensive back and forth between lawyers on both sides, often culminating in court intervention. Once the data have been obtained (to the extent possible), a lot of cleaning and grappling with the extent and form of the data are inevitably going to be necessary. Important to keep in mind is consideration of the strategy of the opposition. Be prepared to look at alternate methods and interpretations.
Let us look at how an actual case might be approached as a cooperative effort between lawyer and statistician.

A company is in trouble and new management is brought in. In an alleged rescue operation, large numbers of employees are "riffed." An older employee alleges that decisions on whom to rif were based on age. What is the legal situation?

- If discrimination occurred is it disparate treatment or disparate impact?
- Is there any anecdotal evidence—was "new blood," "fresh perspectives" or just plain, "Let's get rid of older employees" mentioned in general or with respect to the plaintiff?
- Federal law protects employees 40 or older from being discriminated against in hiring, firing, etc. (exceptions for public safety, management and formerly for tenured faculty).

Statistical issues

- What is the average age of the riffed employees compared to the average age of those not riffed?
- What percentage of the employees aged 40 or older were riffed compared with the percentage of the employees under 40 who were riffed? Should another cutoff point also be used?
- Are only certain categories of employees affected, or affected disproportionately?
- Are there other explanations offered by the defendant for choosing whom to rif and if so, are they legitimate factors and are there relevant data?
  - Performance
  - Job category
  - Location
  - Sex
  - Ethnicity
- What is the relevant period of time?
- What is the relevant point for determining the ages of the employees?
- Two sample *t*-tests
- Logistic regression

How to describe the results?

- The inexorable zero
- *p*-values: 1 in 100 or 0.01
- Standard deviations
- Odds

9. Conclusion

Statisticians and lawyers need to put themselves in the position of their collaborator if the use of statistics is to be useful in a legal context. Communication and clarity as to the role and responsibilities of each are the keys to success.

References

Avery v. Georgia, 345 U.S. 559 (1953).

Baker v. Carr, 369 U.S. 186 (1962).

Castenada v. Partida, 430 U.S. 482 (1977).

Chasen, S.T. (2004). **"**Dilation and evacuation at ≥ 20 weeks: Comparison of operative techniques," American Journal of Obstetrics and Gynecology, vol. 190, p.1180**.**

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

DeGroot, M., Fienberg, S. and Kadane, J.B. (1986). S*tatistics and the Law.* New York: Wiley.

Fienberg, S. (ed.). (1989). *The Evolving Role of Statistical Assessments in the Courts*. New York: Springer Verlag.

Fienberg, S.E., Krislov, S.H., and Straf, M.L. (1995), "Understanding and evaluating statistical evidence in litigation," *Jurimetr. J.,* **36**, 1-32.

Finkelstein, M. O. and Levin, B. (2001). *Statistics for Lawyers*, 2nd edition. New York: Springer Verlag.

Frye v. United States, 293 F. 1013 (1923).

Gastwirth, J.L. (1988). *Statistical Reasoning in Law and Public Policy*, vols. I and II, San Diego: Academic Press.

Gastwirth, J. L. (ed.). (2000). *Statistical Science in the Courtroom*, Springer Verlag, New York.

Gonzales v. Carhart, Gonzales v. Planned Parenthood, Nos. 05-380 & 05-1382 (2006). Brief of *Amici Curiae* Statisticians.

Gray, M. W. (1993), "Can statistics tell us what we do not want to hear? The case of complex salary structures*."* Statistical Sciences, vol. 8**,** pp. 144-179.

Griggs v. Duke Power Company, 401 U.S. 424 (1971).

Hazelwood School District v. United States, 433 U.S. 299 (1977).

Joiner v. General Electric, 522 U.S. 13 (1997).

Kadane, J.B. (2005), "Ethical issues in being an expert witness*,"* Law, Probability & Risk*,* vol. 4, pp. 21-23.

Kaye, D.H. and Freedman, D. A. (2000), "Reference guide for statistics," in *Reference Manual on Scientific Evidence*, Washington DC: Federal Judicial Center.

Kuhmo v. Carmichael, 526 U.S. 137 (1999).

McCleskey v. Kemp, 481 U.S. 279 (1989).

Swain v. Alabama, 408 U.S. 936 (1972).

Wilson v. Maryland, 370 Md. 191 (Ct. of Appeals of Md 2002).

Yick Wo v. Hopkins, 118 U.S. 356 (1986).