Damned Liars and Expert Witnesses

Paul Meier

Damned Liars and Expert Witnesses

PAUL MEIER*

Until recently the applications of inferential statistics in legal proceedings have been minor and limited. With the advent of civil rights legislation, however, the courts have embraced statistical inference with enthusiasm. The needs of the courts are not well matched with the usual practice of statistics, and this mismatch has serious adverse consequences for both fields. The various sources of difficulty are outlined, and tentative proposals for their amelioration are put forward.

1. INTRODUCTION

Although the field of statistics can find its origins in matters pertaining to society and its governance, statistics as a formal discipline has only recently received special recognition in legal proceedings. To be sure, statistics in the sense of numerical summaries are pervasive—in legal settings as in many others. But statistical inference based on probability models is another matter, and in that respect statistics has had only a minor and restricted role in the law.

The Howland will case of 1867 (Meier and Zabell 1980), in which Benjamin Peirce undertook a statistical analysis of handwriting, is a case in point. The analysis was ingenious, and might even have been persuasive, but the court in that instance found a technical excuse to put it aside. From time to time, most notably in the Collins case a century later (*People v. Collins* 1968), statistical analyses of identification evidence have come before the courts, and generally the courts have rejected them, except in the rather special cases of genetic evidence of paternity and of fingerprint evidence.

Following the passage of the Civil Rights Act of 1964, however, the courts have looked to statistical analysis to decide on the substantiality of evidence of illegal discrimination, and by now the statistical expert witness is definitely in the big time.

It might be thought that this is cause for celebration within our profession. Inference is our field, of course, and what could be more appropriate than a long overdue recognition by the courts of our special expertise. But there is room for second thoughts as well, when we pause to consider the consequences for other professions that have come to occupy a similar role. The situation of psychiatry (to choose a not-at-all random example) is notorious. In the case of the most recent would-be presidential assassin, John Hinckley, neither the prosecution nor the defense had any difficulty in finding distinguished psychiatrists, academics and others, to testify that Hinckley was or was not legally sane at the time he fired the shots. Indeed, it is not stretching matters to say that the courts and the bar, and even the public at large, have come to hold the profession of psychiatry in considerable contempt—as a clan of hired guns, available for a price to whichever side first knocks on the door.

That this perception is not altogether fair is beside the point. The statisticians may have cause for congratulation in their newfound status—they also have cause for worry.

Indeed, psychiatry is not alone in its notoriety. The evident ease with which experts in almost any field can be found to testify in support of either side of a case has led to an aphorism in the law that has a familiar ring to statisticians: There are three kinds of liars—liars, damned liars, and expert witnesses. (The origins of this aphorism are uncertain, but it appears in various forms in legal writing during the past century.) Statistics has had a hard time establishing its credibility as a scientific domain, and the credit that it now has may well be threatened by our new-found prominence.

The views expressed here are idiosyncratic, and the interested reader may wish to consult additional sources dealing with the interaction between statistics and law. In particular, the collections edited by Peterson (1983), Monahan and Walker (1985), and DeGroot, Fienberg, and Kadane (1986) present much relevant material and a number of alternative views. Finally, it should be noted that the Committee on National Statistics of the National Academy of Sciences has conducted a three-year study of the applications of statistics in law, and its report is due to appear soon.

The remainder of this article is divided into four sections: the first reviews the different domains in which statistical testimony is sought; the next discusses the environment in which such testimony is given and contrasts that environment with the quite different system prevalent in Europe. The manifold corrupting influences that lead to the unsavory reputation of expert witnesses in American courts are then reviewed, and I close with a very modest proposal for reform.

2. DOMAINS OF APPLICATION OF STATISTICS IN LAW

I start by distinguishing particular domains in which statistical expertise is called upon.

Scientific Sampling

The simplest and, in many ways, the most satisfactory application of statistics in legal proceedings is the use of scientific sampling methods. In this area, W. E. Deming has been the preeminent pioneer (Deming 1954), and he has taken the pains to lay out a clear recipe for satisfactory performance. This consists largely of eschewing any responsibility for choice of population to be sampled or for the evaluation of sampled units. His advice, which I believe to be eminently sensible, is that the sampling expert limit himself to testimony about the inference from the sample to the population, when the same evaluation process is used for both. Deming emphasizes that although the sampling expert may have become familiar with the substantive field and may have given good advice about other

aspects of the study, his *professional* expertise is limited, and he should testify only within that area.

Following in Deming's footsteps, on a number of occasions I have assisted in the sampling of railroad traffic, in connection with studies of the effects of a merger between two railroads or of the effects on railroad A of the abandonment of certain lines by railroad B. I have presented such work before administrative law judges of the Interstate Commerce Commission and have often been cross-examined thereon. However, despite strong controversy and aggressive examination of management personnel, the sampling testimony has generally been accepted with minimum fuss, and the cross-examination has ordinarily consisted solely of emphasizing the limits of my responsibility.

Sampling testimony is not always so cut and dried, however. An early case is that of *Sears Roebuck & Co. v. City of Inglewood* (Sprowls 1957), in which the sales to nonresidents had been erroneously subject to tax and Sears was seeking recovery. The expert retained by Sears sampled 33 days from the 826 business days in the period at issue, and all sales slips from each of those 33 days were examined and assessed. The estimated overpayment was $27,000, subject to a standard error of $2,000. No quarrel with the method of sampling was made, but the judge in the case was uneasy about this unfamiliar technique. He ruled that no recovery could be made for individual sales that had not themselves been individually examined, and Sears had to go back and look at each sales slip. The result is a choice teaching example, of course, because it is one of those exceedingly rare cases in which a well-drawn sample is followed by a complete census, illustrating the ultimate validation of the statistician's art. In this case the complete count was surprisingly close to the estimate—a deviation of less than $300.

But even in the clean world of scientific sampling, a witness may find himself in difficulty. He may be asked to comment on the sample drawn by the opposite party—perhaps by a nonstatistician—and it may turn out to have been a systematic sample, without randomization of any kind. And here a prudent witness has a problem. The failure to randomize opens the way to possible biases, but as all experienced in sampling are aware, for a great many sampling frames (i.e., those with very little internal structure) the bias in the estimate and even in the calculated standard error is not likely to be large. Should one testify that the job was not competently done and the results should therefore not be given credence? (Counsel for one's own side would believe such testimony is entirely proper and the least that is owed him.) Or should one testify that, although the sample does not adhere to the canons, it is not likely that the result is for that reason wide of the mark? One will be tempted to add that when incompetence is manifest in the visible part of the operation, it is suspect in that which is less visible, and therefore the result should be received with caution. Should one yield to that particular temptation, however, he is most likely to be cut off by an objection from the opposing attorney, protesting testimony that is "mere speculation." Since the court is only interested in the result of your judgment about design, one is likely to either overstate the objections to the systematic sample or so understate them as to make one's client wonder why he was put to the trouble of drawing a random sample in his case. (And what should one say when the sample one is called upon to criticize was drawn by Dr. Deming or Professor Cochran, who—in the exercise of professional judgment—decided that the fuss required to randomize was not worth the trouble in the case at hand? Indeed, Cochran was fond of telling of the occasion when he was called on to carry out a sampling study of, I believe, a class of retail stores, and he instructed that the sample consist of every tenth establishment of that type listed in the Yellow Pages. The judge, he said, welcomed his expert testimony as a learning experience and remarked, after Cochran had been sworn, "I am glad to hear and to learn from Professor Cochran about this scientific sampling business, because I know virtually nothing about it. In fact, about the only thing I *do* know is that you should not just start at the beginning and take every 10th one after that.") I confess that, not being Deming or Cochran, I make it a point when drawing a sample to be sure that the design has as much internal credibility as I can give it, and as little dependence on the quality of my own judgment as I can manage.

Paternity and Fingerprints

In the sampling domain, statistical inference works well because we impose the probability model directly on the situation—through randomization—and our testimony has both the appearance and the substance of relative objectivity. We can feel rather sanguine about our contributions to legal proceedings in this domain.

We have somewhat less security when we turn to certain areas of identification evidence. I refer to blood tests for assessing evidence of paternity and to fingerprint evidence. In the former, at any rate, there is a probability element introduced by Mendelian genetics, and the statistical expert may have a real contribution to make. Unfortunately, the ultimate probability calculation depends on population gene frequencies; even where these are known for the population at large, it is often some subset of the population that is at issue, for which the frequency is not well established, and the expert finds himself on doubly uncertain ground. There are controversies aplenty in this domain, but this is not where most of the action lies.

Observational Data

The broad, almost limitless, domain in which the courts have come more and more to look to statisticians for guidance is in the analysis of observational data.

Consider, for example, the association between cigarette smoking and the subsequent development of lung cancer. First identified as an incidental and highly uncertain association, the accumulated evidence today appears overwhelming, although there are no clinical experiments to support it and only indirect support from cellular biology. The primary evidence is indeed statistical, and it *is* convincing, but not as a result of conventional significance testing. Rather it is the robustness of that association over time, place, and population that is convincing. For the most part, probability-based statistical testing is irrelevant to the strength of our conviction.

The apparent negative association between regular aspirin ingestion and heart attacks is a similar problem with a far more ambiguous outcome. An incidental association reported by a Mississippi clinician appeared to be confirmed by two large hospital-based case-control studies. The association is not of

the magnitude of that between smoking and lung cancer, but it appeared to be substantial nonetheless. In this case, a number of clinical experiments have been carried out, and at present we must acknowledge that the degree of protection from heart attacks afforded by aspirin is at best modest, if there is any at all. Again, the probability-based significance level attained by the early case-control studies have no strong bearing on the believability of the aspirin hypothesis.

I do not quite share David Freedman's hard-line position against formal statistical inference for observational data. As explained in a superb elementary textbook (Freedman, Pisani, and Purves 1978), Freedman regards probability-based testing in a situation without a plausible probability model to be at best irrelevant and more likely misleading. I think, in contrast, that such testing serves a useful purpose as a benchmark: If the observed association would not be counted statistically significant had it arisen from a randomized study, it could not' be counted as persuasive, when even that foundation is lacking. If the observed association is highly statistically significant, however, the extent of its persuasiveness depends on many uncertain judgments about background factors, and its persuasive value is not at all reflected in the significance level itself.

But the principles of statistical inference relevant to the courts are not the province of the statistics profession alone.

3. THE COURTS, CIVIL RIGHTS, AND STATISTICS

The Supreme Court has canonized formal statistical inference in a series of decisions beginning with a jury discrimination case, Castaneda v. Partida, decided in March 1977. Having noted that the population of Hidalgo County was 79% Spanish surnamed, but that the jury panels selected in accordance with the prevailing Texas "key man" system averaged only 39% Spanish surnamed (i.e., 339 of 870 jurors), the Supreme Court itself—more likely one of the Justices' law clerks—calculated the familiar critical ratio according to the binomial distribution—that is, the difference (39% minus 79%, or 40%) divided by the standard error (root pq divided by n, which works out to 1.5%), obtaining a critical ratio of 29. The Court then commented, "as a general rule for such large samples, if the difference between the expected value and the observed number is greater than 2 or 3 standard deviations, then the hypothesis that the jury drawing was random would be suspect to a social scientist."

Formal significance testing next appears in an employment discrimination case, Hazelwood School District v. United States (Meier, Sacks, and Zabell 1984), decided three months later. In that case, the proportion of qualified teachers in St. Louis County (excluding the city of St. Louis) who were black was estimated to be 6%, and during the two-year period at issue, only 15 of 405, or 4%, were black. The Hazelwood court now says, "A precise method of measuring the significance of such statistical disparities was explained in 'Castaneda v. Partida' . . . ," and the opinion goes on to paraphrase the earlier two or three standard deviation rule, but with a slight shift; that is, ". . . if the difference exceeds 2 or 3 standard deviations, then the hypothesis that teachers were hired without regard to race would be suspect" (emphasis added). The reference to randomness is now absent, as is the social scientist. Since it is self-evident that the process of selection is not, nor is it desirable

that it be, random, it is far from clear why either the social scientist or the Supreme Court should look upon a standard based on randomness as appropriate to assess the likelihood of purposeful discrimination. To be sure, there was much other evidence in the case, showing explicit discrimination at earlier dates, but the preceding quotation is the only place in the opinion where the relevance of the statistical significance test is in any way explained. Nonetheless, in Hazelwood the court went further, in an obscure remark that pointed clearly to the preeminence of statistics. It said, "Where gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination." Thus, in the space of less than half a year, the Supreme Court had moved from the traditional legal disdain for statistical proof to a strong endorsement of it as being capable, on its own, of establishing a prima facie case against a defendant. (It is sometimes argued that the use of doubtful evidence to support a prima facie—that is, preliminary—finding is a matter of small legal consequence. Such a finding merely shifts the burden of proof from the plaintiff to the defendant. In fact, however, there is nothing at all "mere" about this shifting of the burden, since the difficulty of proving oneself innocent of discrimination turns out to be great indeed.)

The accelerating role of statisticians in employment discrimination cases arises from a combination of the statistical significance testing endorsed in Hazelwood with an earlier decision (Griggs v. Duke Power Company) in 1971. In Griggs, it was found that the requirements of a high school diploma and a certain score on a standardized IQ test for employment in such jobs as maintenance and laboratory work operated to exclude black applicants far more frequently than they did to exclude white applicants. The court concluded that, in the absence of direct evidence that these criteria related to improved performance on the job, the "adverse impact" of those requirements constituted a violation of Title VII, even though there may have been no intent to discriminate on the grounds of race. (To be sure, there was plenty of evidence of intent in the Griggs case, most especially in the facts that the power company had explicitly excluded blacks prior to passage of the Civil Rights Act and that it had put in the new requirements at the same time that the jobs were first made available to blacks. The principle established in Griggs clearly put the issue of intent aside, however, and the doctrine has been widely applied by lower courts in cases in which there was little or no evidence of invidious intent. Once again, proof of "job relatedness" or, as the Supreme Court says, "business necessity," has proved generally elusive, and a great many employment standards and admissions tests have been found to be in violation of the law for lack of such proof.)

The criterion seems reasonable enough until we are faced, as was the Supreme Court, with a case like Washington v. Davis (1976). In this case, Walter Washington, mayor of Washington, D.C., and his police chief had set about to recruit blacks into the D.C. police force, with an aggressive campaign to encourage black applications. The campaign was successful and many blacks were recruited, but among the newly encouraged applicants the written test "operated," in the language of the Griggs decision, "to disqualify Negroes at a substantially higher rate than White applicants." Thus, affirmative action, clearly

intended to recruit blacks, fell afoul of the adverse impact principle developed in Griggs. The trial court dismissed the charge, but the appeals court reversed, citing Griggs. The Supreme Court side-stepped the issue. It supported the District program, but it did so on a technicality that did not require it to comment on the general validity of the Griggs principle. In subsequent cases the Griggs principle has continued to guide the lower courts.

It must be acknowledged that there has been some slackening of the tide in Title VII enforcement in the last two or three years, especially in regards to race and sex discrimination. Under the Reagan administration there has been more emphasis on freeing business from government interference and less on righting the wrongs of the oppressed. In response, the courts appear to give somewhat less weight than before to purely statistical evidence.

4. POSITION OF THE STATISTICAL EXPERT

The result of the preceding and related decisions has been to place the statistical expert witness in a most unaccustomed and exalted position. Despite the moderate recent decline in enforcement, the role of the statistical expert remains critical in the cases that are brought. Lawyers gaze with awe as he examines the entrails of complex multiple regression computer output, and they await breathlessly his conclusion that the coefficient of the variable designating sex is indeed more than twice the standard error. Similar attention attends his calculation of continuity-corrected 2×2 chi squares to see whether they are larger or smaller than 3.84. Indeed, largely on these outcomes the case may be won or lost.

That this position is a false one, none can doubt. Certainly the statistical experts know it, and most of them say so to some extent—or at least they assert that the meaningfulness of their numerical results depends on a number of assumptions that they are unable to verify. The courts, however, are not engaged in academic exercises and, having urgent need to come to some conclusion, turn to the Supreme Court instead of to the witness's cautionary phrases for guidance. Those opinions have, intentionally, a somewhat Delphic quality. They tell us that "gross disparity" in pass rates is evidence of illegal discrimination, and they also tell us that the hypothesis of random selection is made to appear doubtful when a difference is larger than two or three standard errors. They do not quite say that statistical significance at the 5% level constitutes gross disparity, but that is how the lower courts read them. The statistical experts cannot help but find this heady stuff, and we should not be surprised to find ourselves speaking with far more assurance about our conclusions than an objective appraisal of the evidence might warrant.

Thus we are led to the unedifying spectacle of two well-qualified statistical witnesses providing analyses that they interpret oppositely, each supporting the interest of the party who introduces him. Other categories of expert witnesses have been there before us, of course—the psychiatrists, medical internists and surgeons, and structural engineers, among others. The courts have urgent need for the assistance of these experts, but they seem uncommonly ill served by them. The point was made clearly in an editorial in the *British Medical Journal:*

Medical evidence delivered in our courts of law has of late become a public scandal and a professional dishonour. The Bar delights to sneer at and ridicule

it; the judge on the bench solemnly rebukes it; and the public stand by in amazement; and honourably minded members of our profession are ashamed of it. (*British Medical Journal* 1863)

This was printed more than a century ago, but little has changed in the intervening years. Statisticians have escaped comparable condemnation because we have been, until recently, too unimportant in the courts to be noticed, not because of any higher ethical standard of our profession.

One cannot help speculating on the possibility of improvement. In fact, I believe that some of the difficulty is structural and that there are ways in which we could function usefully in legal settings without so large a sacrifice of professional integrity. To this end I now discuss the players in the game and the key influences on them. Among the players or participants in the legal ballet, I distinguish three: (a) the courts themselves, most especially the Supreme Court, who together with the Congress set the rules by which the system operates; (b) the lawyers—collectively the bar—who primarily control the direction of play within those rules; and (c) the expert witnesses on whose performance the integrity of the enterprise ultimately depends.

The Courts and the Expert

The obvious objective of the courts in respect to expert testimony is to optimize the search for truth. The courts would like to get the most well-qualified expert, keep him in a situation in which he can devote his best efforts to analyzing the evidence, and have him testify in an atmosphere free of coercion or bias. The courts also want to be sure that the expert is adequately examined to test and verify his qualifications, the adequacy of his preparation, and his objectivity.

To this end, the courts in Germany and France arrange matters very differently from the English and American courts. In cases in which experts are needed, they are in the first instance appointed by and responsible to the court and not to either party. They are first examined by one of the judges and also cross-examined by him. Attorneys for the plaintiff and defendant may also cross-examine, but the proceedings are not generally adversarial as are our own, and the appearance of neutrality, at least, is the rule. Thus the continental system seeks the best witnesses and seeks to put them in a neutral setting, primarily by putting the major responsibilities in the hands of the judges.

The Anglo-American system, in contrast, is based on the proposition that truth is most likely to emerge through the best efforts of adversaries. No point in favor of the defendant will be overlooked or undervalued, it is thought, if responsibility for bringing it out is assigned to the defendant's advocate.

Nonetheless, whatever the merits of the adversary system may be in general, it is well recognized that it wreaks havoc with expert testimony, and proposals for reform appear regularly. Chief among them is to borrow from the continental system and to have the primary expert witnesses appointed by and responsible to the court. This reform was vigorously advocated by last century's revered commentator on legal evidence, John Wigmore (1940), and model codes have been proposed to this end. Indeed, Rule 706 of the Federal Rules of Evidence (Title 28, U.S. Code, annotated) provides explicitly for court-appointed experts. Regardless of the merits, in practice this power is used extremely sparingly. (There may be some cases of court-appointed statisticians in Title VII cases,

but I have not heard of any.) One can conceive of many reasons for the ineffectiveness of these "reforms," not least the vulnerability to criticism of a judge who appoints an expert later shown to be inadequate, but it is enough for my purposes to observe that such reforms have not taken hold in this country and that they do not seem likely to become influential in the near future.

The Bar and the Expert

The position of legal counsel, although in principle identical with that of the judge, is in fact quite different. Having committed himself to the adversary system as the best method of reaching a just conclusion, the lawyer for the plaintiff now accepts his position in the system, that of advocate, and leaves to the court the responsibility for discerning the path of justice. To him the expert is simply one of the elements that he must fit into place to make the most effective case. To be sure, any lawyer of competence recognizes that it is usually favorable to his case for the witness to appear to be dispassionate and objective. The best lawyers recognize that a witness will make the best appearance of objectivity if he feels that he is indeed free to go where his research and reflection lead him. This is not to say that these excellent advocates are really in the market for unbiased witnesses who may testify to their side's disadvantage.

John C. Shepherd of St. Louis, a distinguished trial lawyer who was president of the American Bar Association in 1984–1985, spoke to a conference for lawyers on relations with the expert witness (Shepherd 1973), and this is what he said:

> Many people are convinced that the expert who really persuades a jury is the independent, objective, nonarticulate type . . . . I disagree. I would go into a lawsuit with an objective, uncommitted, independent expert about as willingly as I would occupy a foxhole with a couple of non-combatant soldiers.
>
> If you find the expert you choose is independent and not firmly committed to your theory of the case, be cautious about putting him on the stand. You cannot be sure of his answers on cross-examination. When I put an expert on the stand, he is going to know which side we are on.
>
> The trial lawyer must make of the expert a convincing, persuasive witness. The lawyer deals in words, and he knows how to put the package together to impress the jury favorably. It is his job to instruct the expert, an exercise requiring great tact and firm conviction. (pp. 21–22)

Keep in mind that the lawyer does not need to make bricks without straw. It is perfectly proper for him to consult a great many potential witnesses but to bring to court only that one whose honest convictions fit well with the lawyer's needs. The phenomenon of "shopping for witnesses" is well recognized by the courts, and it contributes to the wary attitude they have about experts in general. The shopping is done by the lawyers, however, and is thus not subject to exposure in the actual testimony.

Corrupting Influences

As we have just seen, the professional integrity of the expert witness and, through him, of the profession that he represents is not well protected by the courts and hardly at all by counsel. But before we assume too readily that simple morality and personal ethics will be an adequate substitute, we should reflect for a bit on what I call, for lack of a more delicate phrase, *corrupting influences*. Some are inherent in the nature of the situation, and others are special to the adversary situation.

First, there is the fact that the expert witness is playing someone else's game and, inevitably, has to accept the rules as he finds them. His instructor in these matters is, of course, his client's counsel, and the witness is ill-equipped to resist the role of adversary when his lawyer thrusts it upon him. But even supposing the lawyer is less demanding than Shepherd, the expert is beset with temptations.

General

Among the most difficult of the corrupting influences to deal with is what I call *aggrandizement*. In Title VII cases (i.e., those dealing with employment discrimination), the Supreme Court has placed the statistician in the key role. Long ignored and treated with contempt in literature and in the courts, the statistician has been elevated to Olympian levels. Thus the Hazelwood court, quoting its remark in an earlier case, commented:

> We also noted that statistics can be an important source of proof in employment discrimination cases, since, "absent explanation, it is ordinarily to be expected that nondiscriminating hiring practices will in time result in the work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired." Evidence of long lasting and gross disparity between the composition of the work force and that of the general population thus may be significant even though paragraph 703 (j) makes clear that a work force need not mirror the general population. (*Hazelwood School District v. United States* 1977)

Taken together with the court's embrace of statistical significance testing, the statistician is here given a virtual license for intellectual robbery. Indeed, not only the court but a large contingent of fellow academics (economists numerous among them) give strong endorsement to the particularly magical properties of multiple regression analysis. [Two articles in the *Columbia Law Review*—Fisher (1980) and Finkelstein (1980)—are noteworthy in this regard.] All in all, the statistician is strongly tempted to give the definitive rather than a qualified answer to the key questions. He will be tempted to ignore or to minimize those qualifications that he might emphasize in a more academic setting, he may fail to emphasize the existence of schools of thought other than his own, and he may lay claim to overly broad scope for the inferences he draws.

Adversarial

The adversary system adds a host of additional influences, some quite direct, but others indirect:

1. *Bribery.* The witness is paid by his client and, as often noted, he who pays the piper feels a right to call the tune. To be sure, all the client is entitled to is an honest report of the expert's best effort, but an expert who habitually finds evidence against his client will not be much sought after.

2. *Flattery.* Some, of course, are not bought by money or the prospect of future money—either they already have enough of it or they are sufficiently on guard against that particular type of seduction. Other corruptions await them.

I well recall an occasion on which I was asked to consult in a case at a time that was not especially convenient. I explained that I really could not participate on this occasion. The lawyer, with whom I had worked before and for whom I had a great deal of respect, pled the sorry state of statistical testimony in the courts in general, and in the instant case in particular. He read from the transcript some particularly egregious quotes from the statistical expert for the other side, and he urged the importance for the future of statistics in the domain of public affairs of having corrective testimony. That being a viewpoint

I could only share, and tacitly mindful of our shared opinion that I was the ideal candidate to champion the honor of the profession, I reluctantly agreed to testify. Imagine my chagrin when, at a later date, I read some other remarks of the trial lawyer, John Shepherd, whom I quoted earlier. He advises on "Approaching the Expert" as follows:

Almost every one who considers the subject of experts in court will start with the same thought: The first thing you need to get along with your expert witness is money. But the hiring and successful use of an expert may not be that easy—alot of good experts are rich. Although you will eventually be talking about money with your expert, it is wiser to begin on another tack. Tell your expert how justice will be served if he will testify on your side of the case. Remind him that the unfortunate situation in our courts today can be improved if we have people of his caliber to help in the administration of justice. That ploy will impress even the rich expert. (Shepherd 1973, p. 19)

3. *Co-option.* To be sure, effective as this ploy may be, it does not in itself lead the expert away from his duty. It establishes an aura of objectivity and mutual respect, however, which may make the expert especially vulnerable to another inevitably corrupting aspect of the adversary system. That is the simple fact that the expert's introduction to the case comes from the client's counsel and will inevitably tend to appear in the light most favorable to the client. He will be introduced to the principals—perhaps a plaintiff, movingly indignant about years of abuse and low pay, perhaps a defendant who truly believes that his cause is just and is worried sick about the distraction of his institutional resources from their proper role into the defense against a baseless charge. This goes along with co-option into advocacy arising when one is asked to review the other side's testimony, point out flaws therein, and assist in the development of effective cross-examination.

F. Downton of the University of Birmingham has written cogently about this latter difficulty in a symposium on statistics and the law (Downton 1977). Downton had been consulting with the police on games of chance, because the law prescribed strict rules for games that, if violated, would allow the police to close the clubs. Since the clubs were widely regarded as dens of iniquity, this was clearly a public service. Downton wrote:

As in any other consulting situation, a certain amount of identification with the aims of the client is inevitable; it is fortunate that probability and statistics are basically mathematical in content, since the constraints of mathematics act as a brake on overenthusiasm. It cannot, however, be denied that a conscious change of attitude was needed to effect the change-over from helpful consultant to objective expert witness. . . . This ambiguity of roles did create a conflict, which presumably can only be resolved by individual witnesses in their own way. (p. 171)

4. *Gladiatorial Role.* The adversarial environment works against objectivity in yet other ways. The object of cross-examination is not only to expose weaknesses in the expert's analysis but, if possible, to discredit the witness and the weight that should be given to his testimony generally. Thus the cross-examiner may, by adroit framing of questions, force the witness into complex explanations and apparent contradictions. Feeling his credibility slipping away, such a witness may be less likely to give a full and frank answer to a later question that might fairly expose a fact or conclusion operating in favor of the other side. The expert no longer views his interrogator as a fellow searcher for truth, but as an adversary against whom he must defend.

5. *Personal Views.* My final source of corruption is perhaps the most difficult to deal with, and that is the problem of strongly held personal views. Surely there are many cases in which the expert is a priori indifferent between the claims of the contestants, but in other areas, particularly in the great domain opened up by Title VII, there are few of us without strong opinions. In the matter of a contest between a chemical waste disposal company and the residents of a new Love Canal, for example, I would be reluctant to testify on behalf of the company. It might well be, for example, that the evidence of adverse health effects caused by carelessly buried wastes is really nonexistent. Feeling as strongly as I do, however, that such careless behavior is reprehensible and deserving of punishment, I should not like to assist the company's case. I have no problem reconciling my preferences and my professional responsibilities in this case. I am, and should be, free to accept an engagement or not, for whatever personal reason, and reasons of this kind are at least as good as most others.

My problem comes on the other side. Suppose I should be an expert retained by the residents affected by the dump. I find that, in respect of total mortality, there is no evidence of an effect, but in the matter of childhood leukemias there is an excess mortality amounting to 1.8 standard errors greater than the rate in some control group. If I ignore the fact that I am reporting on leukemia because it is the disease category showing the largest difference, and if I adopt the conventional 5% significance level as a standard, and if I urge the relevance here of a one-sided significance test, I may be able to strike a blow for truth and justice, and it would no doubt be tempting to do so. But to paraphrase a major figure in the Watergate investigation, "I could do that, but it would be wrong." I really do not think one-sided 5% level deviations provide convincing evidence one way or the other, and whatever one's views on that, I expect most statisticians would agree with me that it is misleading to the point of dishonesty to quote an unadjusted significance level when I have chosen to present the most extreme of a number of alternative measures.

Perhaps the point can be brought home most forcefully by addressing an even touchier example. There are many of us who view the legacy of slavery as our most appalling and pressing social problem and the effort to explain the low status of the oppressed on the grounds of inherent inferiority as an intolerable offense. Indeed, although the possibility of *some* average difference in intellectual capacity among different groups can never be ruled out, the evidence appears clear that whatever differences there might be in *average* innate ability, they are quite small compared with the variation between individuals. The effects ascribed to race in regression analyses of schoolchild performance, after adjustment for age, years of schooling, mother's socioeconomic status, and the like, are readily explainable as attenuation and other distracting effects that afflict regression analyses generally, and they need not be interpreted as reflecting a real difference due to race.

At the same time, we observe the past and present systematic discrimination against blacks in many areas of employment. Such discrimination has many forms, but its pervasiveness, except where sharply controlled by law, is hardly in doubt. Being confident, then, that a charge of race discrimination likely corresponds to the existence of actual discrimination, what are we to say about a multiple regression in which a salary difference unfavorable to blacks emerges as significant even after adjustment for age, years of schooling, mother's socioeconomic

status, and the like? The problems of attenuation apply with equal force, but we may now be reluctant to dismiss the evidence of bias in pay. This time we may believe that there really is discrimination in the system, but it is by no means clear why we should, as statisticians, take different positions in the two situations.

5. WAYS TO DEFEND THE INTEGRITY OF STATISTICAL TESTIMONY

With the variety of assaults on the credibility of expert statistical testimony, I turn again to the question of what possible defensive measures could be implemented. A change to the apparently more neutral continental system is the one answer that has come from the courts, but there seems little likelihood of its adoption. There have been proposals that experts who testify falsely should be punished for perjury, as is an ordinary witness who testifies falsely about an event. This, too, seems far-fetched, since the essential nature of expert testimony is that it is largely a matter of informed opinion.

Professional Codes

There seems to be only one other direction in which to turn, and I have only slender hopes for it. Seeing that neither the bench nor the bar will help us, the only alternative is to help ourselves—that is, to develop limited codes of ethical behavior in the context of legal proceedings that may help to ameliorate the worst excesses.

I come to this conclusion reluctantly, because I have little taste for collective moral instruction and little confidence in its efficacy in general. And yet one cannot deny that codes of ethics for judges, while they do not eliminate venality, are good to have. Violation of these rules can and does lead to discipline, on occasion, as in the case of a distinguished Supreme Court Justice a few years ago, and reminders such as that help to keep others on the right path. Similarly, codes of medical ethics dealing with the proper relationship between physician and patient serve a useful purpose.

A decade ago, Gibbons (1973) reviewed our society's efforts in the direction of developing such codes, and she clearly laid out some of the problems that such codes might help to solve. Evidently there was some movement in that direction in the early 1950s, but momentum was lost, and nothing came of it. The issue of ethical codes continues to elicit debate, a recent instance being the report of the Ad Hoc Committee on Professional Ethics (1983). I see no sign, however, that this or any other code is likely to be adopted as a guide by any of our major professional organizations. Indeed, although discussion of codes of ethics for statisticians continues, I know of only one instance in which such a discussion has had any noticeable practical effect.

The exception is an interesting one, and it may be instructive. As I mentioned earlier, in the context of sample survey design and analysis, W. Edwards Deming established a code that he provided to clients, explaining the reach and the limitation of his methods. The code is notable for its careful restriction of the role of the statistician. In effect, Deming acknowledges that the statistical consultant may come to have a good deal of knowledge about the subject matter under study and that this knowledge may help him to design an effective sample. He makes clear, however, that *responsibility* for the choice of population to be sampled, and for the processing of each sampled unit, belongs to the client and not to the sampling consultant. The consultant undertakes to say only that, had the entire population been processed in the same way that the sampled elements were processed, the sample estimate for the population would be found to be close to the population value, subject to error limits that can be given in the usual probability sense. It might seem that the scope of the sampling expert's testimony is so narrow according to this code that his contribution will have little weight in the proceedings. In fact, of course, the contrary is true. By not reaching beyond well-stated boundaries, the testimony of sampling experts has achieved an enviable level of credibility.

Deming's code, effective as it is in a specific context, gives only a little guidance for the expert testifying in a Title VII case. I submit that a proper code for the latter expert should copy Deming by being specific to the situation and rather restrictive as to the scope of the testimony. I do not think it will pay to start with an ethical code trying to embrace all statistical activities. Let me try to clarify my proposal by being specific. (Here I borrow from Deming where I can.)

I suggest that a statistican asked to testify in court should require that he be given access to all data thought by the client to be relevant and all previous analyses of that data, and he should demand a commitment on the part of the client to a *"good faith" effort to supply whatever other data the statistician may judge relevant.*

He should advise the client that in his professional role he will remain neutral between the parties (and he should pray for strength when he does this, for he will need it). He undertakes to provide his best effort to analyze the data in ways that seem to him pertinent, and he undertakes further to provide a written report. His report, if it is to be used, must be taken in its entirety.

When testifying, the expert will explain the limitations of his techniques, as seen by a professional statistician, regardless of any statistical principles annointed by the Supreme Court. He will explain the variety of schools of thought within the profession and his place among them.

Doubtless there are a number of other principles to be enunciated, but this is not the time or place for full details. Some will think the principles given are simple and obvious, but they can be assured that they are not obvious to most lawyers. Many lawyers, for example, think it proper to select for attention the principles laid down by the high court as a basis for expert testimony. (I cannot help but wonder if, should a court declare that $\pi = 3$, these lawyers would insist that we accept that too.) Gratuitous testimony about limitations will be especially unwelcome (and in my opinion, especially necessary).

The point is that the expert should be much more his own man and much less the puppet of his client's counsel than is typically the case today.

Consequences

The consequences of such a code, should we adopt and use it, are substantial and not entirely welcome. I do not believe it would help us much in protecting against the influence of our own strongly held social views or against the biases that arise

because we are oriented and informed by just one of the adversaries. Nor would it keep us from reflexive defensiveness under hostile cross-examination.

Adherence to such a code is likely to result in a reduction of the pivotal role that statistical analysis has come to play in discrimination law, and we may see the resulting gap filled by others whose competence and good will we question even more than our own. It is conceivable that a more modest posture might lead the courts to seek greater clarity by adopting the reforms, if such they be, of the continental system with court-appointed experts. It is certain, I think, that adherence to such a code would improve the credibility of statistical witnesses.

6. CONCLUSION

The sorry state of expert testimony might lead one to conclude that no honest person should participate in such a scam, but I feel strongly that this conclusion is wrong. Querulous as this discussion has been about the workings of our legal system, it is, after all, the only one we have and, as the courts are fond of proclaiming in their own self-serving way, every citizen owes his best efforts to serve and to improve the system that ultimately is the protector of those freedoms he cherishes. (Shepherd's appeal to the reluctant witness on these grounds is one that *he* views as a ploy, and for him it is exactly that. It nonetheless expresses an important and commendable sentiment.)

In closing I turn to White's (1964) delightful summary of rude comments about statistics and statisticians. After discussing the familiar aphorisms—lies, damned lies, and whatever, many of which canards White then thought to be reasonably apt—he concludes: "Time passes; the new books are written by other kinds of men, and only an occasional author, poet, or politician, or critic fires a blast at the Statisticians. Who knows what the outcome of all this acceptance is going to be?" (p. 17). Since that was written, acceptance in the courts has grown mightily. As matters now stand, prospects for the outcome of all this acceptance are ominous, not to say grim.

Perhaps we can show ourselves to be wiser and more disciplined than the professions that have been before us. We should at least try.

REFERENCES

Ad Hoc Committee on Professional Ethics (1983), "Ethical Guidelines for Statistical Practice," *The American Statistician*, 37, 1–20.

British Medical Journal (1863), "Medical Evidence in Courts of Law" (editorial), May 2, pp. 456–457.

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

DeGroot, M., Fienberg, S., and Kadane, J. B. (eds.) (in press), *Statistics and the Law*, New York: John Wiley.

Deming, W. Edwards (1954), "On the Presentation of the Results of Sample Surveys as Legal Evidence," *Journal of the American Statistical Association*, 49, 814–825.

Downton, F. (1977), "Experience as an Expert Witness in Gambling Cases," *The Statistician*, 26, pp. 163–172.

Finkelstein, Michael O. (1980), "The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases," *Columbia Law Review*, 80, 737–754.

Fisher, Franklin (1980), "Multiple Regression in Legal Proceedings," *Columbia Law Review*, 80, 702–736.

Freedman, David, Pisani, Robert, and Purves, Roger (1978), *Statistics*, New York: W. W. Norton.

Gibbons, Jean D. (1973), "A Question of Ethics," *The American Statistician*, 27, 72–76.

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

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

Meier, Paul, Sacks, Jerome, and Zabell, Sandy (1984), "What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule," *American Bar Foundation Research Journal*, Winter, 139–186.

Meier, Paul, and Zabell, Sandy (1980), "Benjamin Peirce and the Howland Will," *Journal of the American Statistical Association*, 75, 497–506.

Monahan, John, and Walker, Laura (1985), *Social Sciences in Law: Cases and Materials*, Mineola, NY: Foundation Press.

People v. Collins (1968), 68 Cal. 2d 319, 66 Cal. Rptr. 497.

Peterson, Donald W. (1983), "Statistical Inference in Litigation," *Law and Contemporary Problems*, 46, No. 4, 1–303.

Shepherd, John C. (1973), "Relations With the Expert Witness," in *Experts in Litigation*, ed. Grace W. Holmes, Ann Arbor, MI: Institute of Continuing Legal Education.

Sprowls, R. Clay (1957), "The Admissibility of Sample Data Into a Court of Law: A Case History," *UCLA Law Review*, 54, 222–232.

Washington v. Davis (1976), 433 U.S. 229.

Wigmore, John Henry (1940), *Evidence in Trials at Common Law*, Boston: Little, Brown.

White, Colin (1964), "Unkind Cuts at Statisticians," *The American Statistician*, 18, No. 5, 15–17.