

misperceptions of the implications of changes over time in the proportions women comprised of persons discharged in each branch.

D. The Role of Statistical Significance Testing

In its 1977 decisions in *Castaneda v. Partida*, 430 U.S. 482, *Hazelwood School District v. United States*, 433 U.S. 299, with the focus on jury selection in the former case and employment in the latter, the Supreme Court first discussed the role of statistical significance testing in the appraisal of whether race or other demographic characteristics influenced a decision-making process. Since that time there have been many treatments of issues concerning the statistical significance of differences in outcome rates, particularly with respect to ways in the employment context where tests of statistical significance may yield different conclusions from those yielded by the Four-Fifths Rule.²⁸

By and large, however, those treatments have failed to recognize that the Four-Fifths Rule is a measure of the strength of an association (though, for reasons stated in Section A.3 *supra*, an unsound one) while a test of statistical of statistical significance is aimed at determining the likelihood that an observed difference in outcome rates could have occurred entirely by chance. Though a test of statistical significance is influenced by the strength of association reflected in the difference between outcome rates, it is also a function of the number of observations. Thus, the statistical significance of an observed difference in outcome rates poorly reflects the strength of an association though it often is treated as a measure of association.

To my mind, a preoccupation with statistical significance issues in employment discrimination cases since the *Hazelwood* decision has undermined efforts to appraise the strength of an association and may be part of the reason that little of value has been said about the best ways to appraise the strength of an association reflected by a pair of outcome rates, either with respect to employment discrimination issues or other legal issues involving statistical information on demographic differences.²⁹

It is true that in appraising the implications of a difference in a pair of outcome rates with regard to any issue one has an interest in understanding the extent to which the difference may reflect chance variation. If the Four-Fifths Rule were in fact a sound measure of association, one might well wish to employ it in conjunction with tests of significance. An approach to doing so, however, would seem to be appropriately aimed at determining, not whether a difference between rates that violated the Four-Fifths Rule was statistically significant, but whether the

²⁸ A fair summary of that literature may be found in Scott W. McKinley, *The Need for Legislative or Judicial Clarity on the Four-Fifths Rule and How Employers in the Sixth Circuit Can Survive the Ambiguity*, 172 *Capital University Law Review*, 37:171 (2008).

²⁹ I do not address here the varied misinterpretations of demographic differences arising from the attaching of unwarranted importance to the presence or absence of statistical significance when discrimination is at issue. Discussion of misinterpretations of the absence of statistical significance in epidemiological contexts may be found in the [Statistical Significance Vig](#) subpage of my [Vignettes](#) page.

difference between rates represented a statistically significant departure from situation where the disadvantaged group's favorable outcome rate was in fact four-fifths of the rate of the advantaged group.

In any event, difficult questions may often be involved in the appraisal of a discrimination issue while taking into account both the strength of the association reflected by a pair of outcome rates and the likelihood that the observed difference between rates occurred by chance. In particular, it may at times be difficult to compare situations where the strength of an association reflected by a pair of rates is stronger, but the possibility that the difference occurred by chance is greater, in one setting than in another. But such issues need to be addressed in terms of a sound method of association rather than the unsound measures commonly employed in discrimination cases.

E. Measuring Disparate Impact

Prefatory note: This section addresses a complex issue to which I have given occasional thought for some years. The thinking as of August 2008 is reflected on the [Employment Tests](#) subpage of SR. That page will eventually be conformed to the thinking reflected in this section as modified by further deliberation. But for the workshop at which this paper is being presented, I would probably have thought about this issue a good deal more before attempting to resolve it. In any case, the thinking reflected in this section remains a work in progress.

The Introduction discussed two striking anomalies arising from the failure to recognize that reducing the frequency of an adverse outcome, while tending to reduce the relative differences in the corresponding favorable outcomes, tends to increase the relative differences in adverse outcome rates on which regulators principally rely to measure discrimination. There exist similar anomalies in many situations where differences in the circumstances of two groups are measured in terms of relative differences in adverse outcomes, including situations where what would generally be deemed the most obvious less discriminatory alternative to a practice causing what is perceived to be a dramatic disparate impact would increase the relative difference in adverse outcomes underlying that perception. See the [Less Discriminatory Alternative - Substantive](#) subpage of the [Disparate Impact](#) page.³⁰ But these situations principally reflect errors of understanding, where the anomalous aspects of the situations could be obviated by better understanding of certain statistical issues on the part of those dealing with discrimination issue. There would remain, however, a question of how precisely one might appraise the size of a disparate impact of some criterion, including for purposes of determining

³⁰ My [Getting it Straight When Statistics Can Lie](#), Legal Times (June 23, 1993) discusses a number of misperceptions about less discriminatory alternatives. Warranting mention here is *Fisher v. Transco-Services-Milwaukee*, 979 F.2d 1239 (1992), where the Seventh Circuit overturned summary judgment for defendant in a case challenging the impact of a performance standard on older workers. Noting that “[i]t does not take expertise in differential equations to observe that an adverse ratio of approximately 10 to 1 is disproportionate,, the court based its decision on the view that the stringency of the performance standard had caused the large relative difference in rates of termination for rates of failing to meet it. A less stringent standard, however, would generally yield a larger relative difference in failing to meet it. The *Fisher* case was decided more than two decades ago, but I do not know that there would be any better understanding of the matter in the courts today.