THE ROLE OF PROBATIVE STATISTICS
IN EMPLOYMENT DISCRIMINATION
CASES

ANTHONY E. BOARDMAN* AND AIDAN R. Vining†

I
INTRODUCTION

Over the last thirty years the courts have relied increasingly on statistical evidence in employment discrimination cases. During this period the nature of acceptable statistical evidence has changed markedly. Initially, the courts typically compared the percentage of black employees in a firm with the percentage of blacks in the general population. If the disparity were substantial, the burden of proof would shift to the defense. Subsequently, the concept of statistical significance was introduced in court cases, and recently the results of regression analyses have been introduced. Each case, however, was judged on its merit; there was no simple rule for determining under what conditions the burden shifted to the defense.

Impetus for increased use of more sophisticated statistical methodologies has come from the guidelines promulgated by the federal administrative agencies that have responsibility for employment discrimination. Since Griggs v. Duke Power Co.,¹ these guidelines have been entitled to “great deference.”² Since 1978 an important change has occurred in the administrative guidelines. These agencies have developed an “80% rule”³ whereby an employer will not normally be prosecuted under Title VII if the selection rate or promotion rate of each group is more than 80% of the selection or promotion rate of any other groups.⁴ In practice, the rule and the provision mean that if the selection or promotion rate of a group is more than 80% of the selection or promotion rate of all other groups, an employer cannot be found to be in violation of Title VII.

Unlike previous guidelines and legal decisions, the four-fifths rule is relatively straightforward and easy to implement. Consequently, it is not surprising to find

---

2. Id. at 434.
3. The most recent version of the 80% rule is found in 29 C.F.R. §§ 1607.4(d), 1607.16(r) (1983). See also Adoption of Questions and Answers to Clarify and Provide Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed. Reg. 11,996 (1979).
4. See 29 C.F.R. § 1607.4(c) (1983). This provision states that, as a matter of administrative and prosecutorial discretion, in the usual case the agencies will not take enforcement action upon the disparate impact of any component of a selection process if the total selection process results in no adverse impact.
leading employment discrimination commentators, such as Blumrosen,\(^5\) to be in favor of the rule. In this article it is argued that both the rule and the administrative agencies are misdirected, and that the use of the 80% rule represents a step backwards and diverts the court from its progress towards more well reasoned decisions.

This article is divided into two parts. First, there is a review of the development of employment discrimination case law. This review categorizes cases and presents an analytic framework of the courts’ decisions. Second, there is an analysis of the potential impact and ramifications of the 80% rule. It is concluded, first, that the courts have progressed in their understanding of appropriate statistical methodologies and the situations in which these techniques should be applied, and second, that the 80% rule-of-thumb, while simple, has serious weaknesses.

II

THE DEVELOPMENT OF EMPLOYMENT DISCRIMINATION CASE LAW

A. The Burden of Proof in Discrimination Cases Under Title VII

Title VII of the 1964 Civil Rights Act states “[i]t shall be an unlawful employment practice for an employer . . . to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual’s race, color, religion, [or] sex.”\(^6\) This title, which Professor Schlei has referred to as “a mighty engine that is gradually forcing the alteration of the employment practices of a nation in innumerable ways,”\(^7\) has had an impact far beyond the expectations of its drafters. Originally, the bill was intended primarily to address adverse treatment, that is, malevolent discrimination towards individuals who argued that they had been the victims of overt discrimination. However, since Griggs in particular, Title VII has formed a basis for dealing with discrimination against protected groups, even if there were no overt intention to discriminate on the part of the employer.

In a discrimination case the plaintiff must produce sufficiently damaging evidence of discrimination to shift the burden of proof to the defendant. Over the years the nature of the evidence necessary to establish a prima facie case of discrimination has altered markedly. In particular, the incidence of use of statistical evidence has risen dramatically, and in some suits gross statistical disparities alone have been sufficient to shift the burden of proof.\(^8\) The purpose of this section is to examine the role of these prima facie statistics in employment discrimination cases.

---

7. B. SCHLEI & P. GROSSMAN, EMPLOYMENT DISCRIMINATION LAW xii (2d ed. 1983).
8. See also International Bhd. of Teamsters v. United States, 431 U.S. 324, 339 (1977) (“Statistics are equally competent in proving employment discrimination [as in proving discrimination in jury selection].”); Kaplan v. International Alliance of Theatrical and Stage Employees, 525 F.2d 1394, 1358 (9th Cir. 1975) (“[S]howing wholly disproportionate female membership in a union in comparison to the available female work force in a demographic area . . . is sufficient to establish a prima facie case of sex discrimination.”); United States v. Ironworkers Local 86, 443 F.2d 544, 551 (9th Cir. 1971), cert. denied, 404 U.S. [Vol. 46: No. 4]
Two major issues must be considered. While a variety of different statistical comparisons have been made in different circumstances, the first issue has been to determine whether a particular comparison is appropriate in a particular situation. Once an appropriate comparison has been made, the second issue has been to determine whether the disparity is "substantial." Analysis of the second issue shows that the courts are currently more willing to allow relatively weaker evidence of discrimination to shift the burden of proof than they were in the past; evidence that is now considered to be sufficient would not have been considered as such ten or fifteen years ago.

The courts have considered two types of discrimination cases—adverse treatment cases and adverse impact cases.9 These terms are defined in International Brotherhood of Teamsters v. United States.10 In adverse treatment cases, "[t]he employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin."11 Proof of discrimination in such cases focuses on employers' alleged discriminatory motivation and intent, typically manifested by actions of individual, malevolent discrimination. Alternatively, the plaintiff may follow the guidelines developed in the case of McDonnell Douglas Corp. v. Green.12 An individual may shift the burden of proof by showing:

(i) [T]hat he belongs to a racial minority; (ii) that he applied and was qualified for a job for which the employer was seeking applicants; (iii) that, despite his qualifications, he was rejected; and (iv) that, after his rejection, the position remained open and the employer continued to seek applicants from persons of complainant's qualifications.13

If that showing is made, the employer must then "articulate some legitimate, nondiscriminatory reason for the employee's rejection."14 If the defendant carries this burden, the plaintiff then has an opportunity to show that the employer's stated reason was pretextual, a cover up for racial, sex, or ethnic discrimination. Recently, in Texas Department of Community Affairs v. Burdine,15 the Supreme Court reaffirmed these procedures but emphasized that after the burden shifts to the defendant, the defendant has only to explain the nondiscriminatory reason for its actions and "need not persuade the court that it was actually motivated by the proffered reasons."16

While individual testimony is central to determine intent, statistical evidence is, nonetheless, relevant. In Teamsters the court ruled that in a disparate treatment case "[p]roof of discriminatory motive is critical, although it can in some situations

984 (1971) ("In many cases the only available avenue of proof is the use of racial statistics to uncover clandestine and covert discrimination by the employer or union involved.").
9. The distinction between a disparate treatment case and a disparate impact case is not synonymous with the difference between an individual action and a class action. See, e.g., Davis v. Califano, 613 F.2d 957, 962 (D.C. Cir. 1979). Throughout this article adverse is used synonymously with disparate.
11. Id. at 335 n.15.
13. Id. at 802.
14. Id.
16. Id. at 254.
be inferred from the mere fact of differences in treatment."\(^\text{17}\) Subsequently, in *Furnco Constr. Corp. v. Waters*,\(^\text{18}\) the Supreme Court has ruled that "[p]roof that his work force was racially balanced or that it contained a disproportionately high percentage of minority employees is not wholly irrelevant on the issue of intent when that issue is yet to be decided."\(^\text{19}\)

Statistical evidence may also be relevant to showing that the employer's explanation is merely a pretext. For example, in *McDonnell Douglas* the Court ruled that "statistics as to petitioner's employment policy and practice may be helpful to a determination of . . . a general pattern of discrimination against blacks."\(^\text{20}\) Such statistical evidence is important because, as Schlei and Grossman note, "the great majority of disparate treatment cases turn on the plaintiff's ability to demonstrate that the nondiscriminatory reason offered by the employer was a pretext for discrimination."\(^\text{21}\)

Since *Griggs v. Duke Power Co.*,\(^\text{22}\) a second type of discrimination case has steered attention away from motivation and intent to the effect of an allegedly discriminatory practice, that is, adverse impact. The Supreme Court observed that Title VII is not concerned with employer's "good intent or absence of discriminatory intent" and "Congress directed the thrust of the Act to the consequences of employment practices, not simply the motivation."\(^\text{23}\)

Adverse impact cases "involve employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity."\(^\text{24}\) The procedure in an adverse impact case was developed in *Griggs* and refined in *Albemarle Paper Co. v. Moody*\(^\text{25}\) and subsequent cases. The plaintiff initially has to make a prima facie showing of discrimination which, in general, requires demonstrating that the pattern or practice of hiring or promotion has a disparate impact upon a protected group. The burden of proof then shifts to the defendant who may (1) refute the plaintiff's evidence and argue that the pattern or practice does not, in fact, have a substantial disparate impact on a protected group,\(^\text{26}\) or (2) argue that the existing pattern or practice arises from a "job-related criterion or a policy or practice which is a business necessity."\(^\text{27}\) Where tests or standards are involved "Congress has placed on the employer the burden of showing that any given requirement must have a manifest relationship to the employment in question."\(^\text{28}\)

"If an employer does then meet the burden of proving that its tests are job

\(^{17}\) 431 U.S. at 335 n.15.


\(^{19}\) Id. at 580.

\(^{20}\) 411 U.S. at 805.

\(^{21}\) B. SCHLEI & P. GROSSMAN, supra note 7, at 1317.

\(^{22}\) 401 U.S. 424 (1971).

\(^{23}\) Id. at 432.

\(^{24}\) Teamsters, 431 U.S. at 335, n.15.

\(^{25}\) 422 U.S. 405 (1975).

\(^{26}\) The employer may show that the plaintiff's "proof" is either "inaccurate or insignificant." Teamsters, 431 U.S. at 360.

\(^{27}\) B. SCHLEI & P. GROSSMAN, supra note 7, at 1328.

\(^{28}\) Griggs, 401 U.S. at 432.
related,' it remains open to the complaining parties to show . . . the employer was using its tests merely as a 'pretext' for discrimination." 29 Because of the need to examine effects of hiring and promotion practices, statistical evidence has played an increasingly dominant role in adverse impact suits.

The importance of statistics in discrimination cases has been affirmed by a whole series of cases since 1962. 30 However, the nature of the statistical evidence used by the courts has changed considerably. To clarify this development, a typology of the possible types of parity studies is presented.

B. A Typology of Probative Parity Comparisons

In most cases using parity studies, statistics are presented which compare the percentage of a particular set of actual or potential minority employees with the percentage of that minority group in the available qualified pool. 31 For example, some courts have made population/workforce comparisons in which the percentage of black employees is compared with the percentage of blacks in the general population. Some parity studies have made applicant flow comparisons in which the percentage of protected group members who are hired is compared with the percentage of protected group members in the pool of applicants. There is much debate over what subset of protected group employees in the firm should be used and to what subset of protected group employees in the general population this should be compared. A major issue in discrimination cases is the circumstances under which a particular type of parity study may be probative. Our next task is to present, in a parsimonious way, the set of all possible probative parity studies.

When reading the cases it is important to make sure that one knows the

29. Albemarle, 422 U.S. at 425.
30. See, e.g., Mayor of Philadelphia v. Educational Equality League, 415 U.S. 605, 620 (1973) ("Statistical analyses have served and will continue to serve an important role as one direct indicator of racial discrimination."); Davis v. Calihan, 613 F.2d 957, 962 (D.C. Cir. 1979) ("We have previously indicated, and now explicitly hold, that statistical evidence may establish a prima facie case of employment discrimination in an individual case."); Senter v. General Motors Corp., 532 F.2d 511, 527 (6th Cir. 1976), cert. denied, 429 U.S. (1976) ("Statistical evidence is an important tool for placing seemingly inoffensive employment practices in their proper perspective . . . . An employee is at an inherent disadvantage in gathering hard evidence of employment discrimination particularly when the discrimination is plant-wide in scope. It is for this reason that we generally acknowledge a prima facie case of discrimination under Title VII."); Wetzel v. Liberty Mutual Ins. Co., 508 F.2d 239, 259 (3d Cir. 1975), cert. denied, 421 U.S. 1011 (1975) ("Statistics are particularly applicable in Title VII class actions where it is the aggregate effect of a company’s policy on the class which is important."); Pettway v. American Cast Iron Pipe Co., 494 F.2d 211, 225 (5th Cir. 1974); Brown v. Gaston County Dyeing Machine Co., 457 F.2d 1377, 1382 (4th Cir. 1972), cert. denied, 409 U.S. 982 (1972); United States v. Hayes Int'l Corp., 456 F.2d 112, 120 (5th Cir. 1972) ("While we abjure any desire to become involved in a numbers game, statistics such as these do have some relevance in a Title VII pattern and practice suit."); United States v. Jacksonville Terminal Co., 451 F.2d 418, 442 (5th Cir. 1971), cert. denied, 406 U.S. 906 (1972); Bing v. Roadway Express, Inc., 444 F.2d 687, 689 (5th Cir. 1971) ("We believe it evident that if the statistics in the instant matter represent less than a shout, they certainly constitute far more than a mere whisper."); Farham v. Southwestern Bell Tel. Co., 433 F.2d 421, 426 (8th Cir. 1970) ("We hold as a matter of law that these statistics . . . . established a violation of Title VII."); Jones v. Lee Way Motor Freight, Inc., 431 F.2d 245, 247 (10th Cir. 1970), cert. denied, 401 U.S. 954 (1971); Alabama v. United States, 304 F.2d 583, 586 (5th Cir. 1961), aff'd, 371 U.S. 37 (1961) ("In the problem of racial discrimination, statistics often tell much, and Courts listen."). See also supra note 8.
31. We use the term "minority" for convenience. If the protected group is gender-based it may, of course, not be a numerical minority.
meaning of such expressions as the "percentage of blacks hired." The phrase may mean the percentage of blacks in the set of hired individuals. Alternatively, it may mean the percentage of hires from the group of black applicants. This important distinction is worth explaining in some detail by a simple example. Suppose there are \( n_1 \) protected group members in the population and \( n_2 \) majority group members, and the total population is \( N \) \((N = n_1 + n_2)\). Of the \( n_1 \) protected group members, \( x_1 \) meet some criterion (for example, satisfy a requirement, pass a test, are hired or selected). Of the \( n_2 \) majority group members, \( x_2 \) meet the same criterion, and the total number of people who meet the criterion is \( Y \) \((Y = x_1 + x_2)\). The court may compare the percentage of the protected group selected with the percentage of the protected group in the general population represented by \( x_1 / Y \) vs. \( n_1 / N \). Alternatively, the court may compare the pass rate of the protected group with the pass rate of the majority group represented by \( x_1 / n_1 \) vs. \( x_2 / n_2 \).

These relationships are shown on Table 1.

### Table 1

**NOMENCLATURE**

<table>
<thead>
<tr>
<th></th>
<th>Protected Group</th>
<th>Majority Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected</td>
<td>( x_1 )</td>
<td>( x_2 )</td>
<td>( Y )</td>
</tr>
<tr>
<td>Failed</td>
<td>( n_1 \cdot x_1)</td>
<td>( n_2 \cdot x_2)</td>
<td>( N - Y )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_1 )</td>
<td>( n_2 )</td>
<td>( N )</td>
</tr>
</tbody>
</table>

Parity studies may involve an examination of the work force composition at a particular time, which means that the "stock" of employees is being considered. Other studies examine new hires, assignments, or promotions over time; these studies are concerned with "flows." Data may be presented which concern the total "stock" or total "flow" of employees in an organization, or which concern the "stocks" or "flows" at particular levels or into particular types of jobs. Thus the "percentage of protected group employees" may be the percentage in: (1) the total work force, (2) the work force at a particular level or of a particular type, (3) the total new hires, or (4) the new hires, assignments, or promotions to a particular level or particular set of jobs. The first two comparisons involve "stocks," the last two involve "flows." Where allegedly biased requirements (including tests) are involved, the courts typically consider a fifth comparison: (5) the percentage of protected group employees in the pool of potential employees, that is, those employees who meet the requirement (such as pass the test). These five comparisons constitute the rows in Table 2, a typology of relevant statistical comparisons.32 Every case in the fourth row, for example, concerns the percentage

---

32. The table is comprehensive in the sense that every possible type of comparison is represented by one of the 20 cells. While most major employment discrimination cases are presented in the table, it is not intended to be comprehensive in the sense that it includes every case that has used statistics. In some cases more than one comparison has been made, and in these circumstances, the case name appears in more than one cell.
of recent hires, assignments, or promotions that is protected.

These percentages must be compared with the percentage of some reference group that is protected. Initially, the reference group was the general population. In more recent cases, comparisons are made to the minimum qualified work pool, the set of applicants, or to the set of total employees (or set of employees at a different level or in different jobs); these reference groups constitute the columns in Table 2. For example, a case would appear in cell N if, during the case, the percentage of recent hires, assignments, or promotions that is protected is compared with the percentage of applicants that is protected. This is an example of an applicant-flow comparison.

Cells A and B involve population/work force comparisons. Cell A involves all employees while cell B involves employees at a particular level or set of jobs. The legitimacy of using cell A and B studies was confirmed by the court in *Teamsters*. The Supreme Court endorsed cell A comparisons by saying that:

> Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired.

The actual data in *Teamsters* concerned line drivers, who are a particular set of employees. Thus the actual data pertain to cell B: "With one exception . . . the company and its predecessors did not employ a Negro on a regular basis as a line driver until 1969. And, as the Government showed, even in 1971 there were terminals in areas of substantial Negro population where all of the company's line drivers were white."34

*Parham v. Southwestern Bell Telephone Co.*35 can be placed in cell A because evidence was introduced which showed that, while fewer than 2% of their employees were nonwhite, approximately 22% of the general population of Arkansas was black.36 Since data were also presented which concerned the number of employees in particular categories, *Parham* has also been placed in cell B.

Population/work force comparisons have often been regarded as probative. Nonetheless, the courts are aware of many significant, potential problems. Many of these problems are overcome by the use of alternative probative statistics and are discussed later. One problem that has never been completely resolved is the appropriate definition of "community from which employees are hired" or "relevant geographic region."37

The "stock" comparisons presented above were often sufficient to shift the burden of proof to the defendant. However, such statistics could not differentiate between what may have been discriminatory behavior prior to the effective date of Title VII (July 2, 1965) and subsequent employment practices. In *Teamsters* the
TABLE 2

A TYPOLOGY OF RELEVANT STATISTICAL COMPARISONS

<table>
<thead>
<tr>
<th>Reference Group</th>
<th>General Population</th>
<th>Qualified Work Pool</th>
<th>Applicant Pool</th>
<th>Another Level Within Firm&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protected Group as Percentage of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employers' Total Work Force</td>
<td>Teamsters&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Hazeldwood Mayor of Philadelphia</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parham</td>
<td>Agarwal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>Agarwal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Boston NAACP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work force at a particular level</td>
<td>Teamsters</td>
<td>Agarwal</td>
<td>G</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>Parham</td>
<td>Mecklenburg</td>
<td></td>
<td>Fisher</td>
</tr>
<tr>
<td></td>
<td>Weber</td>
<td>Parham</td>
<td></td>
<td>Griggs</td>
</tr>
<tr>
<td></td>
<td>Fisher</td>
<td>Weber</td>
<td></td>
<td>Dickerson</td>
</tr>
<tr>
<td></td>
<td>Hayes</td>
<td>Roman</td>
<td></td>
<td>Senter</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teamsters</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Baxter</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>James</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Johnson</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Roman</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rowe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Swint</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Davis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mecklenburg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chance&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Recent hires to whole firm</td>
<td>Parham</td>
<td>Hazelwood</td>
<td>H</td>
<td>Ochoa</td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>Agarwal</td>
<td></td>
<td>Hazelwood</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Swint&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Agarwal</td>
</tr>
<tr>
<td>Recent hires, assignments or promotions to particular level, type or division</td>
<td>Teamsters</td>
<td>Furmo</td>
<td>I</td>
<td>Hester&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hester&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>Wade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patterson</td>
<td></td>
<td>NAACP&lt;sup&gt;a&lt;/sup&gt; (Ensley)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agarwal</td>
<td></td>
<td>Agarwal</td>
</tr>
<tr>
<td>Candidates for hire or promotion</td>
<td>Griggs</td>
<td>Albemarle&lt;sup&gt;b&lt;/sup&gt;</td>
<td>J</td>
<td>Chance</td>
</tr>
<tr>
<td></td>
<td>Dothard</td>
<td>Green</td>
<td></td>
<td>James</td>
</tr>
<tr>
<td></td>
<td>Johnson</td>
<td>Kirland</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boston NAACP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Another level within the firm may refer to the whole organization.

<sup>b</sup> This type of parity study was not used but it was endorsed in this case.

<sup>c</sup> In Chance comparison was made to similar positions in different organizations.
Supreme Court ruled that once a plaintiff has established a prima facie case "[a]n employer might show . . . that the claimed discriminatory pattern is a product of pre-Act hiring rather than unlawful post-Act discrimination."\(^{38}\) Rather than focusing on "stock" comparisons, statistics should be presented which consider the "flow" of new hires since the effective date of Title VII. Unfortunately for the Teamsters when such data were analyzed, "the record shows that many line drivers continued to be hired throughout this period [late 1960's], and that almost all of them were white."\(^{39}\)

Pre-Act data may be introduced but they "serve only as a basis for comparison."\(^{40}\) In Parham, population/work force comparisons using data from September 1964 and December 1966 showed no "noticeable increase in the number of blacks employed."\(^{41}\) Implicitly, by comparing data for different time periods, the court was examining "flows" typical of cell C. In fact the defense explicitly introduced data concerning new hires.\(^{42}\) However, the court ruled:

While an employer's more recent employment practices may bear upon the remedy sought, they do not affect the determination of whether the employer previously violated Title VII.\(^{43}\)

The crucial issue in a lawsuit of this kind is whether the plaintiff establishes hiring bias at the time of his rejection for employment and subsequent complaint to the EEOC . . . .\(^{44}\)

Thus, flow data are relevant, but only for the time between the effective date of the Act and the date the complaint is filed.

In Griggs the employer required new employees to possess a high school diploma and to register satisfactory scores on two standardized aptitude tests in order to qualify for placement in any department except labor. Statistical evidence taken from the 1960 census was presented which showed that in North Carolina "while 34% of white males had completed high school, only 12% of Negro males had done so."\(^{45}\) Even more damaging evidence was presented concerning the pass rates on the standardized tests.\(^{46}\) There was, in effect, a substantial disparity between the percentage of black males in the pool of potential employees and the percentage of black males in the general population. Consequently Griggs appears in cell E.

Situations where comparisons should involve the qualified work pool (column 2) rather than the general population (column 1) are described in Hazelwood School District v. United States.\(^{47}\)

In Teamsters, the comparison between the percentage of Negroes on the employer's work force and the percentage in the general areawide population was highly probative because

\(^{38}\) 431 U.S. at 360.

\(^{39}\) Id. at 341.

\(^{40}\) Parham, 433 F.2d at 424.

\(^{41}\) Id.

\(^{42}\) Id. at 425 ("In 1967, 6.1 percent of the Company's new employees were blacks. This figure increased markedly to 17.6 percent, or 135 blacks of 766 new employees, in 1968.").

\(^{43}\) Id. at 426.

\(^{44}\) Id. at 425.


\(^{46}\) Id. ("[A] battery of tests . . . resulted in 58% of whites passing the tests, as compared with only 6% of the blacks.").

the job skill there involved—the ability to drive a truck—is one that many persons possess or can fairly readily acquire. When special qualifications are required to fill particular jobs, comparisons to the general population (rather than to the smaller group of individuals who possess the necessary qualifications) may have little probative value.\footnote{48}

In this case the Supreme Court ruled that "a proper comparison was between the racial composition of Hazelwood’s teaching staff and the racial composition of the qualified public school teacher population in the relevant labor market."\footnote{49}

The importance of job qualifications had been recognized earlier by the Supreme Court in \textit{Mayor of Philadelphia v. Educational Equality League}.\footnote{50} In this case the Third Circuit had found a substantial disparity between the percentage of blacks on a panel that nominated persons for vacancies on the Philadelphia School Board and the percentage of blacks in the general population.\footnote{51} The Supreme Court in reversing the appellate court ruled:

\begin{quote}
[T]his is not a case in which it can be assumed that all citizens are fungible for purposes of determining whether members of a particular class have been unlawfully excluded. . . . Assuming, arguendo, that percentage comparisons are meaningful in a case involving discretionary appointments, the relevant universe for comparison purposes consists of the highest ranking officers of the categories of organizations and institutions specified in the city charter, not the population at large.\footnote{52}
\end{quote}

When qualifications are essential for the job and the data presented to the courts concern hires or promotions rather than the total work force, the courts have argued that comparison should be made with the qualified work force rather than with the general population. In \textit{Hester v. Southern Railway Co.}\footnote{53} the court of appeals ruled:

\begin{quote}
Comparison with general population statistics is of questionable value when we are considering positions for which, as here, the general population is not presumptively qualified. Data Typist applicants were required to prove their ability to type at a minimum speed of sixty corrected words per minute as a prerequisite to consideration by Southern for employment . . . . A more significant comparison might perhaps be between the percentage of blacks in the population consisting of those able to type 60 wpm or better and the percentage hired into the Data Typist position by Southern.\footnote{54}
\end{quote}

Similarly, in \textit{Patterson v. American Tobacco Co.},\footnote{55} the Fourth Circuit stated:

The record discloses that 6.8 percent of the blacks and 1.5 percent of the women in the Richmond [Standard Metropolitan Statistical Area] are placed in a category that includes supervisory personnel. Those percentages furnish a more realistic measure of the company's conduct than the gross percentage of blacks and women in the whole work force, including unskilled labor.\footnote{56}

In \textit{Hazelwood} the Supreme Court certainly regarded as important a comparison of the percentage of new teachers who were black with the percentage of black

\begin{itemize}
  \item \footnote{48} Id. at 308 n.13.
  \item \footnote{49} Id. at 308.
  \item \footnote{50} 415 U.S. 605 (1973).
  \item \footnote{51} Id. at 611 (Initially, the district court "held that differences between the percentage of Negroes in the city’s population (34%) or in the student body of the public school system (60%) and the percentage of Negroes on the 1971 Nominating Panel (15%) had no significance.").
  \item \footnote{52} Id. at 620, 621.
  \item \footnote{53} 497 F.2d 1374 (5th Cir. 1974).
  \item \footnote{54} Id. at 1379 n.6.
  \item \footnote{55} 535 F.2d 257 (4th Cir. 1976), cert. denied, 429 U.S. 920 (1976).
  \item \footnote{56} Id. at 275.
\end{itemize}
teachers in the relevant labor market. Although the Court did not define the relevant labor market it noted that the choice was important and indicated some considerations that should be evaluated.

Consequently, comparisons to the qualified work pool rather than the general population are important for both "stock" and "flow" comparisons. A different type of comparison involves not the whole set of potential employees in the qualified work pool, but the set of applicants. It makes little sense to compare the composition of the work force "stock" with the composition of the applicant pool and so cells K and L are empty. Of much more interest is a comparison of the new "flow" of employees (recent hires, promotions, or assignments) with the applicant pool. These applicant/flow cases are listed in cells M and N.

Although applicant/flow data were unavailable in Hazelwood, the Court realized that such data would "be very relevant" and "might tell a totally different story." Further, the Supreme Court ruled: "It will also be open to the District Court on remand to determine whether sufficiently reliable applicant-flow data are available to permit consideration of the petitioner's argument that those data may undercut a statistical analysis dependent upon hirings alone." Several other cases have suggested that applicant/flow data are more probative than work force/population comparisons. In Ochoa v. Monsanto Co. there was a substantial disparity between the percentage of Mexican-Americans employed and the percentage of Mexican-Americans in the general population, and yet this evidence was found not to constitute a prima facie finding of discrimination, in part because the company's applicant/flow data showed that Mexican-Americans had recently been hired at a faster rate than whites.

Where job requirements or tests are an issue, the Supreme Court's ruling in Albemarle is germane. The Court ruled that the complaining party or class can make out "a prima facie case of discrimination...[by showing] that the tests in question select applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants." This quote proposes a different comparison from that made in Griggs. While Griggs compares the percentage of blacks who were selected for hire with the percentage of blacks in the general population, Albemarle suggests the comparison should be between the percentage of blacks selected for hire and the percentage of black applicants. For this reason we place Albemarle in cell 0 of Table 2. It is important to note, however, that this

57. 433 U.S. at 310 ("The record in this case showed that for the 1972-1973 school year, Hazelwood hired 282 new teachers, 10 of whom (3.5%) were Negroes; for the following school year it hired 123 new teachers, 5 of whom (4.1%) were Negroes. Over the two-year period, Negroes constituted a total of 15 of the 405 new teachers hired (3.7%)." These percentages could have been compared with the percentage of teachers in St. Louis and the city of St. Louis who were black (15.4%) or the percentage of teachers in St. Louis County alone who were black (5.7%).
58. Id. at 311, 311 n.17, 312.
59. Id. at 308 n.13.
60. Id. at 310.
61. Id. at 313 n.21.
63. Id. at 59.
assignment is based on the above statements, not on the actual prima facie evidence presented in *Albemarle*, which concerned seniority systems.

While applicant/flow data may be more probative than population/work force comparisons, there are problems. An important problem is described in detail in *Teamsters*:

If an employer should announce his policy of discrimination by a sign reading 'Whites Only' on the hiring-office door, his victims would not be limited to the few who ignored the sign and subjected themselves to personal rebuffs. The same message can be communicated to potential applicants more subtly but just as clearly by an employer's actual practices—by his consistent discriminatory treatment of actual applicants, by the manner in which he publicizes vacancies, his recruitment techniques, his responses to casual or tentative inquiries, and even by the racial or ethnic composition of that part of his work force from which he has discriminatorily excluded members of minority groups.65

In other words, past discriminatory practices may "chill" the applicant pool.

The fourth column of Table 2 compares the percentage of minority group members actually employed or potentially employed with the percentage of minority employees in the entire firm or at some other hierarchical level of the firm. By definition, cell P contains no entry. As Table 2 suggests, however, comparisons of type Q are often regarded as probative. In *Teamsters*, for example, the evidence concerned systemwide discrimination:

[T]he company had 6,472 employees. Of these, 314 (5%) were Negroes and 257 (4%) were Spanish-surnamed Americans. Of the 1,828 line drivers, however, there were only 8 (0.4%) Negroes and 5 (0.3%) Spanish-surnamed persons, and all of the Negroes had been hired after the litigation had commenced. . . . A great majority of the Negroes (83%) and Spanish-surnamed Americans (78%) who did work for the company held the lower paying city operations and serviceman jobs, whereas only 39% of the nonminority employees held jobs in those categories.66

In the first part of this quote the Supreme Court is comparing the proportion of line drivers who are protected with the proportion of total employees who are protected. In the last part, the proportion of protected employees at a less desirable position is implicitly compared with the proportion of protected employees at a more desirable level.

In *Fisher v. Procter & Gamble Manufacturing Co.*67 a more recent case, the plaintiffs argued that in making cell B and cell Q comparisons they have established a prima facie case of discrimination. The company countered by arguing that a cell G comparison was most relevant. The Fifth Circuit decided that the company would be estopped from using the qualified work force argument in rebutting disparity when the firm itself engages in substantial training for these skilled positions.68

When a company adopts a policy and practice of hiring in at low-level unskilled jobs and promoting to upper-level positions based upon training received and skills developed at the plant itself, it cannot convincingly challenge the prima facie showing under the *Hazelwood* 'qualifications' dicta. Where skills are commensurate with company training, we will

65. 431 U.S. at 365.
66. *Id.* at 338.
approve statistical comparisons between racial make-up in key positions and racial composition in the total work force. 69

Comparisons to a level in the firm or to the whole firm are not made just when the work force at a particular level is examined. It is also relevant when the focus is on flows (cell S) or on candidates for hire or promotion when tests are involved (cell T).

One of the strongest conclusions that emerges from the table is that few cases rely on a single type of comparison. Teamsters, for example, is listed in four cells. The most frequently cited case, Agarwal v. McKee & Co., 70 appears in eight cells. The plaintiff does not rely upon a few statistical comparisons to shift the burden of proof, but upon developing a complete picture of the employer’s practices. Whether the plaintiff is successful depends in part on the magnitude of the disparity. It has to be “significant.”

C. The Meaning of “Significance” in Employment Disparity Cases

The analysis is summarized in Table 3, which presents five “criteria” or measures that have been applied to determine whether the employment disparity is significant, thereby shifting the burden to the defendant. Generally the measures are more sophisticated as one moves down the column. Also, this order approximates the way the courts have changed over time. The change appears to be a function of both the courts’ increasing statistical sophistication and the changing nature of employment discrimination that has come before the courts, that is, from obvious and overt discrimination to subtle and covert discrimination.

While we are concerned primarily with employment discrimination cases, consideration of discrimination in jury selection cases provides a valuable historical perspective in this context. Prior to the 1964 Civil Rights Act, plaintiffs brought cases to the Supreme Court claiming an unconstitutional denial of equal protection. In such cases the magnitude of disparity that was sufficient to shift the burden of proof to the defense was very large. In extreme cases, no protected group member was currently or ever had been in the selected jury group. In effect, there was total exclusion or only token representation of minorities on juries. The use of such evidence in “rule of exclusion” cases dates back over one hundred years to Neal v. Delaware. 71 At that time, over 17% of the population was black but “no colored citizen had ever been summoned as a juror in the courts of the State.” 72 Another example is provided by Norris v. Alabama 73 which dates back to 1935. While blacks comprised over 7.5% of the adult males in the county, “no negro had served on any grand or petit jury in that county within the memory of witnesses who had lived there all their lives.” 74 Similar exclusionary evidence was presented in Hill v. Texas, 75 Patton v. Mississippi, 76 Hernandez v. Texas, 77 Arnold v. North Caro-
Table 3

Criteria to Judge the Significance of Disparity

<table>
<thead>
<tr>
<th>Rule of Exclusion</th>
<th>James Neal Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watkins Norris</td>
<td>Wade Hill Arnold</td>
</tr>
<tr>
<td>Teamsters Pierce</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overwhelming, or gross, disparity</th>
<th>Parham Hazelwood Rouve Kaplan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teamsters Senier</td>
<td>James Griggs</td>
</tr>
<tr>
<td>Dothard</td>
<td>Johnson</td>
</tr>
<tr>
<td>Fisher</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arbitrary % Difference</th>
<th>Castaneda Davis Dothard Chance Sims Test Whitus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate Significance tests</td>
<td>Castaneda Hazelwood Agarwal</td>
</tr>
<tr>
<td>Enslay</td>
<td>Roman Swint</td>
</tr>
<tr>
<td>Wade</td>
<td>Mecklenburg Dickerson</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multivariate</th>
<th>James Stastney Presseisen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mecklenburg Agarwal</td>
<td></td>
</tr>
</tbody>
</table>

Total exclusion of a protected group has also been observed in employment discrimination cases: it carries considerable importance and obviates the need for detailed, narrow attacks. In Teamsters, for example, the Supreme Court observed that:

> In any event, fine tuning of the statistics could not have obscured the glaring absence of minority line drivers. As the Court of Appeals remarked, the company's inability to rebut the inference of discrimination came not from a misuse of statistics but from 'the inexorable zero.'

Other employment discrimination cases in which there was a total exclusion or token representation of protected group members from at least some type of jobs include Franks v. Bowman Transportation Co., Rogers v. International Paper Co., Wade

---

76. 332 U.S. 463, 466 (1947) (“It is to be noted at once that the indisputable fact that no Negro had served on a criminal court grand or petit jury for a period of thirty years created a very strong showing that during that period Negroes were systematically excluded from jury service because of race.”).

77. 347 U.S. 475, 480 (1953) “[P]roof that Negroes constituted a substantial segment of the population of the jurisdiction, that some Negroes were qualified to serve as jurors, and that none had been called for jury service over an extended period of time, was held to constitute prima facie proof of the systematic exclusion of Negroes from jury service.”). 78. 376 U.S. 773 (1964).

79. 251 Miss. 699, 707, 171 So. 2d 129, 133-34 (1965) (“[L]ong continued omission of Negroes from jury service establishes a prima facie case of systematic discrimination.”).


82. 510 F.2d 1340 (8th Cir. 1975), vacated, 423 U.S. 809 (1975).
Total exclusion is sufficient to shift the burden of proof, but such an obvious degree of discrimination is not always necessary. More frequently there is a "substantial" disparity or a "substantial" underrepresentation of a group which is short of total exclusion. Synonyms that have been used by the courts include "gross," "marked," "compelling," and "wholly disproportionate."

In the cases discussed above, a cursory examination of the employers' employment practices suggested obvious discrimination. In these cases, therefore, the relatively gross "rule of exclusion" or "substantial" disparity criterion could be used to shift the burden of proof so that there were few statistical problems. Nonetheless, as will be seen below, some of the courts' early encounters with discrimination data reveal considerable uncertainty and a lack of understanding of the role of statistics. This is illustrated by the cases in row three of Table 3 which are based on arbitrary percentage differences.

In Swain v. Alabama, a jury selection case, "[t]he evidence was that while Negro males over 21 constitute 26% of all males in the county in this age group, only 10 to 15% of the grand and petit jury panels drawn from the jury box since 1953 have been Negroes. . . ." The Court admitted that "[w]e cannot say that purposeful discrimination based on race alone is satisfactorily proved by showing that an identifiable group in a community is underrepresented by as much as 10%." Of course, the Supreme Court should have said "underrepresented by as much as 16%." In any case, the 10-16% range of disparity became the basis for future decisions. For example, the Tenth Circuit in United States v. Test stated:

Disparities were found to be significant in Jones v. Georgia, where the disparity was 14.7% (19.7% of tax list, 5% of jury roles); Sims v. Georgia, where the disparity

83. 528 F.2d 508 (5th Cir. 1976).
84. 559 F.2d 310 (5th Cir. 1977).
87. See, e.g., Griggs, 401 U.S. at 429; Rowe v. General Motors Corp., 457 F.2d 348, 357 n.18 (5th Cir. 1972).
89. Kaplan v. Int'l Alliance of Theatrical and Stage Employees and Motion Picture Mach. Operators, 525 F.2d 1354, 1358 (9th Cir. 1975).
91. Id. at 205.
92. Id. at 208-09.
93. 550 F.2d 577 (10th Cir. 1976).
94. Id. at 587.
96. 389 U.S. 404 (1967).
was 19.7% (24.4% of tax lists, 4.7% of grand jury lists); *Whitus v. Georgia*,\(^97\) where the disparity was 18% (27.1% of taxpayers, 9.1% of grand jury venire); *Turner v. Fouche*,\(^98\) where the disparity was 23% (60% of general population, 37% of grand jury lists); and *Castaneda v. Partida*,\(^99\) where the disparity was 40.1% (79.1% of county population, 39% of persons summoned for grand jury).

Arbitrary percentage comparisons have been introduced in employment discrimination cases attacking allegedly discriminatory requirements. The comparisons, however, are not quite analogous to those used in jury selection cases. Referring back to Table 1, one can see that jury selection cases examine the disparity between the percentage of a protected group in the general population \((n_1/N)\) with the percentage of the protected group on jury lists \((x_1/Y)\). In employment discrimination cases, selection rates are compared as follows: \(x_1/n_1\) vs. \(x_2/n_2\). For example, in *Griggs* 12% of black males had completed high school while 34% of white males had done so.\(^100\) This particular arbitrary percentage difference is important because subsequent cases have used it as a benchmark.\(^101\) Sometimes in employment discrimination cases, however, the courts make an additional step and compute the selection ratio which is compared to some arbitrary number. The selection ratio of whites to the protected group was 2.6 in *Castro v. Beecher*,\(^102\) 3.5 in *Bridgeport Guardians, Inc. v. Members of Bridgeport Civil Service Commission*,\(^103\) and 1.3 in *Chance v. Board of Examiners*.\(^104\) The courts may also compute failure rates: \((n_1 - x_1)/n_1\) vs. \((n_2 - x_2)/n_2\). For example, in *Davis v. Washington*\(^105\) the failure rate was 47% for blacks and 12% for whites so that “[t]he disparity disclosed in this case—more than four to one—is larger than differences held sufficiently disproportionate in other cases.”\(^106\) In *Green v. Missouri Pacific Railroad Co.*,\(^107\) the failure rate for blacks was two and one-half times that of whites.

One problem with computing a ratio should be immediately apparent: the outcome depends upon whether one computes the pass ratio or the fail ratio. In *Davis* the failure ratio was four to one but the pass ratio was 1.66 to one, a major difference. Another factor is the magnitude of the statistics. In *Davis* the failure rate for blacks was four times that for whites. However, if the failure rate for blacks had been 4% and the failure rate for whites had been 1%, the ratio would still have been four to one, but there would have been much less reason for concluding that blacks had been systematically excluded.\(^108\)

The courts' examination of percentage differences and their attempts to relate these differences to some benchmark reflect concern for the magnitude of the dif-

---

\(^{97}\) 385 U.S. 545 (1967).


\(^{100}\) See, e.g., Dothard v. Rawlinson, 433 U.S. 321, 331 (1977).

\(^{101}\) 459 F.2d 725 (1st Cir. 1972).

\(^{102}\) 482 F.2d 1333 (2d Cir. 1973).

\(^{103}\) 458 F.2d 1167 (2d Cir. 1972).


\(^{105}\) Id. at 960.

\(^{106}\) 523 F.2d 1290, 1295 (8th Cir. 1975).

\(^{107}\) See B. SCHLEI & P. GROSSMAN, supra note 7, at 106 n.87.
ferences and the probability that the difference could have occurred by chance. To a statistician these concepts are quite distinct although related: \textit{ceteris paribus}, the larger the difference the lower the probability that it could have occurred by chance, that is, the $p$ value. Consequently, in many situations, it is reasonable for the Supreme Court to argue, as it did in \textit{Castaneda}, "[i]f a disparity is sufficiently large, then it is unlikely that it is due solely to chance or accident, and, in the absence of evidence to the contrary, one must conclude that racial... factors entered into the selection process."\textsuperscript{109} As long as the disparities were sufficiently large or exceeded some arbitrary value, the courts avoided the issue of statistical significance.

In their discussions of chance, which date from the early 1970s, the courts first ruled that plaintiffs and defendants did not have to prove or justify their case "with complete mathematical certainty."\textsuperscript{110} Impetus for the use of the concept of statistical significance came from the 1974 EEOC Guidelines\textsuperscript{111} which, since \textit{Griggs}, are "entitled to great deference."\textsuperscript{112} In 1975 the Supreme Court, in \textit{Albemarle}, accepted the evidence of an expert witness who "regarded as 'statistically significant' any correlation that could have occurred by chance only five times, or fewer, in 100 trials."\textsuperscript{113} Thus the 5% level of significance or 95% level of confidence became one of several criteria for establishing a prima facie case.

An alternative approach to using a significance or confidence level is to calculate the number of standard deviations by which the observed result differs from the result expected if there had been no discrimination. The 5% level of significance approved in \textit{Albemarle} is equivalent to 1.96 standard deviations and so is consistent with the "two or three standard deviations" criterion approved in \textit{Castaneda}.\textsuperscript{114}

Many courts subsequently have mentioned an explicit threshold of significance, such as a 5% level or two or three standard deviations. However, the Fifth Circuit has warned that the significance criterion "must not be interpreted or applied so rigidly as to cease functioning as a guide and become an absolute mandate or proscription."\textsuperscript{115} Similarly, in \textit{United States v. Test}, the Tenth Circuit argued that "[t]he mathematical conclusion that the disparity between these two figures is 'statistically significant' does not, however, require an \textit{a priori} finding that these deviations are 'legally significant'... ."\textsuperscript{116}

One reason the concept of statistical significance is not entirely satisfactory results from the effect of sample size. In large samples, small percentage differences will be statistically significant. To say that a relationship is statistically significant at the 95% level of confidence means that there is a 5% or lower

\textsuperscript{109} 430 U.S. at 494 n.13.
\textsuperscript{110} See Boston Chapter, NAACP, Inc. v. Beecher, 504 F.2d 1017, 1022 (1st Cir. 1974), cert. denied 421 U.S. 910 (1975); Vulcan Soc'y of N.Y. City Fire Dep't, Inc. v. Civil Serv. Comm'n, 490 F.2d 387, 393 (2d Cir. 1973).
\textsuperscript{111} Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1982).
\textsuperscript{112} \textit{Griggs}, 401 U.S. at 433-34.
\textsuperscript{113} \textit{Albemarle Paper Co. v. Moody}, 422 U.S. 405, 430 (1975).
\textsuperscript{114} 430 U.S. at 496 n.17.
\textsuperscript{115} \textit{United States v. Georgia Power Co.}, 474 F.2d 906, 915 (5th Cir. 1973).
\textsuperscript{116} 550 F.2d at 584.
probability of concluding that there is a relationship when, in fact, there is none.\textsuperscript{117} This statement, however, ignores the strength of the relationship; it does not tell us how important it is. A pass rate difference, for example, that is insignificant at the 5\% level may be larger than another case in which it is significant at even the 1\% level. The level of significance indicates only that one can be more confident of the existence of a relationship in the latter case. Consequently, when concluding that a relationship is statistically significant, one may be saying little about its magnitude when the sample size is very large.

When samples are very small, large differentials are necessary to obtain statistically significant results. This fact raises the concern that the test is not as powerful as a test based on larger samples. In other words, when the sample size is small there may be a high probability of concluding that there is no discrimination when, in fact, it exists.\textsuperscript{118}

The courts have been wary of statistics based on small sample sizes, but for different reasons. In \textit{Teamsters} the Supreme Court suggested that an employer could defeat a prima facie showing of discrimination by demonstrating that “during the period it is alleged to have pursued a discriminatory policy it made too few employment decisions to justify the inference that it had engaged in a regular practice of discrimination.”\textsuperscript{119} In \textit{Mayor of Philadelphia} there was a 19\% disparity between the percentage of blacks in the city’s population (34\%) and the percentage of blacks on a panel to nominate members to the school boards (15\%). However, since there were only thirteen members on the panel, the Supreme Court supported the district court’s ruling that the “number of positions on the Panel was too small to provide a reliable sample; the addition or subtraction of a single Negro meant an 8\% change in racial composition.”\textsuperscript{120}

Such numbers are obviously very small, but one may wonder what the critical sample size might be. In \textit{Roman v. ESB, Inc.},\textsuperscript{121} 83 workers were laid off of which 53 (63\%) were black, and 30 (47\%) were white. The plant was 54\% nonwhite. The court ruled “in light of the small statistical sampling involved . . . this figure does not prove a discriminatory layoff . . . Only a small change (8) in the numbers of blacks and whites involved would yield a ratio the same as that of blacks to whites in the plant.”\textsuperscript{122}

In some “small numbers” cases, quite a few people may be involved. For example, in \textit{Ochoa v. Monsanto Co.},\textsuperscript{123} during the relevant period there were 684 applicants, of which 11 were Mexican-American. Fifty-six people were hired, of which one was Mexican-American. Here, Mexican-Americans were hired at a faster rate than other applicants, but the crucial issue is that “the smallness of the numbers demonstrates that the Court was not compelled to allow such statistical showing to set in train the usual presumptions or to make a finding of preference.

\footnotesize
\textsuperscript{117} Statisticians call this mistake a Type I error.
\textsuperscript{118} Statisticians call this mistake a Type II error.
\textsuperscript{121} 550 F.2d 1343 (4th Cir. 1976).
\textsuperscript{122} \textit{Id.} at 1352.
\textsuperscript{123} 473 F.2d 318 (5th Cir. 1973).
thereon.’ In Boston Chapter, NAACP, Inc. v. Beecher, 13 out of 33 self-identified black or Spanish applicants passed a test while 1737 out of 3089 whites passed the test, producing a pass rate difference of 17%. The district court felt that the statistics were “obviously meager” and declined to find that “in themselves” they established a prima facie case. In contrast, a work force of approximately 500 was regarded as “sizeable” in Fisher, and 870 was large in Castaneda.

Many other factors affect the courts’ decisions to consider the plaintiff’s statistical evidence as probative. Theoretical considerations require the samples to be randomly selected from the population. In some situations past discriminatory practices may “chill” the applicant pool, which is then not a random sample. In Boston Chapter the court used this argument to consider statistical evidence which it might otherwise have rejected due to its small size. In Roman, another case in which there was a small sample, the court considered the undisputed economic necessity that compelled the cutback in employees, the reduction in the number of tube loaders needed after the change in the production process, and the uncontradicted evidence as to the selection of those to be laid off on the basis of the affected positions rather than on any personal criteria.

While statistical evidence is sufficient to shift the burden of proof in some situations, it is not always necessary and many other factors are considered. The courts have made this point many times, most clearly in Teamsters:

> We caution only that statistics are not irrefutable; they come in infinite variety and, like any other kind of evidence, they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances.

A major reason for discrediting a parity study is the omission of variables that may explain the disparity. As the court said in United States v. Ironworkers Local 86, the use of statistics must be conditioned on the “absence of variables which would undermine the reasonableness of the inference of discrimination which is drawn . . . .” In the studies discussed above, the importance of considerations such as experience and job-related qualifications have been stressed. In United States v. United States Steel Corp., the court ruled that seniority is relevant in wage disparity cases. In Swint v. Pullman-Standard wage statistics were ruled to be unreliable because they did not consider seniority, overtime, and time worked. Other potential factors include plant hiring needs and whether employees choose to work in that particular plant or job. If any one of these variables is incorpo-

---

124. Id. at 320.
125. 504 F.2d 1017 (1st Cir. 1974).
126. Id. at 1019 n.3.
128. 430 U.S. at 496.
129. 504 F.2d at 1021 n.6.
130. 550 F.2d at 1352.
131. 431 U.S. at 340.
132. United States v. Ironworkers Local 86, 443 F.2d 544 (9th Cir. 1971).
133. Id. at 551.
134. 520 F.2d 1043 (5th Cir. 1975), cert. denied, 429 U.S. 817 (1976).
135. Id. at 1054.
136. 539 F.2d 77 (5th Cir. 1976).
137. Id. at 97.
rated into a parity study in addition to race or another allegedly "discriminatory" variable, the analysis, by definition, is multivariate, and multivariate statistical methodologies must be used.

D. Multivariate Analysis as a Measure of the Significance of Disparity

The multivariate methodology that the courts have used extensively is regression analysis. When discrimination was a less topical issue and evidence of discrimination was more obvious, there was little need for complex statistical methodologies. Discrimination is now a much more sensitive issue and its manifestation is more subtle. For example, firms may hire blacks and whites in proportion to their distribution in the population, but then discriminate between the two groups by paying different wages. Regression analysis enables the courts to identify whether wage differentials are significantly different due to some discrimination factor, such as race, or whether the disparities result from some justifiable causes. A standard model for defining salary discrimination is:

\[
\text{Salary} = f(\text{Demographics, Productivity, Demographics} \times \text{Productivity}).
\]

For example, if \( D \) is a binary demographic variable equal to unity for the protected group and zero for the majority group, and \( P \) is a productivity measure, a statistician may estimate the linear regression model in which salary (\( S \)) is the dependent variable:

\[
S = B_0 + B_1D + B_2P + B_3(DP) + e
\]

Demographic variables (\( D \)) include race, sex, or marital status. Productivity variables (\( P \)) may include experience, qualifications, education, and other on-the-job measures such as ratings and absenteeism. The variable "demographics \( \times \) productivity" (\( DP \)) indicates that some of the discriminatory demographic variables, such as race and sex, interact with some of the productivity variables. It allows for the possibility that productivity has different effects on the salaries of protected and nonprotected group members.

Discrimination is evidenced if the coefficients of the relevant demographic variables indicate a significantly lower mean salary for the protected group such as when \( B_1 \) is negative and significant. In this case the "intercepts" are significantly different. Discrimination is also evidenced if the coefficients of the interaction terms indicate that productivity factors have a significantly smaller effect on the

---

138. Several articles have detailed the relevance of regression analysis to employment discrimination cases. See, e.g., Finkelstein, The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases, 80 COLUM. L. REV. 737 (1980); Smith & Abram, Quantitative Analysis and Proof of Employment Discrimination, 1981 U. ILL. L.F. 33, 34 n.3 (for a complete bibliography of such studies). For an introduction to regression analysis, see Fisher, Multiple Regression in Legal Proceedings, 80 COLUM. L. REV. 702 (1980). If one is interested in a simple comparison between two means, regression has three advantages over a simple \( t \)-statistic. First, the results are easy to obtain using standard computer packages. Second, the regression coefficient itself provides information; typically, it would be the parity difference, an indication of the strength, after controlling for other correlated factors, such as experience and qualifications. Third, the \( t \)-statistic provides a measure of statistical significance, again after controlling for other variables. If there is only one explanatory variable the measure of significance, the \( t \)-statistic, is identical in both methods. The major advantage of regression analysis, however, is that important explanatory factors can be effectively and easily controlled by simply including them in the equation.
protected group’s salary than on the majority group’s salary, that is, when \( B \) is negative and significant. In this case the “slopes” are significantly different.

The courts have recognized three potential problems with regression analyses. First, the effects of discrimination may be underestimated if some of the productivity variables are themselves functions of discriminatory behavior. For example, in \( J a m e s v. S t o c k h a m V a l u e s \) and \( F i t t i n g s C o. \) the court found “the critical factors of ‘skill level’ and ‘merit rating’ were defined in such a way as to incorporate discrimination.” In \( M e c k l e n b u r g v. M o n t a n a S t a t e B o a r d o f R e g e n t s o f H i g h e r E d u c a t i o n \), the defendant’s regression analysis included variables relating to promotion and tenure which the plaintiff’s expert had already found discriminated against women. In \( S t a s t n y v. S o u t h e r n B e l l T e l e p h o n e a n d T e l e g r a p h C o. \),

\[ \text{defendant's regressions use certain independent variables that are themselves apparently the result of sex discrimination . . . . .} \] The independent variable ‘years of schooling’ is also male-biased, since defendant has until recently purposely discriminated against women.

Essentially, courts have argued in these cases that the model of salary discrimination described above is incomplete and that one should add to the model an equation of the form:

\[ \text{Productivity} = g(\text{Demographics}) + u. \] (3)

Plaintiffs may try to establish the importance of the relationships between productivity and demographics while defendant corporations will typically try to show that they are unimportant.

A second common problem in many empirical analyses is measurement error. Productivity is a particularly difficult concept to measure accurately. In \( A g a r w a l v. M c K e e \) \& Co., the plaintiff’s regression analysis included “years of experience” and “years of education” but was criticized for not coding the type of quality of experience and education. In \( S t a s t n y \), the judge argued that “[t]he variable ‘salary class’ is not a productivity factor but rather is the result that, theoretically, productivity factors would explain.” Implicitly, then, the judge postulated:

\[ \text{Salary class} = h(\text{Productivity}) + v. \] (4)

In statistical terms “salary class” is an indicator of productivity. Even if the error term \( v \) has no correlation with the true measure of productivity, the ordinary least squares estimators in equations (1) and (2) are inconsistent and, in the two-variable cases, are asymptotically biased towards zero. Fortunately, recent statistical advances enable the analyst to overcome this problem as long as there are multiple indicators of productivity.

---

139. 559 F.2d at 332.
141. \textit{Id.} at 6496.
143. \textit{Id.} at 323 n.3.
145. \textit{Id.} at 5574.
146. 458 F. Supp. at 323 n.3.
148. Since these methods are quite sophisticated, they will not be discussed here. The reader is referred to the following: Boardman, Hui, \& Wold, \textit{The Partial Least Squares—Fix Point Method of Estimating
A third problem is the omission of crucial variables, productivity again being the most important. In *Mecklenburg* the plaintiff's and defendant's salary studies were criticized because "qualitative variables [teaching ability, research contributions, community contributions, etc.] were not considered." These variables are major measures of productivity. Similar criticisms were raised in *Presseisen v. Swarthmore College*, and in *Pennsylvania v. Local 542 International Union of Operating Engineers*, where the defendant argued that the plaintiff's analysis omitted crucial factors, most importantly skill. In *Agarwal* the plaintiff's regressions were criticized because they "totally excluded certain kinds of information which could have had some bearing upon salary such as job level at McKee, prior salary, and past overseas assignment."

A fourth problem is specification error, which can occur in a number of ways. In one situation a statistician may omit the interaction term \((DP)\) in equation (2). This means that while the "intercepts" differ, the "slopes" are forced to remain the same, thus assuming incorrectly that productivity has the same effect on salary for protected group members and majority group members. Expert testimony identified this problem in *Presseisen*. In another situation the statistician may estimate equation (2) and conclude that there is discrimination when in fact the true model is logarithmic:

\[
\log S = a + b \log P + w. \tag{5}
\]

In this case the statistician may conclude that there is discrimination when, in fact, there is none. Only with the correct formulation of the model will it be concluded correctly that there was no discrimination. (That is, no demographic (D) variable appears in the true model, equation (5).) McCabe, in his article on the use of regression analysis in discrimination cases, provides an example that illustrates when such conclusions may occur.

Many other problems have been identified but they are often case-specific, sometimes incorrect, and sometimes appear to be based more on lack of familiarity with regression than on the regression analyses themselves. For example, in *Dickerson v. United States Steel Corp.*:

---

153. 442 F. Supp. at 615.
The court did strike the regression analyses because they contained evidence outside the limitations period. In that situation, however, the statistical technique was much more sophisticated and sensitive. The court felt that inclusion of the earlier data would so severely influence the regression curves' shapes as to render the analyses irrelevant. Nonetheless, the court did allow summary statistics based on the same data as the regression analyses because "[t]his inactive study, the Court believes, does not have the same problems of delicacy [as does the regression]."157

While statistical methodologies involving the calculation of t-statistics, chi-square statistics or regression coefficients do have problems, the evidence suggests that the courts have become increasingly sophisticated and are sufficiently familiar with the problems to ensure that these methodologies are used appropriately. Indeed, the courts have demonstrated their ability to use the results from more sophisticated methodologies, such as logit analysis, should such results be presented.

In summary, the courts have progressed enormously in their analysis of disparate impact over the last decade. They have moved with the times and become familiar with relatively sophisticated statistical techniques. Given this progress, it is with some concern that we view the administrative agencies' advocacy of reliance on a simple rule-of-thumb or "bottom line" principle such as the 80% rule. Some of the potential impacts of this rule are examined in Part III.

III
THE IMPACT AND RAMIFICATIONS OF THE 80% RULE

A. The 80% Rule

According to the 80% rule:

A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) or (eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.161

To clarify this rule, suppose there are only two groups, as in Table 1. Under the four-fifths rule the protected group may claim adverse impact if \( x_1/n_1 \) is less than

\[ g(\text{Demographics, Productivity, Demographics} \times \text{Productivity}) \]

and can be estimated by logit or probit analysis. Appropriate methodologies are available to estimate models with polychotomous dependent variables.

156. Id. at 79 n.25.
157. Id.
159. See, e.g., Greenspan v. Automobile Club of Mich., 22 Empl. Prac. Dec. (CCH) ¶ 30,812 (E.D. Mich. 1980). Thus far multivariate statistical methods have rarely been used to study hiring and promotion decisions, partially because regression analysis is an inappropriate method when the dependent variable can take only discrete values. In hiring and promotion cases for example, the dependent variable may be dichotomous, \( Y = 1 \) if the individual were hired or promoted and \( Y = 0 \) otherwise. This type of model could be written in a form analogous to equation (1):

\[ P(Y = 1) = g(\text{Demographics, Productivity, Demographics} \times \text{Productivity}) \]

160. For some techniques and methodologies that may be used in the future, see Conway & Roberts, Reverse Regression, Fairness, and Employment Discrimination, 1 J. Bus. & Econ. Statistics 75 (1983); Kaye, Statistical Evidence of Discrimination, 77 J. Am. Statistical A. 773 (1982).
80% of $x_2/n_2$. The majority group could potentially claim "reverse discrimination" if $x_2/n_2$ is less than 80% of $x_1/n_1$.

According to Blumrosen, "Griggs was the stick, threatening an employer with a finding of illegality, an injunction, and financial consequences if his business practices perpetuate the inferior status of minorities and women. The bottom line is the carrot, rewarding the employer whose practices 'mirror' the congressional purpose."\textsuperscript{162} Certainly the four-fifths rule will encourage employers to follow the rule and thus reduce the probability of liability. However, there are several problems that seriously weaken its usefulness. To begin this discussion the four-fifths rule will be evaluated in light of the previous sections.

The four-fifths rule measures the applicant-flow and, therefore, belongs to cells M or N of Table 2. The criterion itself belongs to the arbitrary percentage difference category in Table 3. All of the problems associated with applicant-flow statistics and arbitrary percentage differences discussed in the previous section pertain to the four-fifths rule.

A major conclusion that emerged from an analysis of the cases in Table 2 was that the courts are unwilling to rely exclusively on a single parity study. In order to shift the burden of proof, a plaintiff typically presented the results of many different types of comparisons. After all, the courts were unwilling to rely exclusively on statistical data which they had found could be quite misleading. If only one type of parity study could be presented, there was a preference for studies that used as a reference the composition of the work force at some other level within the firm (column 4) rather than the applicant pool (column 3).

Under the four-fifths rule, there would be a tremendous incentive to "chill" or manipulate the applicant pools in ways that would be detrimental to protected groups. In order to obtain a high selection rate for a protected group, the company has an incentive to discourage minority applications, which is quite contrary to the intent of any reasonable antidiscriminatory policy.

When one considers the four-fifths rule in light of Table 3 one is immediately struck by the arbitrary nature of the 80% rule. Why 80% rather than 85% or 90%? More importantly, the rule ignores the concepts of chance and statistical significance. As a consequence, three problems arise: (1) there is a high probability that an employer will be found to be discriminating under the four-fifths rule, when in fact, he is not discriminating; (2) there is a high probability that an employer will be held harmless due to compliance with the four-fifths rule when, in fact, he is discriminating against a group of employees; and (3) the four-fifths rule and statistical significance criterion indicate discrimination in quite different situations.

To illustrate the first two points consider again Table 1. A company selects employees from a pool of $N$ applicants, $n_1$ of whom belong to a protected group and $n_2$ of whom belong to the majority group. From this pool the company hires a fixed number of employees $Y$, of which $x_1$, belong to the protected group and $x_2$, belong to the majority group. If the company does not discriminate, the probability that it will hire a member from a group equals the proportion of mem-

\textsuperscript{162} Blumrosen, \textit{supra} note 5, at 6.
bers of that group in the applicant pool. The probability of drawing a person from a particular group will not remain constant, and the number of people selected from a group is hypergeometrically distributed. If, however, \( N \) is large relative to \( Y \), then the number of people selected from a group is approximately binomially distributed.\(^{163} \) Making this simplifying assumption, one can easily compute the probability that a perfectly nondiscriminating employer will violate the four-fifths rule. In a recent article, Boardman used the binomial distribution to calculate the probability that the protected group could claim adverse impact when, in fact, the employer is not discriminating, and these results are presented in Table 4.\(^{164} \) One sees immediately that when few people are selected, the probability that a group could claim adverse impact when, in fact, the firm does not discriminate is very high. Even when fifty people are selected the probability of an incorrect claim of adverse impact is greater than 20%.

**Table 4**

<table>
<thead>
<tr>
<th>Probability that the 80% Rule Would Permit a Claim of Discrimination by the Protected Group When in Fact None Exists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Group Size</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>( n_1 = (3/7)n_2 )</td>
</tr>
<tr>
<td>( n_1 = n_2 )</td>
</tr>
<tr>
<td>( n_1 = (7/3)n_2 )</td>
</tr>
</tbody>
</table>


Conversely a company may, in fact, be discriminating but still comply with the four-fifths rule. Suppose, for example, the two groups are of equal size so that the probability of selecting an employee from the protected group is 50% when the firm does not discriminate. But suppose the firm discriminates; instead of hiring protected group employees with a 0.5 probability, it hires them with a lower probability. The company may still satisfy the four-fifths rule. Probabilities that a firm will satisfy the four-fifths rule when, in fact, it is discriminating are given in

\(^{163} \) The binomial distribution is appropriate when people are hired independently of one another and the probability of drawing a member of a particular group remains constant. This probability is likely to remain relatively constant unless there are major changes in the applicant pool mix. In practice, many firms are hiring from the applicant pool which simultaneously is being supplied with new members. The mix of the pool will not change dramatically unless there are major demographic changes or the largest employers make substantial hiring or firing decisions that are inconsistent with the applicant pool mix. In the absence of major changes in the applicant pool mix, the binomial distribution should be used in preference to the hypergeometric distribution.

\(^{164} \) Boardman, *Another Analysis of the EEOC “Four-Fifths” Rule*, 25 MGMT. SCI. 770, 773 (1979). When the work force consists of more than two groups, the probabilities that a group could claim adverse impact where none exists are even higher than those shown in Tables 1 and 2.
Table 5. These probabilities generally decrease as the number of employees selected increases and as the extent of the discrimination increases. Nonetheless, in some instances the probabilities of these incorrect outcomes are quite large.

Table 5

<table>
<thead>
<tr>
<th>Total number</th>
<th>$n_1 = (3/7)n_2$</th>
<th>$n_1 = n_2$</th>
<th>$n_1 = (7/3)n_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>selected (Y)</td>
<td>P₁ = .2 P₀ = .3</td>
<td>P₁ = .2 P₀ = .3 P₁ = .4</td>
<td>P₁ = .2 P₀ = .3 P₁ = .4 P₁ = .6</td>
</tr>
<tr>
<td>5</td>
<td>.263 .472</td>
<td>.058 .163 .317</td>
<td>.007 .031 .087 .187</td>
</tr>
<tr>
<td>10</td>
<td>.322 .617</td>
<td>.033 .150 .367</td>
<td>.001 .011 .035 .172</td>
</tr>
<tr>
<td>15</td>
<td>.352 .703</td>
<td>.018 .131 .390</td>
<td>.0 .004 .034 .151</td>
</tr>
<tr>
<td>20</td>
<td>.196 .584</td>
<td>.010 .112 .404</td>
<td>.0 .0 .006 .058</td>
</tr>
<tr>
<td>25</td>
<td>.189 .659</td>
<td>.002 .044 .268</td>
<td>.0 .0 .002 .032</td>
</tr>
<tr>
<td>50</td>
<td>.186 .777</td>
<td>.000 .012 .234</td>
<td>.0 .0 .0 .016</td>
</tr>
</tbody>
</table>


Being aware of some problems associated with strict adherence to the four-fifths rule, the promulgating agencies attempted to clarify and interpret their employee selection guidelines in March, 1979, in the form of a series of questions and answers. Of particular importance are those situations in which unusually large or small numbers of people are selected. With regard to small number cases, the interpretations state:

[Generally, it is inappropriate to require validity evidence or to take enforcement action where the number of persons and the difference in selection rates are so small that the selection of one different person for the job would shift the result from adverse impact against one group to a situation in which that group has a higher selection rate than the other group.]

In practice, this modification of the four-fifths rule makes little difference.

With respect to large sample cases, the interpretations state that “relatively small differences in selection rates may nevertheless constitute adverse impact if they are both statistically and practically significant.” While practical significance is not defined, the guidelines do suggest that when “large numbers of selections are involved, it would be appropriate to calculate the statistical significance of the difference in selection rates.”

Let us examine if the conclusions differ very much depending on whether one uses the four-fifths rule or the concept of statistical significance. The results of an analysis of this examination are presented in Table 6 and are calculated on the

---


166. Id. at 11,999.

167. Id.

168. Id.
assumption that there are two groups of equal size. The first column contains the number of successful applicants. The second column presents situations in which the protected group could claim adverse impact under the four-fifths rule. For example, if the company selects ten employees, the protected group may claim adverse impact under the four-fifths rule if four or fewer protected group members were hired. Of the situations chosen for exemplification, the results under the four-fifths rule differs from those under the guideline’s small sample size modification only when five applicants are hired.\textsuperscript{169}

### Table 6

**Situations in Which Protected Group Members Could Claim Adverse Impact**

<table>
<thead>
<tr>
<th>Number of Protected Group Members Selected</th>
<th>4/5 Rule</th>
<th>Statistical Significance\textsuperscript{a} (Independence)\textsuperscript{b}</th>
<th>Statistical Significance\textsuperscript{a} ((Y) constant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((Y))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 - 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0 - 4</td>
<td>0 - 2</td>
<td>0 - 1</td>
</tr>
<tr>
<td>15</td>
<td>0 - 6</td>
<td>0 - 4</td>
<td>0 - 3</td>
</tr>
<tr>
<td>20</td>
<td>0 - 8</td>
<td>0 - 6</td>
<td>0 - 5</td>
</tr>
<tr>
<td>25</td>
<td>0 - 11</td>
<td>0 - 8</td>
<td>0 - 7</td>
</tr>
<tr>
<td>50</td>
<td>0 - 22</td>
<td>0 - 19</td>
<td>0 - 18\textsuperscript{c}</td>
</tr>
<tr>
<td>100</td>
<td>0 - 44</td>
<td>0 - 41</td>
<td>0 - 41\textsuperscript{c}</td>
</tr>
<tr>
<td>200</td>
<td>0 - 88</td>
<td>0 - 88</td>
<td>0 - 88\textsuperscript{c}</td>
</tr>
<tr>
<td>400</td>
<td>0 - 177</td>
<td>0 - 183</td>
<td>0 - 183\textsuperscript{c}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Using a 5 percent level of significance and a one-sided alternative.

\textsuperscript{b} Calculated using the normal approximation and assuming \(Y/N\) is not close to unity.

\textsuperscript{c} Calculated using the normal approximation to the binomial.

The last two columns present analogous data using the statistical significance criterion.\textsuperscript{170} In many court cases, statisticians assume that the hiring process for

\textsuperscript{169} In this case, a group could claim adverse impact if none or only one group member were hired.

\textsuperscript{170} It is up to a court to choose a threshold of significance beyond which an imbalance is considered evidence of discrimination. One commonly used threshold has been the 5\% probability level. Operationally, the most commonly accepted maximum chance of error has been 5\%. Thus, we are interested in situations in which the probability of a particular outcome, given that the company has not engaged in discriminatory practices, is less than 5\%. Suppose that \(Y\) people are selected and that two groups are of equal size so that a company does not discriminate when \(P_1 = P_2 = 0.5\). Under the “statistically significant” rule, we are interested in those values of \(x_1\) such that:

\[
P(X_1 \leq x_1 \mid Y, P_1 = 0.5) \leq 0.05
\]

These probabilities are obtained easily from the cumulative binomial tables or by using the normal approximation to the binomial, and are presented in column four. Notice that the absolute size of the applicant pool is irrelevant. Only the relative size is required to determine \(P_1\) and \(P_2\), under the assumption of no discrimination.

It is more common in court cases to test for significant differences between two selection rates using the formula:

\[
z = \frac{x_1 / n_1 - x_2 / n_2}{\sqrt{\pi(1-\pi)} \left(1/n_1 + 1/n_2\right)}
\]
one group is quite independent of the hiring process for the other group and, consequently, the total number of people selected is a random variable. Such an assumption is appropriate in an employment setting in which all applicants who score above some predetermined level on a test are hired. Under this assumption, the proportion of applicants from each group which is hired is approximately normally distributed if the applicant pools are large. Situations in which the protected group may claim that it has been discriminated against under this criterion are presented in the third column of Table 6. In most employment situations, however, the employer, in attempting to maximize profit, will initially determine the number of employees to be selected which is subsequently considered as a predetermined constant. In this situation, the number of protected group members selected is distributed approximately binomially or hypergeometrically. Situations in which the protected group may claim that it has been discriminated against, assuming a binomial distribution, are presented in the last column.

Table 6 shows that there is a substantial difference between the situations in which a group of employees may claim adverse impact under the statistical significance rules and those under the four-fifths rule. When few individuals are selected the probability that the protected group might claim adverse impact is much higher under the four-fifths rule than under the statistical significance rules. When 200 people are selected, the rules are identical, while for larger selections the statistical significance rules are more stringent for the employer than is the four-fifths rule.

An additional problem with the four-fifths rule is that it ignores everything besides the selection rate; the courts have long recognized that the simple parity studies presented in Table 2 are appropriate only when no other variable is important. According to the preceding discussion of regression analysis, it is clear that the courts and employers recognize that skill and productivity are important and legitimate factors in hiring and promotion decisions. In contrast, the guidelines ignore these variables.

The rule also encourages undesirable corporate responses. Suppose that a firm does not discriminate but that productivity variables are important for a particular position. A correctly performed multivariate analysis would reveal that demographic variables are insignificant and productivity variables are significant so that a discrimination case would be won by the defendants under current legislation. However, a corporation would have a strong incentive to comply with the four-fifths rule in order to avoid potentially large legal and administrative costs.

\[ z = \frac{x_1 - x_2}{\sqrt{Y \cdot (Y^2/N)}} \]  

where \( Y = x_1 + x_2 \) and \( \pi = Y/(n_1 + n_2) \). If \( n_1 = n_2 \) the formula reduces to

where \( y = x_1 + x_2 \).

The test statistic \( z \) is asymptotically normally distributed. Since we are likely to have prior information that the selection rate of group 1 is less than group 2 we will use a one-sided alternative and we require to know the values of \( x_1 \) such that

\[ P(z < -1.645) = .05 \]  

These numbers are presented in column three of Table 6.
In such situations average work force quality may decline. Conversely, some firms may have desired to discriminate in the past but were discouraged from doing so because of ignorance relating to the decision rules used by the courts or the administrative agencies. Now they may push the four-fifths rule to its limit. As we have seen, the rule allows considerable latitude for discrimination.

The problem is more complex with some junior positions. Most large organizations attempt to create a pool of junior employees from which future promotions will be made. Although certain skills and qualifications may be unnecessary for the immediate position, they are rationally required by the organization.

"Warm-body" comparisons almost completely ignore this point. This type of rule, therefore, is appropriate only when skills and qualifications are not required for the present position and there is little likelihood that individuals from within the organization will be promoted to positions requiring greater skill, education, and productivity. Indeed, "warm-body" rules, to the extent that they discourage the hiring of workers with such potential, will encourage organizations to recruit for more skilled or senior positions from outside the organization.

Finally, it is important to remember that the four-fifths rule ignores wage discrimination. Under the rule a corporation may pay majority group workers more than protected group workers whether majority group workers are more, less, or equally productive than protected group workers.

IV
Conclusion

This analysis has attempted to demonstrate two major points: (1) over time the courts have become increasingly competent in dealing with employment discrimination cases, including the use of appropriate statistical comparative frameworks and sophisticated statistical methodologies; and (2) recent federal guidelines have introduced a four-fifths rule which has serious weaknesses. Our conclusion is that the introduction of such a "warm-body" rule could prove to be a serious impediment to the progress of the courts. While the administrative agencies' yearning for such a simple and straightforward rule is understandable, we have argued that the outcomes are likely to be highly unsatisfactory. The weaknesses of the four-fifths rule can be summarized as follows: (1) in many cases the rule will signal the prima facie existence of discrimination when, in fact, it does not exist; (2) in many cases the rule will not detect discrimination when, in fact, it does exist; (3) the rule ignores the central role that future and potential skill plays in our modern society; (4) the rule ignores the emerging centrality of wage discrimination; and (5) the rule downplays the use of statistical significance despite its advantages.

The courts' progress in dealing with statistical comparisons and methodologies has been substantial, and that progress should not be interfered with lightly. For-
Unfortunately, the courts appear to view the guidelines with increasing skepticism. For example, the Second Circuit recently noted:

The [Supreme] Court appears to have applied the Guidelines only to the extent that they are useful, in the particular setting of the case under consideration, for advancing the basic purposes of Title VII . . . . To the extent that the Guidelines reflect expert, but non-judicial opinion, they must be applied by the courts with the same combination of deference and wariness that characterizes the proper use of expert opinion in general . . . . They are entitled to deference, not obedience.¹⁷²

¹⁷² Guardians Ass'n of the N.Y. City Police Dep't v. Civil Service Comm'n. 630 F.2d 79, 91 (2d Cir. 1980), cert. denied, 452 U.S. 940 (1981).