TESTING FOR DISCRIMINATION IN EMPLOYMENT PRACTICES

ROBERT FOLLETT*
AND FINIS WELCH†

I
INTRODUCTION

The last twenty years have witnessed a revolution to the extent government statutes and agencies require employers and other institutions to ensure that their policies and practices are nondiscriminatory. Title VII of the Civil Rights Act of 1964[1] proscribes discrimination in employment practices on the basis of race, color, religion, sex, or national origin. Title IX of the Education Amendments of 1972[2] prohibits sex discrimination in education programs or activities that receive federal financial assistance.

These and related developments have spawned a vast amount of litigation and regulation in a very short time. The budget for the Equal Employment Opportunity Commission grew from $3 million to $142 million between 1966 and 1981. In 1971, 3,970 civil rights cases were tried in the federal courts. By 1981, this number increased to 13,750. Although the number of cases tried under Title VII is not ascertainable, 41% of the civil rights cases filed in 1981 were brought under Title VII.

This article illustrates the use of statistics to test for discrimination in employment practices. A fairly abstract discussion of the general problem is followed by a series of examples that are restricted to promotion practices. Although the issues discussed apply to many other aspects of employment, the discussion does not extend to some important questions. For instance, should one use applicant flow or external availability norms for comparison with the race or sex composition of new hires? Further, what statistical tests are appropriate when the number of persons affected by a particular practice is not fixed? One can easily imagine a fixed number of slots into which persons might be promoted, but except for occasional RIFs (reductions in force), firms do not have slots for terminations. Finally, this article does not discuss the appropriate techniques for analyzing continuous events (pay for example) which take on a range of outcomes, as opposed to binary events (hiring, promotion, and termination) which either do or do not occur. Despite

* Economist and data analyst at Welch Associates.
† President of Welch Associates; Professor of Economics, University of California at Los Angeles.
1. 42 U.S.C. § 2000e (1976 & Supp. V 1981), *amended by* General Accounting Office Personnel Act of 1980, § 8(g), 42 U.S.C. § 2000e-16 (Supp. V 1980) (originally enacted as Civil Rights Act of 1964, §§ 1971, 1975a-1975d, 2000a-2000h-6, Pub. L. No. 88-352, Title VII, § 701, 78 Stat. 253 (1964).
2. 20 U.S.C. § 1681 (1976).

these limitations there are significant similarities between the tests for analyzing promotion and other employment practices. Restricting the discussion to promotion is sufficient for the points to be made in this article.

The last section discusses the natural tensions that arise between statistics, as the pure application of laws of chance, and a litigation process which targets for scrutiny only those practices which have disparate outcomes.

II

COMPARISONS BETWEEN GROUPS

Presumably the days of blatant discrimination, when only whites or men need apply, are gone. Increasingly, the emphasis in fair employment litigation is on the disparate effect of "patterns of practice." This is the area where statistics and notions of chance variability emerge.

Inferences of discrimination always involve comparisons of two or more groups; throughout this discussion only two groups are considered, one legislatively protected group which we call "black" and a reference nonprotected group which we call "white." These groups might be women versus men, blacks or Hispanics versus whites, or other numerous identifiable groups. There is always an employment practice at issue, such as promotion or pay, as well as a set of facts: higher proportions of whites are promoted or whites receive higher average pay. These facts, together with the number of blacks and whites involved, are the essential elements of statistical tests.

If all whites were treated identically in terms of job characteristics, comparing blacks to whites would be a simple matter. The two groups either would or would not be identical. Ambiguity occurs because all whites are not treated the same; some are more successful than others. Similarly, some blacks are more successful than other blacks. Almost every black will be treated worse than some whites but better than others; hence, this is where ideas of statistics or of chance become relevant. The issue is whether there is a pattern favoring one group over another. We begin by comparing averages (the facts). Is the proportion of whites who are successful greater than the proportion of successful blacks? Normally, even in the most evenhanded firm, success rates for blacks and whites will not be identical; consequently, the next question becomes whether the observed difference is so small that it may reasonably be attributed to chance. Developing an answer requires information about the way chance differences occur.

Before examining the logical underpinnings of statistical tests, two issues require clarification. First, the facts are almost always confounded with questions of qualifications. Are the two groups equally qualified? In practical application, this is often the key question. The nature of qualifications, the ways they can be and are measured, the question of their relevance (are they really a sham?), and the way qualifications are introduced into statistical computations are often crucial. For illustrative purposes, however, qualifications can be ignored to highlight the role of chance variability. We do not assume that all blacks and whites are necessarily equally qualified. Rather, we assume that any particular randomly selected white is as likely as any randomly selected black to be the most qualified, or the next most qualified, and that this result is true as to all rankings down to the least qualified. That is, we assume that group designation is irrelevant to rankings based on qualifications.

The second issue to be clarified is the old truism that statistics prove nothing. Regardless of the facts, a statistician cannot say with certainty that a firm either does or does not discriminate. Statistics is like civil law in that it considers the weight of the evidence. A statistician might comfortably say that if the XYZ company did not discriminate the observed facts would not be surprising, and on this basis, the presumption of equal treatment "agrees" with the facts. Conversely, for a company that did not discriminate the facts may be surprising, and if the facts are very surprising, the presumption that the company does not discriminate may be untenable to reasonable people. At issue is the conformability of the facts with the presumption of nondiscrimination.

A. Testing for Discrimination in a Single Employment Practice

When qualifications can be ignored, nondiscrimination means only one thing: group membership is irrelevant to chances of success. This establishes the basis for all statistical tests of discrimination. For concreteness, assume an explicit set of facts: one hundred employees were considered and fifty were promoted. Six of twenty black employees were promoted and forty-four of eighty whites were promoted; that is, 55% of the white employees and 30% of the black employees were promoted. If blacks are as likely to be promoted as whites, would this pattern be surprisingly disparate?

To answer this question, think of an experiment in which one takes 100 slips of paper, writes "promoted" on fifty of them, "not promoted" on the other fifty, thoroughly mixes them in a hat and has a blindfolded referee draw twenty. These twenty slips are "pseudo-blacks" and the remaining eighty are "pseudo-whites." Calculate the proportion of slips in each of these two groups that say "promoted." The observed black and white promotion rates in this company differ by 25%. After selecting the samples of 20 and 80 slips of paper and calculating the proportion of promotion successes of pseudo-blacks and pseudo-whites, two questions are asked: do the success rates of the two pseudo-groups differ by 25% or more and does the difference favor pseudo-whites by as much as 25%?

The slips are again placed in the hat, mixed, drawn, and the answers to these two questions are recorded. After a large number of repetitions, the fraction of yes's for the two questions is counted. The results should show that the first question is answered "yes" approximately twice as often as the second.

These two proportions, the fractions of yes's, are the essential ingredients of statistical tests. The first is known as a two-tailed or unsigned test which measures the difference in success rates irrespective of whether the pseudo-whites or pseudo-blacks are favored. The second, a one-tailed test, counts cases when only the pseudo-whites are favored. The first proportion is used for tests of inequality while the second is used when the nature of the inequality is also at issue. For present purposes, we concentrate only on the first.[3] Next, a criterion is needed for determining whether the observed 25% differential is statistically significant. Assume that the random selection process was such that 80% of the draws resulted in a differential that was as large or larger than 25%. In this event, if treatment were colorblind, one expects to see a difference as large as the one that actually occurred four times in five. The observed differential would be unsurprising. If the trial differentials exceeded 25% only one time in five, the observed differential would be larger than average, but not uncommon. Alternatively, if the repetitive sampling process yields a difference as large as 25% only once in a thousand, the observed difference would be surprising if blacks and whites truly had equal chances of promotion.

The process of mixing and drawing is tedious and unnecessary since the laws of chance governing such processes are known. Statisticians, if given the facts described, would simply use this information and the appropriate formulas to do one of two things: either compute the exact probability of a difference between groups as large as 25% based on a hypothetical process, or appeal to approximations known to be adequate in such cases where differences in success rates follow normal or bell curve distributions. Whatever the mechanics, the conceptual experiment is the same: contrast what is—the observed difference—with what would occur from chance if group membership were irrelevant to chances for success.

The first step in the statistical test is to introduce an hypothesis on which probability statements are made. For analyses of discrimination, the hypothesis is that protected groups (blacks) are treated the same as the reference group (whites). The next step is to use the observed difference, 25% in the above example, to compute the probability that as large a difference could have occurred if the two groups in fact had equal promotion chances. The third step requires interpretation of results. If the computed chance probability is so low that equal treatment is implausible, then factors other than chance were probably involved.

How low is low? If the probability computed under the null or equal treatment hypothesis is below 0.05 then the observed differential is statistically significant at the 5% level. This "critical value" is frequently used by scientists. In large number cases it is equivalent to the two standard deviation rule alluded to in *United States v. Hazelwood School District*[4] and many subsequent cases.

It is important to recognize that the 0.05 rule does not say that when a difference as large as two standard deviations occurs the probability that the two groups are treated equally is 5% or less, nor does it say that the probability of unequal treatment is 95% or more. It only says that if the two groups are treated equally, one would expect a difference as large as the one observed 5% of the time. One presumes that the groups either are or are not treated equally. The choice of a critical value represents a convention that describes the chance that the "weight of the evidence" will be viewed as damning of a nondiscriminating employer. It is obvious that a conservative rule, one with a smaller critical value, carries both

3. The courts appear to favor a two-tailed over a one-tailed test, but to the authors' knowledge, the courts have never explicitly addressed the issue of which is appropriate.
4. 433 U.S. 299 (1977).

greater risks of conclusions of insignificant difference when treatment is truly disparate and smaller risks of convicting innocent parties. This tradeoff accompanies all judgments when uncertainty persists.

B. Testing for Discrimination in Several Employment Practices

The above discussion describes tests for discrimination in a single employment practice. When multiple practices are contested, that procedure should not be followed repetitively as though each of the tests being conducted was analyzing an isolated event.

Although ideas of multiple tests are beyond the scope of this article, the central issue can be illustrated by a simple example. If one were to toss fifty dimes and thirty landed heads, that result in isolation might not be surprising. If one were to toss fifty dimes and fifty quarters, it would be less surprising to find that one of the two sets of tosses resulted in thirty heads. If one were to toss fifty dimes and fifty quarters it would be more surprising to discover that both resulted in thirty heads.

If the 0.05 rule were applied separately to two independent employment practices in tests for discrimination, the chance that one of the two would fail—even with neutral treatment—is about 0.10. In fact, if the 0.05 rule were applied to fourteen independent practices, the chance is about 50-50 that one or more would fail, even with neutral treatment.

If a firm's employment practices are neutral with respect to group membership, random chance will result in some practices which produce disparities in outcomes to the detriment of a given group, while other practices produce disparities to that group's benefit. An overall evaluation of a firm's practices requires a broad view that considers the complete range of observed outcomes. The preceding paragraph demonstrates that an individual, statistically significant disparity to the disadvantage of a particular group may not be statistically surprising within the context of a broader view, if the set of practices under scrutiny includes other outcomes that favor the same group of employees. One expects an equal number of favorable and unfavorable outcomes, some of which may appear extreme if viewed in isolation. Conversely, a set of outcomes that are unsurprising individually may lead to inferences of statistical significance if viewed jointly. If every one of a large number of independent employment practices produces a small disparity to the detriment of a particular group, the weight of the evidence may be sufficient for a statistician to pronounce the set of observed outcomes as collectively unlikely to be attributable to chance. A set of practices is jointly significant if the observed pattern of outcomes is not consistent with group-neutral behavior by the firm. A large individual disparity may not be surprising; a set of small disparities that are all to the detriment of the same group may be very surprising.

III

PROMOTIONS

This section works through an extended example of tests for discrimination between two groups in a hypothetical firm, Ajax Computer Games. Although the logic of the calculation of chance probabilities remains that of conceptual experiments when group membership is randomly determined, this example will show that the context in which promotions occur can affect calculations of chance probabilities used in tests. A point that is frequently misunderstood is also illustrated; namely, in adjusting for factors which may be correlated with qualifications and are correlated with group membership, the adjustment does not necessarily result in inferences of less significant differentials between the two groups.

The first context assumes that promotions occur to fill a fixed number of vacancies. This assumption might be true if promotions were to managerial or supervisory positions where it is known in advance that vacancies exist and will be filled. The conceptual experiment for such cases is a race where one's chances of winning depend on the qualifications of all contestants. The alternative context involves cases where promotion depends only on one's own qualifications. This would include proficiency promotions where a specified level of competence guarantees advancement. Mixed contexts where the best are selected only if they satisfy minimal requirements are not discussed.

The first hypothetical can be presented in the following way:

TABLE 1

AJAX EMPLOYEES

	Promoted	Not Promoted	Total
Whites	71	109	180
Blacks	14	66	80
All	85	175	260

In this table, the first row refers to white candidates and the second row refers to black candidates. The first column refers to successful candidates, and the second refers to unsuccessful candidates. Thus of 180 white employees, 71 were promoted and 109 were not, a success rate of 39.4%. The corresponding figures for black employees show that 14 were promoted, 66 were not, giving a success rate of 17.5% among the 80 black candidates. The bottom row indicates that overall 85 of 260 employees were promoted. This presentation is what statisticians refer to as a 2 x 2 contingency table, meaning two groups (blacks and whites) and two outcomes (promoted and not promoted).

The null hypothesis of equal chances for promotion success, which is used to test whether the difference in promotion rates (17.5% for blacks and 39.4% for whites) is statistically significant, is that a randomly selected black candidate has an even chance to be more qualified than a randomly selected white candidate, and vice versa. Since blacks account for 30.8% (80 of 260) of all candidates, the probability that the most qualified candidate is black is 0.308. This probability is also the chance that the least qualified candidate is black.

The first context of a fixed number of promotion slots refers to a conceptual experiment in which all candidates are ranked and the best 85 are chosen. The process can be thought of as though the 260 candidates are examined to select the best, the remaining 259 are then examined to select the second best, and this procedure is continued until all the promotion slots are filled. This process of selecting one candidate and then selecting a second from the remaining pool is called sampling without replacement. If the two groups are equally qualified, then the chance probabilities associated with contingency tables like the one depicted above follow the hypergeometric distribution.

The second context of selecting all those who are qualified refers to a somewhat different conceptual experiment in which only a candidate's own qualifications matter; that is, the qualifications of other candidates are irrelevant. Each candidate is examined and either passes or fails, independently of the number of other candidates who pass or fail. The process is considered as though the 80 black candidates are themselves each a random draw from a super group. Either the super group is so large that selecting any number up to a maximum of 80 does not measurably alter the chance that the next chosen will pass, or each of the 80 drawn is replaced before the next is selected. What matters is that each of the 80 has an equal chance of passing the evaluation. In such a case, the number passing among the 80 drawn follows the binomial distribution. Similarly, each of the 180 whites is viewed as drawn from a super group and each has an equal chance of passing. The number who pass among the 180 drawn is again presumed to follow the binomial.

The null hypothesis of equal promotion chance is the same as assuming that the two super groups are one. In this case, the probability of observing a particular set of facts—14 of 80 blacks are promoted versus 71 of 180 whites—depends on something that is unknown, the chance of promotion within the super group. Strictly speaking, statisticians cannot calculate true chance probabilities in such cases. But if we are willing to add one fact, that 85 of the 260 total draws pass the evaluation, this unknown probability does not affect calculations.[5] This process, known as conditioning on the full result, gives an identical calculation for the two contexts.

The fact that context does not matter for the simple 2 x 2 comparison can be examined by assuming that the 260 candidates are lined up in rank order of qualifications. If the difference between passing and not passing is the same as for being in the top 85 versus the bottom 175, then context is irrelevant. In a single event, when the competition is fixed, it does not matter whether the context is that of selecting the best, however many, or of selecting only those who surpass some hypothetical norm. It would indeed matter if the 85 promotions were viewed as a series of contests, each consisting of a smaller number of contestants.

Given only the facts of Table 1, the chi-square statistic is appropriate for testing the neutrality of Ajax promotions. The probability of observing a disparity as large as 17.5% versus 39.4% is only 5 in 10,000 (0.0005), much less than the 5 in 100 (0.05) generally regarded as indicating statistical significance. Thus, on the

5. The null hypothesis that the chances of success in the super groups are equal seemingly makes this a fairly innocuous assumption, and it is almost always so viewed. There are exceptions, however, and different techniques based on Bayes Law are sometimes used. *See* A. ZELLNER, AN INTRODUCTION TO BAYESIAN INFERENCE IN ECONOMETRICS 10-11 (1971).

basis of the facts presented in Table 1, it is highly unlikely that chance alone resulted in the observed black/white differential.

The preceding discussion and test assumed that the distribution of job relevant attributes is the same for each group. If this is not the case, any available information on qualifications should be incorporated so that valid inferences can be made regarding the relationship between group membership and the employment practice under scrutiny.

For simplicity, assume that the probability of being promoted depends on how long one has been employed. Some promotions occur during a person's first year, but most are reserved for those with greater seniority. This factor may help to explain the racial discrepancy in promotion rates *if* whites are less likely than blacks to be in their first year of employment. Controlling for a factor in which the groups being compared differ, such as seniority, will not help to explain differences in outcomes *between* groups unless they help explain differences in outcomes *within* groups. For example, even if the black employees in a firm have less seniority on average than white employees, including seniority as a control factor will not eliminate the measured effect of race on promotion probability unless differences in seniority help to explain why some whites are promoted and others are not, or why some blacks are promoted and others are not. This can be best demonstrated by returning to our example.

First, assume that dividing Ajax's candidate pool into high and low seniority generates the following pair of tables.

TABLE 2

AJAX EMPLOYEES BY SENIORITY

A. High Seniority

	Promoted	Not Promoted	Total
Whites	47	68	115
Blacks	4	16	20
All	51	84	135

B. Low Seniority

	Promoted	Not Promoted	Total
Whites	24	41	65
Blacks	10	50	60
All	34	91	125

Note that 115 of the 180 white machinists, or 63.9%, are in the high seniority class compared to only 20 of 80 blacks (25.0%).[6] Note also that promotions are

6. This fact raises the following question: why do white machinists have more seniority than black machinists? It may be that blacks in this firm are more likely to be terminated during their first year than are whites, so that a smaller proportion of blacks than whites ever reach the high seniority category. Alternatively, it may be that this firm has recently instituted an affirmative hiring policy, so that many of its recent hires (persons with low seniority) are black.

more common among those with high seniority (37.8% of the high seniority group is promoted versus 27.2% for the low seniority group).

In this context, controlling for seniority implies that separate comparisons be made between high experience whites and blacks and between low seniority whites and blacks. This differentiation addresses the question whether whites and blacks with similar amounts of experience have similar chances of being promoted. If so, then the large unadjusted disparity in success rates is attributable to experience, rather than to race.

The table for high seniority shows that 47 of 115 whites and 4 of 20 blacks received promotions to supervisor. These numbers translate into success rates of 0.409 for whites and 0.200 for blacks. Among employees in their first year, 24 of 65 whites and 10 of 60 blacks were promoted. The corresponding success proportions are 0.369 for whites and 0.167 for blacks.

Whites on average are more experienced than blacks and a larger proportion of those with high seniority are promoted. However, among whites alone, the success rates of high and low seniority machinists are similar (0.409 and 0.369, respectively). The success rates among high and low seniority blacks are 0.200 and 0.167. Although the probability of promotion for a white or a black increases with seniority, this alone is too weak an explanation for the large unadjusted disparity between white and black success rates.[7] In this example, the disparity in unadjusted success proportions is *not* attributable to seniority, and, after a control for seniority is made, the disparity in promotion rates remains statistically significant.

Next, consider an alternative example. Again the candidates are divided into those with high and low seniority, but now the resultant contingency tables are as follows:

TABLE 3

AJAX EMPLOYEES BY SENIORITY

A. High Seniority

	Promoted	Not Promoted	Total
Whites	64	51	115
Blacks	9	11	20
All	73	62	135

B. Low Seniority

	Promoted	Not Promoted	Total
Whites	7	58	65
Blacks	5	55	60
All	12	113	125

7. In this example, the difference in success rates of high and low seniority employees would occur by chance (assuming a seniority-neutral promotion policy) approximately 7 times in 100 (0.07). Thus, seniority is not a statistically significant factor in determining promotion success.

These tables are similar to the preceding pair in several ways.[8] The number of whites and blacks within each seniority category remains the same. As in the previous case, whites are both more likely to be experienced and more likely to be promoted than are blacks, and, on average, those with more seniority are more likely to be promoted.

The two pairs of tables differ in one significant respect. In this second example, seniority is closely related to promotion probability *within* race. Among whites, the success rates for high and low seniority are 0.557 and 0.108, respectively. For blacks, the corresponding rates are 0.450 and 0.083. There are much greater differences between persons of the same race but different seniority level than between persons of different race but the same seniority level. The disparity in unadjusted success proportions is attributable to the large differences in seniority between white and black candidates. The remaining differential between white and black success rates is not statistically significant.

This discussion assumes that each employee's seniority is known with certainty. In practical application, some of the information deemed necessary to properly differentiate among employees' qualifications may be incomplete or inaccurate. In such cases, controlling for factors which are legitimate predictors of promotion success may fail to eliminate an observed disparity in outcomes even if the firm is behaving in a neutral manner. The ability of control factors to eliminate a disparity is limited by the accuracy with which these factors are measured and recorded. Errors in data tend to mask the relationship between qualification and success; if qualifications seem not to matter, group membership remains as a statistically significant predictor of success. Thus, data inaccuracies are generally to the advantage of plaintiffs.

Before returning briefly to the issue of adjusting for potentially confounding characteristics, reconsider the question of context. The discussion to this point assumes that the 85 promotions occurred on the same day. More often than not, one analyzes the outcome of a firm's employment practices over a relatively prolonged period of time. For practices like promotion, there is a legitimate question concerning the construction of eligibility pools.

Again, assume that the distribution of qualifications is the same among blacks and whites. Ajax's promotions, however, occurred on as many as 85 different days. How does this alter the methodology for comparing the two groups' relative success rates?

Unlike the presumption surrounding the discussion of Table 1, where context did not matter, in this situation the choice of approach depends on whether promotions occur to fill fixed openings or in response to candidates surpassing some qualification criterion. Passing a fixed criterion is a contest, but the contestants are not well specified. Suppose that the pass/fail outcome is based on a test where one passes if the score achieved exceeds a specified cutoff. When a test is administered to different people on different days, there is little if any significance to whether those taking the test were predominately white or predominantly black on

8. *See supra* Table 2, at 178.

promoted when the pool is 90% black and that a black will not be promoted when the pool is 10% black, than there would be if the pools were always 50-50. With the two 90-10 mixtures, the probability that exactly one black is promoted is 0.82, whereas the probability would be 0.50 if both pools were 50-50. The promotion process leaves less room for chance when the composition of pools varies than when the composition is stable.

In the previous discussion of seniority, the division of Ajax employees into high and low seniority groups was arbitrary. A more reasonable approach would be to assume that seniority affects chances of promotion in a more continuous way than in this simple high-low dichotomy. There is a family of techniques which are loosely called regressions that can be used to estimate effects of many factors, including seniority and race or sex, on the outcome being analyzed. "Ordinary least squares" is the most common statistical technique used in litigation, but for EEO matters this technique is usually restricted to analysis of pay. For binary events such as promotion, related techniques are used.

One such appealing technique is probit; the idea underlying this analytical scheme is a cutoff that distinguishes between true and false, pass and fail, and other similar outcomes. The standard probit is designed for situations in which an individual's chances for success are related solely to his or her own characteristics. This procedure is well suited for studying a single event or for studying cases in the proficiency cutoff context. Probit is not directly applicable to cases in which a fixed number of promotion slots are filled by the best candidates available on a particular day. In such cases, each individual's characteristics must be compared to the characteristics of all of that day's contestants. Someone who is likely to win one contest may be unlikely to win another, not because the person is different, but because the competition is stronger.

To our knowledge, techniques employing this broader view of contests have not been used in litigation, but the introduction of such analytical tools is perhaps only a matter of time.

IV

CONCLUSION

The preceding discussion has attempted to provide an intuitive feel for the logical underpinnings of statistical tests used in EEO litigation. It has also tried to illustrate that the contexts in which real world data are produced can importantly affect the tests used. In doing so, issues of correcting for potentially confounding factors have received only cursory treatment.

The authors' experience in these matters is fairly extensive and leaves the impression that differences which emerge in statistical perspective during litigation are more often the product of incomplete and faulty data or of something less than careful scrutiny of the data, than of differences in statistical philosophy. To the extent that philosophical differences occur, they most often focus on issues involving the legitimacy of controlling for factors such as education or credentials that may be viewed either as the vehicle through which discrimination operates or as crucial determinants of productivity. For hiring studies, there are often legitimate questions concerning the choice of a norm: should applicant flow or external availability be used, and, if external availability is appropriate, how is the scope of the market to be defined? These are not unimportant issues. They are, however, issues for another day.

Hopefully, the restriction of scope to promotion has not given the impression that analyses of other employment practices are either not equally rich or are fully redundant. For some purposes, studies of promotion are the simplest and the most straightforward. Promotions were selected as illustrative because they are conducive to the central points of this discussion.

In closing, there is an issue surrounding litigation in general, and EEO litigation in particular, that can be addressed neither by reference to a single employment practice nor by examination of a set of practices. This issue is the dynamics of litigation: the way that cases are selected for filing, the winnowing process that accompanies the steps through class certification, the resolutions of motions for summary judgment that trim away parts of a case and, barring settlement, the restricted issues finally addressed at trial.

The fundamental idea of statistical tests is that differences emerge by chance in neutral environments. The section on comparisons between groups briefly discussed methods for simultaneously testing more than one statistical hypothesis. Recall that if an employer engages in as many as fourteen independent employment practices, each with a 0.95 probability of producing an insignificant disparity (under the 0.05 rule), the chances are better than 50-50 that at least one practice will result in a statistically significant disparity due solely to random chance. Thus, analysis of an individual practice in isolation can lead to a mistaken inference of discrimination if the practice that is scrutinized is chosen because it presents the statistics that are least favorable to an employer.

This selection of the most unflattering practices for analysis is exactly what happens in many EEO law suits. The process begins when plaintiffs file charges and continues through the search for attorneys willing to assume part of the risk. Initially, a suit may charge a company with discrimination in a wide range of areas (hiring, placement, promotion, pay, discipline, termination, etc.). However, once plaintiffs have possession of the data necessary to assess the validity of such charges, they may withdraw complaints against practices which clearly show no disparity, and, if plaintiffs do not withdraw these complaints, defense attorneys are able to file for summary judgment to dismiss areas where differences seemingly do not arise.

There are obvious economies to be realized from reducing the scope of a case, both for plaintiffs and defendants. From a statistical perspective, however, such a procedure may be counter to correct inference. Statistical tests and their associated significance levels are predicated on the notion of randomness. Criteria for assessing statistical significance that would be correct if applied to a practice chosen at random are overly severe if applied to a practice chosen because it is atypically unfavorable to the defendant. Judging a company's guilt or innocence based on its statistically most damning employment practice is similar to estimating a family's average height by measuring its tallest member.

This reduction in the scope of the case has at least two ramifications. First, although narrowing the class to persons affected by only a subset of an employer's practices is likely to reduce the maximum award for which the employer might be held liable, it also allows attention to be focused only on those practices which produce the most unfavorable statistics. Again, it would be statistically surprising if a large employer comprised of many partially autonomous departments, each behaving neutrally toward persons of different race, sex, or ethnicity, did not include some practices which produce unfavorable statistics even in the absence of discrimination. Second, given that an innocent employer may be statistically guilty somewhere in its operation, it behooves one to monitor each employment practice by department to avoid even the appearance of discriminatory policies. The information needed to allow such ongoing monitoring of EEO performance should be generated and analyzed regularly, as a matter of course, before a law suit is filed.

Executive Order 11,246 added affirmative action to our vocabulary, requiring that federal contractors "take affirmative action to eliminate the continuing effects of past discrimination."[11] This directive is anything but a request for neutral, (color- or sex-blind) treatment. Rather, it is a mandate that cognizance of race, sex, and ethnicity be taken to ensure equal treatment. Leaving matters to chance is not a good business practice for federal contractors. Given the dynamics of litigation, it also is not good business practice for noncontractors.

11. Exec. Order No. 11,246, 3 C.F.R. 179 (1965).