and it has hardly eradicated employment discrimination. Furthermore, race, religion, national origin, and gender discrimination lack the tremendous economic incentives for employers to discriminate that exist in health-based discrimination. Simply focusing on more enforcement also will be of questionable fairness to the employers who will be forced to pay more than their share of the escalating healthcare costs. The better alternative involves more fundamental changes in both our employment system and in our health care system, but the specifics of those changes is another issue for another day. At the very least, the adverse consequences and the potential for discrimination raised by the Human Genome Project must be considered carefully and dealt with; the time is already at hand.

METHODS FOR ASSESSING THE SENSITIVITY OF STATISTICAL COMPARISONS USED IN TITLE VII CASES TO OMITTED VARIABLES

Joseph L. Gastwirth*

INTRODUCTION

Several recent cases—*Wards Cove Packing Co., Inc. v. Antonio,*[1] *City of Richmond v. J.A. Croson Co.*,[2] and *Stuart v. Roache*[3]—emphasized the importance of ensuring that similarly qualified minority and majority applicants are considered in statistical comparisons used in litigation concerning the fairness of hiring and promotion practices or the validity of a consent decree arising from a claim of discrimination.[4] As the Court noted in *Bazemore v. Friday*, plaintiffs need only incorporate the major job-related factors in establishing a *prima facie* case.[5] It is easy to raise the possibility that another variable or factor could explain an observed difference between minority and

*Joseph L. Gastwirth is Professor of Statistics and Economics at George Washington University. This research was supported in part by a National Science Foundation grant.

[1]Wards Cove Packing Co., Inc. v. Antonio, 490 U.S. 642 (1989).

[2]City of Richmond v. J. A. Croson Co., 488 U.S. 469 (1989).

[3]Stuart v. Roache, 57 Fair Empl. Prac. Cas. (BNA) 902 (1st Cir. 1991). The decision upheld the consent decree as it reported data supporting a *prima facie* case. *Id.* at 905. For example, at the time of the complaint only one of 222 police sergeants was black, although blacks formed 4.5% of the eligible police officers. *Id.* at 905. The opinion notes that the "Decree compared the number of black sergeants, not with the Boston population in general, but with those police officers with the minimal qualifications needed to become sergeants." *Id.* at 906.

[4]In *Croson*, the Court did not accept a comparison of the minority proportion of contractors to the minority proportion of the general population. *Croson*, 488 U.S. at 501-04.

[5]Bazemore v. Friday, 478 U.S. 385, 400 (1986). Although the Court was discussing a regression analysis, the principle should apply to any statistical presentation.

majority hiring and promotion rates. To ascertain whether an omitted factor could completely explain an observed difference in the success rates of two groups, we apply and extend an approach used by Cornfield.[6]

After Cornfield's result is stated in section I, then how to incorporate sampling error when the result is used to analyze applicant flow data is shown.[7] The method is subsequently used to demonstrate that the comparison accepted in *Allen v. Seidman*,[8] which showed a disparate impact of an exam, probably could *not* be explained by an alleged omitted factor, while the comparison rejected in *Waisome v. Port Authority*[9] might well be. When data on actual applicants are unavailable or seriously flawed, courts rely on a demographic method, which compares the minority fraction of actual hires to their fraction of an external labor pool of qualified and available potential employees. While factors such as education, work experience, salary of the position, and geographic proximity to the job site can be considered in developing the relevant external labor pool, it is impossible to include personal characteristics such as interest or motivation in the demographic approach.[10]

In section II, we develop the analog of the result of section I and apply it to the data used to assess the impact of the practice of hiring relatives (nepotism). This practice was found *not* to have a disparate impact on Native Alaskans in *Wards Cove*.[11] Our analysis questions the Court's conclusion. Several data sets from recent cases are also analyzed by this technique. The article concludes with a general discussion of the role of sensitivity analysis in evaluating issues of causation, and its relationship to other basic statistical concepts.

I. APPLICANT FLOW ANALYSIS

Provided that the recruitment process is fair, the comparison of test pass rates is the appropriate statistical analysis to assess whether an employment

[6]The issue originally arose in studying whether smoking causes lung cancer. Further background and references are given in Samuel W. Greenhouse, *Jerome Cornfield's Contributions to Epidemiology*, 38 BIOMETRICS 33 (1982).

[7]Applicant flow analyses compare the hire or success rates of minority applicants to those of majority applicants. Because this information relates to the individuals who actually applied for the jobs at issue, when the data are reliable courts have indicated a preference for its use. *See, e.g.*, Payne v. Travenol Lab. Inc., 673 F.2d 798, 824 (5th Cir. 1982); Mister v. Illinois Cent. Gulf R.R. Co., 832 F.2d 1427, 1435 (7th Cir. 1987).

[8]Allen v. Seidman, 881 F.2d 375 (7th Cir. 1989).

[9]Waisome v. Port Auth., 55 Fair Empl. Prac. Cas. (BNA) 731 (S.D.N.Y. 1991).

[10]See Joseph L. Gastwirth, *Estimating the Demographic Mix of the Available Labor Force*, 101 MONTHLY LAB. REV. 26 (1981) and Chapter 4 of DAVID C. BALDUS & JAMES W.L. COLE, STATISTICS AS PROOF OF DISCRIMINATION (1980 & Supp. 1987) for the statistical methods used in creating the appropriate labor market.

[11]*Wards Cove*, 490 U.S. at 657.

practice has a disparate impact on a particular group.[12] If we denote the pass rate of the minority group by p_1, the pass rate of the majority group by p_2, and let R be their ratio (where $R = p_2/p_1$), then we can assess the sensitivity of our statistical inference to another factor (X) that was not incorporated in our analysis, using a variant of *Cornfield's Lemma*.[13] In order for factor X to explain a disparity between two rates, the factor must multiply one's chance of passing by at least R *and* the proportion of majority group members possessing the factor must be at least R times the proportion of minority group members possessing the factor.

The ratio R is the reciprocal of the selection ratio, which government agencies have used to assess the disparate impact of tests.[14] Indeed, the four-fifths rule corresponds to R exceeding 1.25. The ratio of the fraction of the majority group possessing the factor (f_2) to the corresponding fraction of the minority group (f_1) is called the prevalence ratio (Θ).

The cited proof of Cornfield's lemma yields a stronger result when the prevalence (f_1) of factor X in the minority group is known. With f_1 known, one can determine the amount by which the prevalence ratio (Θ) must exceed R before the omission of factor X in the analysis can fully explain the disparity. Letting R_x denote the ratio of pass rates of persons possessing factor X to those who do not, then $\Theta = f_2/f_1$ must be at least

(1.1a) $$R + \left(\frac{R-1}{R_x-1}\right) \cdot \frac{1}{f_1}$$ or equivalently

(1.1b) $$f_2 \geq Rf_1 + \frac{R-1}{R_x-1} .$$

[12]Courts classify equal employment cases into two categories: disparate treatment and disparate impact. Disparate treatment cases focus on whether or not an employer intended to treat similarly situated minority and majority employees or applicants differently. *See* Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248 (1981). Disparate impact cases focus on whether a specific employment practice such as a test or job requirement eliminates a disproportionate share of minority group members relative to majority group members. If it does, the practice has to be shown to be job-related. The *Wards Cove* opinion describes the burdens of the parties. 490 U.S. 642, 651–60. Under the *Wards Cove* standard, if plaintiffs demonstrate that a practice has a disparate impact, the defendant needs to produce evidence showing that the practice serves a legitimate business purpose. Title VII of the Civil Rights Act, 42 U.S.C. § 2000e (1991), reversed this aspect of the *Wards Cove* opinion. Under § 2000e-2(k)(l)A of the Act, defendants will need to demonstrate that a practice with a disparate impact "is job related for the position in question and consistent with business necessity."

[13]Our statement is a variant of Cornfield's original result and is proved in JOSEPH L. GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY 296 (1988).

[14]The selection ratio is the ratio of the minority pass rate to the majority pass rate. The *Uniform Guidelines on Employee Selection Procedures*, 43 Fed. Reg. 38295, 38297 (1978) state that when the selection rate of any protected group is less than four-fifths that of the race-sex-ethnic group with the highest pass rate, federal agencies will generally regard the practice as having an adverse impact. The job-relatedness of the practice must then be demonstrated.

To illustrate the implication of this strengthened version of Cornfield's lemma, suppose that 60% of whites pass a test while only 40% of a minority group pass. Thus, R = 1.5. Suppose that a certain type of training or education was job related, but the applicant data did not report this information—this training is our factor X. Assume that persons with such training had twice the chance of passing, i.e., $R_x = 2.0$. If one believed that at least 10% of the minority group had this training, then equation (1.1b) implies that at least

$$f_2 = (1.5)\,(.10) + \frac{(1.5-1)}{(2-1)} = .65$$

or 65% of the whites would need to possess this background to explain the disparity in pass rates. However, if the training tripled one's chance of passing, then the training would explain the disparity if 40% or more of the whites had it.

Cornfield's original result did not consider sampling error; therefore, a *statistically conservative* approach is to construct a 95% confidence interval (C.I.) for R and use the lower end (R*) in place of R when determining whether a suggested factor could reduce an observed disparity to an insignficant amount.[15] Furthermore, it is not good practice to assess the sensitivity of a statistical analysis to an omitted factor without considering the precision of the estimated value of R.[16] Another advantage of forming the confidence interval is that one can determine whether the value 1.25, corresponding to the four-fifths rule, falls within it. If R* exceeds 1.25, then a statistical test of R = 1.25 . . . i.e., the employment practice has a "borderline" disparate impact of exactly 1.25 . . . also would be rejected. This provides stronger evidence for the need to validate the practice.[17]

Denoting the observed pass rates of the minority group and the majority group by p_1 and p_2, respectively, standard statistical theory yields the following formula for the 95% C.I. for R[18]:

[15] It is important to recall that a 95% confidence interval has the following property: in a long sequence of independent repetitions of the situation generating the data, it should contain the true value of the parameter of interest (R here) 95% of the time. We cannot say that any particular C.I. has probability 0.95 of containing the true parameter. See David H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54 (1987) for further discussion of these concepts as well as examples of their misuse in the legal literature.

[16] *See* Paul R. Rosenbaum, *Sensitivity Analysis for Matched Case Control Studies*, 47 BIOMETRICS 87, 93 (1991) (stating that the most extreme point estimates will typically come from studies with the smallest sample sizes).

[17] The usual test of equality of pass rates determines whether R = 1 or equivalently whether 1.0 lies in the confidence interval. Here we are testing whether R = 1.25, or equivalently whether the selection ratio is 0.8. If this hypothesis is rejected we conclude that R exceeds 1.25 or the selection ratio is statistically significantly less than 0.8.

[18] The proof uses the well-known fact that the sample proportions p_1 and p_2 are approximately normally distributed and the fact that, under suitable conditions, functions of normal random variables also have a normal distribution. For technical details see C.R. RAO, LINEAR STATISTICAL

$$\text{(1.2)} \qquad \exp\left\{lg\,(p_2/p_1) \pm (1.96)\sqrt{\left(\frac{(1-p_1)}{p_1 n_1} + \frac{(1-p_2)}{p_2 n_2}\right)}\right\},$$

where lg denotes the natural logarithm, exp means take *e* to the power of the term in the brackets, and n_1 and n_2 are the sample sizes (numbers) of minority and majority persons in the data set. It should be noted that because formula (1.2) is based on large sample theory, its validity needs to be checked on small samples.[19]

A. *Allen v. Seidman*[20]: The Disparate Impact of a Test

The plaintiffs in this case were black bank examiners employed by the FDIC who failed the "Program Evaluation" test used by the defendant in deciding whether to promote examiners to the GS-11 level from the GS-9 level. Although the opinion discusses several related data sets, we will concentrate on the pass-fail data. It showed that 14 blacks out of a total of 36 passed, while 329 whites out of 391 passed.[21] This difference in pass rates was highly significant.[22] For this set of data, R = 2.16 and the 95% confidence interval is (1.98, 2.36).

As 1.98 is essentially 2.0, Cornfield's result implies that in order for an omitted factor to explain the observed difference in pass rates, the factor must double one's probability of passing the test and the proportion of whites possessing the factor must be twice the corresponding fraction of blacks. The appellate opinion, written by Judge Posner, notes that the defendant suggested that the black candidates *might* have lesser qualifications, e.g., education. However, the FDIC did not introduce evidence supporting this assertion nor any other difference between the groups with regard to a job-related characteristic. The opinion noted that all the examinees had worked for the defendant for between 5 and 15 years, had worked at least one year at the GS-9 level, and were required to obtain a recommendation from their regional director in order to take the exam. Therefore, the pool of test-takers appeared to be reasonably homogeneous with respect to job-related background characteris-

INFERENCE AND ITS APPLICATIONS (2d ed. 1973). Although formula (2.2) is given for a 95% C.I., replacing (1.96) by the appropriate point ($z_{\alpha/2}$) of the normal distribution yields a 100 $(1-\alpha)$ % C.I.

[19] Although Fisher's exact test for comparing two sample proportions is available in several statistical packages, e.g., SAS and Stat Xact, exact tests and confidence intervals for the ratio p_1/p_2 are not readily available. See Thomas J. Santner and Mark K. Snell, *Small Sample Confidence Intervals*, for p_1-p_2 and p_1/p_2 in 2×2 *Contingency Tables*, 75 J. AM. STAT. ASS'N 386, 389 (1980) for further discussion. However, Stat Xact yields results for the odds ratio.

[20] *Allen*, 881 F.2d 375 (7th Cir. 1989).

[21] The data come from the opinion of the district court in Allen v. Isaac, 39 Fair Empl. Prac. Cas. (BNA) 1142, 1149 (N.D. Ill. 1986).

[22] The opinion, *id.*, reports a difference of 6.5 standard deviations, well in excess of the two to three standard deviation criteria the court adopted in Castaneda v. Partida, 430 U.S. 482 (1977). An R of 2.16 implies that the selection ratio used to assess disparate impact by various government agencies equals 0.46, which is substantially less than 0.8. *See supra* note 14.

tics.[23] Our sensitivity analysis implies that the court concluded it was unlikely that an additional factor could double one's probability of passing the exam and be twice as prevalent among white GS-9 employees as black GS-9 employees. The court emphasized that the FDIC did not offer evidence of an imbalance in any relevant characteristic of the examinees[24] and that the defendant could have submitted a regression analysis incorporating an appropriate factor that explained the disparity.[25] Moreover, the strengthened version of Cornfield's lemma implies that the omitted factor would require an R_x greater than 2.0 to explain the disparity if a modest fraction of blacks had also possessed the qualification. For example, if $R_x = 3.0$ and only 10% of blacks possessed the relevant factor, then 70% of the whites would need to have it. If a factor that tripled one's probability of success existed, is it plausible to believe that the employer would not determine which employees possessed it?

In conjunction with the lack of evidence of an imbalance in the qualifications of the examinees, the sensitivity analysis indicates that the observed statistical disparity in pass rates could only be explained by a factor the employer was unaware of. This factor must have been strongly related to passing the test. Moreover, a far greater fraction of white examinees must have possessed it than black examinees. While a sensitivity analysis cannot prove that such a factor did not exist, it clearly suggests that the court's assessment is likely to be correct.

B. *Waisome v. Port Authority*[26]: The Disparate Impact of a Written Test

In *Waisome*, black applicants for the position of sergeant alleged that a written exam had a disparate impact. Although the difference in pass rates (78.12%, 89.57%) was statistically significant, the four-fifths rule was not satisfied. The court found that a difference of 2.68 standard deviations (equivalent to a p-value of 1 in 100) did not have practical significance, in part, because had *two more* black candidates passed, the difference in rates would not have been significant.[27] The practice of changing data has been questioned by several commentators.[28] Since courts rarely decide that a *prima facie* case would be established had one *less* black passed (or been hired), a sensitivity analysis of the data may lead to a more substantive justification for the decision.

The written exam was taken by 508 whites and 64 blacks. Of these, 455 whites and 50 blacks passed.[29] The selection ratio was 0.872 ($R = 1.147$). A 95% confidence interval for R, based on formula (1.2), is (1.004, 1.310). Thus, an omitted variable, X, which increased an applicant's probability of passing by the multiple of 1.01 and was 1.01 times more prevalent among White applicants could reduce the observed difference in rates to a statistically *insignificant* one.

In view of the large difference in educational background of blacks and whites in society as a whole,[30] and the likelihood that more experience with written tests increases an applicant's chance of doing well, it is quite plausible that the two groups could differ by a sufficient amount with respect to educational background. This type of analysis may also aid in deciding how much evidence a party should be required to submit to substantiate an alleged flaw in the statistics offered by the opposing party. In *Waisome*, it might have been reasonable to ask the defendant to demonstrate that the groups differed with regard to educational background or grades in school, if such information was required of the applicants. Otherwise, one might rely on general knowledge about society at large, as the increase in prevalence of a higher level of education among white relative to black applicants needed to satisfy Cornfield's criteria is minimal. The main difference between our analysis of the data in *Waisome* and *Allen* is that a very *slight difference* in the prevalence of a trait that *marginally increased* one's probability of passing the written test could explain the observed disparity in *Waisome*. In contrast, in order for an omitted factor to explain the disparity in *Allen*, that factor had to more than *double* a person's probability of being successful and be at least *twice* as prevalent among white applicants than among blacks. The fact that nearly 80% of black applicants passed the test in *Waisome* also indicates that the test may not be a substantial barrier to their advancement.

II. THE DEMOGRAPHIC COMPARISON

A. The Analog of Cornfield's Lemma

In cases where the demographic approach is appropriate, one determines the minority fraction (π) of the *relevant* labor force, i.e., persons qualified and available for the positions at issue, and compares the minority fraction (s) of hires to π using the binomial model.[31] When the statistical test indicates a

---

[23] *See Allen*, 881 F.2d at 379.

[24] *Id.* at 379–80.

[25] *Id.* at 380.

[26] *Waisome*, 55 Fair Empl. Prac. Cas. (BNA) 731 (S.D.N.Y. 1991).

[27] *Id.* at 735.

[28] *See* DAVID C. BALDUS & JAMES W. L. COLE, STATISTICS AS PROOF OF DISCRIMINATION (1980 & Supp. 1987); Joseph B. Kadane, *A Statistical Analysis of Adverse Impact of Employer Decisions*, 85 J. AM. STAT. ASS'N 925 (1990); David H. Kaye, *Improving Legal Statistics*, 24 LAW & SOC'Y REV. 1255 (1990).

[29] *Waisome*, 55 Fair Empl. Prac. Cas. (BNA) at 735.

[30] For the younger age groups, where the applicants were likely to come from, approximately 74% of blacks and 86% of whites completed high school. The proportions of those 25 to 29 years old who had a college degree were 11.4% for blacks and 23.9% for whites. U.S. Bureau of the Census, *U.S. Census of Population: 1980*, Table 246.

[31] Following Castaneda v. Partida, 430 U.S. 482 (1977), courts often use the normal approximation to the binomial model, i.e., if n persons are hired we expect $n\pi$ to be of some minority. The standard deviation of the number of minority hires is $\sqrt{n\pi(1-\pi)}$. The test compares the number (S) of minority hires with their expected number—the difference $S-n\pi$ divided by $\sqrt{n\pi(1-\pi)}$ measures this difference in standard deviation units. If one compares the fraction (s) of minority hires to the minority fraction (π) of the appropriate labor pool, one obtains the

shortfall of minority hires (i.e., s is significantly less than π) the analog of the strengthened version of Cornfield's lemma becomes *Lemma 2.1*: Let Θ be the ratio of the prevalence of a job-related factor (X) in the majority group to its prevalence in the minority group. Assume that this factor was omitted in developing the external labor pool. For factor X to explain a disparity between s and π, it must increase a person's probability of being hired by the multiple R_x, and R_x must be at least R, where

$$\textbf{(2.1a)} \qquad R = \frac{\pi/(1-\pi)}{s/(1-s)}.$$

Moreover, Θ must exceed:

$$\textbf{(2.1b)} \qquad R + \frac{(R-1)}{(R_x-1)f_i}.$$

The proof is given in the Appendix. The relations (2.1a) and (2.1b) are the analogs of (1.1a) and (1.1b). As before, if factor X increases the probability of passing a test by *just* R, then the ratio (Θ) must substantially exceed R. We note that R can be regarded as an odds ratio—the ratio of the odds of selecting a minority from the relevant labor pool to the odds of selecting a minority from the individuals who were hired.[32]

As the Lemma does not incorporate chance effects or "sampling error," a cautious approach is to replace the observed fraction (s) of minority hires by the upper end (u) of the confidence interval for π derived from the data. Essentially, one assumes that π is not known and asks: What is the largest value π could have which would be statistically consistent with the observed data?[33] Because we are only concerned with assessing the influence of an omitted factor on an analysis indicating a statistically significant shortfall of minority hires, the upper end of the confidence interval (u) will be *less* than π. We then use

same result because the standard deviation of the fraction (s) is 1/n times the standard deviation of the sum, i.e., the statistic $Z = \frac{s-\pi}{\sqrt{\pi(1-\pi)/n}}$ is approximately a standard normal variate and is used for the test.

[32]This interpretation is similar to one given in David Kairys et al., *Jury Representativeness: A Mandate for Multiple Source Lists*, 65 CAL. L. REV. 776 (1977).

[33]In this article we assume all tests are two-sided following the Court's decision in *Castaneda*, 430 U.S. 482. A discussion of one- and two-sided tests and their relevance to the establishment of a *prima facie* case is given in Palmer v. Shultz, 815 F.2d 84, 92–97 (D.C. Cir. 1987). In situations where a one-sided test is appropriate, e.g., where the defendant has a prior history of discrimination, one can use a one-sided test. Then the value (u) is the upper end of a one-sided upper confidence interval for π obtained from the data. For purposes of calculation one need only consider one-sided tests because one can adjust the confidence level appropriately, i.e., the upper endpoint of a 95% one-sided confidence interval is the upper endpoint of a 90% two-sided confidence interval.

$$\textbf{(2.2)} \qquad R^* = \frac{\pi/(1-\pi)}{u/(1-u)}$$

in place of R in our sensitivity analysis.

To apply the result, we recall that, in large samples, the value (u) of the upper end of the 95% confidence interval for π is given by

$$\textbf{(2.3)} \qquad u = s + (1.96)\,\frac{\sqrt{s(1-s)}}{\sqrt{n}},$$

where n is the number of hires. We now illustrate the method using data from three cases.

B. The Practice of Favoring Relatives of Employees in *Wards Cove*[34]

One aspect of *Wards Cove* concerned the possible disparate impact of defendants' practice of hiring relatives of employees (nepotism) on Native Alaskans. The *defendants'* labor market analysis set the minority share (π) of the labor pool at 0.101, while their share of the 349 nepotistic hires was 0.0487 (17 out of 349).[35] Using formula (2.2), we find u = 0.072, thus R* = 1.44. Thus, an omitted factor such as motivation or reliability would need to increase the probability of hire by at least 1.44 and would need to be 1.44 times more prevalent among relatives of current employees than among Native Alaskans. Because the defendants' labor pool considered persons employed in similar occupations in the labor market area, is it plausible to believe that:

(a) There would be an important job-related factor that increases one's chance of being hired by nearly 50%, yet the employer does not obtain information about it from applicants *and*

(b) The factor would be nearly 50% more prevalent amongst relatives of current employees than amongst other persons employed in similar occupations, who live in the firm's labor market area?

So far our analysis has relied on the defendants' availability figure (π) being equal to 0.101. Based on possible flaws in the defendants' development of their geographic labor market and on fragmentary postcharge applicant data indicating that 25.6% of the applicants were black, it has been argued that a more reasonable estimate of π would be in the range of 0.147 to 0.187.[36] Using

[34]*Wards Cove*, 490 U.S. 642.

[35]*Id.* at 678 n.28 (Stevens, J., dissenting).

[36]*See* Joseph L. Gastwirth, Wards Cove Packing Inc. et al. v. Antonio: *How Inadequate Evidence and Reasoning Contributes to Questionable Decisions*, Technical Report, Dep't of Statistics, George Washington University (1992).

Using $\pi = 0.147$ yields an R^* of 3.87, which virtually precludes any subjective factor such as motivation or reliability from explaining the disparate impact of nepotism, especially because the labor market availability calculation incorporated the occupational backgound of the potential labor force.

The above conclusions are strengthened when we incorporate the relationship (2.1b) between the prevalence ratio and the measure (R_x) of the effect of the omitted factor, X. Observe that (2.1b) implies that the prevalence (f_2) of factor X in the majority group must be at least 1.44 times its prevalence in the minority group. If only 10% of the minority group had the qualification (X), and $R_x = 2.0$, then (2.1b) implies that 59% of the relatives of the majority group employees would need to possess the qualification. It is very unlikely that such a job-related qualification exists and no relevant data are collected by the employer or accounted for in its availability calculation.

C. *EEOC v. O&G Spring Wire Co.*[37]: Discrimination in Hiring

A charge of hiring discrimination against blacks was filed in 1984 and the EEOC's attempt at conciliation failed in July 1985.[38] For the 1979-1985 period, *no* blacks were among the 85 persons hired for a low-skilled job operating a kick-and-punch press.[39]

Because applicant data for the 1979-1985 period were not available, the plaintiffs' expert developed a weighted labor market based upon census data of people employed as machine operators in the area and the residence of applicants in 1986 and 1987.[40] Because blacks comprised 23% of this relevant labor pool, the fact that no blacks were among the 85 hires is highly significant.[41]

In assessing the sensitivity of our inference to a possible relevant factor that was not included in developing the labor pool, we note that formula (2.3) is not applicable because there were *no* black hires, i.e., the factor $[s(1-s)]^{1/2}$ is zero. Thus, one needs to use an exact confidence interval for π based on the binomial distribution,[42] which yields $u = 0.0425$. Using this value for u in formula (2.2) gives $R^* = 6.73$. Therefore, for another job-related factor, which was not incorporated in the occupational and geographic determination of the available labor pool, to "explain" the statistically significant shortfall of black hires, the factor should increase a person's probability of on-the-job success sevenfold, and the factor must be nearly seven times more prevalent amongst whites than amongst blacks in the appropriate labor pool. No relevant characteristic that 15% of blacks possess can satisfy this condition.

D. *Stuart v. Roache*[43]: Justifying an Affirmative Action Plan

In this case, white police officers challenged a consent decree requiring the Boston Police Department to favor minority officers. The plaintiffs contended "that the statistical disparities in the Department cannot be sufficient to justify an affirmative action plan" because the Supreme Court had rejected the statistical data used by the City of Richmond in *Croson*.[44] Judge Breyer's opinion in *Roache* distinguished the general population data used by the City of Richmond from the statistical comparison in the Boston decree. Only one of 222 police sergeants was black, although blacks formed 4.5% of police officers who were eligible for promotion. The appellate court found this to be a "strong basis in evidence" to support the city's claim that the decree served a proper remedial purpose to correct for prior discrimination.[45] The standard statistical test shows that the probability of such an outcome occurring when sergeants were randomly chosen from the pool of eligibles is only 0.0004. This result is highly significant, and it supports the appellate court's position.[46]

We now explore the possibility of a job-related factor explaining the statistical disparity between the one sergeant in the department and the 10 that would be statistically expected if appointments were randomly selected from the pool of eligible officers. The opinion noted that seniority might explain the shortfall, but since the greater seniority of white officers was due to prior discrimination it was not a legally proper factor. Because of the low availability fraction (0.045), the normal approximation to the binomial distribution is not appropriate. Thus, equation (2.3) will not be used. The upper end of a 95% confidence interval can readily be calculated with existing software and is equal to 0.0248, i.e., the observed data on sergeants would not reject an hypothesized minority availability less than 0.0248.[47] Using this value for u in (2.2) yields $R^* = 1.853$.

Suppose a factor, such as having a college education, doubled one's probability of passing a valid promotion exam. Condition (2.1b) on the preva-

[37]EEOC v. O&G Spring Wire Co., 48 Fair Empl. Prac. Cas. (BNA) 1540 (N.D. Ill. 1988).
[38]*Id.* at 1541 (finding 5).
[39]*Id.* at 1542 (finding 10).
[40]*Id.* at 1543 (finding 20).
[41]The probability of observing 0 blacks among 23 hires when they form 23% of the relevant labor pool is less than one in a million.
[42]When no minorities are among n hires, u satisfies the equation $(1-u)^n = 0.025$ as we create a one-sided 97.5% confidence interval for u rather than a 95% two-sided one. When minorities form a small fraction of hires, the normal approximation upon which formula (2.3) is based may not be accurate. However, exact confidence intervals are readily available from the recent STATXACT (Version 2) program.

[43]Stuart v. Roache, 57 Fair Empl. Prac. Cas. (BNA) 902 (1st Cir. 1991).
[44]*Id.* at 905-06.
[45]*Id.* at 905.
[46]This probability is based on the binomial distribution. The usual significance level accepted by courts is 0.05; however, there is no absolutely fixed threshold.
[47]The calculation was obtained from the STATXACT program.

lence ratio implies that the prevalence of the factor among whites must exceed 0.85 + 1.85f, where f is the prevalence of the job-related factor among black officers. This implication is not plausible. If only 10% of the black officers possessed the factor, it could not explain the disparity even if all white officers possessed the factor. Assuming the factor *tripled* one's chances of promotion, its prevalence among white officers would need to exceed 0.427 + 1.85f to explain the disparity. If 10% (20%) of the black officers had the qualification, then 61.2% (79.8) of the white officers would need to have it. These results indicate that it is quite unlikely that a legally proper job-related factor exists but was unaccounted for in the data, which explains the observed disparity between black police officers and sergeants. Thus, the sensitivity analysis provides further support for the decision.

III. DISCUSSION

Because virtually all data sets are not perfect, inevitably the opposing party can raise questions concerning omitted or mismeasured job-related factors.[48] Using a sensitivity analysis, the factfinder may assess the potential impact of the alleged missing factor on the ultimate inference. Once the value of R^* is obtained, a court can assess whether the influence of a factor omitted from the analysis could be sufficient to explain the observed disparity. When using a sensitivity analysis it is important to note that even when R^* is relatively small, e.g., $R^* = 1.2$, indicating that an omitted variable could explain the statistical disparity, no such factor may exist.[49]

Sensitivity analysis can also be regarded as an approach to assessing causation, because it provides an estimate of the properties an "unobserved variable would need to provide an explanation of the effect under study."[50] Of course, other criteria such as temporal ordering and subject matter knowledge play a role in the determination of whether one variable caused an observed effect.[51] In light of several recent decisions of the Court (indicating that in

disparate impact cases plaintiffs need to show that an employment practice causes discrimination in the work force,[52] or in school desegregation cases that the vestiges of prior segregation need to have a causal link to current conditions[53]), sensitivity analysis should prove to be a useful tool for judicial decisions. The proper use of sensitivity analysis still requires the decision-maker to form a judgment concerning the potential effect of an omitted or unobserved factor. Thus, courts should insist that the party questioning the data provide information indicating that the omitted factor meets both requirements of Cornfield's Lemma.[54]

Courts should also determine whether a seemingly valid factor, omitted from the analysis, played a role in the employment decisions at issue. In a recent case[55] concerning the subjective method used by the defendant in selecting leadmen and foremen, plaintiffs showed a statistically significant disparity in promotions to leadmen.[56] The defendant asserted that the relevant eligible pool should not be all employees in the hourly work force because seniority and membership in the craft of open leadmen should be incorporated in a proper analysis.[57] Judge Cudahy's opinion indicated that seniority could not explain the disparity, because there were minority employees who were senior to the promoted whites in 8 of the 43 promotions.[58] Similarly, many employees outside the first class of the relevant craft had been selected for upgrades.[59]

---

[48] See Arthur P. Dempster, *Employment Discrimination and Statistical Science*, 3 STAT. SCI. 149 (1988) for an overview of the statistical literature relating to regression models and the effect of mismeasured and omitted variables. A less technical discussion is found in M. FINKELSTEIN & B. LEVIN, STATISTICS FOR LAWYERS (1990). This issue arises in many applications. In assessing the geographic model used by the government in an antitrust case, the court noted that "it is always possible to take pot shots at a market definition." United States v. Rockford Memorial Corp., 898 F.2d 1278 (7th Cir. 1990).

[49] Paul R. Rosenbaum & Abba M. Krieger, *Sensitivity Analysis for Two-Sample Permutation Inferences*, 85 J. AM. STAT. ASS'N 493 (1990).

[50] David R. Cox, *Causality: Some Statistical Aspects*, 155 J. ROYAL STAT. SOC'Y A 291, 293 (1992).

[51] *Id.* at 293. Cox also notes, *id.* at 297, that while the sensitivity approach is a useful guide for summarizing observational data (the type of data occurring in the Title VII cases), it may be too far from an underlying scientific explanation to use the term "causal." Sensitivity analysis can be regarded as an example of practical reasoning. For a general discussion of the relationship between practical and legal reasoning, see RICHARD A. POSNER, THE PROBLEMS OF JURISPRUDENCE 71–123 (1990).

[52] *See* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 997 (1988) (stating that "the flaw in the plaintiff's proof was its failure to establish the required causal connection between the challenged practice (testing) and discrimination in the work force") (citing Carroll v. Sears Roebuck & Co., 708 F.2d 183, 189 (5th Cir. 1983)).

[53] *See* Freeman v. Pitts, 112 S. Ct. 1430 (1992) (holding that a district court can relinquish its supervision of those aspects of a school system in which there has been compliance with a desegregation decree if other aspects of the system remain in noncompliance). The opinion, *id.* at 1448, notes that the vestiges of segregation must be sufficiently real that they have a causal link to the current *de jure* violation being remedied. Justice Souter's concurring opinion discusses the type of inquiry a district court should apply, *id.* at 1454, and cites statistical correlations, *id.* at 1455, showing lower pupil expenditures and overrepresentations of black administrators in schools with high percentages of black students as indicating a causal relationship with respect to the consequences of the district court's relinquishing supervision over a remedial aspect of a decree.

[54] This suggestion is consistent with the approach to imperfect statistical evidence stated by Baldus and Cole, *supra* note 10, Supp. 1987 at 196 ("the ultimate question is not whether a proof has a flaw, but how the flaw is likely to bias the results"). By quantifying the required relationship of an omitted factor to on-the-job success and its relative prevalence in the two groups, a sensitivity analysis enables the factfinder to assess the potential consequences of an alleged flaw on the ultimate inference.

[55] Mozee v. American Comm. Marine Serv. Co., 940 F.2d 1036 (7th Cir. 1991).

[56] *Id.* at 1043. The opinion mentions several issues, including the propriety of combining data for several years prior to the relevant time period (beginning 300 days prior to the data of the complaint), and the interpretation of statistical results just at the borderline of statistical significance. *Id.* at n.7. As they were not raised on appeal, the court did not treat them in detail.

[57] *Id.* at 1045.

[58] *Id.* at 1046.

[59] *Id.* at 1046.

In our discussion of *Waisome*[60] we relied on census data to raise the possibility that an imbalance in the educational levels of the two groups might explain the disparity, especially if R* was relatively small. In general, it is preferable to rely on data concerning the actual employees or applicants because it will usually be based upon a self-selected subset of the population with qualifications and interests appropriate to the particular job(s) at issue. Hence, the data should not be expected to reflect the characteristics of the general population. For example, it is often stated that men have more work experience than women of the same age. In *Vuyanich*, Judge Higginbotham rejected the plaintiff's regression analysis, which used age as a proxy for years of experience, after the defendant showed that male employees had an average of two years more experience than females of the same age.[61] However, in *Denny v. Westfield State College*[62] the use of actual prior experience instead of its proxy, age, reduced the estimated gender effect.

Although we applied sensitivity analysis to the type of data that typically arises in hiring and promotion cases, related methods exist for use on stratified data,[63] comparing two sets of wage data[64] and regression analyses.[65] By utilizing sensitivity analyses in conjunction with the concepts of the power of a test to detect a meaningful difference,[66] courts can assess the meaning of a statistical presentation and avoid the questionable practice of totally relying on whether or not a statistical test reaches a pre-set level of significance.[67]

[60] Waisome, 55 Fair Empl. Prac. Cas. (BNA) 731 (S.D.N.Y. 1991).

[61] Vuyanich v. Republic Nat'l Bank, 505 F. Supp. 224, 315 (N.D. Tex. 1980), *vacated on other grounds*, 723 F.2d 1195 (5th Cir. 1984).

[62] Denny v. Westfield State College, 669 F. Supp. 1146 (D. Mass. 1987).

[63] Paul R. Rosenbaum and Donald B. Rubin, *Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome*, 45 J. ROYAL STAT. SOC. B 212 (1983).

[64] *See* Rosenbaum and Krieger, *supra* note 49.

[65] Daniel W. Schafer, *Measurement Error Diagnostics and the Sex Discrimination Problem*, 5 J. BUS. AND ECON. STAT. 529 (1987).

[66] *See* Richard Goldstein, *Two Types of Statistical Errors in Employment Discrimination Cases*, 26 JURIMETRICS J. 32 (1985); JOSEPH L. GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY 180–84, 252–59 (1988).

[67] See David H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333 and STEPHEN E. FIENBERG, THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN COURTS (1989) for further discussion about the limits of hypothesis testing and problems courts have in interpreting the results of statistical tests. The relationship between statistical significance and the power of a test in the context of toxic tort cases is clearly presented in Michael D. Green, *Expert Witness and Sufficiency of the Evidence in Toxic Substance Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 NW. U. L. REV. 643, 682–94 (1992).

APPENDIX
The Analog of Cornfield's Lemma for the Binomial Model

We now present the proof of Lemma 2.1. We assume that the labor market consists of N_1 minority members and N_2 majority members. Let $\pi = N_1/(N_1 + N_2)$. Both N_1 and N_2 are large and n hires are considered a random sample from the pool. Of the n hires, x are minority. Thus, $s = x/n$ has expected value π when the distribution of qualifications is the same in both groups. Suppose another job-related factor (X) exists that multiplies one's probability of success by R, but was not accounted for in determining the labor pool. Let f_1 (f_2) be the respective fractions of the minority (majority) members of the labor pool possessing factor X. The analog of Cornfield's result is

Lemma 2.1: For factor X to fully explain the disparity between *s* and π, both R and Θ must exceed ψ, where:

$$\text{(A.1)} \qquad \Psi = \frac{\pi/(1-\pi)}{s/(1-s)} \text{ and } \theta = f_2/f_1.$$

PROOF

Denote the probability an individual without factor X is successful by p, so that Rp is the corresponding probability for someone with X. The expected number of minority hires (successes) is N_1 {2 $(1-f_1)p+Rf_1p$} and the expected number of majority hires is N_2 {$(1-f_2)p+Rf_2p$}. For the observed fraction (s) to be the expected fraction of hires,[68] the following must hold:

$$s = \frac{N_1\{(1-f_1)\,p + Rf_1p\}}{N_1\,\{(1-f_2)\,p + Rf_1p\} + N_2\,\{(1-f_2)\,p + Rf_2p\}},$$

$$= \frac{\pi\{1 + (R-1)\,f_1\}}{\pi\{1 + (R-1)\,f_1\} + (1-\pi)\,\{1 + (R-1)\,f_2\}}.$$

Therefore, $$\frac{s}{1-s} = \frac{\pi\{1 + (R-1)\,f_1\}}{(1-\pi)\,\{1 + (R-1)\,f_2\}} \quad \text{or}$$

$$\text{(A.2)} \qquad \Psi = \frac{\pi/(1-\pi)}{s/(1-s)} = \frac{1 + (R-1)\,f_2}{1 + (R-1)\,f_1} = \frac{1-f_2 + Rf_2}{1-f_1 + Rf_1}$$

[68] Recall that Cornfield's lemma ignored sampling error and required the new factor to fully explain the data, i.e., the observed data should be at its expected value.

Because $s < \pi$, $f_2 > f_1$, or $\Theta > 1$. The largest possible value of the right side of (A.2) is R, which is achieved when $f_2 = 1$ and $f_1 = 0$. Thus, $\Psi \leq R$, solving (A.2) for Θ yields:

$$\text{(A.3)} \qquad \Theta = \frac{(\Psi - 1)}{(R-1)\, f_1} + \Psi \text{ which exceeds } \Psi \ (\Psi > 1 \text{ because } s < \pi).$$

Comment: Equation (A.3) yields more than Cornfield's original result, since it gives the amount by which the prevalence ratio (Θ) must exceed Ψ in order for the observed data to be at its expected value. Equivalently, it means that the prevalence (f_2) of the characteristic in the majority group should be $\Psi f_1 + (\Psi - 1)/(R-1)$ in order for the disparity to be fully explained. This indicates that when a modest fraction (f_1) of the minority group has the trait and Ψ exceeds two or three, it will be difficult for the prevalence condition to be satisfied unless the effect (R) of the omitted factor is very large. Under these conditions, once the major job-related characteristics have been used to define the labor pool, it is virtually impossible for an omitted characteristic to "explain" a statistically significant disparity.

VIRTUAL REALITY: COPYRIGHTABLE SUBJECT MATTER AND THE SCOPE OF JUDICIAL PROTECTION

Andrew H. Rosen*

Less than ten years ago, before word processing replaced typing pools, before spreadsheets turned business people into computer jockeys, before telecommunications linked the electronic cognoscenti into a packet-switched Worldnet, a small but fanatic subset of computer enthusiasts understood that the personal computer would shape the way businesses worked, the way scientists operated, the way students learned, over the coming decades: the infonauts, who helped change the way we live today. An even smaller and at least equally fanatic subset of that subculture began to think not only about what lay beyond the horizon but about what might be found beyond that: the cybernauts, who are changing the way we live in the future.[1]

I. INTRODUCTION

A. Coming Attractions of an Emerging Technology

Pacing across the courtroom during his closing argument, Duncan saw the juror to the far left in the second row begin to doze. He knew he had to either place himself directly in front of that juror and address her personally, or tone down the legal rhetoric. He walked toward her and chose the latter. Just at that moment, a dim light in his right peripheral vision began to blink. Duncan raised his right hand and pointed to the light; in a small box to the top right of his vision appeared his secretary. The courtroom scenario paused.

*Member, California State Bar, B.A. University of Michigan, J.D. Southwestern University School of Law, candidate for LL.M. in Taxation, New York University School of Law. The author wishes to thank Professors Lawrence Sullivan and Robert C. Lind, Southwestern University School of Law, for their encouragement and suggestions. This article has been entered in the Nathan Burkan Competition.

[1] HOWARD RHEINGOLD, VIRTUAL REALITY 135 (1991).