

TEACHING STATISTICS TO AN AUDIENCE OF ONE:
THE JUDICIAL RECEPTION OF STATISTICAL EVIDENCE

Mary W. Gray, American University, USA

*Until recently, in United States federal courts the standard for the admissibility of any scientific evidence was whether it was based on generally accepted methodology. "Generally accepted" in turn usually meant in practice "published in a peer-reviewed journal". Thus it was the opinion of other experts that determined the validity of the methodology. With the decision of the U.S. Supreme Court in *Daubert v. Dow Pharmaceutical*, the standard has changed, placing the burden of deciding whether to hear scientific evidence squarely on the judge. Moreover, in *General Electric v. Joiner*, the Court not only reaffirmed *Daubert*, but held that judges may assess the conclusions of experts as well as their methodology, ruling that the two "are not entirely distinct." Thus statistical experts, like other scientific experts, must teach judges enough about probability and statistics to convince them of the acceptability of the full gamut of the evidence being presented.*

The role of statistics in educating judicial decision makers, particularly, but by no means exclusively, in the United States, has become increasingly prominent in the past quarter century. Admission of evidence in U.S. federal courts is governed by the Federal Rules of Evidence, which evince a preference for admitting expert evidence. For many years, judges operated under the *Frye* rule (1923), which stated "the things from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." "Gained general acceptance" in turn usually meant in practice "published in a peer-reviewed journal." The idea was to assure that the expert testimony was not just an individual's crackpot idea or so-called "junk" science, but rather met a professional standard in the field. Thus it was on the opinion of other experts that decisions about the validity of scientific methodology were to be based.

With the opinion of the U.S. Supreme Court in *Daubert v. Dow Pharmaceutical* (1993), the standard has changed, placing the burden of deciding whether to hear the evidence squarely on the judge. Under *Daubert* a judge may now decide whether (s)he finds the particular statistical technique worthy of admission with only the guidance of a checklist, offered by the Court with the *caveat* that none of its items is to be considered determinative and that in its entirety it cannot be deemed definitive:

1. whether the theory or technique can be, and has been, tested;
2. whether the technique has been published or subjected to peer review;
3. whether actual or potential error rates have been considered;

4. whether the technique is widely accepted within the relevant scientific community.

Thus in seeking to present statistical evidence before a court, the expert must first convince the judge of its relevance, where to be relevant means that it “tends to make existence of a consequential fact more or less probable” than it otherwise might have been. In making this determination, the judge needs to evaluate the methodology used, but, according to the recently decided *General Electric v. Joiner* (1997), may also assess the conclusions drawn by the expert, for the two “are not entirely distinct.” The *Joiner* Court went on to say that “[N]othing ... requires a district court to admit opinion evidence which is connected to existing data only by the ipse dixit of the expert. ...A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.”

In the *Joiner* case the question was whether studies connecting stomach cancer in mice with polychlorinated biphenyls (PCBs) were relevant to whether PCBs might have caused lung cancer in the plaintiff, whereas it is frequently the case that statistical experts face a skepticism that statistical evidence can “prove” anything. Of course, in some sense it cannot, but it can certainly make the existence of some fact “more or less probable.” However, the difficulty with statistical evidence arises in establishing not only the acceptability of the methodology, but also just why certain conclusions derive validly from the evidence whereas others do not. This has, of course, always been the conundrum in the presentation of statistical evidence. When faced with conflicting conclusions of experts, judges have often thrown up their hands, as, for example, when judges have been unable to sort out the merits of rival regression models of faculty salaries (*Presseisen* 1978; *Ottaviani* 1989).

The process of evaluating evidence once it has been admitted relates to its probative value, or what is often called its “weight.” To put it in statistical terms, “good” evidence has a substantial effect on the probability of the existence of a material fact. Although judges, whether under the *Frye* rule or the *Daubert* rule, must decide whether evidence is admissible, it is not always they who decide on its weight, for in a jury trial it is the jury who are the finders-of-fact. One might justify the *Daubert* ruling in jury cases on the grounds that the judge ought to be allowed broad discretion to decide whether complex and possibly invalid statistical evidence should go before a jury; after all, should

the standards for publication in a journal where conclusions are subjected to further close scrutiny by technically trained experts be the same as for submitting the evidence to the judgment of a lay jury?

On the other hand, is a judge any better able to screen out “junk” statistics, whether methodologically or otherwise flawed, than a jury? On the contrary, it is the author’s experience that juries are often much more receptive to and capable of evaluation of such evidence than are judges. Most juries know they do not know much about statistics as indeed they know little about many aspects of what is being presented to them, but they also recognize their responsibility to be instructed. Judges, on the other hand, all too frequently are so accustomed to being the authority figure that they do not want to be instructed, particularly about a subject with which they have never felt comfortable. In a concurring opinion in *Joiner*, Justice Stephen G. Breyer suggested that judges rely more on scientific panels to help sort out evidence. Such proposals have been around for a long time, but only rarely have judges used either panels or court-appointed experts to help with the evaluation of technical evidence.

This brings us to how to present statistical evidence in a way to get it admitted in the first place and to have it properly understood in the second. Keeping in mind that the question of the acceptability of a technique has a guideline for admissibility has not necessarily disappeared with the advent of *Daubert*, the expert will generally find that a journal is a better forum for the introduction of radically new methodology than is the courtroom. In an age discrimination case involving the U. S. Federal Marshals Service, the judge discounted defense evidence based on a technique known only to their expert, who failed, in any event, to provide a clear explanation of what he was doing (*Sandy v. Reno*).

If one is studying whether men and women who have the same qualifications and are doing the same work, it might seem appropriate to formulate a regression model based on the men alone and then to look at the salaries predicted for women under this model. The average residual for women might then be taken as a measure of the sex-based disparity; this might be refined by modeling the men’s salaries on those of the women and comparing their average residual to that of the women under the men’s model. But what are the chances of the judge or jury following this analysis? On the other hand, if one simply throws all the men and women into the model, including a variable for sex along

the way, the sex coefficient is a nice, convenient way to show the disparity, if any. That is why that is the technique generally used in sex discrimination cases (Gray 1993).

Simplicity needs to be balanced with completeness, theoretical purity with practical application. It is important to note that while experts are generally accorded wide latitude in their testimony--only they are allowed to express opinions and to respond to hypothetical questions--nonetheless the evidence not only ought to have a substantial effect on the probability of the existence of a material fact but it should be apparent why that it the case. In a landmark case involving charges of sex discrimination in college athletics, the defense expert--at great expense to the defendants--instituted telephone surveys of student interest of questionable methodology, extensive canvassing of the policies and practices of comparable institutions, and data mining of student application questionnaires when the question to be addressed by statistical evidence was merely whether the percentage of women in the student body was substantially the same as the percentage of women among the student athletes (*Cohen* 1995).

It is important to remember that statistical--and other--experts must couch their evidence in terms that do not pronounce upon the ultimate legal question. Thus, statistically significant sex-based disparities in salaries that remain once a number of legitimate factors have been taken into account cannot prove the existence of discrimination; they can only make it very unlikely that these differences happened by chance.

Several years ago in the U.S., public attention was drawn in a highly-charged atmosphere to the DNA evidence in the O.J. Simpson case. Although few, if any, believe that the case turned on the DNA evidence, or in particular on the statistical aspects of that evidence as opposed to the questions of care in handling the evidence and the laboratory techniques involved, it is certainly the case that many TV viewers mesmerized by the trial took the opportunity of the presentations by opposing statistical experts for welcome breaks. Perhaps more clarity in presentation might have educated many about the nature of DNA evidence if not have contributed to a decision as to the ultimate legal issue: i.e., does the DNA evidence show that O.J. was there?

The key is to accept that the judge and then, perhaps, the jury are a class whom it is necessary to convince of the validity of statistical techniques and the results they produce. Then the general concepts need to be connected in a clear and convincing way to the case at hand. For example, courts are partial to the concept of "statistical

significance,” frequently without any understanding of what it means. A coin-tossing example, accompanied if permitted by a demonstration, may prove helpful--ten heads in a row is much clearer to most laypeople than the abstract concept of one in a thousand. The same pedagogical guidelines apply in teaching these courtroom classes as in the ordinary classroom. Graphs and other visual aids may help—but graphs can also be misleading (Tufté 1997); although there is no specific code of ethics binding statisticians, general ethical principles would certainly preclude some statistical presentations that have found their way into courtrooms.

The statistical expert must convince the courtroom audience that the evidence being presented can be relied upon. Reliability in a forensic concept should be in essence what we usually think of:

Sensitivity—can our methodology be relied upon to produce results from the quantity and quality of data being examined?

Quality control—are factors affecting outcomes understood and controlled for?

Discriminating power—can we distinguish between possible outcomes?

plus good old-fashioned honesty. Is the presentation such that the judge or jury is convinced that the expert is telling the truth and not tailoring the evidence in order to earn a fee? But often this hinges on nothing more than what every good teacher knows: giving the listeners a sense of contact and confidence combined with absence of evasion and not undermined by superciliousness or arrogance.

However, few cases are won or lost on statistics alone, and it is usually a skillful blending of statistical evidence with other evidence to tell a believable tale that prevails. In a class action case, no matter how overwhelming the statistical evidence for the plaintiffs’ case, the appearance of individual witnesses who are unconvincing can easily negate its effect.

REFERENCES

- Aitken, C. G. G. and Stoney, D. A. (eds.) (1991). *The Use of Statistics in Forensic Science*. Ellis Horwood, Chichester.
- Cohen v. Brown University*. (1995). 879 F. Supp. 185 (D.R.I.)
- Daubert v. Dow Pharmaceutical*. (1993). 509 U.S. 579 .
- Frye v. United States*. (1923). 293 F. 1013.

- Gastwirth, J. L. (1988). *Statistical Reasoning in Law and Public Policy. Volumes I and II*. Academic Press, Boston.
- General Electric Co. v. Joiner*. (December 15, 1996). U.S. Supreme Court 96-188, reversing 78 F.3d 524 (11th Cir. 1996).
- Gray, M. W. (1993). Can statistics tell us what we do not want to hear? The case of complex salary structures (with discussion). *Statistical Sciences*, 8, 144-179.
- Presseisen v. Swarthmore College*. (1978). 442 F. Supp. 593 (E.D.Pa.), *aff'd without opinion*, 582 F.2d 1275 (3d Cir. 1978).
- Robertson, B. and Vignaux, G. A. (1995) *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. John Wiley and Sons, Chicester.
- Roeder, K. (1994). *Statistical Sciences* 9, 222-243.
- Sandy v. Reno*. (1992). (U.S. District Court, District of Columbia).
- Tufte, E. R. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Chesire CT.