Use and Misuse of Statistics in Employment Discrimination Litigation

The National Law Journal, April 19, 2001

By Jay W. Waks and Gregory R. Fidlon

For the practitioner, statistical evidence, a mainstay of employment discrimination litigation, remains largely misunderstood. Its purpose of proving workplace bias stems from the recognition that discrimination is often subtle and there is rarely any "smoking gun" evidence of discriminatory intent. A recent decision of the U.S. Supreme Court, *Reeves v. Sanderson Plumbing Products*, 530 U.S. 133 (2000), may serve to increase the already prevalent use of statistical evidence in these cases. In Reeves, the Court held that once a plaintiff shows that a defendant's proffered reason for a challenged employment decision is false, the plaintiff does not always need to come forward with direct evidence of discrimination. Therefore, after *Reeves*, it may be easier for plaintiffs to rely solely on indirect evidence, such as statistics, to demonstrate intent. Although statistical evidence may be highly probative, it has inherent weaknesses and, if misused, can be of little or no value.

THE STATISTICAL MODELS USED DEPEND ON SEVERAL FACTORS

Although a variety of useful statistical models exist, the one chosen will often depend on the issues in the case and the data made available by the employer. The most common statistical models approved by the Supreme Court are demographic statistics, comparative statistics and regression analyses. *See, e.g., Teamsters v. U.S.*, 431 U.S. 324, 337-38 (1977); *Dothard v. Rawlison*, 433 U.S. 321, 329-31 (1977); *Bazemore v. Friday*, 478 U.S. 385, 400 (1986).

Demographic statistics compare the percentage of workers in the "protected" category at issue in the employer's workforce with the percentage of qualified members from that same protected category in the relevant labor market. This technique is often used when applicant-flow data (identifying actual applicants for the position at issue) is unavailable. *See, e.g., EEOC v. Chicago Miniature Lamp Works*, 947 F.2d 292, 300 (7th Cir. 1991).

Comparative statistics weigh the percentage of persons hired or promoted from the protected category against that of the "nonprotected" category and are usually utilized in disparate impact cases. *See, e.g., Bullington v. United Airlines Inc.*, 186 F.3d 1301 (10th Cir. 1999); *EEOC v. Joint Apprenticeship Commission*, 186 F.3d 110 (2d Cir. 1999); *Pietras v. Board of Fire Commissioners*, 180 F.3d 468 (2d Cir. 1999).

Regression analysis is a statistical model designed to estimate the effect of several independent variables (e.g., an employee's education, experience and gender) on a single dependent variable (e.g., salary) in order to determine whether the employee's membership in a protected category was a factor considered in decision-making. Regression analysis is often used in pay discrimination cases. *See, e.g., Bazemore*.

Earlier this year, the U.S. Court of Appeals for the 2nd Circuit extended its use in the Equal Pay Act context, holding that it was appropriate for the female plaintiff to compare herself to a statistical composite of male faculty members created using a multiple regression analysis, and observing: "It is undisputed that multiple regression analysis, which was used by the experts here, is a scientifically valid statistical technique for identifying discrimination." *Lavin-McEleney v. Marist College*, 2001 WL 91733, at *6 (2d Cir. Feb. 2, 2001).

DISPARATE IMPACT AND DISPARATE TREATMENT CASES

Statistical evidence is commonly used by plaintiffs in disparate impact cases to show the number of employees in a protected group who satisfy a selection procedure required to be hired or promoted. To blunt the plaintiff's statistical analysis, defendants will challenge the relevant labor pool, the geographic scope of the comparison, the time period or the sample size, and occasionally offer their own statistics.

There are several methods of measuring the adverse impact element in disparate impact cases. The Equal Employment Opportunity Commission, in its uniform selection guidelines, uses an 80% guideline. In other words, the EEOC will find that an adverse impact exists if members of a protected class are selected at a rate less than four-fifths, or 80%, of that of another group. Although the 80% test is easy to calculate, and often is of considerable concern to human resources departments, most courts do not rely on it, or they consider it as being no more than "a rule of thumb."

Courts generally prefer other methods such as the 2- to 3-standard-deviation test, approved in *Hazelwood School District v. U.S.*, 433 U.S. 299 (1977). In *Hazelwood*, the Supreme Court held that an adverse impact would exist if the difference between the number of members of the protected class actually selected and the number that would be anticipated under random selection is more than two or three standard deviations, or an outcome with less than a 5% chance of occurrence.

Adverse impact determination may also turn on regression analysis as well as the "t-test method." *See, e.g., Endres v. Helms*, 617 F. Supp. 1260, 1267 (D.D.C. 1985). The t-test first assumes a "null hypothesis" that postulates the relationship between observed data and sampled data. In discrimination cases, the null hypothesis generally presumes that the employee's protected status had no effect on the adverse employment action. A level of "statistical significance," which measures the probability that the null hypothesis will be rejected incorrectly, is then set -- typically, at 5%. Data is then analyzed to determine whether the null hypothesis can be rejected, given the selected level of statistical significance. If the null hypothesis is rejected, the test presumes that discrimination may exist.

The role of statistics in individual disparate treatment cases has been less significant than in disparate impact cases and class actions. This is largely due to the distinction between disparate impact cases or class actions, where the focus is on the relationship between the protected class and the overall pattern of outcomes, and disparate treatment cases, where the focus is on the individual factual situation.

Although some courts have held that statistical evidence may establish a prima facie case of individual disparate treatment, *see, e.g., Davis v. Califano*, 613 F.2d 957, 962 (D.C. Cir. 1979), the majority position is that statistics alone cannot prove discrimination. *See, e.g., Kidd v. Illinois State Police*, 167 F.3d 1084, 1102 n.16 (7th Cir. 1999). Following the majority rule, however, plaintiffs still may offer statistics in conjunction with anecdotal evidence of the defendant's discriminatory intent, especially as proof of pretext.

CLASS ACTIONS AND PATTERN OR PRACTICE CASES

Most class actions, whether brought by private plaintiffs or the EEOC, turn in part on statistical evidence. In pattern or practice lawsuits -- the EEOC's version of the class action -- the EEOC is not subject to the traditional four requirements of Rule 23 class actions: numerosity, commonality, typicality and adequacy of representation. Instead, the EEOC's initial burden is to demonstrate that unlawful discrimination has been the company's "standard operating procedure -- the regular rather than the unusual practice." *Teamsters*, 431 U.S. 336. Typically, the EEOC must present reliable statistical evidence, buttressed by direct evidence of a general discriminatory policy and specific instances of discrimination.

Class action defendants, utilizing their own statistical analyses, will often attempt to cast doubt on the plaintiffs' statistical evidence to defeat the class action, by showing, in essence, that the plaintiffs' statistics were not linked to the relevant departments, decision-makers or job categories.

MISUSE OF STATISTICS CAN OCCUR IN SEVERAL WAYS

Courts have recognized several common weaknesses with statistical evidence in all types of discrimination cases. The foremost example is the use of a statistical sample that is too small or otherwise incomplete. Statistical evidence derived from small samples "has little predictive value and must be disregarded." *Harper v. Trans World Airlines Inc.*, 525 F.2d 409, 412 (8th Cir. 1975), *cert. denied,* 429 U.S. 1050 (1977). This is because, when dealing with small samples, slight changes in the data "drastically alter" the inferences to be drawn. *Sengupta v. Morrison-Knudson Co.*, 804 F.2d 1072, 1076 (9th Cir. 1986).

Another weakness is that statistical proof may be derived from inadequate statistical models. As the 8th Circuit has noted, statistical evidence must be based on a "meaningful statistical comparison." *Miller v. Weber*, 577 F.2d 75, 77 (8th Cir. 1978). While it is unclear what types of analyses qualify as "meaningful," plaintiffs who use statistical evidence must compare similarly situated workers from the relevant labor pool. For example, except in cases involving jobs requiring no skills -- or requiring only readily obtainable skills -- raw comparisons between an employer's workforce and the general population are generally recognized to be inappropriate. *See, e.g., Gay v. Waiters' & Dairy Lunchmen's Union*, 489 F. Supp. 282, 307 (N.D. Cal. 1980), *aff'd*, 694 F.2d 531 (9th Cir. 1982).

Additionally, statistical calculations "are not probative of anything without support from an underlying statistical theory." *Frazier v. Consolidated Rail Corp.*, 851 F.2d 1447, 1452 (D.C. Cir. 1988). The typical theory advanced by plaintiffs is that discrimination exists when the observed representation of female, minority or older workers in the employer's workforce is lower than the representation that would be expected if employment decisions were made randomly with respect to sex, race or age. This theory, however, is based on faulty logic, which one commentator has termed the "statistical fallacy" or "transposition fallacy." See Kingsley R. Browne, "Statistical Proof of Discrimination: Beyond 'Damned Lies,' " 68 Wash. L. Rev. 477, 484-503 (1993); Kingsley R. Browne, "The Strangely Persistent 'Transposition Fallacy': Why 'Statistically Significant' Evidence of Discrimination May Not Be Significant," 14 Lab. Law. 437 (1998).

The fallacy is the assumption that statistical analyses can reveal the probability that an observed workforce disparity was produced by chance; whereas, in reality, statistical tests merely provide the probability of a certain observed disparity when randomness, or chance, is assumed. They do not and cannot say anything about causation. Nevertheless, as

Browne observes, this mistake has been made by numerous courts, statistical expert witnesses and both legal and statistical commentators.

Finally, even assuming that a plaintiff demonstrates a statistically significant probability that a given outcome is not due to chance, such a showing is insufficient by itself to prove that unlawful discrimination is the cause. Significant disparities in the workplace may result from a variety of other, nonrandom causes. For example, different sex, race and age groups may display very different interests in certain jobs.

PARTIES MUST ENSURE THAT THEIR THEORY IS SOUND

As the Supreme Court has recognized, statistics "come in an infinite variety and ... their usefulness depends on all of the surrounding facts and circumstances." *Teamsters*, 431 U.S. 340. Although statistics, when combined with anecdotal evidence, may indicate discriminatory intent, there are weaknesses in this type of evidence.

Therefore, in order to utilize statistics effectively, the party proffering such evidence must be certain that its theory is sound, its models and assumptions are fine-tuned to the particular context of the comparison, and its expert testimony is *Daubert*-proof. See *Daubert v. Merrel Dow Pharmaceuticals*, 509 U.S. 579 (1993).