Rejoinder

Mary W. Gray

Professor Roberts may be correct in his assessment of our disparate views on the potential for using statistics to "prove" the existence of discrimination. Fortunately for plaintiffs, the standard required in civil cases is not the "beyond reasonable doubt" to which he refers. Instead, all that is required is that a preponderance of the evidence shows that it is more likely than not that discrimination has occurred. However, I thoroughly agree with his analysis of where the roots of the problem, if there is one, may lie—namely in initial placements, a factor also highlighted by Professor Gastwirth.

In looking at regression analyses of faculty salaries and underlying data at more than three dozen institutions—at the behest of faculty or of administration, in litigation or in the hope of avoiding litigation, I have found that gender-based differences in salaries almost always relate in whole or in part to placement policies. Historically the difficulty arose from disparate starting salaries for the same position; whereas such practices have become less prevalent, the disparity in starting *ranks* for similarly qualified men and women continues to be a problem, particularly for those who come with several years of post-doctoral teaching experience and a research record that could conceivably qualify them either for a senior (untenured) assistant professorship or a (possibly tenured) associate professorship. Clearly, those men or women offered the latter will have an essentially insurmountable lead in salary even though the rank disparity may eventually disappear; more importantly, those offered the lower rank may lose out at the tenure-decision stage, sometimes because standards have increased, sometimes because of financial problems at the institution, and sometimes through sheer perversity or bad luck. When women are disproportionately in this group, subsequent salary differences as well as the paucity of women at upper ranks are easily explained.

One reason for this disparate placement relates to Professor Roberts' interesting discussion of homogeneous job groupings. He mentions the possibility of overqualified women ending up, for whatever reason, in the clerical category. In academe the equivalent is the "nontenure-track" or temporary position. Approximately 20% of all full-time women faculty, as compared with less than 8% of all full-time men faculty are in such positions (AAUP Salary Survey, 1992). That women are more likely to have had prior experience in temporary positions is one reason cited for the disparities in rank placement discussed above.

Women also disproportionately occupy part-time positions, some by choice, but more because they cannot get full-time academic employment. The salaries, as well as job security, for all of these "irregular" positions are simply not comparable with the salaries of full-time tenure-track and tenured faculty. In fact, as I noted, faculty occupying such positions are usually excluded from regression analyses of salaries, although in one case plaintiffs attempted unsuccessfully to prove that the disproportionate tracking of women into these positions in itself constituted disparate-impact discrimination (*Griffin v. Board of Regents,* 1986).

The pay-for-knowledge plans mentioned by Professor Roberts are becoming more widespread as employers come to value flexibility in the workplace. I am more sanguine than he about the application of a variant in academe, especially with the worry about stagnation in light of the coming abolition of mandatory retirement. Of course, to one who has taught a broad range of undergraduate mathematics and statistics courses; graduate algebra, applied statistics and mathematics history courses; courses on programming and computer law; a women's studies course on politics; and a political science course on civil rights and civil liberties in the past few years, such a scheme has a great deal of appeal.

I have little faith in the use of matching studies as proposed by Professor Gastwirth, for I believe that they are even more subject to manipulation than are regressions. I do agree, however, that to be effective regression studies ought to be quite robust to reasonable changes in variables and additions or subtractions of specific data points. Moreover, Professor Roberts' assessment of the importance of anecdotal evidence, concurred in by Professor Gastwirth, is right on target. Bringing the numbers to life is crucial, no matter how good the statistical evidence and how open-minded the judge. This depends on having sympathetic plaintiffs; unfortunately many victims of discrimination, possibly as a result of having been victimized, are not effective on their own behalf. Moreover, even the best qualified, most presentable witness can be vulnerable to accusations from disgruntled students or colleagues or inappropriate comparisons. With the passage of the 1991 Civil Rights and Women's Equity in Employment Act effective witnesses will be even more crucial if the Act's provisions for jury trials are invoked. However, good case history testimony depends to some extent on the presence of a "smoking gun," which colleagues and administrators in this day and age are unlikely to provide. Providing an alternative means of redress, the

disparate-impact model of discrimination is based on the theory that if the effect of an action is in fact discriminatory we need not assess the motive behind it.

All in all, Professor Roberts' lessons provide excellent instruction. If only statisticians, lawyers and courts could be convinced!

Professor Gastwirth also sounds some appropriate warnings, including the possibility that statistical experts – and thus the court – may be unaware of the existence of data that do not support their conclusions. It is not uncommon for a party to litigation to hire one expert to do preliminary work to find out what data is supportive and then provide to the experts to be used at trial only what is favorable. Experts can, of course, inquire about the existence of additional data. In one recent tenure case, the defendant "accidentally" burned a substantial amount of subpoenaed evidence; eschewing sanctions on the defendant, the judge left the plaintiff severely crippled in proceeding with her case (*Jackson v. Harvard University*, 1989). In general, Gastwirth's emphasis on the need for reliable data is all too often ignored.

I found Professor Gastwirth's analyses of several recent cases, particularly of Judge Posner's views, very instructive and in fact reinforcing of my own view on the action of courts to the extent that I have never felt that faith in the efficacy of statistics is correlated in any way with political beliefs. I fear we disagree on the meaning of the *Ward's Cove* case, but the passage of the 1991 Civil Rights Act has in any event remade the landscape. Particularly useful is Professor Gastwirth's more detailed exposition of the issues involved in choosing between the "males-only" and the "total-population" regression models and his suggestions regarding the combining of data.

On the other hand, I have little confidence that Bayesian methods can be made understandable to judges or to juries. One of the problems we currently face is the superficiality of the use of statistics, particularly the mystique surrounding the use of a 0.05 significance level. In *Ottaviani* the judge totally discounted any descriptive statistics – even those with the inexorable zero denoting total exclusion – which is what led to the excluded analysis referred to by Professor Gastwirth. If we seek new approaches, perhaps exploratory data analysis, rather than Bayesian methods, will be our salvation.

Professor Conway perceptively notes that the problems encountered in constructing an appropriate regression model are interrelated; her specific examples are enlightening, particularly with respect to grouping and treatment of outliers. The question she raises of when enough is enough has, of course, no easy answer, but Professor Gastwirth's observation that the most effective model is one which is quite robust provides guidance. The goal is to create a model essentially impervious to the addition of other variables that might be reasonably thought to be influential and unbiased. Part of this process is, as Professor Conway suggests, to anticipate the rebuttal of the opponent.

Ideally one's best statistical judgment and knowledge of the actual situation being modeled need to be combined to give a sense of confidence that the model makes sense. Thus it is important that a statistical expert have a good understanding of the real-world process being studied, as well as of the underlying statistical concepts.

Part of the basis of the expert's judgment as to what feels right must be how the model can be explained to the finders of fact – judge or jury. Certainly the use of graphs and charts provides major assistance. I always think of the judge or jury as students in an elementary statistics course, students whose attention must be engaged and held. They need to be caught up in the explanation in such a way as to guarantee that they will make the effort to understand complex concepts with which they are probably uncomfortable. All the tricks of the teaching trade have to be employed, such as relating complicated concepts to more readily understood ideas from everyday life. One may still fail, particularly if the anecdotal evidence is less than compelling and/or the judge is not disposed to believe that statistics really "prove" anything that he or she does not already believe.

The cautions of Meier, Fisher, and Dempster, referenced by Conway, concerning the difficulties of reconciling limitations with the legal system's preference for definitive, unqualified judgments are well taken. Professor Conway's own exposition of the problems encountered with statistical evidence is very much on point; certainly the complexity of faculty compensation structures would make it difficult to make a convincing case even in a sympathetic forum. However, the fact remains that the traditional reluctance of courts to intrude into an academic setting has allowed colleges and universities license that most defendants are not accorded.

Finally, although each comment has provided useful cautions and helpful advice, neither my view that the courts are in general unlikely to rely on statistics that do not fit their view of the world, nor my belief that they should, has been altered.

ADDITIONAL REFERENCES

AAUP SALARY SURVEY. (1992). *Academe* 78 7–89.

BHATTACHARYA, P. K. (1989). Estimation of treatment main effect and treatment-covariate interaction in observational studies using band-width matching. Technical report, Div. Statistics, Univ. California, Davis.

BHATTACHARYA, P. K. and GASTWIRTH, J. L. (1989). Estimation of the odds-ratio in an observational study using band-width matching. Technical report, Div. Statistics, Univ. California, Davis.

BLOOM, D. E. and KILLINGSWORTH, M. R. (1982). Pay discrimination research and litigation: The use of regression. *Industrial Relations* 21 318–339.

CAIN, G. G. (1986). Labor market discrimination. In *Handbook of Labor Economics* (O. Ashenfelter and R. Layard, eds.) 1 chap. 13. North-Holland, Amsterdam.

CECIL, J. S., HANS, V. P. and WIGGINS, E. C. (1990). Citizen comprehension of difficult issues: Lessons from civil jury trials. *American University Law Review* 40 727–774.

CONWAY, D. A. and ROBERTS, H. V. (1983). Reverse regression, fairness and employment discrimination. *Journal of Business and Economic Statistics* 1 75–85.

CONWAY, D. A. and ROBERTS, H. V. (1984). Rejoinder to comments on "Reverse regression, fairness, and employment discrimination." *Journal of Business and Economic Statistics* 2 126–139.

CONWAY, D. A. and ROBERTS, H. V. (1986a). Regression analysis in employment discrimination cases. In *Statistics and the Law* (M. H. Groot, S. E. Fienberg and J. B. Kadane, eds.) 107–108. Wiley, New York.

CONWAY, D. A. and ROBERTS, H. V. (1985). Homogeneous job groups, candidate pools, and study of employment discrimination. In *Proceedings of the Survey Research Methods Section* 77–86 Amer. Statist. Assoc., Washington, DC.

DEMING, W. E. (1986). *Out of the Crisis.* Center for Advanced Engineering Study, MIT.

DEMPSTER, A. P. (1984). Alternative models for inferring employment discrimination from statistical data. In *W. G. Cochran's Impact on Statistics* (P. S. R. S. Rao and J. Sedransk, eds.) 309–330. Wiley, New York.

EHRENBERG, D. G. and SMITH, R. S. (1985). *Modern Labor Economics, Theory and Public Policy*, 2d ed., chap. 14. Scott, Foresman, Glenview, IL.

FIENBERG, S. E. and STRAF, M. L. (1991). Statistical evidence in the U.S. courts: An appraisal. *J. Roy. Statist. Soc. Ser. A* 154 49–60.

FISS, R. (1979). A theory of fair employment laws. *University of Chicago Law Review* 235 237–238.

GASTWIRTH, J. L. (1992a). Statistical reasoning in the legal setting. *Amer. Statist.* 46 55–69.

GASTWIRTH, J. L. (1992b). *Ward's Cove Packing Inc. v. Atonio:* How inadequate statistical evidence and reasoning contributes to questionable decisions. Technical report, Dept. Statistics, George Washington Univ.

GOLDBERGER, A. S. (1984). Redirecting reverse regression. *Journal of Business and Economic Statistics* 2 114–116.

GOLDBERGER, A. S. (1988). Comment on "Employment discrimination and statistical science." *Statist. Sci.* 3 165–166.

GOLDSTEIN R. (1989). Power and sample size via MS/PC-DOS computers. *Amer. Statist.* 43 253–260.

GRAYSON, C. J., JR. and ODELL, C. (1988). *American Business: A Two-Minute Warning.* The Free Press, New York.

GREENHOUSE, S. W. (1982). A tribute to Jerome Cornfield. *Biometrics* 38 33–45.

GRIFFIN V. BOARD OF REGENTS. (1986). 795 F.2d 1281 (7th Cir. 1986).

HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* 81 935–960.

JACKSON V. HARVARD UNIVERSITY (1990). 721 F.Supp. 1397 (D.Mass. 1989), aff'd, 900 F.2d 464 (1st Cir. 1990).

KADANE, J. B. (1990). A statistical analysis of adverse impact of employer decisions. *J. Amer. Statist. Assoc.* 85 925–933.

KAYE, D. H. and KOEHLER, J. J. (1991). Can jurors understand probabilistic evidence? *J. Roy. Statist. Soc. Ser. A* 154 75–81.

MALKIEL, B. G. and MALKIEL, J. A. (1973). Male-female pay differentials in professional employment. *American Economic Review* 63 693–705.

McCABE, G. P. JR. (1980). The interpretation of regression analysis results in sex and race discrimination problems. *Amer. Statist.* 34 212–215.

PETERSON, D. W. (1986). Measurement error and regression analysis in employment cases. In *Statistical Methods in Discrimination Litigation* (D. H. Kaye and M. Aickin, eds.) 107–131. Dekker, New York.

PRATT, J. W. and SCHLAIFER, R. (1984). On the nature and discovery of structure (with discussion). *J. Amer. Statist. Assoc.* 79 9–33.

ROSENBAUM, P. R. and KRIEGER, A. M. (1990). A sensitivity analysis for two-sample permutation inferences in observational studies. *J. Amer. Statist. Assoc.* 85 493–498.

ROSENBAUM, P. R. and RUBIN, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. Roy. Statist. Soc. Ser. B* 45 212–218.

SAKS, M. J. and KIDD, R. (1980). Human information processing and adjudication: Trial by heuristics. *Law and Society Review* 15 123–156.

SCHAFER, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika* 74 385–391.

SCHONBERGER, R. J. (1990). *Building a Chain of Customers: Linking Business Functions to Create the World Class Company.* The Free Press, New York.

SWAMY, P. A. V. B. and TAVLAS, G. S. (1992). Is it possible to find an econometric law that works well in explanation and prediction? *Journal of Forecasting* 11 17–33.

THEIL, H. (1971). *Principles of Econometrics.* Wiley, New York.