

1993

Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims

Joseph L. Gastwirth

Follow this and additional works at: <http://scholarship.law.umn.edu/lawineq>

Recommended Citation

Joseph L. Gastwirth, *Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims*, 11 LAW & INEQ. 151 (1993).

Available at: <http://scholarship.law.umn.edu/lawineq/vol11/iss1/3>

Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims

Joseph L. Gastwirth*

Introduction

Employment discrimination is an enduring problem in this country, preventing the employment or impeding the advancement of certain members of society. While Title VII has done much to eliminate practices which are openly discriminatory, many requirements for employment and promotion function with discriminatory effect. "Disparate impact" describes the inhibitory effect of facially neutral employment and promotion practices on protected groups.¹ Disparate impact claims pose complicated problems to plaintiffs, employers and the courts.

The current methods of statistical analysis used by courts to evaluate disparate impact claims are insufficient to yield fair results. These methods often fail to examine the persuasiveness of

* Joseph Gastwirth is a professor of statistics and economics at George Washington University. He has a B.S. in mathematics from Yale University, 1958, and a Ph.D. in mathematical statistics from Columbia University, 1963. Professor Gastwirth has published a number of articles on statistical methodology and its application in the legal setting and is a member of the American Statistical Association's Committee on Law and Justice Statistics. He wishes to express his appreciation to the editors of *Law and Inequality* for their helpful suggestions.

1. Title VII prohibits discrimination due to membership in a protected class (race, color, sex, religion or national origin). 42 U.S.C. § 2000e-2(a) (1988). When the Supreme Court interpreted Title VII in *Griggs v. Duke Power Co.*, it distinguished cases focusing on the consequences of specific employment practices from cases concerning whether the employer intended to discriminate. *Griggs*, 401 U.S. 424, 432 (1971). A practice, typically a job requirement, which eliminates a substantially higher fraction of a protected class than other applicants or employees is said to have a disparate impact. In *Griggs*, the Court stated that practices having a disparate impact need to be shown to be related to successful performance of the job. *Id.* at 431. Cases concerning whether an employer was motivated, at least in part, by discriminatory factors are called disparate treatment cases. As the Court noted in *Int'l Bro. of Teamsters v. United States*, the issue in disparate treatment cases is whether "[t]he employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin." *Int'l Bro. of Teamsters*, 431 U.S. 324, 335 n.15 (1977). "Proof of discriminatory motive is critical, although it can in some situations be inferred from the mere fact of differences in treatment." *Id.* The use of statistics pertaining to similarly qualified individuals in disparate treatment cases is described in CHARLES A. SULLIVAN, MICHAEL J. ZIMMER AND RICHARD F. RICHARDS, *EMPLOYMENT DISCRIMINATION*, 251-264 (2d ed. 1988).

an employer's asserted reasons for disparities in its work force. One case which illustrates the problems with the current use of statistical methods is *Wards Cove Packing Co., Inc. v. Atonio*.² While parts of the holding in *Wards Cove* have been changed by the 1991 amendments to Title VII, the statistical analysis is still used by courts to illuminate the basic issues. With a more scrutinizing statistical analysis, the Supreme Court might have reached different conclusions, finding in favor of plaintiffs on some of their claims. A number of other recent cases further demonstrate the need for improved statistical analysis of disparate impact claims.

This article examines current statistical methods used by courts to evaluate disparate impact claims, and suggests an additional method for improving statistical analysis in this context. Part I describes the two predominant methods currently used by courts and examines their application in several recent cases. Part II introduces Cornfield's method, a superior means for courts to evaluate an employer's asserted reason for observed disparities in hires or promotions. Part III applies this new method to cases focusing on two types of data to demonstrate how the outcomes of disparate impact claims can depend on statistical analysis. Part IV suggests several ways to improve the quality of recordkeeping to the benefit of all parties. The article concludes that the fair evaluation of disparate impact claims requires a more scrutinizing approach to analysis of data, more thorough recordkeeping by employers and more refined labor market data from the government.

I. Statistical Analyses Most Used by Courts

Presently, courts use two types of statistical analysis when confronted with a claim of disparate impact. These methods are "intuitive comparisons" and "significance measures."

A. *Intuitive Comparisons*

When the raw data presents a gross imbalance in the work force, courts may find discrimination through intuitive comparison.³ This method is applied only when the disparity appears quite large. For example, in *Fisher v. Transco-Services Milwaukee*,⁴ only one of 25 employees under 40 was terminated while 10 of 27 employees over 40 were.⁵ The difference in the termination rates

2. 490 U.S. 642 (1989).

3. MACK A. PLAYER, EMPLOYMENT DISCRIMINATION LAW, 344 (1988).

4. 979 F.2d 1239 (7th Cir. 1992).

5. *Id.* at 1245.

of younger (4%) versus older (37%) workers was nearly ten-fold. The court found this difference was clearly disproportionate.⁶

In *Griggs v. Duke Power Company*, the U.S. Supreme Court accepted the use of an intuitive comparison of raw data.⁷ The employer had implemented two new requirements for promotion on the day Title VII took effect.⁸ These requirements had a significant disparate impact on African Americans residing in North Carolina at the time.⁹ Neither standard was related, by design or intention, to an applicant's ability to learn or perform a particular job.¹⁰ The Supreme Court held that because the two requirements operated to inhibit promotion of a *markedly disproportionate* number of African Americans and were not job related, they were unlawful under Title VII.¹¹

The Court concluded: "Under [Title VII], practices, procedures, or tests, neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices. . . . If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."¹²

One major problem with using intuitive comparison of raw data to determine disparate impact is that the outcome may depend on whether the observed data measures the qualification or disqualification rate. Take, for example, a test which excludes 5 percent of African American applicants, but only 2.5 percent of White applicants. If disqualification rates are examined, a court may find disparate impact because twice the number of African Americans fail as Whites; however, if the focus of the inquiry is qualification rates, African Americans would qualify 95 percent of the time, compared to Whites' 97.5 percent — a ratio which passes

6. In reversing the lower court's summary judgment decision for defendant, the court stated, "it does not require expertise in differential equations to observe that an adverse ratio of approximately 10 to 1 is disproportionate." *Id.* at 1245. The court may be alluding to Judge Easterbrook's use of a partial derivative in *Branion v. Gramley*, 855 F.2d 1256, 1265 n.7 (7th Cir. 1988). See also D.H. Kaye, *Statistics for Lawyers and Law for Statistics*, 89 MICH. L. REV. 1520, 1538 (1991) (showing that the formula can be derived using only high school algebra).

7. 401 U.S. 424 (1971).

8. The requirements were a high school diploma and passage of two aptitude tests. *Id.* at 427-28.

9. The EEOC found, of one battery of such tests, that 58% of Whites passed the test, compared to only 6% of Blacks. *Id.* at 430, n.6. The Court noted that basic intelligence needed a means to manifest itself fairly in a testing process and Blacks had received an inferior education in segregated schools. *Id.* at 430.

10. *Id.* at 427-28.

11. *Id.* at 426 (emphasis added).

12. *Id.* at 430-31.

the government's four-fifths rule.¹³

B. Significance Measures

A significance measure is based on the probability that an observed difference in rates would occur by chance.¹⁴ If this probability is low, usually less than .05, we say that a party's data is "statistically significant." If there is a disparity in hires, but that disparity has a moderate probability of occurring by chance, then a court may find that the employer's practice doesn't have a disparate impact. Significance tests do not have the problem of qualification versus disqualification rates mentioned above.¹⁵ Nevertheless, results of significance tests require careful interpretation.¹⁶

C. Why These Methods Are Sometimes Inadequate

Intuitive comparisons and significance measures are inadequate to resolve all disparate impact claims fairly. Many practices which adversely affect minority groups will not involve a level of disparity that is outrageous on its face. For these cases, intuitive comparisons of raw data alone may not make plain a disparate impact. Furthermore, neither procedure is adequate when the employer does not attribute the disparity to chance, but to the

13. The Uniform Guidelines on Employee Selection Criteria focus only upon qualification rates. 29 C.F.R. § 1607.1 et seq (1992). The Supreme Court has examined both, but has not indicated when use of either is appropriate. *PLAYER*, *supra* note 3 at 364.

14. Examples of "significance" measures are the Z-score and the Fisher exact test. The Z-Score measures the number of standard deviations by which a particular observed outcome differs from its expected value, assuming random selections from the relevant pool. The Fisher exact test is useful when the sample size is small; this test calculates the probability of obtaining by chance a result that is at least as extreme as the actual observed result. For a thorough explanation of the use of these tests, see *PLAYER*, *supra* note 3 at 196-504.

15. The importance of this issue is illustrated by the decision in *Council 31 v. Ward*, reversing a district court's award of summary judgment to a defendant. *Council 31* 978 F.2d 373 (7th Cir. 1992). In *Council 31*, the employer's decision to reduce its work force at certain locations had a disparate impact on Blacks. Indeed, 8.6% (130 of 1512) of Blacks were terminated in contrast to only 3% (87 of 2900) of Whites. *Id.* at 375. Plaintiffs argued that the ratio of termination rates, 3.5 (3/8.6), clearly failed the EEOC's four-fifths rule while the defendants noted that the ratio of the retention rates, .94, satisfied it. *Id.* at 376. Applying Fisher's exact test to the data yields a p-value of less than one in a million. More importantly, if all 287 employees who were laid off were Black, 82.34% of the Black employees would have been retained. In such a case a court would not find disparate impact by applying the EEOC's four-fifths rule to retention rates, despite extreme facts.

16. See D.H. Kaye, *Is Proof of Statistical Significance Relevant?* 61 WASH. L. REV. 1333 (1986); *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN COURTS* (Stephen E. Feinberg ed. 1989) (discussing the limits of hypothesis testing and problems courts have in using the results of statistical tests).

presence or absence of a particular quality which the test is designed to measure. The employer then argues that it is this quality, measured by the test and present/absent in the applicants, which accounts for the disparity in those accepted or rejected; finally, the employer argues that the presence/absence of this quality is a business necessity. While intuitive comparison and significance measures cannot analyze the validity of these claims by defendants, the next section details a method which can.

II. Cornfield's Method — Explanation

A. How It Works

Cornfield's method quantifies the degree to which a "missing factor" must be present in majority groups in order to account for observed disparities between majority and minority groups in the pass rates of a hiring, firing or promotion system.¹⁷ Because it asks whether a particular "missing factor" could account for observed disparities — as opposed to asking whether the disparity could arise from chance alone — Cornfield's method is responsive to an individual employer's defense; it is more particularized to the facts of a case than an intuitive comparison or significance measure.¹⁸

17. The method was developed by the late Professor Jerome Cornfield to explore whether the difference in cancer rates between smokers and non-smokers could be due to some characteristic, other than smoking, which was more prevalent among smokers. For further discussion of Professor Cornfield's analysis, see 1 JOSEPH L. GASTWIRTH, *STATISTICAL REASONING IN LAW AND PUBLIC POLICY*, 296-98 (1988).

18. Adaption of Cornfield's Lemma to Pass or Hire Rates: let p_1 be the pass (or hire) rate of minority applicants, while p_2 denotes the pass rate of majority applicants. Let $R = p_2/p_1$ be the ratio of the rates. Note that R is the inverse or reciprocal of the selection rates used in the Uniform Guidelines.

Lemma: In order for a job-related factor, x , to explain disparity between the two rates, the factor must multiply one's chance (probability) 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 it.

Remark: It is intuitively clear that the missing factor (in our analysis) must cause an R -fold increase in the probability of passing as otherwise it simply could not fully explain the disparity.

Proof: Let R_x be the actual effect of the missing factor and assume it fully explains the disparity, i.e., sufficiently many more majority members have factor x . Thus, a person without x has probability p of passing, while a person with x has probability $R_x p$ of passing.

Let f_1 and f_2 be the fractions of minority and majority members possessing factor x ; these fractions are the prevalence of the factor and it is clear that $f_1 < f_2$.

The probability that a random member of the minority group will pass the test is

$$p_1 = f_1 R_x p + (1 - f_1) p$$

as f_1 minorities have factor x and $(1 - f_1)$ don't.

III. Cornfield's Method Applied

In this article's examination of the statistical data relied on by courts, two variations of Cornfield's approach are employed to assess whether a possible defect or missing factor in data offered by a party could actually explain an observed disparity. These variations are applicable to two types of data comparisons currently recognized by the courts in Title VII actions.¹⁹ The first type of data is "demographic" or "applicant pool" data used when the actual percentage of minority hires is compared to a potential or proxy labor pool derived from an external data source, e.g., census data on persons employed in similar jobs.²⁰ The second type of data is "applicant flow" data; this is used when the majority and minority pass rates of actual applicants or employees are compared with respect to a hiring or employment practice.²¹

Similarly, the probability that a random majority member passes the test is:

$$p_2 = f_2 R_x p + (1 - f_2) p$$

Our observed R_0 is of course p_2/p_1 or

$$1) R_0 = p_2/p_1 = \frac{f_2 p R_x + (1 - f_2) p}{f_1 p R_x + (1 - f_1) p} = \frac{R_x f_2 + (1 - f_2)}{R_x f_1 + (1 - f_1)}$$

[Note: the actual value of p drops out.]

In equation (1), we ask, what is the largest possible value our observed data R_0 can have? This occurs when the numerator is at its maximum, i.e., $f_2 = 1$ [$R_x > 1$ as x explains the disparity], and when the denominator is at its minimum, i.e., $f_1 = 0$ [no minority has x]. Then the numerator is R_x , denominator is 1, so R_x is largest possible value, or $R_0 \leq R_x$. This is the intuitive part, R_x must exceed the observed ratio p_2/p_1 .

Now consider the prevalence ratio $\theta = f_2/f_1$ or $f_2 = \theta f_1$ [e.g. if 30% of majority have x and 15% of minority do, then $\theta = 2$] Substituting θf_1 for f_2 in equation (1) yields:

$$(2) R_0 = \frac{\theta f_1 (R_x - 1) + 1}{f_1 (R_x - 1) + 1}$$

Cross multiplying (2) yields

$$R_0 f_1 (R_x - 1) + R_0 = \theta f_1 (R_x - 1) + 1$$

or

$$R_0 - 1 = f_1 (R_x - 1) (\theta - R_0)$$

Now one can observe that $\theta - R_0 = \frac{R_0 - 1}{f_1 (R_x - 1)}$ is a positive number

as both R_0 and R_x must exceed 1 (remember $R_0 > 1$ to begin with as we are trying to assess whether factor x explains the disparity, R_0) since $\theta - R_0$ is positive; $\theta > R_0$ and the prevalence condition is satisfied.

19. See PLAYER, *supra* note 3 at 357-61.

20. *Id.* at 357.

21. *Id.* at 359.

A. Demographic Data

Applied to demographic data, the method determines how small the percentage of minority members of the proxy labor pool (f) would have to be in order to reflect the percentage of minority members actually hired. Given the minority share (s) of actual hires, one determines the smallest fraction (f) of the proxy labor pool that minority members would have to form in order for the employer's hiring data to be statistically consistent with fair (random) selection from that pool. This minimal fraction (f) can be compared to the minority percentage (p) of the proxy labor pool to assess whether it is likely that an extra factor could reduce the minority share to f.²²

For example, in *Wards Cove Packing Co., Inc. v. Atonio*, decided in 1989, plaintiffs claimed disparate impact and disparate treatment.²³ In support of their disparate impact claim, plaintiffs relied on the composition of different classifications of workers within the employers' work forces to prove discrimination in at-issue jobs;²⁴ defendants made their arguments from census data.²⁵ Part of the Supreme Court's majority opinion focused on which proxy labor pool was more appropriate — plaintiffs' or defendants'.²⁶ The Court also described the "appropriate" shift in the burdens of production and persuasion, and established the "legitimate business purpose test."²⁷

The legal significance of *Wards Cove* is diminished by the 1991 amendments to the Civil Rights Act.²⁸ This act overturned the *Wards Cove* holdings regarding burden of proof,²⁹ and the "le-

22. This method was used persuasively in *Capaci v. Katz & Besthoff Inc.*, 711 F.2d 647 (5th Cir. 1983). In *Capaci*, plaintiffs developed several plausible proxy labor pools from which the defendant could select manager trainees. *Id.* at 651-52. Plaintiffs then showed that the chance that there would be no women among the 267 hires made between mid-1965 and January, 1973 was less than one in a billion, whichever proxy pool one used. *Id.* The appellate opinion also noted that as long as women comprised at least 2% of the appropriate labor pool, the employer's hiring data would be statistically significant. *Id.* at 653-54. As the female fraction of the proxy pools ranged from 16% to 29%, the court did not believe that factors other than the sex of the applicant (such as willingness to work weekend and night hours) that the defense claimed were not accounted for in these proxy pools, could reduce the female fraction of qualified persons to less than 2%. *Id.* at 652-53.

23. 490 U.S. 642, 648 (1989).

24. *Atonio v. Wards Cove Packing Co., Inc.*, 34 Empl. Prac. Dec. (CCH) ¶ 34,437, p. 33,841 (W.D. Wash. 1983) *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

25. *Id.* at 33,843.

26. 490 U.S. at 650-55.

27. *Id.* at 656-61.

28. Civil Rights Act of 1991, Pub. L. No. 102-166, § 3, 1991 U.S.C.C.A.N. (105 Stat.) 1071.

29. *Id.* at § 5, 1991 U.S.S.C.A.N. (105 Stat.) 1074.

itimate business purpose" test,³⁰ restoring the standards set out in *Griggs v. Duke Power Co.*³¹ *Wards Cove* may remain legally significant for cases brought after it but before the 1991 amendments.³² Importantly, however, the statistical issues arising in the case are not affected by the 1991 Act.

Though the factual basis for the *Wards Cove* decision rested on statistical analysis, the Supreme Court did not discuss how lower courts should evaluate statistical data. The Court did not indicate which job qualifications must be included in a proxy applicant pool, nor how to analyze multi-year data containing some duplicate applications from the same persons — two alleged flaws in plaintiffs' data set. Instead, the Court accepted defendants' data set without analyzing the degree to which these flaws may have affected the accuracy of plaintiffs' statistical analyses. As virtually no data set is perfect, some methods are needed to assess whether an alleged flaw in the data — e.g., an omitted factor or a modest fraction of multiple applications — could explain a large disparity in hiring or promotion rates. This section will apply an analysis similar to Cornfield's to the demographic data in *Wards Cove* to demonstrate how this method could have been used to evaluate the alleged flaws in plaintiffs' data.

1. Application of the Method to *Wards Cove*

In *Wards Cove*, nonwhites, including Chinese, Filipinos, Japanese and Native Americans, alleged that the hiring, promotion, pay and housing policies of three Alaskan salmon companies discriminated against them.³³ The canneries and fish camps were located in remote, widely separated areas of Alaska.³⁴ Each was a self-supporting installation which operated only during the summer salmon run, a period of three to four months.³⁵ The first charge of discrimination by a plaintiff was filed in October 1973.³⁶

The nonwhites alleged that the companies' hiring practices had denied them employment as "noncannery" workers on the ba-

30. *Id.*

31. *Id.* at § 3, 1991 U.S.C.C.A.N. (105 Stat.) 1071.

32. *Id.* at § 402, 1991 U.S.C.C.A.N. (105 Stat.) 1099. See also Daniel V. Kinsella, *The Civil Rights Act and the Retroactivity Muddle*, 36 TRIAL LAWYER'S GUIDE 460 (1993) and Kristine N. McAlister, *Retroactive Application of the Civil Rights Act of 1991*, 45 VAND. L. REV. 1319 (1992).

33. *Wards Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642, 663, n.4 (1989).

34. *Id.* at 646.

35. *Id.* at 646-47.

36. *Id.* at 647. The Equal Employment Opportunity Commission (EEOC) had filed its own charge in June 1972. *Id.*

sis of race.³⁷ The noncannery positions typically paid more than the "cannery" jobs.³⁸ Most of the noncannery jobs required some specialized skill (e.g. mechanical, accounting, etc.) but there were some relatively unskilled jobs (e.g. kitchen help, dock crew, etc.).³⁹ Hiring for the noncannery jobs occurred at the defendants' offices in Washington and Oregon, and the hires were predominately White.⁴⁰ Cannery workers processed the salmon; these positions were filled by nonwhites, primarily Alaskan Natives residing in villages near the canneries and Filipinos hired through a Union (Local 37).⁴¹

The plaintiffs alleged that the following practices constituted disparate treatment and had a disparate impact on minorities:

- 1) Separate hiring channels for the cannery and noncannery positions and the use of word of mouth recruitment.
- 2) The use of nepotism in hiring.
- 3) Preferences given to past employees (rehires).
- 4) A lack of objective employment criteria.
- 5) Segregation of housing and eating facilities.⁴²

In support of their claim, plaintiffs introduced statistics showing that nonwhites were concentrated in the lower paying cannery jobs, while Whites held the vast majority of noncannery jobs.⁴³ By comparing the employers' noncannery work force to a "proxy applicant pool" consisting of cannery employees, plaintiffs sought to show that the cannery work forces were not representative of a fair (random) hiring process. Plaintiffs' proxy labor pool data incorporated a person's willingness to work in the remote locations of the canneries.⁴⁴ However, it failed to account for special abilities needed to perform the skilled noncannery jobs.

To rebut this evidence, defendants developed a geographic labor market from which they claimed to have recruited.⁴⁵ This

37. *Id.* at 647-48.

38. *Id.* at 642.

39. *Id.* at 647, n.3 (quoting *Atonio v. Wards Cove*, 768 F.2d 1120, 1123 (9th Cir. 1985) *rehearing granted and opinion withdrawn by* 787 F.2d 462 (9th Cir. 1985, *rev'd* 490 U.S. 642 (1989)).

40. *Id.* at 648.

41. *Id.*

42. *Id.* at 647-48.

43. *Id.* at 650.

44. Deposition of Prof. Flanagan at A-376, filed May 17, 1982, *Atonio v. Wards Cove Packing Co., Inc.*, 54 Fair Empl. Prac. Cas. (BNA) 1623 (W.D. Wash. 1983), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989). Plaintiffs' data covered employment in the entire Alaska salmon canning industry as measured by the U.S. Bureau of Fisheries (from 1906 to 1939) and the U.S. Fish and Wildlife Service (from 1941 to 1955). *Id.* This data showed the racial composition of persons who actually chose to work in these seasonal, migrant jobs. *Id.*

45. Affidavit of Dr. Rees at A-255, filed May 17, 1982, *Atonio v. Wards Cove*.

market consisted of Alaska, Washington, Oregon and California. "Availability fractions" were obtained from census data from those states.⁴⁶ These fractions are the minority proportions of all workers in the state with the skills and other qualifications needed to perform a particular job. They are determined from census data by matching the occupational categories used in the census to the jobs at issue.⁴⁷ Since an individual's willingness to travel or commute to the job site is not accounted for in census data, the defendant weighted the minority availability in each sub-area by the fraction of its employees who lived there at their time of hire.⁴⁸

The district court found that, even though the census data introduced by defendants was dominated by persons preferring full year fixed location employment, this data was appropriate for defining the labor supply for seasonal work such as in the canneries.⁴⁹ The court also found that Whites comprised about ninety percent of the available labor force for cannery and other non-skilled jobs.⁵⁰ The court did not define the labor market for the more skilled jobs but indicated that defendants had compared the fraction of White employees in various departments to the "relevant labor supply" figures.⁵¹ The district court found that no general pattern of over or underrepresentation emerged from that comparison.⁵²

The district court further found that about 48% of all employees in the Alaska salmon cannery industry were nonwhite and that nearly all nonwhites were employed in the cannery jobs.⁵³ Since nonwhites formed about 10% of the "laborer" job category in the Census, the court concluded that nonwhites were over-represented in these jobs.⁵⁴ It noted also that Alaskan Natives formed a high percentage of the population living near four of the canneries.⁵⁵ Hence, they comprised a high fraction of the resident can-

46. *Id.* at A-256-58.

47. The determination of the availability fraction for a job requires knowledge of the skills needed to perform the job and their relationship to the occupational categories used in the Census. An individual's willingness to travel or commute to the job site as well as the wage paid relative to similar jobs in the area should also be considered.

48. Rees affidavit at A-255.

49. *Atonio v. Wards Cove Packing Co., Inc.*, 34 Empl. Prac. Dec. (CCH) ¶ 34, 437, p. 33,829 (W.D. Wash. 1983) (finding 120), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

50. *Id.* at 33,828 (findings 106 and 107).

51. *Id.* at 33,829 (finding 123).

52. *Id.* at 33,829-30 (finding 123).

53. *Id.* at 33,829 (finding 121).

54. *Id.* at 33,828 (findings 105 and 107).

55. *Id.* (finding 109).

nery labor force.

On appeal, the Ninth Circuit held that the difference in the fraction of nonwhites holding cannery and noncannery jobs established a *prima facie* case of disparate impact in both unskilled and skilled noncannery jobs.⁵⁶ The Court of Appeals gave more weight than had the district court to plaintiffs' evidence of housing segregation, and held that the employers' practices of nepotism had a disparate impact as there were only 17 nonwhites among 349 nepotistic hires between 1970 and 1975.⁵⁷

The Supreme Court held that the cannery work force in no way reflected the pool of *qualified* applicants or the *qualified* labor force appropriate for the skilled noncannery workers.⁵⁸ It noted that in one cannery about fifteen percent of the new hires for office work were nonwhite and that if nonwhites formed no more than fifteen percent of the qualified applicants or relevant labor force it was "hard to see how the plaintiffs could make their case."⁵⁹

The Supreme Court's decision also rejected the Ninth Circuit's holding that the different racial proportions of employees in cannery and noncannery jobs established a *prima facie* case of discrimination with respect to the unskilled noncannery positions.⁶⁰ The court stated,

Racial imbalance in one segment of an employer's work force does not, without more, establish a *prima facie* case of disparate impact with respect to the selection of workers for the employer's other positions, even where workers for the different positions may have somewhat fungible skills (as is arguably the case for cannery and unskilled noncannery workers). As long as there are no barriers or practices deterring qualified nonwhites from applying for noncannery positions, if the percentage of selected applicants who are nonwhite is not significantly less than the percentage of qualified applicants who are nonwhite, the employer's selection mechanism probably does not operate with a disparate impact on minorities. Where this is the case, the percentage of nonwhite workers found in other positions in the employer's labor force is irrelevant to the question of a *prima facie* statistical case of disparate impact.⁶¹

Statistically speaking, both plaintiffs' and defendants' proxy

56. *Atonio v. Wards Cove Packing Co., Inc.*, 827 F.2d 439, 444 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

57. *Id.* at 445, 447-49.

58. *Wards Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642, 651 (1989).

59. *Id.* at 652.

60. *Id.* at 653.

61. *Id.*

applicant pools were inadequate. Plaintiffs' pool failed to account for skill differences between skilled and unskilled noncannery jobs; defendants' pool failed to incorporate as a job qualification, willingness to relocate temporarily to Alaska. In choosing between the two proxy applicant pools, the Court simply preferred one weakness over another. This result was not inevitable. Reexamining the data highlights three important errors in the courts' analysis of this case: 1) defendants' estimate of the minority "availability fraction" for the unskilled noncannery jobs was too low, 2) nonwhites were seriously underrepresented in defendants' nepotistic hires and 3) plaintiffs' data concerning the Bumble Bee and Red Salmon facilities supported a *prima facie* case against Bumble Bee by Native Alaskans.

Defendant's estimate of the minority "availability fraction" for the unskilled, noncannery jobs was based on the assumption that the percentage of minorities in the unskilled labor pool was the same as that of the overall labor pool, skilled and unskilled.⁶² This assumption is illogical. Nonwhites (especially Native Alaskans) comprised a lower fraction of the highly skilled, noncannery workers than the nonwhite percentage of the overall labor force in the four-state region. Common sense implies that they will therefore form a larger percentage of the labor pool available for unskilled jobs, both cannery and noncannery. A comparison of the percent of nonwhites employed as managers with the corresponding percent of laborers in Table 1 illustrates this point. Notice that the relationship between minority availability fractions for the occupational categories is similar in all the geographical regions. Thus, overrepresentation in one (unskilled) category and underrepresentation in another (skilled) are inextricably linked.

An examination of the affidavit of the defendants' expert indicates that the availability fractions were from a four-state weighted labor market developed using a three-step process.⁶³

62. Affidavit of Dr. Rees at A-255, files May 17, 1982, *Atonio v. Wards Cove Packing Co., Inc.*, 54 Fair Empl. Prac. Cas. (BNA) 1623 (W.D. Wash. 1983), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

63. Step 1. For each county with a population of 250,000, or each Census group of counties aggregated to have at least that population, the defendants' expert determined the proportion of all employees who resided there at the time of their hire. *Id.* at A-254. (The Census file used by defendants ((the 1% sample)), does not report data for counties with fewer than 250,000 residents. It combines adjacent small counties in order to reach that threshold. As a result, the state of Alaska is a single county group in this data file. *Id.*) This proportion estimates the share of hires that should come from that sub-area of the labor market, assuming that there are no barriers to all potential applicants.

Step 2. Using the Census 1% Public Use Sample, one determines the minority fraction of the relevant labor force, i.e., persons in each county group possessing the

TABLE 1 The Minority* Percentage of Persons Employed in 1980 in Various Occupations by Geographic Region.

Region	OCCUPATION**		
	Manager	Service	Labor-Handler
All Four states	9.00	16.76	13.51
Alaska	9.53	26.82	18.49
Rural Alaska	7.13	36.67	32.72

Source: Author's calculations from Table 78 of the General Social and Economic Characteristics Volume for each state from the 1980 Census.

* The Minority Category includes African Americans, Native Americans and persons of Asia-Pacific origin. The relative fraction of these groups varies within the states. Asians form the largest minority in California; Native Americans the largest in Alaska.

** The percentages in the table are used to assess whether an employer's work force can be regarded as a sample from the relevant labor pool. If one believes that the appropriate labor market for service workers is the four-state area, one compares the minority percentage of such employees to 16.76%. If one felt that the rural Alaska area was most appropriate because most people would not travel far for such low paying jobs, one would compare the minority percentage of service employees to 36.67 to assess whether hiring seemed fair.

relevant skill. (The Census Bureau issues a 1 in 100 random sample of individual records for research use. *Id.* at A-253-54.) The defendants calculated availabilities for many job groups but this analysis will focus on the overall figure.

Step 3. The minority share of the labor force for each type of job is a weighted average of their availability fractions in each county group. The weights are the county group fractions obtained in step 1. *Id.* at A-255.

The Census 1% Public Use Sample is much smaller than the 15 to 20% sample actually used in the Census publications reporting occupational, labor force and educational characteristics. Thus, the availability fractions used by the defendants' expert were subject to greater variability than usual in these cases. This variability was properly incorporated in the statistical tests relied on by the defendants' expert. *Id.* at A-269.

However, there remain serious statistical questions concerning the defendants' proposed labor market. Because of the population constraint in the creation of county groups by the Census Bureau, the state of Alaska is just one county group. *Id.* at A-254. Thus, the weighting procedure did not give areas near the canneries their proper weight. More importantly, it is preferable to base the weights used in step 1 on the applicant data, rather than on the residence pattern of employees, as the employees reflect the hiring practices of the defendant. See DAVID C. BALDUS & JAMES W.L. COLE, STATISTICAL PROOF OF DISCRIMINATION 61-67 (1987 Supp.); see generally SHELDON E. HABER & JOSEPH L. GASTWIRTH, *Specifying the Labor Market for Individual Firms*, 101 MONTHLY LAB. REV. 26 (Aug. 1978). This is especially important when the effect of practices such as "word of mouth" recruiting, nepotistic hiring or a preference for rehiring former employees is at issue, as here. The difference between minority availability percentages obtained from employee versus applicant weights can be substantial. See *Markey v. Tenneco Oil Co.*, 635 F.2d 497 (5th Cir. 1981), *aff'd*, 707 F.2d 172 (5th Cir. 1983). In an action for discriminatory hiring practices under Title VII, the court in *Markey* affirmed the use of weights based on the percentage of minority applicants from a particular area rather than the percentage of minorities from the areas in which employees re-

From this weighted labor market, defendants estimated the overall minority availability for the skilled noncannery jobs at 10.1%.⁶⁴ Based on the questionable assumption regarding the relationship between availabilities for skilled and unskilled jobs mentioned above,⁶⁵ defendants estimated that the minority availability for the unskilled noncannery jobs was also 10.1%.⁶⁶ A more accurate estimate would have been higher.⁶⁷

Applying Cornfield's approach to limited applicant data given in the decision on remand⁶⁸ also indicates that defendants' availability figure is too low. Of 1,966 applicants for the at-issue jobs, 539 identified their race.⁶⁹ Of these, 138 were nonwhite and 401 were White, i.e., nonwhites formed 25.6% of racially identified applicants.⁷⁰ The district court discounted this data because only a quarter of the applications had racial information.⁷¹ In fact, the court concluded that the 138 minority applications formed only 7% of the total applications and may have considered this percentage as their availability.⁷² This conclusion is based on an unreasonable assumption: that the 1,427 non-racially identified applicants were all White.

In order for the racially identified applicant data to be consistent with the defendants' 10.1% availability figure, nonwhites must be far more likely to identify their race on an application than Whites. Cornfield's method shows that nonwhites would need to racially identify three times more often than Whites, a rather

sided. The court reasoned that using applicant data would deter a business from hiring its employees from a largely White area in an effort to limit the percentage of minorities in the potential employee pool. *Id.* at 500. Weighting by applicants' residences yielded an availability of 42%, while weighting by the residence of employees at their time of hire yielded an availability of 32.6%. Compare *Markey v. Tenneco Oil Co.*, 707 F.2d 172, 173 (5th Cir. 1983) with *Markey v. Tenneco Oil Co.*, 635 F.2d 497, 500 n.3 (5th Cir. 1981). This difference is quite noticeable.

64. Rees Affidavit at A-276 (Subtracting 89.9% from 100% Whites). The defendant's questionable weighting process implies that the minority availability figures accepted by the court were lower than they should have been.

65. *Supra* note 62 and accompanying text.

66. *Atonio v. Wards Cove Packing Co., Inc.*, 34 Empl. Prac. Dec. (CCH) ¶ 34,437, p.33,828 (W.D. Wash. 1983) (finding 107), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

67. The weighting system utilizing residence of employees at time of hire rather than applicants is the most serious statistical problem in a case where fair recruitment is a major issue and the racial mix in the different sub-areas of the labor market is so varied.

68. *Atonio v. Wards Cove Packing Co., Inc.*, No. C5-74-145-JLQ, 1991 WL 67529 (W.D. Wash. 1991).

69. *Id.* at *2.

70. *Id.*

71. *Id.* The employers could have been more diligent in obtaining this information.

72. *Id.*

large factor.⁷³ A more conservative assumption of 1.5:1 or 2:1 would estimate the minority share of the 1,966 applicants at significantly higher percentages: 18.7% and 14.7% respectively.⁷⁴

Combined with a careful analysis of the assumptions upon which defendant's "availability fractions" were based, Cornfield's approach creates serious doubt about the accuracy of defendant's estimate and suggests that plaintiffs' data, though flawed, may have been preferable, especially in regard to the lesser skilled non-cannery jobs. At the least, the Court should have given both estimates weight in its determination of whether defendants had discriminated, rather than rejecting one estimate as inaccurate in favor of another estimate which is equally flawed.

A similar technique can be used to assess the impact of nepotism on nonwhites. The nonwhite share of nepotistic hires, 4.87%,⁷⁵ is statistically consistent with a nonwhite availability of 7.2% or less using the 95% level of confidence.⁷⁶ In order for the nonwhite hire rate of 4.87% to fall within two standard deviations⁷⁷ of what would be expected given fair hiring, the nonwhite availability must be no more than 7.2%. The criteria for three standard deviations yield a corresponding figure of 8.4%.⁷⁸ Since the defendant's availability figure, 10.1%, is greater than 8.4%, the nepotistic hires cannot have been drawn from a pool with the same nonwhite availability figures as the four-state labor pool accepted by the court. This indicates that the defendant's use of nepotism

73. If minority members formed 10.1% of the 1,966 applicants, there would be 199 minority and 1,767 majority applicants. This implies that 138/199 or 69.3% of the minority members answered the question while 401/1767 or 22.7% of the Whites did. Such a large difference in response rates should at least require a careful explanation. A ratio of 1.5:1 to 2:1 seems more reasonable than 3:1.

74. Suppose minorities are twice as likely as Whites to respond to the question about race. Let M and W denote the number of minority and White applicants, respectively, and let p be the probability that a White applicant responds. The following equations then hold: $M + W = 1,966$; $pW = 401$; $2pM = 138$. Their solution is: $p = .239$, $W = 1,677$, and $M = 289$. Hence, the minority fraction of applicants is $239/1966 = .147$.

75. $17/349 = 4.87\%$.

76. The upper end of the two sided 95% confidence interval is given by $p + 2p(1-p)/n$ or $(.0487) + 2(.0487)(.9513)/343 = .072$. See GASTWIRTH, *supra* note 17 at 164.

77. The Supreme Court generally assumes that the line of statistical significance lies at two standard deviations. See *Castenada v. Partida*, 430 U.S. 482 (1977); *Hazelwood School District v. U.S.*, 433 U.S. 299 (1977).

78. The upper end of the two sided confidence interval corresponding to the three standard deviation criteria is given by

$$p + 3 \frac{p(1-p)}{n} = (.0487) + 3 \frac{(.0487)(.9513)}{343} = .0836$$

as a hiring practice reduced the nonwhite fraction of hires below their availability in defendants' four-state labor pool.

A significance test shows that the number of actual nepotistic minority hires (17) is 3.24 standard deviations from the number expected (35.25) assuming a process of fair (random) selection from an applicant pool with a nonwhite availability fraction of 10.1%.⁷⁹ This result is generally considered highly statistically significant. Moreover, the above calculation assumed defendants' availability figure of 10.1%. Had the court used a more realistic figure, obtained by assuming that nonwhites were perhaps twice as likely as Whites to racially identify themselves, the disparity would be even more significant. Given the significance of the impact of nepotism, defendants should have been required to provide a stronger rebuttal than simply showing that only qualified relatives were hired.⁸⁰

Finally, the statistics concerning three skilled jobs at Bumble Bee and Red Salmon support a *prima facie* case against the Bumble Bee facility. Because Bumble Bee and Red Salmon were both located on Bristol Bay, an area with a high fraction of Native Alaskans,⁸¹ the two canneries faced the same labor market and should have hired from a common pool of potential employees. However, the fraction of nonwhites hired by Bumble Bee, most of whom presumably were Native Alaskan, was .68% (3/443) while the corresponding fraction among Red Salmon's hires was 7.9% (42/530).⁸² Applying Fisher's exact test to the difference between the two fractions of minority hires yields a statistically significant result at the .05 level.⁸³ Indeed, the p-value⁸⁴ is less than one in a billion.⁸⁵ This result is highly significant and should have provided strong support for plaintiffs' case against Bumble Bee.

79. Clearly, this exceeds the two to three standard deviation test accepted by the Court in *Castenada*. *Castenada v. Partida*, 430 U.S. 482, 496 n.17 (1977).

80. *Atonio v. Wards Cove Packing Co.*, 54 Fair Empl. Prac. Cas. (BNA) 1623, 1626 (W.D. Wash. 1983) (finding no nepotistic practice for hiring "at-issue" employees because the applicants were qualified for the positions), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989).

81. *Atonio v. Wards Cove Packing Co., Inc.*, 34 Empl. Prac. Dec. (CCH) ¶ 34,437, p. 33,828-29 (W.D. Wash. 1983) (finding 109), *rev'd*, 827 F.2d 439 (9th Cir. 1987), *rev'd*, 490 U.S. 642 (1989). Eighty-eight percent of the plaintiff class members were Native Alaskan. *Id.* at 33,829 (finding 123).

82. *Wards Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642, 677 n.24 (1989) (Stevens, J., dissenting).

83. A similar test of the Red Salmon data against the 10.1% figure does not yield a significant result at the .05 level.

84. The p-value is the probability that such a disparity in actual hires would occur given the same hiring rates at the Bumble Bee and Red Salmon plants.

85. This use of comparative data for firms in the same or similar labor markets is discussed in BALDUS and COLE, *supra* note 63. They note that defendants have also used such data to rebut allegations of discrimination. *Id.* at 50-51.

Cornfield's approach supports the inference that Bumble Bee discriminated against Native Alaskans. The smallest availability fraction consistent with the nonwhite fraction of Bumble Bee's hires at the two standard deviation or 95% confidence level criteria, is 1.5%. This is far below any of the availability figures in the record.⁸⁶ In fact, even using the 10.1% availability fraction accepted by the district court, Bumble Bee's nonwhite hires were 6.5 standard deviations below what would be expected by fair (random) hiring from the defendants' labor market.⁸⁷ It is surprising that none of the opinions realized this gross disparity nor required Bumble Bee to justify its hiring procedures.⁸⁸

Wards Cove illustrates the difficulty courts have when the relevant data is not available. Often, data on actual applicants ("applicant flow" data) including their background qualifications can be used. However, federal guidelines do not require the retention of personnel records for seasonal jobs as is required for full-time, permanent jobs.⁸⁹ In this case, too, the employers' hiring practices — nepotism and hiring at sites other than the plants themselves — were such that data on actual applicants were likely to be nonrepresentative of the pool of potential applicants. Moreover, both plaintiffs' and defendants' proxy labor pool data were nonrepresentative. Cornfield's method highlights the degree of the flaws in both data sets, and would have enabled the court to give appropriate weight to both parties' data in rendering its decision.

The preceding *Wards Cove* analysis shows how Cornfield's approach and related techniques could be used by plaintiffs. The following decision indicates how this new method can also serve defendants.

2. Demographic Data in *Wooden v. Board of Educ.*

Though it frequently may be used to give greater weight to plaintiffs' data in the face of defendants' assertions of flaws, Cornfield's approach should not be considered a plaintiff's tool. Indeed, the method may sometimes be used to strengthen defendants' re-

86. The lowest availability fraction was the 7% figure based on the assumption that all applicants who did not identify their race were White. *Supra* note 72 and accompanying text.

87. This is equivalent to a probability or p-value of less than 1 in 100,000.

88. *But see Ward's Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642, 677-78 (1989)(Stevens, J., dissenting).

89. 29 C.F.R. § 1602.49(b) (1992) (exempting "seasonal" employees from some record keeping requirements).

buttals because it highlights the degree to which an alleged flaw or missing factor may have affected a data set.

In *Wooden v. Board of Educ.*,⁹⁰ plaintiff alleged he was discriminated against on the basis of age. He had been a teacher in the area from 1959 to 1972 and had pursued private business for ten years.⁹¹ In 1983, he returned to teaching as a substitute teacher.⁹² He was hired as a part-time teacher in 1987.⁹³ He alleged that the Board's policies of giving a maximum of fourteen years credit for prior experience and giving greater weight to recent teaching experience had a disparate impact on applicants over forty.⁹⁴ Though there were over 2000 applicants,⁹⁵ plaintiff did not present any statistical data demonstrating the adverse impact of these practices.⁹⁶

The Board asserted that it was able to hire better qualified teachers because the applicant pool was so large, that the plaintiff had some unfavorable evaluations in his file and that his interviews were unimpressive.⁹⁷ Moreover, it demonstrated that 41% (36/86) of its 86 hires were over 40 years of age.⁹⁸ The district court granted summary judgment for the defendant and the Sixth Circuit Court of Appeals affirmed.⁹⁹

Obtaining a confidence interval for the fraction of the available labor pool over 40 shows that defendant's 41% figure is statistically consistent with fair (random) hiring as long as applicants over 40 comprised no more than 52.3% of the applicant pool.¹⁰⁰ The court did not determine a labor pool, but applicants typically are younger than the overall labor force.¹⁰¹ Census data indicates the over-40 percentage of the labor force for similar teaching positions would be less than 50%.¹⁰² Therefore, it is unlikely that ap-

90. 931 F.2d 376 (6th Cir. 1991).

91. *Id.* at 377.

92. *Id.*

93. *Id.*

94. *Id.*

95. *Id.* at 378.

96. *Id.* at 380.

97. *Id.* at 378.

98. *Id.*

99. *Id.* at 377.

100. As previously, we use the .05 level or two standard deviation criteria to define statistical significance. Smaller deviations are considered consistent, i.e., the probability that they would occur by chance exceeds .05.

101. See GASTWIRTH, *supra* note 17 at 252-53.

102. In 1980, the median age of secondary school teachers in the nation was 36.3 years for males and 34.8 years for females. The corresponding median ages for elementary school teachers were 36.2 years for males and 35.8 years for females. The detailed data is given in 1980 CENSUS REPORT, DETAILED POPULATION CHARACTERISTICS, U.S. SUMMARY SECTION A: UNITED STATES, Table 280. The data refers to

plicants over 40 formed over 52.3% of the applicant pool. The district court's finding is strengthened by the application of Cornfield's approach.

B. Applicant Flow Data

Cornfield's method is designed to evaluate whether an observed disparity of minority versus majority pass or success rates could be due to or explained by an omitted factor.¹⁰³ The method is relevant for disparate impact analysis as government guidelines use the "four-fifths" or "80%" rule.¹⁰⁴ The 80% rule states that whenever the selection ratio — the ratio of the minority pass rate to the majority pass rate — is less than four-fifths, the employer's requirement or test has a disparate impact and should be shown to be job related.¹⁰⁵ In order for a missing variable to account for a disparity meeting the 80% rule, the fraction of majority applicants possessing the factor must be at least 1.25 times the corresponding fraction of minority applicants and the factor should increase one's chance of passing the test or meeting the requirement by 1.25. If the selection ratio is .5 (50%) then the missing factor would have to double one's chance of passing the test, and the fraction of majority applicants possessing it must be twice the corresponding fraction of the minority applicants. Hence, the smaller the selection ratio is, the greater the difference must be between majority and minority groups with respect to qualifications that were not accounted for in a statistical comparison; in order for the disparity to be due to these qualifications rather than race, sex, etc., majority groups must be much more likely to possess the quality.

1. Applicant Flow Data in *Allen v. Seidman*

In *Allen v. Seidman*, plaintiffs were a class of African American bank examiners who failed the "Program Evaluation" test used by the FDIC to decide whether to promote examiners from the GS-9 level to the GS-11 level.¹⁰⁶

The district court found that the program evaluation test had a disparate impact because 39% of the Black candidates passed compared to 84% of Whites; this pass rate was highly significant.¹⁰⁷

employed persons, so recent graduates, the majority of whom are under 40, would not be included.

103. For further discussion, see GASTWIRTH *supra* note 17 at 296-98.

104. Uniform Guidelines for Employer Selection Procedures, 29 C.F.R. § 1607.4(D) (1978).

105. *Id.*

106. *Allen v. Seidman*, 881 F.2d 375, 378 (7th Cir. 1989).

107. *Id.*

The FDIC claimed that the data did not demonstrate a disparate impact because: passing the test was not equivalent to being promoted as 56% of the Black candidates and 92% of the Whites were promoted to GS-11 within one year after they took the exam; the plaintiffs' analysis was simplistic in that it did not account for other possible factors, such as education, that may have affected performance on the test.¹⁰⁸

The Seventh Circuit, in an opinion written by Judge Posner, dismissed the first argument, noting that the relationships between the exam pass rates and promotion success rates for Blacks and Whites were approximately the same;¹⁰⁹ virtually everyone who passed the test was promoted.¹¹⁰ Moreover, only 27% of Blacks who failed the test were promoted within one year, compared to 53% of Whites.¹¹¹

Judge Posner's opinion observed that if the two groups, Black and White GS-9 bank examiners, had substantially different skill characteristics, as in the skilled and unskilled jobs in *Wards Cove*,¹¹² then the plaintiffs would need to take them into account.¹¹³ In the present case, all the examiners had worked for the FDIC for between 5 and 15 years, served one or more years at the GS-9 level and obtained a recommendation from their regional director in order to take the exam.¹¹⁴ As the pool of test takers seemed reasonably homogeneous with respect to relevant job characteristics, the large disparity in success rates suggested racial bias. The opinion mentioned a few possible ways in which the Black candidates might have been less qualified.¹¹⁵ However, it emphasized that the FDIC did not demonstrate an imbalance in any of the relevant background characteristics.¹¹⁶ The court observed that it is easy to take "pot shots" at a statistical analysis and indicated that the defendant needed to do more than raise a potential flaw.¹¹⁷ Defendant should show that the alleged defect had a major effect on the ultimate inference by performing an appropriate reanalysis.¹¹⁸

108. *Id.*

109. *Id.*

110. *Id.*

111. *Id.*

112. 490 U.S. 642 (1989).

113. *Allen v. Seidman*, 881 F.2d at 379.

114. *Id.*

115. *Id.* at 379-80.

116. *Id.*

117. *Id.* at 380.

118. *Id.* The opinion correctly observed that a regression analysis, which included the alleged omitted variables, might have strengthened plaintiffs' case instead of reducing the racial disparity. *Id.*

Applied to applicant flow data, Cornfield's method supports the appropriateness of the simple comparison of pass rates when the candidate pools are approximately comparable with respect to background qualifications. For example, in *Allen*, the ratio of the White pass rate to the Black pass rate was $84/39 = 2.15$. In order for some other factor to explain a disparity of this magnitude, the factor, by itself, must increase one's chance of passing by at least 2.15 and 2.15 times as many White candidates as Black candidates must possess the factor.¹¹⁹ Thus, a modest imbalance in, say, whether the candidates had a master's degree in a business subject, could not fully explain the disparity between the pass rates in *Allen* even if it doubled one's chance of passing the test because the second condition (prevalence) would be unlikely to be met; in other words, though possessing a master's degree in a business subject might have doubled an employee's chance of passing the "Program Evaluation" test, it was unlikely that twice as many Whites had Master's degrees in a business subject given the homogeneity of the class of candidates for promotion.

Of course, to obtain the best estimate of the effect of race, one would have to use an appropriate statistical technique that would incorporate the additional factor which defendant alleged created the racial disparity. Cornfield's method, however, shows that the disparity in pass rates could not be fully explained unless the new factor satisfied the strength of relationship and prevalence requirements. From the information in the opinion, it is difficult to imagine that the proportion of White candidates possessing a factor so highly related to passing — i.e., a factor which makes its possessor 2.15 times more likely to be promoted — was twice the proportion of Black applicants possessing the factor, particularly when that factor could not be identified by the FDIC.

2. Applicant Flow Data in *Hill v. Miss. State Employment Serv.*

In *Hill v. Miss. State Employment Serv.*, the plaintiff, Ms. Hill, challenged the state's job referral system's reliance on subjective traits and out of code job referrals.¹²⁰ In order to support her claims of disparate treatment and disparate impact, plaintiff

119. The general result given in GASTWIRTH, *supra* note 17 at 296, states that to explain a ratio of R between two rates, a factor must create an R-fold increase in the rate (of passing a test) and must be R times more prevalent in the group with the higher rate.

120. 918 F.2d 1233, 1235 (5th Cir. 1990). Out of code referrals were referrals to jobs other than that or those to which the applicant specifically applied.

showed statistically significant racial disparities in referrals for the waitress and cashier jobs.¹²¹ Defendant's expert criticized plaintiff's analysis because it did not incorporate education and experience factors; however, he noted that these changes would only reduce the number of standard deviations calculated by the plaintiff's expert.¹²² Nonetheless, the district court found for defendant.

The majority decision upheld the lower court, in part relying on the plaintiff's failure to factor in the effect of education and experience.¹²³ Judge Rubin's dissent noted that on several occasions rather unusual referral patterns occurred; for example, when filling a request for waitresses, the Employment Service sent 22 Whites but no Blacks for an interview.¹²⁴ Ten of the Whites were "out of code," three failed to satisfy the minimum experience requirement and five lacked the requisite education.¹²⁵ Data from the employment service showed that nonwhites formed about 40% of all job applicants.¹²⁶

Using the methods described in this article, one can respond to defendant's criticism of plaintiff's data. The largest fraction of the eligible pool which Blacks could comprise in order for the observed data, p , (none out of 22) to have a reasonable probability of occurring by chance would be $p = .156$.¹²⁷ Since the fraction of Black applicants for such jobs was more than twice that figure, i.e., $p \geq .40$, it is unlikely that an even-handed application of education and prior experience criteria would have reduced the Black fraction of persons qualified for waitress referrals to less than .156. Furthermore, it was illogical to require plaintiffs to account for factors that defendants merely said they used when the actual data shows the criteria were not used systematically. These considerations would have supported Judge Rubin's dissent, especially as he identified several similar referral patterns indicating that this ex-

121. *Id.* at 1235 n.5. Statistical comparisons of the pool of applicants in the job codes and the Service's referrals for waitress and cashier positions yielded standard deviations of 8.78 and 3.41 from their expected values. Both of these test results are highly statistically significant.

122. *Id.* at 1237. From the opinion, it appears that the defendant did not show that incorporating these factors would reduce the standard deviation of the shortfall in Black referrals to less than two. Indeed, the dissent indicates that defendant's expert conceded that a significant disparity would remain after these factors were included. *Id.* at 1243.

123. *Id.* at 1240.

124. *Id.* at 1242.

125. *Id.*

126. *Id.* at 1236.

127. Setting this probability at .025, corresponding to the two-sided two standard deviation criteria adopted by the Supreme Court, $p = .156$. See *supra* note 77.

ample was not an isolated one.¹²⁸

3. Applicant Flow Data in *Rose v. Wells Fargo*

In *Rose v. Wells Fargo*, two managers filed a charge of age discrimination after they were terminated following the merger of the Crocker and Wells Fargo banks.¹²⁹ The district court granted defendant's request for summary judgment and the plaintiffs appealed.¹³⁰ Of the 153 employees in Crocker's Special Asset Division, about half lost their jobs, including four of eight managers.¹³¹ The plaintiffs submitted data showing that 73.5% of the division's employees over 50 were terminated, while the corresponding figures were 34.1% for employees between ages 40 and 49 and 28.2% for those under 40.¹³² Another statistical analysis of the termination decisions showed that age was the most important factor in predicting whether an employee was discharged.¹³³

To respond to plaintiff's data, the bank showed that the average age of Wells Fargo and Crocker employees before the mergers was 35.80 and 36.61, respectively, while twenty-two months after the merger the average age of employees was 36.77.¹³⁴ Other data showed that the percentage of employees age 40 or over remained the same.¹³⁵ Moreover, the bank claimed that a greater proportion of discharges occurred at the management level — positions typically held by older employees — than in other jobs.¹³⁶ The Ninth Circuit accepted this explanation without requiring defendant to submit data, by age group, for the different job categories.¹³⁷ As

128. *Id.* at 1243.

129. 902 F.2d 1417, 1419 (9th Cir. 1990).

130. *Id.* at 1419-20.

131. *Id.* at 1420.

132. *Id.* at 1423.

133. Unfortunately, the nature of this analysis was not described. *Id.* at 1423. Most likely it was a logistic regression relating the probability of being discharged to various factors, including age.

134. *Id.* at 1423 n.5. The court's decision to accept this data on average age of employees two years after plaintiffs' terminations is logically questionable given that the same circuit, in *O'Brien v. Sky Chefs*, did not accept defendant's data from a period after plaintiff's complaint. *O'Brien*, 670 F.2d 864, 867 (9th Cir. 1982). Also, in a recent sex discrimination case, the Seventh Circuit said that evidence of harassment which occurred 13 months before a woman quit was not admissible because it could not have been the cause of her decision. *Tobey v. Extel/JWP Inc.* 985 F.2d 330, 333-34 (7th Cir. 1993).

135. *Rose*, 902 F.2d at 1423.

136. *Id.*

137. *Id.* at 1425. If the data for the different categories were available, the differences in termination rates in each could be combined as in GASTWIRTH, *supra* note 17, at 229-37, to assess whether there was a common age-related differential. Alternatively, variables indicating the type of job held could be included in a suitable logistic regression analysis.

workers over 50 were nearly twice as likely to lose their jobs as employees between 40 and 49, and nearly three times as likely as those under 40, it is not clear that holding a management post fully explains the disparity. Cornfield's method indicates that managers would need to have been fired at two to three times the rate of other employees and that employees over 50 would have to be twice as likely to hold managerial jobs as persons between 40 and 49.¹³⁸ Although these conditions might have been satisfied, the defendant should have been required to document them with relevant statistics. It is far from clear that employees over 50 would be twice as likely to be employed as managers as persons age 40 to 49.

4. Applicant Flow Data in *Gilty v. Village of Oak Park*

*Gilty v. Village of Oak Park*¹³⁹ illustrates how defendants can prevail using reasoning similar to Cornfield's by focusing on whether the practice at issue caused the plaintiff to lose a job opportunity. Plaintiff applied for promotion to police sergeant and took part in the evaluation process, which consisted of a written test (50%), performance evaluation (30%) and oral interview (20%).¹⁴⁰ When the eligibility list was issued, he ranked thirty-third out of fifty-one candidates.¹⁴¹ While he did well on the written test (the 11th highest score), he received the lowest score (22.977 of 30) on the performance assessment.¹⁴² In August 1987, he filed a charge of discrimination.¹⁴³

Plaintiff's case was based on the fact that no Black person had ever been a police sergeant or lieutenant in the Village; however, plaintiff did not submit data on the number of those positions or the number of such jobs which were filled during the relevant time period.¹⁴⁴ Plaintiff also noted that the other four Black applicants received low performance evaluations,¹⁴⁵ but did not show they were statistically significantly lower than those of the White applicants.

The Seventh Circuit found that even if plaintiff's weak statistical showing established a *prima facie* case, i.e., that the perform-

138. Similarly, employees over 50 would have to be three times as likely to hold managerial jobs as persons under 40.

139. 919 F.2d 1247 (7th Cir. 1990).

140. *Id.* at 1248.

141. *Id.* at 1249.

142. *Id.*

143. *Id.*

144. *Id.* at 1253.

145. *Id.* at 1249.

ance evaluation portion of the promotion process had a disparate impact, defendant showed that dropping that part of the exam would not have altered the plaintiff's ultimate failure to obtain the job.¹⁴⁶ Using an argument similar to Cornfield's, the court observed that in order for the plaintiff to have received one of the two promotions made from the list, he would have to justify an increase in his total score of at least 8.04 points.¹⁴⁷ Even had he received a perfect score of 30 his score would have been raised by only 8.023 points.¹⁴⁸ Moreover, no candidate had a perfect evaluation (the highest was about 28.5).¹⁴⁹ The opinion also found that even if the performance evaluation scores were eliminated, i.e., the applicants were ranked only on the written and oral exams, Gilty would move from thirty-third to twentieth.¹⁵⁰ Hence, the performance evaluation aspect of the promotion process could not have caused the plaintiff's failure to be promoted.

IV. Suggestions for Improving Employment Discrimination Data

The 1991 Act resolves many of the issues raised by the *Wards Cove* decision;¹⁵¹ however, the Act did not address the statistical analysis or the types of evidence upon which plaintiffs may rely when an employer has not maintained records which facilitate a narrowly focused analysis of its practices.¹⁵² This article demonstrates that inadequacies in the statistical data and analyses submitted to the court in *Wards Cove* forced it to choose between two labor pools — one incorporating skill level but not willingness to relocate, and the other the reverse.¹⁵³ The Courts' failure to use statistical tests and other tools to assess the soundness of the inferences drawn from the available data caused it to miss the strength of plaintiffs' case on the issues of nepotism and hiring for the

146. *Id.* at 1253 n.10. In their assessment that the statistical case was weak, both the district court and the Seventh Circuit felt that the sample size of 66 officers — only 5 of whom were Black — was too small. If the five Black officers received the five lowest evaluations, standard statistical procedures would find this significant. Indeed, the probability that this would occur by chance, i.e., assuming the groups were equally qualified, is less than one in a million. Courts would gain additional insight if they allowed the parties to analyze samples of smaller sizes. Nevertheless, because the defendant could have rebutted a *prima facie* case in *Gilty*, the lack of an appropriate statistical analysis of the evaluation scores probably did not affect the ultimate outcome. *Gilty*, 919 F.2d at 1253.

147. *Id.* at 1253 n.10.

148. *Id.*

149. *Id.*

150. *Id.* at 1254.

151. *Wards Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642 (1989).

152. See *supra* note 62 and accompanying text.

153. See *supra* notes 62-89 and accompanying text.

lesser skilled, noncannery jobs, especially at the Bumble Bee facility.¹⁵⁴

This article shows that many courts are not requiring plaintiffs to make a "perfect" statistical presentation, nor do they carefully scrutinize the defendants' criticisms to ensure that an asserted defect is serious enough to alter the ultimate inference of a statistically significant disparity. This article introduces two methods which should assist courts to assess imperfect data.

The following suggestions are intended to ensure that a sufficiently reliable database is available to examine charges of discrimination and the business necessity of practices having a disparate impact:

A. Record Keeping

The EEOC (or a new statute) should require all employers to preserve personnel records, including test results and personnel evaluations, for at least three years; these records should be retained longer if a charge is brought. If the data needed to establish a *prima facie* case is missing because of defendant's failure to comply with these requirements, courts should assume the information would have been sufficiently favorable to plaintiffs to establish a *prima facie* case.¹⁵⁵ As defendants would need these records to rebut a plaintiff's case, they will have an incentive for careful recordkeeping. Courts could no longer decide that an employer who destroyed applications 30 days after hiring¹⁵⁶ acted in good faith.

The labor law and industrial relations community should take steps to inform the Census Bureau of their need for reliable occupational, wage and educational data on a countywide basis. Unfortunately, the 1980 Census EEO tape only reported this information for Metropolitan Statistical Areas which include several counties. This over-aggregation makes it difficult to develop an appropriate weighted labor pool for the various occupations, especially for employers at a single site in a large metropolitan area. Employers would benefit from such information. They could develop more accurate affirmative action plans using their own appli-

154. See also Bill Shaw, Gary A. Moore and Michael K. Braswell, *Ward's Cove Packing Company v. Atonio*, 41 LABOR L.J. 183 (1990).

155. Of course, courts would ascertain whether the other criteria needed to establish a *prima facie* case, such as the availability of the job, were satisfied by the plaintiff.

156. See EEOC v. Alton Packaging Corp., 52 Fair Empl. Prac. Cas. (BNA) 1734 (11th Cir. 1990). For a general survey of the incentives to destroy evidence, see Charles R. Nesson, *Incentives to Destroy Evidence in Civil Litigation: The Need for Vigorous Judicial Action*, 13 CARDOZO LAW REV. 793 (1991).

cant data to create proper weights, thereby monitoring their recruitment and hiring policies to ensure fairness.

B. Statistical Analysis

Courts should take a more decision-theoretic approach to disparate impact claims and consider the economic and public policy implications of an employment or hiring practice at issue. Courts might consider the following policy questions when developing the legal standards:

- a. the magnitude of the statistical disparity,
- b. the number of individuals affected by the practice,
- c. whether similar jobs in similar industries have validated the practice in question,¹⁵⁷
- d. the difficulty of conducting a validation study for a particular position,
- e. the risk and/or cost to the public of having a somewhat less able person on the job.

If a practice has a large statistical disparity, say a selection ratio less than .6, meets the three standard deviation criteria and a large number of individuals are affected, then courts should require a greater correlation between the employment or hiring practice and a major productivity characteristic than in situations where a practice has a disparity just reaching statistical significance, a selection ratio near .8 and affecting only a modest number of individuals.

The above considerations did play a part in the early EEOC guidelines,¹⁵⁸ but this role was diminished when the Uniform Guidelines were established in 1978.¹⁵⁹ The new guidelines are more concerned with assuring that validation studies follow technical standards for the validation of psychological tests than with developing decision rules balancing the degree of job relatedness with the effect the practice has on protected groups, the economic value the practice creates and the health and safety of the public.¹⁶⁰

Statistical significance depends on the magnitude of the disparity and the sample size. Thus, when a large number of people

157. The 1978 guidelines mentions the possibility of relying on other validity studies in § 7 and discusses the possibility of cooperative studies in § 15 F. Uniform Guidelines on Employer Selection Procedures 29 C.F.R. §§ 1607.7 and 1607.15 (1992). The EEOC or Department of Labor might assume a major role in creation of industry-wide validation studies.

158. 35 Fed. Reg. 12,333 (1970).

159. 43 Fed. Reg. 38,290 (1978).

160. See NATIONAL RESEARCH COUNCIL COMMITTEE ON THE GENERAL APTITUDE BATTERY, FAIRNESS IN EMPLOYMENT TESTING: VALIDITY GENERALIZATION, MINORITY ISSUES, AND THE GENERAL APTITUDE TEST BATTERY, 209-34 (1989) (John A. Hartigan and Alexandra K. Wignor, eds.).

take a test, even small differences in pass rates can be statistically significant. This may have motivated the four-fifths rule. It makes sense to link the strength of a relationship between a practice and on-the-job performance to the magnitude of the impact and the number of people affected.¹⁶¹

While the increased record-keeping recommended earlier would increase costs, it should be fairly minimal in this computer era.¹⁶² Furthermore, the data on the validity of tests and related practices would aid employers in improving the productivity of their work force as well as eliminating pre-employment criteria shown to be unrelated to job performance.

In adopting some aspects of cost-benefit analysis in their analysis of employment practices, courts will need to ensure that protected groups are not disadvantaged in seeking upper-level positions because of the relatively small number of them. Plaintiffs should be allowed to consider data from a variety of similar positions, perhaps even in other locations, in order to examine the effect on protected groups of subjective practices typically used to fill upper-level jobs.¹⁶³ Finally, courts have incorporated risk to the public sector in cases involving police officers and airline pilots.¹⁶⁴

Statistical analysis should be a component of the legal decision-making process. The above considerations might allow an employer to use, for example, a word-of-mouth recruitment process to ensure a *small* readily available pool of potential employees, especially when a business is subject to relatively sudden, non-seasonal changes in the demand for its products. If a substantial fraction of employees were hired in that fashion and the employer had few minority employees, then courts could continue their policy of not accepting word-of-mouth recruitment as legitimate absent strong justification for the practice.¹⁶⁵

These recommendations will help to establish a more systematic and reliable database so that courts can assess business practices and their justifications in the same way in similar cases. If an

161. Otherwise, employers may need to spend large sums to validate criteria having a minimal impact on job opportunities for protected groups.

162. Similarly, the cost of more detailed Census tabulations should also be small as the data has been collected and the same program used to obtain the Detailed Characteristics information for metropolitan areas could be run for each county.

163. See Elizabeth Bartholet, *Application of Title VII to Jobs in High Places*, 95 HARV. L. REV. 947 (1982).

164. See *Spurlock v. United Airlines*, 475 F.2d 216 (10th Cir. 1972); *Pennsylvania v. O'Neill*, 465 F. Supp. 451 (D. Pa. 1979).

165. See *Thomas v. Washington Cty. Sch. Bd.*, 53 Fair Empl. Prac. Cas. (BNA) 1754 (4th Cir. 1990).

employer who preserves employment records can have them used to prove discrimination while employers who do not preserve them are excused, obviously there is no incentive to keep records. The majority noted in *Wards Cove* that an employer's justification of a practice having a disparate impact is subject to a reasoned review and stated: "A mere insubstantial justification in this regard will not suffice, because such a low standard of review would permit discrimination to be practiced through the use of spurious, seemingly neutral employment practices."¹⁶⁶ If employers are required to preserve their records and the government provides reliable detailed labor market data, a proper analysis of employment practices will be possible. With modern statistical techniques applied to a fairly comprehensive database, courts should be able to undertake a reasoned review of employment practices to ensure that prohibited factors are not playing a role in employment decisions, while encouraging employers to reward qualifications that increase productivity and efficiency.

166. *Wards Cove Packing Co., Inc. v. Atonio*, 490 U.S. 642, 659 (1989).

