The Mismeasure of Discrimination

James P. Scanlan

Paper presented at University of Kansas School of Law
Faculty Workshop
September 20, 2013*

I have recently treated the subject of this article with respect to fair lending issues in The Perverse Enforcement of Fair Lending Laws. Mortgage Banking (May 2014). Key points of this paper appear in a somewhat abbreviated form in an amicus curiae brief filed November 17, 2014 in Texas Department of Housing and Community Development, et al. v. The Inclusive Communities Project, Inc., Supreme Court No. 13-1731 (which develops the argument outlined in Is HUD's Disparate Impact Rule Unconstitutionally Vague?, American Banker (Nov. 10, 2013)). A recent, extensive treatment of the underlying concepts, with limited attention to discrimination issues, may be found in my Race and Mortality Revisited, Society (July/Aug. 2014)].

* The PowerPoint presentation accompanying this paper may be found here. The presentation covered only the subjects addressed in Sections A and B of this paper.

Table of Contents

ABSTRACT

There are four standard measures by which demographic differences in outcome rates are generally measured in the law and the social and medical sciences: (a) relative differences in a favorable outcome, (b) relative differences in the corresponding adverse outcome, (c) absolute differences between outcome rates, and (d) odds ratios. Those relying on these measures, however, have generally failed to recognize the ways the measures tend to be systematically affected by the prevalence of the outcome. For reasons inherent in the shapes of the underlying distributions of factors associated with the likelihood of experiencing an outcome, the rarer the outcome, the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. Thus, as the prevalence of an outcome changes, relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes tend to change systematically in opposite directions. For example, relaxing mortgage lending criteria or public school discipline standards tends to increase relative differences in adverse lending and discipline outcomes while reducing relative differences in the corresponding favorable outcomes. Absolute differences between rates and odds ratios tend also to be affected by the prevalence of an outcome, though in a more complicated way than the two relative differences. Broadly, as uncommon outcomes become more common absolute differences tend to increase; as already common outcome outcomes become even more common absolute differences tend to decrease. As the prevalence of an outcome changes, differences measured by odds ratios tend to change in the opposite direction of absolute differences. This paper explains these patterns and the way the failure to understand them undermines efforts to interpret data on group differences with respect to discrimination issues involving both disparate treatment and disparate impact. It also describes a method for appraising the differences in the circumstances of two groups reflected by a pair or outcome rates, or, put in a way more pertinent to discrimination issues, the strength of the forces causing the outcome rates to differ. The method involves deriving from a pair of outcome rates the difference between the means of the underlying distributions. The method is imperfect, but it is nevertheless greatly superior to the common reliance on a standard measure of difference between outcome rates without consideration of the way the measure is affected by the prevalence an outcome and without recognizing that a different standard measure would commonly yield a different conclusion as to the comparative size of differences in outcome rates in different settings. The paper also addresses certain related issues pertaining to statistical proof of discrimination.

INTRODUCTION

Few things in law enforcement would seem more perverse than that the government should pressure or encourage people to do things that make it more likely that the government will sue them. But two matters lately receiving considerable media attention involve precisely such situations and highlight an incongruous civil rights policy that has persisted for almost two decades. Further, the misunderstanding of statistics underlying that policy has long undermined virtually all efforts in the law and the social and medical sciences to interpret data on differing rates at which advantaged and disadvantaged demographic groups experience favorable or adverse outcomes.

Between December 2011 and June 2012, the Department of Justice (DOJ) settled claims of racial and ethnic discrimination in mortgage lending with recoveries totaling over half a billion dollars. The largest involved Countrywide Financial Corporation ($335 million) and Wells Fargo Bank ($175 million). The complaints in both cases fault lenders for failing to implement less discriminatory alternatives to practices believed to cause minorities to receive subprime rather than prime loans at higher rates than whites, as well as for various practices that led generally to greater frequency of subprime loans.

To put the matter in context, one must look back to the 1990s when there existed great concern about severalfold racial and ethnic differences in mortgage rejection rates. In 1994, belief that a substantial part of those differences resulted from the greater difficulty minorities had meeting standard lending criteria prompted ten federal agencies monitoring fair lending laws to issue an Interagency Policy Statement advising that lenders could be held liable for unnecessarily stringent criteria that disqualified minorities at higher rates than whites.

The approach accorded with federal policy in the employment discrimination context where lowering test cutoffs had for decades been universally regarded as reducing a test's disparate racial impact because lowering cutoffs tends to reduce relative (percentage) differences in pass rates. For example, suppose that at a particular cutoff pass rates are 80% for whites and 63% for a disadvantaged minority group. At this cutoff the minority pass rate would be 79% of the white pass rate. If the cutoff is lowered to the point where 95% of whites pass the test, assuming normal test score distributions, the minority pass rate would be about 87%. With the lower cutoff, the minority pass would be 92% of the white pass rate. The former situation would violate the Four-Fifths Rule the federal government uses to identify an adverse impact (*i.e.*, adverse impact will be found where the pass rate of the lower-scoring group is less than 80% of the pass rate of the higher-scoring group), while the latter would not.

Lending criteria operate just like test cutoffs, and, as with lowering cutoffs, relaxing lending criteria tends to reduce relative differences in meeting the criteria. Lenders that responded to pressures to relax lending criteria would tend to show smaller relative differences between mortgage approval rates of whites and minorities than lenders that did not.

But, while lowering cutoffs tends to reduce relative differences in pass rates, lowering cutoffs tends to increase relative differences in failure rates. In the situation above, the minority failure rate was initially 1.85 times the white failure rate (37%/20%). With the lower cutoff, the minority failure rate would be 2.6 times the white failure rate (13%/5%).

The pattern of contrasting directions of changes in relative differences in a favorable outcome and relative differences in the corresponding adverse outcome as the outcome changes in overall prevalence is not peculiar to test score data or the numbers I chose to illustrate it. Inherent in the shapes of other than highly irregular distributions of factors associated with experiencing an outcome is a pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. Thus, as any outcome changes in overall prevalence relative differences in experiencing it and relative difference in avoiding it tend to change systematically in opposite directions.

Relaxing mortgage lending standards in any manner (including reducing minimum loan amounts or minimum income requirements, to use examples the federal government has cited as practices potentially having a disparate impact on disadvantaged minorities), while tending to reduce relative differences in rates of securing mortgages, tended to increase the relative differences in mortgage rejection rates that had prompted concerns about lending disparities in the first place. Unaware of such pattern, however, regulators and others concerned about lending disparities continued to measure those disparities in terms of relative differences in mortgage rejection rates. Thus, lenders responding to encouragements to relax criteria – hence, tending to reduce relative differences in approval rates while increasing relative differences in rejection rates – increased the likelihood that they would be thought to discriminate on the basis of race or ethnicity. The situation is comparable to pressuring or encouraging employers to reduce test cutoffs and then suing the employers who reduce their test cutoffs the most.

The same pattern exists when the adverse lending outcome is assignment to subprime loan status. And the more lenders follow the suggestions in the *Countrywide* and *Wells Fargo Bank* complaints to reduce rates of such assignment, the larger will tend to be the relative differences in adverse outcomes that regulators monitor.

The second situation lately in the news where the government encourages conduct that makes one more likely to be sued involves racial disparities in public school discipline rates. In March 2012 the Department of Education (DOE) released data showing large relative differences between rates at which minority and white students are suspended or expelled. As both DOE and DOJ had previously done, observers attributed the size of the differences to "zero tolerance" discipline standards in effect in recent decades and called for relaxing those standards. Various jurisdictions have taken or are considering measures to do that.

As in the testing and lending contexts, however, relaxing standards and otherwise reducing discipline rates will tend to increase, not reduce, relative difference in discipline rates. And, as in the lending context, federal investigations will focus on the school systems with the largest relative differences in discipline rates, which frequently will be the systems most responsive to encouragements to relax standards.

But, while these situations involve striking incongruities, the misunderstanding they reflect is pervasive among those analyzing demographic differences in outcome rates. Moreover, that misunderstanding is but a part of a larger failure to recognize that all standard measures of differences between outcome rates tend to be systematically affected by the prevalence of an outcome. As a result of that larger failure, to date little that has been said about the size of the difference in the circumstances of two groups reflected by a pair of outcome rates, whether in the law or in the social and medical sciences, has had a sound statistical foundation.

Section A of this paper explains the patterns by which standard measures of differences between outcome rates tend to be systematically affected by the prevalence of an outcome as well as the fact that the relative difference is an illogical measure of association.

Section B addresses the fallacy of the notion that choice of a measure of differences between outcome rates involves a value judgment or that measures providing contrary interpretations as to the size of a disparity may each be valid in its own way. In doing so, the section explains a theoretically sound means of appraising the strength of an association reflected by a pair of outcome rates, which is also the soundest method of appraising the likelihood that an entity made biased decisions in employment or other contexts.

Section C discusses that is not possible to appraise the likelihood or degree of discrimination in a particular setting based solely on information as to the proportion a group comprises of persons potentially experiencing an outcome and the proportion the group comprises of those experiencing the outcome, though many discrimination analyses are based on such information.

Section D discusses the role of statistical significance testing in the appraisal of whether a difference between outcome rates resulted from biased decision-making.

Section E explores whether, given the pattern whereby changing the frequency of an outcome tends to cause relative differences in experiencing the outcome and relative differences in avoiding the outcome to change in opposite directions, modifying a criterion can reasonably be deemed to affect the disparate impact of the criterion.

Section F discusses the statistical flaws in analyses of discrimination issues that fail to examine the universe of persons subject to the process at issue.

A. Patterns by Which Standard Measures of Difference between Outcome Rates Tend to be Systematically Affected by the Prevalence of an Outcome and the Illogic of the Relative Difference as a Measure of Association

This section describes the principal patterns by which standard measures of differences between outcome rates (proportions) tend to be systematically affected by the prevalence (frequency) of an outcome and certain corollaries to those patterns. It also addresses the illogic of the relative difference between outcome rates as a measure of association.

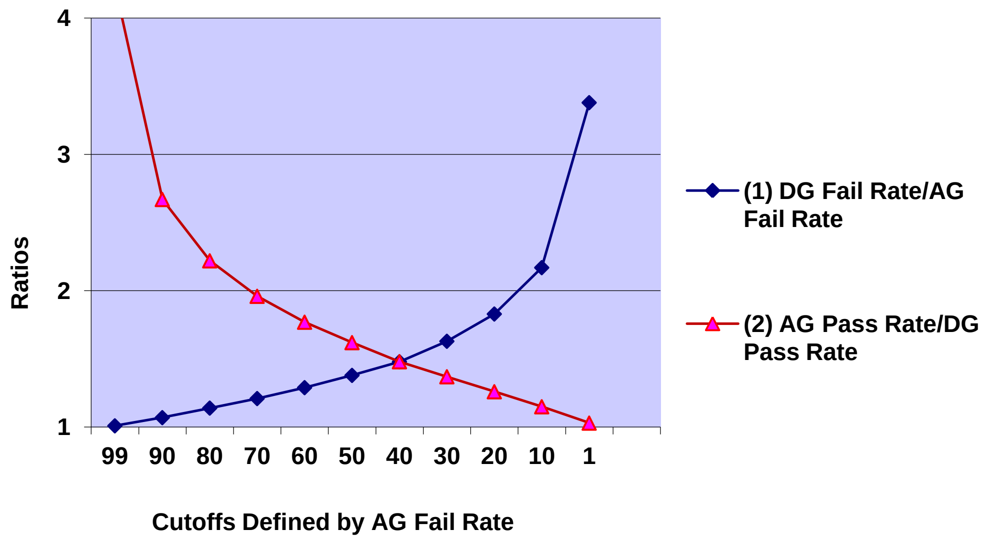
1. The Principal Patterns

This subsection discusses the patterns by which the four most standard measures of differences between outcome rates – (a) relative differences in a favorable outcome, (b) relative differences in the corresponding adverse outcome, (c) absolute differences between outcome rates, and (d) differences measured by odds ratios – tend to be systematically affected by the prevalence of an outcome. As somewhat reflected by the situations described in the Introduction, the most important of these patterns in legal contexts is that whereby the rarer an outcome, the greater tends to be the relative (percentage) difference in experiencing it and the smaller tends to be the relative difference in avoiding it. And, while I will give particular attention to relative differences in this paper, I will also discuss the patterns by which absolute differences and odds ratios tend to be affected by the prevalence of an outcome.

The patterns I describe here can be illustrated with virtually any data showing the proportions of two groups falling below (or above) various points on a continuum of factors associated with the likelihood of experiencing an outcome or any data that simply show the rates at which two groups experience an outcome at various overall prevalence levels. Illustrations using income data, and showing, for example, that reducing poverty will tend to increase relative differences in poverty rates while reducing relative differences in rates of avoiding poverty, may be found in my 2006 Chance editorial *Can We Actually Measure Health Disparities?*. Illustrations based on credit score data, National Health and Nutrition Survey data on systolic blood pressure and folate level, published life tables, and online calculators from the Framingham Study may be found by means of the Collected Illustrations subpage of the Scanlan's Rule page of jpscanlan.com (SR). But in order to illustrate the patterns in their most essential form, and to provide the framework for a sound method of appraising the strength of forces causing a pair of outcome rates to differ, I base the illustrations below on normally distributed test score data.

Figure 1 is based on a situation where the means of the test scores of an advantaged group (AG) and a disadvantaged group (DG) differ by half a standard deviation and where the two distributions of tests scores have the same standard deviation (which are also the specifications underlying the test score hypothetical in the Introduction). A difference of half a standard deviation means that approximately 31% of the lower-scoring group scores above the mean score of the higher-scoring group. The numbers at the bottom of the figure are the failure rates of AG. These are used as benchmarks for the overall prevalence of an outcome and its opposite.[1]

Figure 1. Ratios of (1) DG Fail Rate to AG Fail Rate and (2) AG Pass Rate to DG Pass Rate at Various Cutoff Points Defined by AG Fail Rate



The blue line with the diamond marker represents the ratio of the failure rate of DG to the failure rate of AG at various cutoff points. The red line with the triangle marker represents the ratio of the pass rate of AG to the pass rate of DG at the same cutoff points.[2] Moving from left

---

[1] The prevalence or frequency of an outcome is, strictly speaking, a function of the outcome rate of each group in a population and the proportion each group comprises of the total population. Thus, technically-minded epidemiologists may question my use of the word "prevalence." But I do not expect my usage to confuse readers. Further, some may recognize that, strictly speaking, it is not the prevalence of an outcome that is the determinative factor, but that extent to which the outcome is restricted toward either end of the overall distributions. But I have been describing the patterns of relative differences discussed here with respect to prevalence or frequency of an outcome for a few decades and discussions by others of my treatments of the subject have been in those term. So I am inclined to continue with that usage.

[2] The relative difference is a function of a fraction termed the rate ratio, risk ratio, or relative risk (RR). Depending on which group's rate is used in the numerator, RR can be above 1 or below 1. The relative difference is RR minus 1 where RR is above 1 and 1 minus RR where RR is below 1. It is the somewhat more common practice to use the disadvantaged group's rate as the numerator in the RRs for both the favorable outcome and adverse outcome, in which case the RR will be above 1 for the adverse outcome and below 1 for the favorable outcome, and in the former case the larger the RR the larger the relative difference and in the latter case the smaller the RR the larger the relative difference. This is the approach used in for the Four-Fifths Rule of the Uniform Guidelines on Employee Selection Procedures for appraising the size of differences in favorable outcome rates, for which reason I used that approach in discussion of the pass rates in the Introduction. I also used that approach in the 2006 Chance editorial mentioned *supra*. But for reasons involving both ease of presentation and ease of interpretation, I now prefer always to use the larger rate as the numerator, in which case both RRs will be above 1 and the larger is each RR the larger is the relative difference. While the RR with the larger rate in the numerator will be the reciprocal of the RR with the smaller rate in the numerator, choice of numerator can affect the way a relative difference is characterized –

to right we observe the consequences of serially lowering a cutoff from a point where almost everyone fails to a point where almost everyone passes. And we see that as test failure becomes less common and test passage becomes more common, the relative difference in failure rates increases while the relative difference in pass rates decreases. Thus, we observe the common pattern whereby as the prevalence of an outcome changes relative differences in experiencing it and relative differences in avoiding it tend to change in opposite directions.

That the two relative differences tend to change in opposite directions as the prevalence of an outcome changes may seem counterintuitive at first sight. In fact, however, each element of the pattern is implied in the other, if, indeed, the two elements are not exactly the same thing. For if it is so that the rarer an outcome the greater tends to be the relative difference in experiencing the outcome, it follows that the more the common an outcome the smaller tends to be the relative difference in experiencing it. And if one outcome is decreasing in overall prevalence (hence, increasing the relative difference in experiencing it) the opposite outcome is necessarily increasing in overall prevalence (hence, reducing the relative difference in experiencing that outcome).

The described pattern by which relative differences in favorable and adverse outcomes are affected by the prevalence of an outcome does not merely involve situations where overall prevalence changes over time. The patterns will also be observed in comparisons of demographic differences in different settings or among different subpopulations where the overall prevalence of an outcome varies. For example, relative racial and socioeconomic differences in adverse outcomes will tend to be larger, while relative differences in the corresponding favorable outcomes will tend to smaller, in comparatively advantaged subpopulations (or settings) where the adverse outcomes are less common than in comparatively disadvantaged subpopulations (or settings) where the adverse outcomes are more common. By way of more concrete example based on situations where demographic differences are almost universally misinterpreted, racial differences in mortgage rejection rates and other adverse lending outcomes will tend to be larger, though relative differences in the corresponding favorable outcomes will tend to be smaller, among higher-income than lower-income mortgage loan applicants simply because adverse lending outcomes tend to be less common among the former than the latter. Racial differences in infant mortality tend to be larger, though relative differences in infant survival tend to be smaller, where parents are well educated than when they are less educated. Relative racial, gender, and socioeconomic differences in rates of experiencing any adverse health outcome tend to be larger, while relative differences in avoiding those outcomes tend to be smaller, among the young than among the old. Similarly, relative socioeconomic differences in most adverse outcomes will tend to be larger, though relative differences in the corresponding favorable outcomes will tend to smaller, among whites than among blacks.[3]

---

e.g., 80% is 20% less than 100% while 100% is 25% greater than 80%. But the choice of numerator is of no consequence to any material point in this paper.

[3] Many of the web pages on jpscanlan.com discussed *infra* at 14 and many of the articles referenced in note 13 *infra* discuss the widespread misunderstandings concerning large relative racial or socioeconomic differences in adverse outcomes among advantaged subpopulations. A fair summary may be found at pages 15 to 17 of the October 9, 2012 Harvard University Measurement Letter discussed *infra* at 14.

Appraisals of the comparative size of differences between outcome rates measured in absolute (percentage point[4]) terms or in terms of odds ratios are unaffected by which outcome one examines. In the case of the hypothetical test scores discussed in the Introduction, the absolute difference between rates – 17 percentage points at the initial cutoff and 8 percentage points with the lower cutoff – is the same regardless of whether one examines pass or failure rates. Similarly, regardless of which outcome is examined, the odds ratio is either 2.35 or its reciprocal (0.43) with the initial cutoff and 2.84 or its reciprocal (0.35) with the lower cutoff.[5]

But in order for a measure to effectively quantify the difference between the circumstances of two groups reflected by a pair of outcome rates (or, put another way, to quantify the forces causing the rates to differ) a measure must remain constant when there occurs a general change in the prevalence of an outcome akin to that effected by the lowering of a test cutoff. And, like the two relative differences, the absolute difference and the difference measured by the odds ratios tend to change systematically as the prevalence of an outcome changes. They do so, however, in a more complicated way than the two relative differences.

Roughly, as uncommon outcomes (less than 50% for all groups being compared) become more common, absolute differences between rates tend to increase; as common outcomes (greater than 50% for all groups being compared) become even more common, absolute differences tend to decrease. In cases where the outcome is either common or uncommon, the

---

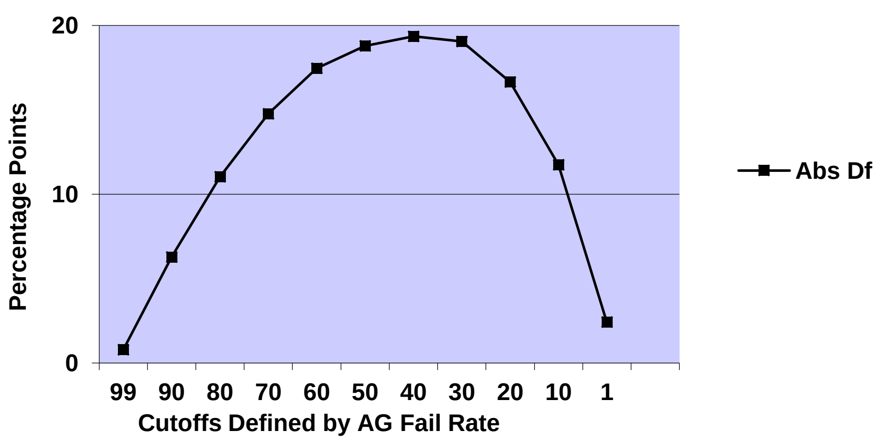
[4] See the Percentage Points subpage of the Vignettes page of jpscanlan.com regarding the extent to which observers use the word "percent" when referring to "percentage points" and the confusion that can result from such usage.

[5] It would be correct to say both (a) that the size of the absolute difference is unaffected by which outcome one examines and (b) that the comparative size of differences measured in absolute terms is unaffected by which outcome one examines. While it is sometimes said that the size of the odds ratio or the difference measured by the odds ratio is unaffected by the prevalence of an outcome, neither is precisely correct. A group's odds of experiencing an outcome is the group's rate of experiencing the outcome divided by its rate of failing to experience the outcome. There are four possible ways to calculate the odds ratio. These include (1) AG's odds of experiencing the favorable outcome to DG's odds of experiencing the favorable outcome; (2) DG's odds of experiencing the adverse outcome to AG's odds of experiencing that outcome; (3) AG's odds of experiencing the adverse outcome to DG's odds of experiencing that outcome; and (4) DG's odds of experiencing the favorable outcome to AG's odds of experiencing that outcome. Methods (1) and (2) reach the same result as each other, which would be 2.35 with the initial cutoff and 2.84 with the lower cutoff. Methods (3) and (4) also reach the same result as each other, which would be .43 with the initial cutoff and .35 with the lower cutoff, and which are the reciprocals of the odds ratios calculated with Methods 1 and 2. It is because an odds ratio calculated by any of the methods will be the same as, or the reciprocal of, odds ratios calculated by the other methods that observers commonly say that the odds ratio is unaffected by which outcome one examines. But since whether the odds ratio is above 1 or below 1 is affected by which outcome is examined (and which group's odds is used as the numerator), I have commonly stated that "the difference measured by the odds ratio is unaffected by whether on examines the favorable or the adverse outcome." That statement also is incorrect in the same sense that it would be incorrect to say that the size of the relative difference is unaffected by which rate is used in the numerator of the rate ratio (see note 2 *supra)*, since, as with the 2.35 and .43 odds ratios with the higher cutoff, in the first case one odds is 135% greater than the other while in the second case one odds is 57% less than the other. It is true, however, that the determinations as to the comparative size of differences between rates, as measured by the odds, will be unaffected by the method by which the odds ratio is calculated. In the case of the hypothetical in the text, for example, the difference measured by odds ratio is greater with the lower cutoff regardless of which of the four methods one uses to calculate the odds ratio. But these nuance, which I note merely for precision, do not affect any material point in this paper.

pattern of direction of changes in absolute differences as the prevalence of an outcome changes will tend to track the pattern of direction of changes of the smaller relative difference. Where the rate of either outcome is less than 50% for one group and more than 50% for the other group, the prevalence-related pattern is difficult to predict. Similarly, such patterns may be difficult to predict when a group's outcome rate crosses either of the points defined by a rate of 50% for an advantaged or disadvantaged group. A more detailed discussion of the pattern by which absolute differences tend to change as the prevalence of an outcome changes, including the way that the size of the difference between means of the underlying distribution can affect that pattern, may be found in the Introduction to the Scanlan's Rule page of jpscanlan.com.

Figure 2 charts changes in the absolute difference between outcome rates as cutoffs are lowered according to the same specifications underlying Figure 1.

Figure 2. Absolute Differences between Rates of AG and DG Pass (or Fail) Rates at Various Cutoff Points Defined by AG Fail Rate

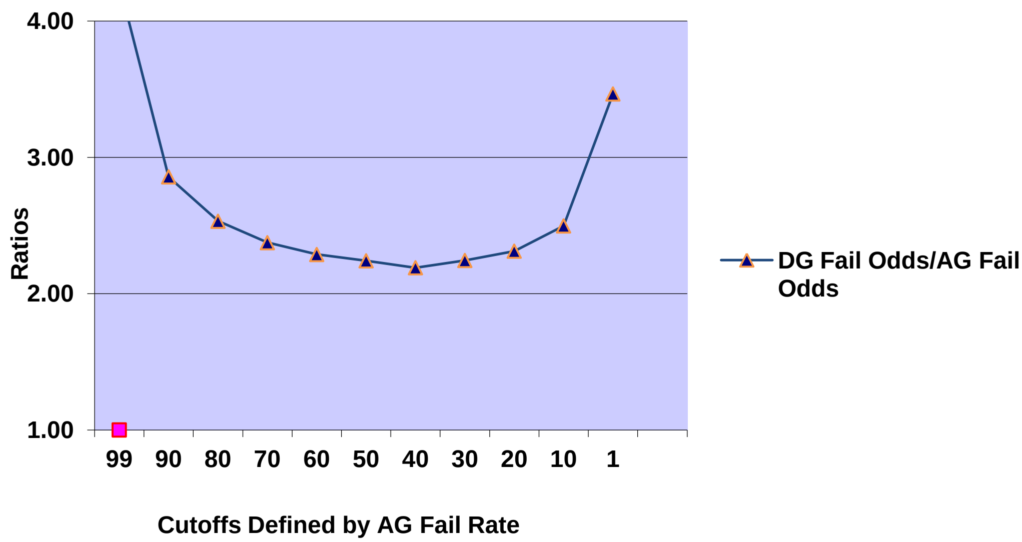


As with the above-described pattern concerning the two relative differences, the pattern by which absolute differences between rates of advantaged and disadvantaged groups tend to be affected by the prevalence of an outcome is observed with respect both to changes in the prevalence of an outcome over time and to differences in prevalence among different subpopulations or settings. For example, for healthcare procedures where rates are low, general increases in those procedures will tend to increase absolute differences between rates of advantaged and disadvantaged groups; for healthcare procedures where rates are high, general increase will tend to reduce those absolute differences. For healthcare procedures where rates are low, hospitals with generally higher rates for the procedure will tend to show larger absolute differences between the rates of advantaged and disadvantaged groups than hospitals with generally lower rates; for healthcare procedures where rates are high, hospitals with generally higher rates will tend to show smaller absolute differences than hospitals with generally lower rates.[6]

[6] For discussion of the way failure to understand these patterns led to misperceptions about the implications of pay-for-performance on racial differences in healthcare and to the inclusion by Massachusetts of a disparities element in

As the prevalence of an outcome changes, the difference measured by the odds ratio tends to change in the opposite direction of the change in the absolute difference. Figures 3 charts the odds ratio as it would be calculated based on either AG's odds of passing to DG's odds of passing or DG's odds of failing to AG's odds of failing (see note 4 *supra*) according to the specifications underlying Figures 1 and 2.

Figure 3. Ratio of DG Odds of Failure to AG Odds of Failure at Various Cutoff Points Defined by AG Fail Rate



An illustration of these patterns in tabular form using the same specifications underlying Figures 1 though 3 may be found in Table 1 of my 2006 British Society for Population Studies (BSPS) paper *The Misinterpretation of Health Inequalities in the United Kingdom*. An illustration showing the shapes of the underlying distributions may be found in Figure 1 of my 1994 *Chance* article *Divining Difference*.[7]

The illustrations immediately above are based on perfectly normal distributions. I note, however, that the patterns whereby the relative differences change in opposite directions as the prevalence of an outcome changes, as illustrated in Figure 1, would hold so long as the distributions are not highly irregular. They would exist, for example, when distributions are

---

its Medicaid pay for performance program that may well increase healthcare disparities, see pages 21 to 24 of the Harvard University Measurement Letter.

[7] A graphic of the distributions themselves helps to reveal the reasons for the significance of the 50% rates for the groups being compared with respect to the described patterns of changes in absolute differences. For it shows that, in circumstances where an outcome rate is increasing, up to the 50% point any given change along the x-axis is associated with an increase in the incremental portion of the population experiencing the outcome as a result of the change, but beyond the 50% point any given change along the x-axis is associated with a decrease in the incremental portion of the population experiencing the outcome as a result of the change.

uniform (rectangle-shaped). The described patterns of absolute differences and odds ratios will not necessarily hold when distributions are not normal. With uniform distribution, for example, absolute differences would remain the same as the prevalence of an outcome changes up to the point where one group's rate for the outcome or its opposite reaches 0%/100%. Examples of the ways odds ratios change when the distributions are truncated parts of normal distributions may be found in Figures 8 and 10 (slides 13 and 15) of my 2008 International Conference on Health Policy Statistics presentation *Measuring Health Disparities* and in Figure 3 accompanying the Credit Score Illustrations subpage of SR.

But, as shown in most of the other illustrations referenced at the outset of the section, the underlying risk distributions for most outcomes do tend toward the normal. And, while the possibility that the underlying distributions may not be normal is an important consideration with respect to appraising the strength of an association reflected by a pair of outcome rates in a way unaffected by the prevalence of an outcome, that possibility does not detract from the importance of recognizing that standard measures of differences between outcome rates will almost invariably be affected by the prevalence of an outcome in some manner. For the fact that a measure of differences between outcome rates is in *any manner* affected by the prevalence of an outcome renders it an unsound measure of association, even when (or especially when) we do not know just how it tends to be affected.[8]

Further, one will of course find many departures from the described patterns even when the distributions are perfectly normal. Observed patterns of differences between rates at which two groups experience or avoid an outcome are invariably functions of (a) the strength of the forces causing the rates to differ and (b) the prevalence-related/distributionally-driven forces described above. As a rule one's interest in examining pairs of outcome rates involves understanding (a). But only with a firm understanding of (b) can one understand (a).

2. Corollaries to the Principal Patterns

There are two important corollaries to the above-described pattern of correlations between the two relative differences and the prevalence of an outcome. First, reducing the frequency of an outcome will tend to increase the proportion the group most susceptible to the outcome comprises of both (a) those who continue to experience it and (b) those who no longer experience it. For example, assuming groups of equal size, the lowering of the cutoff discussed in the Introduction would cause the proportion minorities comprise of those who pass to increase from 44.1% to 47.8% and the proportion minorities comprise of those who fail to increase from 68.8% to 72.2%. Table 1 of the 2006 Chance editorial shows how reducing poverty will cause blacks to comprise a larger proportion of both the poor and the non-poor.[9]

[8] Examples of the way other measures of differences between outcomes rates tend to be affected by the prevalence of an outcome may be found in the Gini Coefficient and Concentration Index subpages of the Measuring Health Disparities page and Sections A.13 (Phi Coefficient), A13a (Cohen's Kappa Coefficient), and A.14 (Longevity) of SR.

[9] For discussion of the mistaken significance attributed to increases in the proportion a disadvantaged group comprises of the poor during times of declining poverty (always without recognition either of the extent to which the increases were functions of declining poverty, including the poverty of disadvantaged groups, or of the corresponding increases in the proportion such groups comprise of the non-poor), see my *The 'Feminization of*

The pattern whereby reducing the frequency of an outcome tends to increase a group's representation both among those experiencing it and among those failing to experience it may seem even more counterintuitive than the contrasting pattern of relative differences discussed earlier. Again, however, one element is implied in the other. For, if reducing the frequency of an outcome tends to increase the proportion the group most susceptible to the outcome comprises of those who continue to experience the outcome, it follows that increasing the prevalence of an outcome tends to reduce the proportion the group most susceptible to the outcome comprises of those experiencing it. And if one outcome is decreasing (hence increasing the proportion the more susceptible comprises of those experiencing the outcome and correspondingly reducing the proportion the less susceptible group comprises of those experiencing the outcome), it follows that the opposite outcome is increasing (hence decreasing the proportion the group more susceptible to that outcome comprises of those experiencing that outcome and correspondingly increasing the proportion the group less susceptible to that outcome comprises of those experiencing that outcome).

Second, when an outcome changes in overall prevalence, the group with the lower baseline rate for experiencing the outcome will tend to experience a larger proportionate change in its rate of experiencing the outcome than the other group, while the other group will tend to experience a larger proportionate change in experiencing the opposite outcome. The lowering of the cutoff described in the introduction, for example, caused the minority pass rate to increase by 38% while the white pass rate increased by only 18.8%; but it caused the white failure rate to decrease by 75% while the minority failure rate decreased by only 64.9%.[10]

The described patterns of correlations between absolute differences and differences measured by odds ratios with the prevalence of an outcome have their corollaries as well. As

---

Poverty' is Misunderstood" Plain Dealer (Nov 11, 1987) (reprinted in Current (May 1988) and Annual Editions: Social Problems 1988/89, Dushkin (1988)); *The Perils of Provocative Statistics*, Public Interest (Winter 1991); and *Race and Mortality*, Society (Jan./Feb. 2000) (reprinted in Current (Feb. 2000)). These items explain that the more poverty is concentrated in disadvantaged groups, the better off tend to be both society at large and the disadvantaged groups. But see my *Comment on "McLanahan, Sorensen, and Watson's 'Sex Differences in Poverty, 1950-1989"*, Signs 16 (2):409-13 (1991) concerning the varying forces underlying observed patterns. An example of a mistaken emphasis on the concentration of an adverse outcome in a disadvantaged population in a different context may be found in Isaiah, *Widening the Gap: How the Housing Crisis Deepened Racial Disparities in St. Paul and Hour to Fix it* (2010), which discusses the perceived effect of the housing crisis on racial disparities concerning various housing issues in St. Paul and emphasizes the disproportionate concentration of vacant building in disadvantaged areas. It is doubtful that the disproportionality grew, rather than decreased, with the housing crisis. In any case, typically the more vacant buildings are concentrated in the disadvantaged areas of a community, the smaller tend to be the numbers of such buildings in those areas.

[10] Discussion of some of the varied mistaken interpretations of patterns where different demographic groups experience different proportionate changes in their favorable or adverse outcome rates, including proffered explanation for the different proportionate changes and proffered explanations of their implications – all without recognizing that the observed patterns were to be expected simply because of the different baseline rates or that opposite patterns of the comparative size of proportionate changes would be observed for the opposite outcome – may be found in the Reporting Heterogeneity subpage of the Measuring Health Disparities page and the Subgroup Effects, Explanatory Theories, and Criminal Record Effects of SR. An article discussed in the last item is treated in Section B *infra*.

with the patterns of correlations themselves, however, such corollaries do not lend themselves to summary description. It suffices here to note that when uncommon outcomes become somewhat more common higher baseline rates tend to show larger percentage point increases than lower baseline rates; when common outcomes becomes even more common lower baselines rate tend to show larger percentage point increases than higher baseline rates. Because the hypothetical in the Introduction involved high pass rates, we observed the latter pattern.

3. The Illogic of the Relative Difference as a Measure of Association

The rate ratio (with its attendant relative difference, see note 2 *supra*) is the most widely used measure of association in every setting where observers attempt to quantify group differences in outcome rates, including in circumstances where the groups are comprised of treated subjects and control subjects in clinical trials. Yet irrespective of the above-described pattern by which relative differences tend to change systematically in opposite direction as the prevalence of an outcome changes, the rate ratio is an illogical measure of association.

The point can be best illustrated with respect to the expectation, explicit in the clinical context and explicit or implicit in varied other contexts, that, absent the occurrence of a subgroup effect (also termed differential effect/effect heterogeneity/interaction), a factor that affects the likelihood of experiencing an outcome will cause equal proportionate changes in different baseline rates for that outcome. Yet, if a factor causes equal proportionate changes in different baseline rates for an outcome, it will necessarily cause unequal proportionate changes in the corresponding opposite outcome rates. That is, if Group A has a baseline rate of 5% and Group B has a baseline rate of 10%, a factor that reduces the two rates by equal proportionate amounts, say 20% (from 5% to 4% and from 10% to 8%), would necessarily increase the opposite outcome rates by two different proportionate amounts (95% increased to 96%, a 1.05% increase; 90% to 92%, a 2.2% increase). And since there is no more reason to expect that two groups would undergo equal proportionate changes in one outcome than there is to expect them to undergo equal proportionate change in the opposite outcome, there is no reason to regard it as somehow normal that the two groups would undergo equal proportionate changes in either outcome. See the Subgroup Effects, Illogical Premises, and Inevitability of Interaction subpages of SR. The same, of course, holds for expectations about the ways two groups' rates change over time as the prevalence of an outcome changes over time. See note 10 *supra*.

The same reasoning applies with respect to regarding a rate ratio of, say, 0.8 (or 1.25 when the larger figure is used in as the numerator), as in the Four-Fifths Rule of the Uniform Guidelines on Employee Selection Procedures, to mean the same thing when the rates are 8% and 10% that it means when the rates are 40% and 50%. See the Illogical Premises II subpage of the SR and the Four-Fifths Rule subpage of the Disparate Impact page of jpscanlan.com.

So far as I can tell, however, the illogic of the rate ratio as a measure of association in the context of the appraisal of demographic differences, including differences in outcome rates of treated subjects and control subjects in clinical trials, does not undermine a rule whereby, in circumstances where evidence establishes that a drug (or other agent) more than doubles the risk that a certain category of persons will experience some adverse outcome, and a person in that category who takes the drug experiences that outcome, it will be appropriate to infer that more likely than not the drug caused the outcome.[11] But that it is illogical to assume that a factor that is observed to cause a certain proportionate change in a particular baseline rate will cause the same proportionate change in other baseline rates may importantly affect analyses of particular cases (as explained in the next section).

Despite the many types of data illustrating the patterns described above, in mainstream literature on demographic differences in the law and the social and medical sciences one finds almost no recognition of the way standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome (or even of the problematic nature of the relative difference as a measure of association). Indeed, there is very little recognition that different measures sometimes yield contrary results, much less that they tend systematically to do so. The scope of the misinterpretation of data on demographic differences arising from the failure to understand these issues is reflected in the materials in, or made available by, the Measuring Health Disparities (MHD), Scanlan's Rule, Mortality and Survival, Discipline Disparities, Feminization of Poverty, Immunization Disparities, Educational Disparities pages of jpscanlan.com (along with their varied subpages), as well as the simple fact that neither the Department of Justice nor the Department of Education has yet shown an understanding that lowering standards will increase relative differences in failing to meeting them. The most comprehensive description of the scope of the misunderstanding of these issues in one place may be found in my October 9, 2012 Harvard University Measurement Letter. That letter also shows that health and healthcare disparities research at Harvard Medical School and Harvard School of Public Health reflects no better an understanding of the issues than that reflected in the civil rights enforcement policies of the Departments of Justice and Education.[12]

Section E.7 of MHD discusses the extent of scholarly agreement with my descriptions of the patterns by which standard measures tend to be affected by the prevalence of an outcome and shows that there has been some increased understanding to these issues in recent years. But, save for the misguided decision of the National Center for Health Statistics (NCHS) to address the issues simply be recommending that health and healthcare disparities always be measured in

[11] Though not challenging this rule, I am not certain how to reconcile it with the fact that a factor that causes a baseline rate of 10% to increase to 21% (to use the example discussed in the next section), hence more than doubling the chance of that outcome, decreases the rate of experiencing the opposite outcome by only 12.2% (90% reduced to 79%).

[12] Published treatments of the misunderstandings of these issues in particular contexts, many of them legal, may be found in my *Race and Mortality Revisited*, Society (___ 2013) (in press); *The Paradox of Lowering Standards*, Baltimore Sun (Aug. 5, 2013); *Regulators Need Schooling on Measuring Lending Bias*, American Banker (June 14, 2013); *Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies*, Amstat News (Dec. 2012); *Racial Differences in School Discipline Rates*, Recorder (June 22, 2012); *"Disparate Impact": Regulators Need a Lesson in Statistics*, American Banker (June 5, 2012); *The Lending Industry's Conundrum*, National Law Journal (Apr. 2, 2012); *Can We Actually Measure Health Disparities?* Chance (Spring 2006); *Understanding Racial Difference in Infant Mortality*, PrenatalEd Update (October 2000); *Race and Mortality*, Society (Jan./Feb. 2000), *Mired in Numbers* Legal Times (Oct. 12, 1996); *When Statistics Lie*, Legal Times (Jan. 1 1996); Divining Difference, Chance (Spring 1994); *Getting it Straight When Statistics Can Lie*, Legal Times (June 23, 1993); *Bias Data Can Make the Good Look Bad*, American Banker (Apr. 27, 1992); *The Perils of Provocative Statistics*, Public Interest (Winter 1991); *Comment on "McLanahan, Sorensen, and Watson's 'Sex Differences in Poverty, 1950-1989"* Signs 16 (2):409-13 (1991); *An Issue of Numbers*, National Law Journal (Mar. 5, 1990); *The "Feminization of Poverty" is Misunderstood*, Plain Dealer (Nov 11, 1987).

terms of relative differences in adverse outcomes (see Harvard Letter at 29-31), that increased understanding has yet to be reflected in changes in the way research is conducted even among researchers and institutions who have evidenced that understanding.

B. Appraising the Likelihood That a Difference between Outcome Rates Resulted from Biased Decision-Making

This section addresses, in light of the patterns of correlations between the prevalence of an outcome and standard measures of differences between outcome rates described above, an approach by which one might soundly appraise the likelihood that an entity engaged in biased decision-making with respect to some favorable or adverse outcome.

I illustrate the main point of this section with the four rows of data in Table 1. The table presents four situations that we might initially regard as the hiring patterns of four employers who draw from the same labor market and where we are required to rank the employers in descending order of the likelihood that it has made biased hiring decisions or the degree of bias in those decisions. The principles I intend to elucidate, however, would apply equally in a range of circumstances, including, among varied others: (a) where the employers hire from different labor markets and we are required to rank the employers by the size of the difference in the qualifications of the applicants that would be necessary to explain the observed disparities as other than a result of biased hiring decisions; (b) where the figures represent the hiring pattern of a single employer from year to year and we are required to determine the comparative likelihood that the employer made biased decisions in the various years or the degree bias from year to year; (c) where the figures represent the hiring pattern of a single employer for different types of jobs and we are required to determine with respect to which jobs the employer was more likely to have engaged in biased decision-making or the comparative degree of bias for each type of job; or (d) where the figures represent the hiring pattern of a particular employer for a particular job broken down by level of qualifications of the applicants for the job and we are required to determine with respect to which level of qualifications the employer is more likely to have engaged in biased decision-making or the comparative degree of bias for each level of qualification. The four columns following the hire rate columns contain the four standard measures of differences between outcome rates discussed in Section A.

Table 1. Hypothetical Patterns of Hiring Rates of Applicants from an Advantaged Group (AG) and a Disadvantaged Group (DG) at Four Employers [ref b4510 a1]

Employer	AG Hire Rate	DG Hire Rate	Rate Ratio Hire	Rate Ratio Rej	Abs Diff	Odds Ratio
A	20.0%	9.0%	2.22 (1)	1.14 (4)	0.110 (4)	2.53 (1)
B	40.0%	22.6%	1.77 (2)	1.29 (3)	0.174 (2)	2.28 (3)
C	70.0%	51.0%	1.37 (3)	1.63 (2)	0.190 (1)	2.24 (4)
D	80.0%	63.4%	1.26 (4)	1.83 (1)	0.166 (3)	2.31 (2)

There are four principal ways observers might rank the employers according to the likelihood that their hiring decisions were biased or the extent of such bias. Those who measure disparities in terms of relative differences in favorable outcomes – as might be done in an employment discrimination case involving hiring or promotion – would rank the employers A,B,C,D.

Those who measure disparities in terms of relative differences in adverse outcomes – as would commonly be done in a lending discrimination case or in an investigation of disparities in school discipline, and as might also be done in an employment discrimination case where the favorable outcome is retention and the adverse outcome is termination – would rank them D,C,B,A, the opposite of the first approach.

Those who measure disparities in terms of absolute difference between rates – as has been done in studies of lending disparities by the Federal Reserve Board[13] and as is increasingly done in studies of public school proficiency disparities and healthcare disparities – would rank them C,B,D,A.

And those who measure disparities in terms of odds ratios – as might be done in analyses that attempt to adjust for differences in characteristics of the two groups by means of logistic regression – would rank them A,D,B,C, the opposite of the ranking based on absolute differences.

In health and healthcare disparities research, it has been asserted that even when various measures lead to contrary conclusions about such things as whether a disparity has increased or decreased over time, each measure may be valid in its own way and observers must make value judgments in choosing among the measures. I have used this illustration to refute that assertion by pointing out that it would be absurd to maintain that one employer is more likely to be biased than another as to selection while another is more likely to be biased as to rejection. It would be likewise absurd to say that contrasting interpretations as to likelihood of bias based on either of the two relative differences and the absolute difference (or odds ratio) could all be sound or that determining which employers are the more biased involves a value judgment. Rather there can exist only one reality as to the comparative bias of the employers reflected in the data. The same holds for the alternative formulations of the hypothetical.[14]

---

[13] See Marvin M. Smith and Christy Chung Hevener, *Subprime Lending Over Time: The Role of Race*, (Federal Reserve Bank of Philadelphia (October 2010) and Robert B. Avery, Kenneth P. Brevoort, Glenn B. Canner, *The 2006 HMDA Data*," Federal Reserve Bulletin (December 2007).

[14] A similar point could be made in any circumstance where one seeks to explain the forces causing an observed difference to be larger in one circumstance than another or to draw inferences based on the perception that one difference is larger than another. For example, if the C and D rows of Table 1 pertained to highly-credentialed applicants who tend generally to have higher success rates than applicants with limited credentials, and the A and B rows pertained to applicants with limited credentials, some observers might maintain that the smaller hire rate ratios for rows C and D reflect a pattern whereby employers tend to treat applicants of different demographic groups more equally when objective indicators of qualification are present, but will rely more heavily on stereotypes when objective indicators of qualifications are not present.[14] See Seik Kim. Statistical Discrimination, Employer Learning, and Employment Differentials by Race, Gender, and Education. Population Association of America 2013 Annual Meeting, New Orleans, LA (Apr. 11-13, 2013. But the comparative size of the relative differences as to rejection would support an opposite inference. Both inferences cannot be correct.

What then would be the soundest ranking of the employers with regard to the likelihood or degree of bias in its hiring decisions or the issues in the other formulations of the hypothetical? Each row of information is based on the specifications underlying Figures 1 through 3 at different cutoff points. There thus is no rational basis for asserting that the strength of the forces causing the observed differences in hire rates varies among any of the four situations reflected in the table, and any measure that suggests the strength of those forces does vary from situation to situation is a flawed measure.

Implicit in the above illustration is that the most theoretically sound way to appraise the strength of the forces causing rates of advantaged groups and disadvantaged groups to differ is to derive from pairs of outcome rates the difference between the means of the underlying distributions of factors associated with experiencing the outcome at issue. That approach exists in the form of procedure known as the probit. I commonly call the value generated the EES, for estimated effect size.[15]

Thus, just as we are able on the basis of a particular difference between means of the underlying distribution – half a standard deviation in each of the above illustrations – to determine the rate of DG corresponding to any AG rate (or vice versa), the probit allows us to estimate from any pair of rates the difference between the means of the underlying distributions. For example, when favorable outcome rates for AG and DG are 30% and 10% we can estimate that the difference between the means is .757 standard deviations; when those rates are 15% and 5% we can estimate that the differences between the means is .608 standard deviations.

To put these figures in perspective, the .757 standard deviation difference reflects a situation where, in testing terms, approximately 22% of DG score above the mean for AG; the .608 standard deviation difference reflects a situation where approximately 27% of DG score above the mean for AG. Similar illustrations, by tenths of a standard deviation, based on situations at different levels of overall prevalence where the disadvantaged group's favorable outcome rate is four-fifths of the advantaged group's favorable outcome rate may be found in Table 1 of the Four-Fifths Rule subpage of the Disparate Impact page of jpscanlan.com.[16]

A more concrete illustration, both of the patterns by which relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes tend to change in opposite directions as an outcome changes in overall prevalence, and of a sound method of appraising the strength of the forces causing the rates to differ, is set out in Table 2,

[15] After much thought on how one might measure disparities in health or healthcare outcomes in a way not affected by the prevalence of an outcome, in 2007 I developed a mechanical means of deriving from a pair of rates the difference between means and made that procedure available through the Solutions and Solutions Database subpages of the Measuring Health Disparities page of jpscanlan.com. I gave a number of presentations on the procedure at health policy, demographic, or statistical conferences over the next few years. But in 2010 I came to recognize that the procedure has long existed in the form of the probit.

[16] A principal purpose of the Four-Fifths Rule subpage is to show that the Four-Fifths Rule of the Uniform Guidelines on Employee Selection Procedures is an illogical measure of association, a point made with respect to all rate ratios/relative differences in Section A.3 *supra*.

which is based on the rates at which black and white male public school students received various levels of discipline. The table presents data on three categories of severe discipline – (1) Suspension – Out of School, (2) Total Expulsion, (3) Expulsion – Total Cessation – ordered according to increasing severity. The table shows the rate of receiving any form of the discipline ((1), (2), or (3)), the two most severe forms ((2) or (3)), and the most severe form ((3)), along with the black/white ratios of rates of receiving the discipline and the white/black ratios of rates of avoiding the discipline.[17]

The table shows that, consistent with the pattern of relative differences described earlier, as the level of discipline becomes more severe and hence less common (or, more precisely, as it is increasingly restricted toward the tail of the overall distribution, see note 1 *supra*), the relative difference in discipline rates increases while the relative difference in rates of avoiding discipline decreases. But the EES figure shows that, notwithstanding that many would read the increasing relative differences in disciplines rates at each increasingly severe level of discipline as in some manner reflective of bias of decision-makers, to the extent that we can measure the forces causing black and white rates to differ, such forces appear to be smaller for expulsion than for suspension.

Table 2. Black and White Male Rates of Discipline with Ratios of Rates of Experiencing and Avoiding Discipline and Estimated Effect Sizes [ref b2813 c 4]

Cat	B	W	B/WAdvRatio	W/BFavRatio	EES
All	19.50%	6.92%	2.82	1.156	0.63
2&3	0.83%	0.26%	3.18	1.006	0.36
3	0.19%	0.06%	3.26	1.001	0.40

Further illustrations of EES values involving discipline rates, including in situations where the standard measures behave according to the patterns described at the outset and in situations where there occur departures from those patterns, may be found on the Discipline Disparities page and its Los Angeles SWPBS subpage. The latter item also shows that following implementation of a program aimed at reducing overall discipline rates in the public schools of Los Angeles, relative differences in discipline rates not only increased, but did so to a much greater degree than would be expected to result solely from the statistical forces described here.

Table 3 illustrates several pertinent issues in the employment setting. The table is based on a study[18] that attempted to determine whether a criminal record had greater effect on the employment prospects of black or white job applicants by analyzing callback rates of tester applicant pairs comprised of persons of the same race, one of whom noted a criminal conviction

---

[17] I present the data in terms of rates of experiencing each type of discipline or a more severe type of discipline (rather than rates of experiencing each type of discipline) for reasons discussed in Section A of the NEPC Colorado Study subpage of the Discipline Disparities page of jpscanlan.com. Related issues are discussed in the Life Tables Illustrations subpage of the SR and the Life Table Illustrations document.

[18] Debra Pager, The Mark of a Criminal Record, American Journal of Sociology, Vol. 108, No. 5, pp. 937-975 (2003). The study is also treated in a 2007 book by the same author, *MARKED: Race, Crime, and Finding Work in an Era of Mass Incarceration.* Chicago: University of Chicago Press.

on the application and one of whom did not. The table shows callback rates for blacks and whites with and without convictions and, for each race, (a) the ratios of the callback rates for those without a conviction to the callback rates of those with a conviction, (b) the ratios of the rates of not receiving a callback for those with a conviction to the rates of not receiving a callback for those without a conviction, and (3) the EES for the effect of the conviction on callback prospects. Based on fact that the ratio of callback rates for those failing to note a conviction to the callback rates of those noting a conviction was larger for blacks than for whites (2.8 versus 2.0), the author concluded that effect was larger for blacks and posited various not implausible explanations for the difference.

Table 3. Rates of Receiving Callbacks for Testers Indicating a Conviction and Not Indicating a Conviction on their Applications by Race of Tester, with Rate Ratios for Receipt and Non-Receipt of Callbacks and Estimated Effect Size

Race	Conviction	NoConviction	RateRatioCall	RateRatioNoCall	EES
W	17%	34%	2.00	1.26	0.542
B	5%	14%	2.80	1.10	0.565

But the observed pattern of larger relative differences between callback rates for those with and without convictions among blacks than whites on which the author based her conclusions about the comparative size of the effect of convictions was of the nature that would typically occur in the situation given that blacks without convictions had lower hire rates than whites without convictions. On the other hand, the relative difference in rates of failing to receive a callback was larger for whites than for blacks, and those relying on that measure would have been in a situation of positing different explanations for the pattern. The EES, however, shows that effects of convictions on the callback situation of blacks and white were very similar

Table 4 then treats the study's callback rates in terms the differences between the callback rates of blacks and whites among persons who indicated a conviction and those who did not. From this perspective, too, we observe the common patterns where in the advantaged subpopulation where the favorable outcome is more common (those without convictions), the relative difference between blacks and whites is smaller for the favorable outcome, but larger for the adverse outcome, than among the disadvantaged subpopulation.[19] The EES then shows that, consistent with the fact that the racial difference in effects of a criminal conviction was small, racial difference among those with a conviction differs only slightly from the racial difference among those without a conviction.

Table 4. Black and White Rates of Receiving Callback by Whether Applicant Indicated a Conviction, with Rate Ratios for Receipt and Non-Receipt of Callbacks and Estimated Effect Size

Conviction	W	B	RateRatioCall	RateRatioNoCall	EES
N	34%	14%	2.43	1.30	0.668
Y	17%	5%	3.40	1.14	0.691

[19] One may note that Tables 3 and 4 illustrations the way that one will observe the same regardless of how on

One may note that the two EES figures in Table 4 are not particularly close to the EES figures in Table 3. There is no reason they should be since they are measuring different things (effect of criminal conviction by race in Table 3 and effect of race by conviction status in Table 4). But the difference between the two EES figures in one table will always equal the difference in the other, which in this instance is .023 standard deviations in both cases.

As discussed on the Criminal Record Effects subpage of SR, the author conducted another study that did seem to identify a greater effect of convictions on employment prospects of black than white applicants. But, as shown above, standard measures ordinarily are problematic for identifying the differences in such effects and are always problematic for quantifying size of the effects or the size of the difference between the effects.

Illustrations of EES values, along with patterns of standard measures, pertaining to proficiency rates of different racial and ethnic groups in public schools in the United States may be found in the tables of the Educational Disparities page of jpscanlan.com and its Disparities by Subject, Harvard CRP NCLB Study, and New York Proficiency Rate Disparities subpages.[20]

At the end Section A.3., I mentioned an issue regarding the fallacy of the assumption of a constant rate ratio reflecting the effect of a factor across different baseline rates as it might bear on the determination of causality in a tort context. The measurement approach described in this section bears on that issue in the following respect. Assume that a study had shown that for persons with a 10% baseline rate of experiencing a certain type of adverse outcome a drug caused an increase of that rate to 21%. That would be sufficient to allow one to infer causality in a situation where persons with a 10% baseline rate who took the drug experienced an adverse outcome. As noted, however, it would not be logical to assume that the observed 2.1 rate ratio (110% increase in risk) would apply to all baseline rates.

But one can on the basis of the increase in the adverse outcome rate from 10% to 21% found in the study derive an estimate of the effect of the drug of .48 standard deviations (as in Table 3 *supra*). And from that figure, one can estimate that the drug would increase a baseline rate of 20% only to 36% (an 80% increase that would not satisfy the doubling standard) but would increase a baseline rate of 5% to approximately 12.3% (a 144% increase that would meet the doubling standard). Similarly, if a study showed that a drug increased a 10% baseline rate to 18%, the fact that the drug failed to double the rate in the study ought not to preclude use of the study to show causality with a lower baseline rate. The increase from 10% to 18% reflects a change of .37 standard deviations. That would correspond to an increase in a 5% baseline rate to just over 10%, thus meeting the doubling standard.

---

[20] Illustrations of EES values in studies of demographic differences in health or healthcare outcomes (including vaccination, cardiac procedures, mortality) where one observes, with minor exception, the standard, contrasting patterns of the two relative difference, may be found in Tables 3 to 7 (slides 11 to 15) of my 2008 16th Nordic Demographic Symposium Presentation Measures of Health Inequalities that are Unaffected by the Prevalence of an Outcome. Like illustrations, based on life table information, for racial and gender differences with respect to mortality at various ages, or with respect to surviving to various ages, may be found the Interactions by Age and Life Tables Illustrations subpages of SR and the Life Table Illustrations document.

Illustrations of this method in the application of a risk reduction observed in a clinical trial in order to estimate the absolute risk reduction and corresponding number-needed-to-treat in order to avoid one adverse outcome in circumstances involving baseline rates different from that in the clinical trial may be found in the Subgroup Effects subpage of SR.

One nuance of the above-described approach that is of potential importance to the analysis of particular discrimination claims warrants mention here. In an employment discrimination case where disparities are analyzed in terms of relative differences in selection rates, it typically does not matter when applicants of different demographic groups applied or whether all applications were actually examined by the employer, so long as there are no group-related patterns concerning time of applications relative to openings or otherwise concerning likelihood that an application will actually be reviewed. That is, suppose that, consistent with the figures in the first row of Table 1, an employer had 1000 AG applicants of which it hired 200 (20%) and 1000 DG applicants of which it hired 90 (9%), a hire rate ratio of 2.22. But suppose also that, because the employer had far more applicants than it needed or because it only examined applications submitted near the time that it needed to fill positions, the employer in fact examined only half of those applications; that the determination of which applications to examine was uninfluenced by the group membership of the applicants; and that both groups then experienced the same 50% reduction in applications examined. While the proportion the hires from each group comprise of the examined applications from the group (*i.e.*, the hire rates based on the applications) would be double the proportion those hires comprise of total applications from the group, the hire rate ratios would remain the same as in Table 1, as shown in Table 5. That is why, when disparities are measured in terms of relative differences in hire rates, the timing of the applications and the number of applications actually examined ordinarily would not affect the appraisal of the size of a difference in hiring rates.

Table 5. Illustration of the Implications of the Fact that Only Half of an Employer's Applications Actually Receive Attention With Respect to Appraising the Likelihood that the Employer Made Biased Hiring Decisions [ref b4510 a 2]

Employer	AG Hire Rate	DG Hire Rate	Rate Ratio Hire	EES
A – all apps examined	20.00%	9.00%	2.22	.50
A – half of apps examined	40.00%	18.00%	2.22	.68

As also shown in Table 5, however, with the smaller number of applications, the AG and DG selection rates translate into an EES figure of .66 standard deviations compared with the .50 standard deviations that one would derive from the hire rates calculated on the basis of gross application figures. And since an appraisal of whether a decision-making process is biased must be based on applications about which decisions were actually made, the .66 standard deviation figure is the pertinent one. Hence whether all applications are actually examined in circumstances where such examination might lead to a selection can be important regardless of whether applications of one group are more likely to be examined that those of another group.[21]

[21] I address this issue at somewhat greater length in the Case Study II subpage or SR. Table 2 of that subpage shows the ways in which the fact that only a certain proportion of applications examined, while not affecting the hire rate ratio, will affect the other standard measures of differences in outcome rates. But, since, like the hire rate ratio,

The above-described approach to appraising the strength of the forces causing outcome rates to differ is inexact in a number of respects. For example, it relies on an assumption that the underlying distributions of factors associated with experiencing an outcome are normal. Rarely can we be sure that the underlying distributions are normal and sometimes we will know that they are not normal, as, for example, when the distributions are truncated part of normal distributions. There also exists a range of more subtle issues.[22] But an approach of this nature is clearly superior to reliance on standard measures of differences between outcome rates without consideration of way the measure tends to be affected by the prevalence of the outcome at issue. For it provides a benchmark for appraising the strength of the association reflected by any pair of rates and for comparing the strength of association reflected by two or more pairs of rates when standard measures would yield varying interpretations as the comparative size of differences between rates. And it can at least spare us from wrongly concluding, on the basis of one preferred standard measure or another, that there is reason to distinguish among the employers in Table 1 and then mistakenly devoting resources to exploring the reasons for the perceived differences, drawing inference based on the perceived differences, or making decisions of consequence based on the perceived differences.[23]

---

those measures are ineffective means of appraising the strength of an association, I do not here discuss the specifics of how those measures are affected.

[22] For a fuller explanation of the varied problems with the described approach, see the Solutions sub the Solutions, , Irreducible Minimums, and Cohort Considerations subpages of MHD and the Truncation Issues subpage of SR.

[23] I recently found a paper of which the lead author was University of Kansas Economics Professor Donna K. Ginther, Race, Ethnicity, and NIH Research Awards, Science 333: 1015-1019 (2011), where a probit analysis was used to adjust for difference in characteristics in an analysis of racial and ethnic differences in NIH research awards, and where the probit coefficient was then converted to an absolute difference between rates. The procedure was essentially the same as that discussed above for applying the results of a study of a factor's effect on a particular baseline rate to estimate the factor's effect on a different baseline rate (save that in the procedure I described the probit coefficient is first derived from one pair of rates while in the Ginther study the probit coefficient was derived from a probit analysis that attempted to adjust for demographic differences in outcome-related characteristics). In confirming my interpretation of her approach, Professor Ginther also directed me to lecture notes by Notre Dame Sociology Professor Richard Williams, Marginal Effect: Discrete and Continuous Changes (undated) , which provide a guide to deriving that absolute difference from the probit coefficient and reflecting, I believe, views about measurement akin to those expressed here (though perhaps not necessarily reflecting the same view as the superiority of the probit over related methods). Similar views may also underlie the use of the probit in the Ginther paper.

But the Ginther paper reflects a different perspective from mine in an important respect. For it translates the probit coefficient into an absolute difference, presumably to make it more understandable to the reader. By contrast, I would question whether that the absolute difference would be useful to readers, given, for example, that a 10 percentage point difference means different things when the rates at issue are 10% and 20% from when the rates at issue are 30% and 40%, and it is the probit that tells us what those different meanings are (.44 in the former case and .27 in the latter case). Further, the approach in the Ginther paper would seem to suggest that a correct ranking of the employers in Table 1 is that based on absolute differences even though they are all reflections of the same probit coefficient. Nevertheless, I think that the basic approach underlying the Ginther article and Williams lecture notes is consistent with the reasoning in this paper. That consistency, however, is most evident with respect the matters discussed immediately above regarding the use of information secured in a study as to the effect of a factor on a particular baseline rate in order to estimate the percentage point change the factor would cause to different baseline rate, for the purpose either (a) to determine whether the factor more than doubles that baseline rate (as is pertinent to

Some observers have objected to this approach, mainly in the context of discussions of health or healthcare disparities, on the basis of its complexity. I am uncertain that the concern is in fact that the approach is more complex than standard approaches rather than it entails thinking about things in terms in which we are unused to thinking about them. In any case, those citing the complexity of the approach have stressed the need to describe disparities issues for policy makers in terms that policy makers can readily understand. But standard measures of disparities in health and healthcare disparities are not merely inexact. Rather they commonly communicate false information, often when an unstated measure that is no less legitimate, according to conventional lights, supports an exact opposite interpretation.[24] Even when an observer's preferred method yields conclusions about such things as the directions of changes in disparities over time that are broadly correct in the sense that a sound measure would yield the same conclusion, use of a standard measure misleadingly implies that it effectively quantifies the size of the difference in the circumstances of two groups reflected by a pair of rates. Thus, the contention that reliance on standard measures is necessary to inform policy is an argument that policy makers benefit more from false information that they can readily understand than from true information that may be very difficult for them to understand. Few policy makers would agree with that view. The same goes for courts.

The reasoning above regarding the appraisal of the strength of the forces causing outcome rates to differ would broadly apply to appraisals of the size of a disparate impact of some criterion or procedure in any context where the size of the disparate impact is at issue. But questions of precisely how one might appraise the size of a disparate impact, and whether and how determinations as to cutoff scores or the stringency of a criterion might affect that determination, raise some additional issues. For that reason, issues concerning the measurement of a disparate impact are deferred to Section E.

C. The Problematic Appraisal of Demographic Differences Based on the Proportion a Group Comprises of Persons who Could Experience an Outcome and the Proportion the Group Comprises of Persons who Experience the Outcome

In many settings analyses of discrimination issues are based on comparisons between the proportion a group comprises of persons potentially experiencing a favorable or adverse outcome and the proportion the group comprises of persons who experience the outcome. Such analyses are quite common in employment discrimination cases. The Supreme Court's seminal decision

---

the tort context) or (b) to determine in circumstances of that baseline rate how many persons it will be necessary to treat with an intervention in order to avoid one adverse outcome (as is pertinent in the clinical context).

[24] In health disparities research commentators have lately stressed the importance of reporting both relative and absolute differences especially when they yield different conclusions as to directions of changes over time. Yet those stressing this importance commonly do so without apparent recognition that there even exist two relative difference, much less that the two relative differences tend commonly to change in opposite directions as the prevalence of an outcome changes or that anytime a mentioned relative difference and the absolute difference have changed in opposite directions the unmentioned relative difference will have changed in the same direction as the absolute difference and the opposite direction of the mentioned relative difference. See my November 8, 2012 BMJ comment *Discussions of Relative and Absolute Differences Cannot Ignore that There are Two Relative Differences* and my November 3, 2012 PLoS Medicine comment *Clarity in the Reporting of Health Equity Issues Requires Addressing Measurement Issues*.

in *Hazelwood School District v United States*, 433 U.S. 299 (1977), was based on a comparison of the proportion blacks comprised of the relevant market and the proportion they comprised of the defendant's teacher hires, and many other employment cases are based on the proportion a group comprises of persons who applied for a job and the proportion the group comprises of those who were hired or the proportion a group comprises of employees available to experience a favorable or adverse employment action and the proportion the group comprises of those experiencing the outcome.

Analyses of many outcome differences in educational contexts, including those concerning low minority representation in advanced placement courses and gifted programs and high minority representations among students required to repeat a grade, placed in special learning programs, or suspended and expelled are based on such comparisons. Racial profiling analyses are invariably based on a minority group's representation in a population potentially stopped by police and its representation among those stopped.

All these analyses are problematic, however. For they are based on insufficient information to appraise the strength of the forces causing the rates to differ.

Table 6 presents information on two settings where the proportion the disadvantaged group (DG) comprises of the persons eligible to experience some favorable outcome is larger than the proportion it comprises of those who experience the outcome – 40% versus 20% in setting A and 15% versus 5% in Setting B. The three columns after the representation rates show the ways in which the differences reflected in these figures might commonly be quantified by those attempting to analyze the situation, where, for example, discrimination in selection is at issue. Some would rely on the absolute (percentage point) difference between the proportion DG comprises of the eligible pool and the proportion it comprises of selections. Others would rely on the fact that the proportion DG comprises of selections is a certain percentage lower than the proportion it comprises of the pool (Rel Diff 1) or that the proportion it comprises of the pool is certain percentage higher than the proportion it comprises of selections (Rel Diff 2).[25]

Each of these figures, while derived by means of the same sort of elementary calculations used to derive the measures in Table 1, involve relationships between proportions that are somewhat different from those involved in Table 1. We can, however, derive from the proportions DG comprises of the pool and the proportion it comprises of selections the ratio of the selection rate of the advantaged group (AG) to the selection rate of DG, which, as noted, is the most common measure used in cases involving selection disparities.[26]

[25] The difference between these percentages is simply a function of the whether the higher or lower figure is used as the numerator of a fraction (the subject of note 2 *supra*). Such difference should not be confused with differences related to whether one is examining the favorable or the adverse outcomes. In this situation we do not know the proportion DG comprises of persons experiencing the adverse outcome.

[26] Assuming all persons not in DG are in AG, where DGP is the proportion DG comprises of the pool and DGS is the proportion it comprises of selections, the ratio of AG's selection rate to DG's selection rate would be ((1-DGS)/(1-DGP))/(DGS/DGP).

Table 6. Proportions Disadvantaged Group Comprises of Persons in a Pool and Persons Selected in Two Settings, with Standard Measures of the Differences Reflected by Those Proportions

Setting	DG Prop Pool	DG Prop Selection	Abs Diff	Rel Diff 1	Rel Diff 2	AG/DG Select Ratio
A	40%	20%	0.20	50.00%	100%	2.67
B	15%	5%	0.10	66.67%	200%	3.35

Given these figures for the proportion DG comprises of the pool and the proportion it comprises of selections, all the values in the table would hold regardless of the underlying selection rates for AG and DG. But from the data on the proportion DG comprises of the pool and the proportion it comprises of selections, we are not able to determine what those selection rates are. And without those selection rates, it is not possible to soundly appraise the strength of the forces causing the proportion DG comprises of the pool and the proportion it comprises of selections to differ.

Table 7 illustrates the potential implications of the absence of information on the selection rates of the two groups (or, put another way, the implications of what those selection rates might in fact be). The table presents the same information on the proportion DG comprises of the pool and the proportion it comprises of selections shown in Table 6. But it also shows the actual selection rates for AG and DG in circumstances of differing overall selection rates, and, based on the AG and DG selection rates, the EES. The table also includes the AG/DG Selection Rate Ratio in order simply to show that the ratio does not change as the actual selection rates change and that the ratio is consistent with a smaller EES when selection rates are low than when they are high.

The key figure is the EES. The variation in the EES with different overall selection rates demonstrates that in order to determine whether the forces causing the proportion DG comprises of selections to differ from the proportion it comprises of the pool are stronger in one setting than the other we must know the actual selection rates in each setting. We must also know those selection rates in order to appraise whether in either setting those forces should be deemed strong or weak.

Table 7. Proportions Disadvantaged Group Comprises of Persons in a Pool and Persons Selected in Two Settings, with Actual Selection Rates and Estimated Effect Size of the Differences Reflected by Those Proportions in Circumstances of Differing Overall Selection Rates [ref b4505a5]

Setting	DG Prop Pool	DG Prop Selections	Overall Selection Rate	DG Selection Rate	AG Selection Rate	EES	AG/DG Selection Ratio
A	40.00%	20.00%	1.00%	0.50%	1.33%	0.35	2.67
A	40.00%	20.00%	10.00%	5.00%	13.33%	0.53	2.67
A	40.00%	20.00%	20.00%	10.00%	26.67%	0.66	2.67
A	40.00%	20.00%	30.00%	15.00%	40.00%	0.78	2.67
A	40.00%	20.00%	40.00%	20.00%	53.33%	0.92	2.67
A	40.00%	20.00%	50.00%	25.00%	66.67%	1.11	2.67
B	15.00%	5.00%	1.00%	0.33%	1.12%	0.43	3.35
B	15.00%	5.00%	10.00%	3.33%	11.18%	0.60	3.35

Setting	DG Prop Pool	DG Prop Selections	Overall Selection Rate	DG Selection Rate	AG Selection Rate	EES	AG/DG Selection Ratio
B	15.00%	5.00%	20.00%	6.67%	22.35%	0.74	3.35
B	15.00%	5.00%	30.00%	10.00%	33.53%	0.86	3.35
B	15.00%	5.00%	40.00%	13.33%	44.71%	0.97	3.35
B	15.00%	5.00%	50.00%	16.67%	55.88%	1.11	3.35

The same situation exists when all that is known is the proportion the disadvantaged group comprises of persons who may experience the adverse outcome and the proportion it comprises of those who experience that outcome. In those situations, it is the relative difference in the adverse outcome that can be derived from the available data, but not the actual rates of experiencing the adverse (or corresponding favorable) outcome that are required to soundly appraise the strength of the forces causing the proportion the group comprises of those experiencing an outcome to differ from the proportion it comprises of the pool.

The notable thing in the contexts where adverse outcomes are examined, however, is that reductions in the frequency of the adverse outcome – as is the nearly inevitable consequence of relaxing discipline standards and that is a common consequence of most efforts to address findings of racial profiling – will tend to increase the proportion the disadvantaged group comprises of those experiencing the adverse outcome while leaving the proportion it comprises of the population potentially experiencing that outcome unaffected. Thus, according to each of the methods by which observers commonly measure the disparity, it will seem to have increased.

The remedial order in *Floyd v. City of New York*, No. 09 Civ 1034 (SAS) (S.D.N.Y., Aug. 12, 2013), issued in conjunctions with a finding that race and ethnicity improperly influenced stop decisions under New York City's stop and frisk policy, is certain to substantially reduce the frequency of stops, which will tend to increase the proportion blacks and Hispanics comprise of persons who are stopped. The extent to which the order also reduces any racial/ethnic influence on stop decisions will counter somewhat the tendency for the reduction in number of stops to increase that proportion. But the reduction of that influence will have to be very substantial for the reduction in the number of stops not to cause an increase in the proportion blacks and Hispanics comprise of those who are stopped. In any case, in order to determine whether the order actually reduces the influence of race or ethnicity on stops and the extent of such reduction, it will be necessary to secure data on actual stop rates of different racial/ethnic groups.[27] Such data are not easy to collect and the determination of the numerators and denominators for those rates raise a number of complex issues affecting the size of the rates themselves (and hence the EES) even when such issues do not affect the relative difference between the rates.

See the Gender differences in DADT subpage of the Vignettes page of jpscanlan.com for discussion of misperceptions regarding the varying likelihood of (or degree of) bias among the different military branches based on branch by branch comparison of the proportions women comprised of member of the branch and the proportions they comprised of members of the branch discharged for violations of "don't-ask-don't tell" policies. The subpage also discusses

[27] See my *The Profiling Conundrum* (unpublished) (2001) regarding the way reductions in the numbers of searches for contraband conducted by U.S. Customs increased the proportion minorities comprised of those searched.

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 or rates is stronger, but the possibility that the difference occurred by change 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.

whether there exists a less discriminatory alternative to a challenged practice, as contemplated by the Civil Rights Act of 1991, 42 U.S.C. § 2000e-2(k) (2000), and the Department of Housing and Urban Development's recently released rule on the discriminatory effects standard under the Fair Housing Act, 78 Fed. Reg. 11460 (February 15, 2013). For example, given that lowering a test cutoff increases the relative difference in failure rates at the same time that it reduces the relative difference in pass rates, does lowering a test cutoff in fact reduce the impact of a test on which minorities do not perform as well as whites?

In appraising disparities in things like mortality/survival or nonreceipt/receipt of some type of health care, in the common situations where the adverse outcome has declined in overall prevalence while the favorable outcome has increased in overall prevalence – and the standard measures have changed as they usually do in the circumstance and there is no other indication that the forces causing the rates to differ have grown larger or smaller – there does not seem to be a rational argument as to why a demographic difference has changed in any meaningful sense. I think the same reasoning holds in any legal setting where a criterion such as a test score entirely dictates the outcome of a process. That is, where all persons who pass a test experience the favorable outcome and all who fail it experience the corresponding adverse outcome, as would typically be the case with bar exams and other certification procedures as well as teacher competency and high school exit tests, the disparate impact of the test would seem unaffected by the cutoff.

But employment tests do not generally operate that way. Rather, they merely restrict the pool or persons eligible for further consideration. Usually in such circumstances there is reason to expect that lowering a test cutoff will in fact reduce the impact of the test, or at least the impact of the test on the selection process of which is a part.

Table 8 illustrates this point. It is based on a situation where 1000 members of AG and 1000 members of DG apply for a job and an applicant must pass a test in order to be further considered for 200 positions, and where the difference between mean scores of AG and DG is the half a standard deviation difference underlying the illustrations in Sections A and B. It is assumed that among persons who pass the test at each cutoff, members of AG and DG are equally likely to be selected.[31]

Table 8. Illustration of the Effect of Lowering Cutoffs on Differences between Process Outcome Rates of Members of an Advantaged and Disadvantaged Group – Hiring

AG Pass	DG Pass	Test EES	AG Hire Rate	DG Hire Rate	Hire EES
20.00%	9.01%	0.50	13.79%	6.21%	0.44
30.00%	15.39%	0.50	13.22%	6.78%	0.37
40.00%	22.66%	0.50	12.77%	7.23%	0.31
50.00%	30.85%	0.50	12.37%	7.63%	0.27
60.00%	40.52%	0.50	11.94%	8.06%	0.22

[31] The table, which is based on Table 1 of the 2006 BSPS paper mentioned in Section A, is limited to situations where, under the terms of the hypothetical, there are more test passers than there are hires.

AG Pass	DG Pass	Test EES	AG Hire Rate	DG Hire Rate	Hire EES
70.00%	50.80%	0.50	11.59%	8.41%	0.17
80.00%	63.31%	0.50	11.16%	8.84%	0.13
90.00%	78.23%	0.50	10.70%	9.30%	0.09
95.00%	87.29%	0.50	10.42%	9.58%	0.04
97.00%	91.62%	0.50	10.29%	9.71%	0.02
99.00%	96.56%	0.50	10.12%	9.88%	0.01

The first two columns show the pass rates for AG and DG at each cutoff along with the EES for test passage (which I include for reference even though it is always the .50 standard deviation figure underlying the hypothetical). The table then shows the actual hire rates for each group (*i.e.*, the proportion the hires of each group comprise of the total applicants from each group) at the cutoff, along with the EES based on those hire rates. As the cutoff is lowered, in contrast to the test pass rates that increase for both groups, the hire rate decreases for AG but increases for DG. Correspondingly, the EES based on these rates decreases. Thus, whether or not we regard the lowering the cutoff as reducing the impact of the test itself (and we could not so regard it if everyone who passes the test is hired and the selection rates are hence the same as the pass rates), lowering the cutoff reduces the impact of the test on the process as a whole.[32]

Of course, it would not necessarily be the case that members of DG and AG who pass the test have equal chances of selection. Among test passers, members of DG could be selected at either higher or lower rates than members of AG. Assuming that there exists some correlation between test scores and factors associated with selection among persons who pass the test, there would be reason to expect that among persons who pass the test, members of DG would be hired at lower rates than AG. But as a rule in the employment context, the overall difference in selection rates (properly measured) would tend to be smaller with the lower cutoff than with a higher cutoff.

It does not follow, however, that in the employment setting a lower cutoff will always show a smaller impact as to the total process than a higher cutoff. Consider a reduction in force where an employer excludes from consideration for termination all employees with certain ratings, and where the difference between the ratings of AG and DG is the same as the difference in test scores underlying Table 8. Initially, it might seem that the more one limits the protected status to the highest ratings, the more the protecting of highly-rated rated employees disadvantages the group with average lower ratings. But Table 9 illustrates why that would not necessarily be the case.

[32] In *Connecticut v. Teal*, 457 U.S. 440 (1982), with respect to a situation where a test showed a disparate impact adverse to black applicants but where the test did not result in a disparity in hiring rates, the Supreme Court held that the black applicants could nevertheless challenge that impact of the test with respect to the way its limitation on the opportunity to compete further in the process. (In *The Bottom Line Limitation to the Rule of Griggs v. Duke Power Co.*, 18 U. Mich. J. Law Reform 705 (1985), I argued that the *Teal* result was correct, but for somewhat different reasons from those on which the Court had relied.) While the Teal ruling focuses on the impact of the test itself with respect to its limiting of the pool eligible for further consideration, I do not read it as affecting the reasoning in this section with respect to the interpretations as to the impact of the test, at different cutoffs, on the selection process as a whole.

Table 9. Illustration of the Effect of Lowering Cutoffs on Differences between Process Outcome Rates of Members of an Advantaged and Disadvantaged Group – Termination

AG Prot Rate	DG Prot Rate	Prot EES	AG Retention Rate	DG Retention Rate	Retention EES
1.00%	0.24%	0.50	90.04%	89.96%	0.00
3.00%	0.87%	0.50	90.11%	89.89%	0.01
5.00%	1.62%	0.50	90.17%	89.83%	0.01
10.00%	3.75%	0.50	90.34%	89.66%	0.03
20.00%	9.01%	0.50	90.64%	89.36%	0.08
30.00%	15.39%	0.50	90.95%	89.05%	0.10
40.00%	22.66%	0.50	91.26%	88.74%	0.14
50.00%	30.85%	0.50	91.61%	88.39%	0.18
60.00%	40.52%	0.50	91.96%	88.04%	0.22
70.00%	50.80%	0.50	92.42%	87.58%	0.28
80.00%	63.31%	0.50	92.94%	87.06%	0.33
90.00%	78.23%	0.50	93.70%	86.30%	0.44

The table is based on a situation where with a workforce comprised of 1000 AG employees and 1000 DG employees an employer decides to eliminate 200 employees. The employer then implements that reduction by protecting a certain proportion of employees on the basis of their meeting a ratings threshold, while eliminating 200 employees not so protected on the basis of seniority or some other factor as to which there is no group-related difference among employees. The first three columns show the rates at which AG and DG achieve protected status, the same figures as the pass rates in Table 8,[33] along with corresponding .50 EES related to achieving that status. The next three columns then show the retention rates of AG and DG and the EES based on those retention rates.[34] In contrast to Table 8, where lowering cutoffs decreased the hire rate of AG while increasing hire rate of DG, Table 9 shows that lowering the cutoff decreases the retention rate of DG while increasing the retention of AG and correspondingly increases the EES. So here differences between AG and DG prospects are smaller with higher than lower cutoffs.

The reconciliation of the patterns reflected in Tables 8 and 9 lies in the fact that the disparate impact of the criterion on the entire process is decreased by reducing the proportion of total outcomes dictated by the criterion. That is accomplished in the former situation by lowering the cutoff and in the latter situation by raising the cutoff.

[33] In this case, the table is limited to the situations where there are more than 200 employees who fail to achieve protected status.

[34] I chose to use the retention rates rather than the termination rates in this table in order to make it consistent with the rates in Table 8. The use of termination rates would make the differences between AG and DG seem larger, particularly at the lower cutoffs. But the principal point of this paper is that perceptions of the differences in the circumstances of two groups based on standard measures are mistaken. The EES, of course, is the same whether one examines the retention rates or the termination rates.

I have not fully considered these issues with respect to varied other types of employment practices that disparately affect minority groups. It would seem that in the case of policies requiring termination for rules violations or failure to meet performance standards (as in *Fisher* case discussed in note 29 *supra*), and where reducing the stringency of the standard will increase the relative difference in failing to meet the standard while reducing the relative difference in meeting it, the impact would be unaffected by the stringency of the standard. On the other hand, limiting the scope of a rule disqualifying from consideration applicants with criminal records would seem to reduce the impact of the rule on the selection process as a whole in the same way lowering a cutoff reduces the impact of a test on that process. But these issues warrant further thought.

I am uncertain of how the above reasoning or the patterns reflected in Tables 8 and 9 would apply in a lending context. Consider the situation where it might be necessary to have a minimum income or credit score in order to qualify for consideration for some preferred type of loan. Lowering the requirement would initially seem to have the same implications as lowering a test cutoff. But in a setting where decision are as data-driven as they are in lending, there is less reason than in the employment setting to believe that the criteria that determine outcomes among persons meeting the designated minimum will have a smaller disparate effect on the disadvantaged group than the criterion whose impact at issue. In addition, I remain uncertain of the implications of the fact that in the lending process, there typically will not be a defined number of the favorable outcomes among persons meeting a minimum criterion.

In sum, there exist a number of matters that warrant further consideration. But such consideration should be premised on a complete understanding of the contrasting ways in which altering a standard affects relative differences in meeting it and relative differences in failing to meet it.

F. The Problematic Nature of Analyses of Discrimination Issues That Fail to Examine the Entire Universe of Persons Subject to the Decision-Making Process at Issue

A number of quite successful employment discrimination cases have been based on claims, usually termed assignment discrimination or job segregation, that an employer's hires from a particular demographic group are disproportionately assigned to jobs deemed to be less desirable. Commonly, however, these claims have involved situations where one group would tend to comprise a larger proportion of the persons hired for the putatively less desirable jobs than the putatively more desirable jobs regardless of whether the group is discriminatorily excluded from the latter jobs. Many of these cases involved the grocery industry, where women once tended to comprise a quite high proportion of cashiers regardless of whether they were discriminatorily excluded from stocker positions.

More fundamentally – which is to say, irrespective of the accuracy of the described pattern whereby one group would tend to comprise a large proportion of certain jobs at an employer regardless of whether they are excluded from the employer's other jobs – statistical analyses cannot determine whether persons have been excluded from the putatively more desirable jobs without consideration of all the persons seeking those jobs, including those who were not hired at all because there were not offered positions or declined the positions offered.

The problems in assignment analyses that are focused solely on persons who are hired thus differ from the problems with analyses based on comparisons of the proportion a group comprises of an applicant pool or labor market and the proportion it comprises of persons hired that were discussed in Section C in the following respect. In the case of the analyses based on differences between the proportion a group comprised of the pool and the proportion it comprised of those hired, one is able to determine from the available information that a difference exists between hire rates of members of different demographic groups (even though one is not able in such circumstances to appraise the strength of the forces causing the rates to differ and hence determine the likelihood that the difference resulted from discrimination or compare such likelihood in one setting with that in another setting). In assignment analyses, by contrast, the absence of information on the applicants make it impossible to determine whether there exists a difference in selection rates at all. See illustrations in my Illusions of Job Segregation, Public Interest (Fall 1988).[35]

Some may suggest that the greater willingness of certain groups to accept jobs deemed less desirable may be a function of discrimination in society. Even assuming that all of the greater willingness is a function of discrimination in society, the employer does all that the law can require of it when it treats applicants of different demographic groups seeking/willing to accept either job on an equal basis. If there in fact exists widespread discrimination against certain groups (and even if there merely exists a perception that discrimination is widespread), that merely provides greater reason to expect members of the group subject to the discrimination to comprise a larger proportion of applicants seeking/willing to accept a less desirable job at a particular employer than they comprise of those seeking/willing to accept only a more desirable job at the same employer.[36] A pattern of seeming disproportionate representation of certain

[35] In *Illusions of Job Segregation* and other articles on this subject discussed in note 35 *infra*, I occasionally stated or suggested that to determine whether a group was discriminatorily excluded from a particular job one needed to know the proportion the group comprised of persons seeking a position and proportion it comprised of persons securing it. As indicated in Section C *supra*, however, I have since come to recognize that to effectively appraise the strength of the forces causing any selection disparity one must also know the actual selection rates.

[36] The retail industry is, or at least used to be, a setting where women would tend to comprise a very high proportion of noncommission sales positions regardless of whether they were discriminatorily excluded from commission sales positions. The Equal Employment Opportunity Commission's (EEOC's) nationwide gender discrimination case against Sears, *EEOC v. Sears, Roebuck and Co*., 839 F.2d. 303 (7th Cir. 1988), which was tried in 1984-1985 and which has been the subject of much commentary, was based on a 1977 EEOC administrative decision that emphasized that female sales workers at Sears comprised a much higher proportion of noncommission than commission sales jobs. See Peter Milius, *EEOC: Job Bias "at All Levels" of Sears*, Washington Post (Feb. 25, 1974), discussing the EEOC finding that women held 77 percent of noncommission sales jobs compared with 23 percent of commission sales job. The suit the EEOC in fact brought against Sears, however, sought to prove that women were discriminatorily denied hire into commission sales jobs based on differences between the female proportion of sales applicants, adjusted by type of sales jobs applied for (and varied other factors), and the female proportion of commission sales hires. Thus, while the EEOC analysis was subject to the issues addressed in Section C *supra*, it was not subject to the partial picture issues addressed in this section. Analyses of applications in the case revealed that women did indeed comprise a much higher proportion of noncommission sales applicants and undifferentiated sales applicants than commission sales applicants. Information developed in the litigation also revealed that when Sears dramatically increased the female proportion of commission sales hires, the pattern whereby noncommission sales workers were overwhelmingly female went unchanged. For the forces that created a largely female pool of workers interested in noncommission sales positions at Sears and other retailers were little

groups in less desirable jobs at a particular employer would thus exist regardless of whether that employer discriminatorily excluded certain demographic groups from its more desirable jobs. Nevertheless, in the 1990s, these cases proved to be quite successful even after the Supreme Court, in *Wards Cove Packing v. Atonio*, 490 U.S. 642 (1989), appeared to reject analyses of the nature underlying assignment claims.[37]

Claims of discrimination in pay involve continuous measures and hence are not subject to the measurement issues pertaining to dichotomous/binary measures discussed in Sections A through C. Indeed, the EES/probit approach described in Section B is a means of translating such measures into continuous measures.

But pay discrimination claims implicate the same partial picture issues that undermine assignment claims. This is most evident in analyses that fail to adjust for starting salary while ignoring the universe of persons who preferred to look elsewhere rather than accept a particular salary. Once again, that there exists (or is believed to exist) widespread discrimination against a group will tend to have an effect, in particular, to diminish the bargaining position of members of a disadvantaged group. To the extent that employers, aware of such pattern, take a harder bargaining position on the basis of a protected characteristic, such action would be unlawful discrimination. But as long as applicants of one group are more willing to accept a particular salary than similarly qualified applicants of another group, there is reason to expect that greater willingness to translate into a starting salary disparity in any circumstance where there is flexibility in starting salaries of persons with similar credentials. In any event, one cannot appraise the fairness of a process for setting starting salaries without examining the salary request/offer situations of all persons who have been subject to the process.

The same issue applies to analyses of salaries of longer term employees that adjust for starting salary or that involve situations where all starting salaries are the same. For all employees who consider their salary progression to be unsatisfactory or not consistent with their worth (usually employees who tend toward the lower end of salary ranges) will have greater incentives to seek employment elsewhere than employees who feel they are being well compensated (usually employees toward the higher end of salary ranges). And in situations where among the less satisfied employees different demographic groups have (or believe they have) unequal prospects with other employers, it is the group with greater optimism about those prospects that will disproportionately seek opportunities elsewhere. The result is that an analysis that examines only the persons who remain will tend to show a disparity in compensation

---

affected by Sears' hiring practices regarding commission sales positions. Issues involved in the Sears case, in which I represented the EEOC, have affected my thinking about many issues addressed in this paper. See the Sears Case page and the Sears Case Illustration subpage of SR.

[37] These cases are discussed in my *Unlucky Stores: Are They All Guilty of Discrimination?*, San Francisco Daily Journal (Jan. 29, 1993), *Multimillion-Dollar Settlements May Cause Employers to Avoid Hiring Women and Minorities for Less Desirable Jobs to Improve the Statistical Picture*, National Law Journal (Mar. 27, 1995), and *Women Employees' Case against Publix, Built on Wrong Data, Doesn't Compute*, Miami Daily Business Review (Aug. 2, 1996).

adverse to the disadvantaged group whether or not the employer makes biased compensation decisions.

Again, however, irrespective of whether these things are likely to happen, as a purely statistical matter, in order to determine whether a decision-making process treats members of different demographic groups differently, one must examine how that process treated all persons who have been subject to it.

Similar factors affect the validity of analyses of demographic differences in rates of assignment of loans to subprime (rather than prime) status and demographic differences in loan costs that have been the subject of recent litigation (including the *Countrywide* and *Wells Fargo Bank* cases mentioned in the Introduction) and many recent lending studies. As noted earlier, claims of discrimination in assignment to subprime status implicate the same anomaly found in claims involving rejection/approval of mortgage applications whereby reducing the frequency of an adverse outcome, as the government encourages lenders to do, increases the likelihood that a lender would be deemed to have discriminated. But analyses of racial differences in loan rejection rates at least satisfied the fundamental requirement of a statistical analysis that an examination of the fairness of a process must consider what happened to all persons subject to the process. By contrast, analyses of claims of discriminatory assignment to subprime status, like analyses of claims of discriminatory assignment to putatively less desirable jobs, consider only those persons who were offered and accepted particular loan packages, while ignoring applicants offered no loans and applicants who refused loans because they had or believed they had better options elsewhere. Analyses of differences in loan terms (apart from assignment to subprime status) similarly suffer from failure to consider the entire universe of persons subject to the processes under examination even if they do not implicate the anomaly whereby acceding to government pressures tends to increase disparities as they are measured by the government.[38]

With respect to the problem of the failure of analyses to consider entire universe of persons subject to the process, issues in the lending context differ from those in the employment context in two respects. Analyses in the lending context are not affected by demographic differences in interest that generally are important contributors to differences in the gender composition of sometimes very disparate jobs. While there may be minor issues respecting the type of loan for which one formally applied, in general one can assume that all applicants are simply seeking the best terms available. On the other hand, the extent of beliefs that there exists widespread employment discrimination against certain groups, and the effects of such beliefs on the willingness of employees of certain groups to more readily accept certain job or

[38] Sometimes things that appear to involve continuous measures are importantly affected by dichotomies and that possibility may be an issue in some analyses of loan costs, if, for example, meeting a certain credit score threshold puts one in a certain rate range. In such cases, issues concerning the effects on different groups of the relaxing of a qualifying criterion may be implicated, though appraisal of the effects the modification of the criterions can be complex. See Table 2 of the 2006 BSPS paper concerning the effects of general changes in mortality on longevity differentials. See also my BMJ comment *Recognizing Why Dichotomous and Continuous Measures May Yield Contrary Results* (June 11, 2007) regarding the way continuous scores for self-rated health may be affected by categorical responses. But these complexities would appear to be somewhat different from the complexities involved in determining, given that altering a cutoff will tend to cause relative differences in adverse outcomes and relative differences in the corresponding favorable outcomes to change in opposite directions, whether and how modifying cutoffs may in fact affect the disparate impact of a criterion (the subject of Section E).

compensation situations than members of other groups, may be uncertain. But in the lending context, the large settlements in cases like *Countrywide* and *Wells Fargo Bank*, as well as the numerous studies finding what often are characterized as huge lending disparities, give minority loan seekers substantial reason to believe that their options in the loan market are limited by discrimination. A belief of such nature can translate into substantial differences in loan terms regardless of whether lenders take advantage of that belief and regardless of whether the statistical analyses underlying the settlements or studies are sound.

Related Materials:

Listed below are some materials available on jpscanlan.com addressing issues closely related to those directly addressed in this paper.

The Disparities – High Income subpage of the Lending Disparities (LD) page discusses misperceptions about the fact that relative differences in adverse lending outcome tend to be greater among higher-income than lower-income applicants.

The Underadjustment Issues subpage of the LD addresses reasons adjustments for differences in characteristics in analyses of discrimination issues are almost invariably inadequate.

The File Comparison Issues subpage of LD discusses the problematic nature of efforts to identify lending discrimination by means of comparisons of files of rejected and approved applicants. The discussion reflects the broader point that, save in circumstances where decision-makers actually express discriminatory motivations, rarely will efforts to determine whether individuals were discriminatorily treated shed light on whether there exists a pattern of widespread discrimination.

The Disparate Treatment subpage of the Discipline Disparities (DD) page addresses issues concerning interpreting data as to the likelihood that some part of observed racial and ethnic differences in suspension and expulsion rates resulted from discrimination

The Offense Type Issues of DD discusses the validity of perceptions the fact that relative differences between discipline rates of whites and blacks public school students are larger for subjectively identified (or less serious) misconduct than for objectively identified (or more serious) misconduct is evidence of racial discrimination in the imposition of discipline.

Description of various other pertinent pages may be found on the Home Page of jpscanlan.com.

Efforts to explain patterns by which standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome to the Departments of Education and Justice and the Federal Reserve Board as those patterns bear on the particular agency's activities regarding discrimination issues include:

Department of Education Measurement Letter (Apr. 18, 2012)

Department of Justice Measurement Letter (Apr. 23, 2012)

Federal Reserve Board Measurement Letter (March 4, 2013)

The agencies have yet to indicate an understanding that measures they employ for quantifying differences in outcome rates tend to be affected by the prevalence of an outcome. See the Holder/Perez Letter and Federal Reserve Letter subpage of LD and the Duncan/Ali Letter subpage of DD. The failure to understand these patterns and their bearing on the missions of government agencies appears to be universal among federal agencies, with the exception of the

National Center for Health Statistics (NCHS). NCHS's understanding of the issues, mentioned *supra* at 13-14, is discussed further in the references provided on those pages.

ADDENDUM
(Nov. 17, 2014)

This page expands on note 38 (at 35) regarding the potential for cutoffs for particular loan terms to affect differences between loan rates of advantaged and disadvantaged groups. It also qualifies the statement at 35 that pay issues are not subject to the measurement issues pertaining to dichotomous/binary measures discussed in Sections A through C.[39] As discussed in the referenced note, the point made here is akin to the point made with respect to longevity differences at 6-7 (and illustrated in Table 2 at 28) of the 2006 British Society for Population Studies paper *The Misinterpretation of Health Inequalities in the United Kingdom* and made more generally about the ways that seemingly continuous measures may be affected by dichotomies in *Comment on Chandola BMJ 2007*.

Table 1 illustrates the point of note 38 by showing the effects lowering a cutoff for a credit score that enables a loan recipient to receive a loan at 4% rather than 8%. The statistical patterns illustrated would generally apply regardless of the particular advantages accorded to those meeting the threshold for that preferred treatment. They would apply in the pay context where, for example, an employer offered salary enhancements for employees/hires meeting some criterion where members of an advantaged group were more likely to meet the criterion than members of a disadvantaged group.

Addendum Table 1 below is based on the credit score data that underlie the Credit Score Illustrations subpage of the Scanlan's Rule page, which data are drawn from Table 4 (at 35) of Class Certification Report of Howell E. Jackson submitted in support of class certification in *In re Wells Fargo Mortgage Litigation*, No. 8-CV-01930-MMC (JL) (M.D. Cal.). It presents the implications of lowering the cutoff that determines whether borrowers receive an 8% rate or a 4% rate with respect to the differences between the average interest rates of black and white borrowers in the pool.

The table shows for each of the credit score cut points, (a) the proportion of each racial group that receives the favored loan rate, (b) the average rate for each racial group, (c) the ratio of the black average rate to the white average rate, and (d) the absolute difference between the black and white average rates.[40] As the cutoff is increased from the point where very few receive the favored rate to one where everyone receives the favored rate, the relative and absolute differences increase for a time, and then decline.

[39] That the implications of the patterns by which measures of differences between outcome rates are affected by the prevalence of an outcome are not limited to lending claims specifically involving dichotomies might also be inferred from language in *Fair Lending Studies Paint Incomplete Picture*, American Banker (April 24, 2013).

[40] The table shows the absolute difference in terms of a percent rather than percentage points because it is a difference between rates rather a difference between proportions. I am uncertain whether the usage raises the issue addressed on the Percentage Points subpage of the Vignettes page of jpscanlan.com. See also the *Comment on McWilliams Ann Int Med 2009* .

Addendum Table 1. Illustration of Effects on Relative and Absolute Differences between Black and White Average Loan Rates at Various Credit Score Cutoffs for Receipt of Favored Loan Rate

CUT	PropWRecFavRt	PropBRecFavRt	AvWRt	AvBRt	RatioB/Rt	RtDiff
800	5.15%	1.65%	7.79%	7.93%	1.02	0.14%
780	18.90%	6.32%	7.24%	7.75%	1.07	0.50%
760	33.97%	12.50%	6.64%	7.50%	1.13	0.86%
740	46.80%	19.50%	6.13%	7.22%	1.18	1.09%
720	57.77%	27.01%	5.69%	6.92%	1.22	1.23%
700	67.83%	35.67%	5.29%	6.57%	1.24	1.29%
680	76.73%	45.42%	4.93%	6.18%	1.25	1.25%
660	83.90%	55.70%	4.64%	5.77%	1.24	1.13%
640	89.24%	65.80%	4.43%	5.37%	1.21	0.94%
620	93.28%	75.67%	4.27%	4.97%	1.17	0.70%
600	95.89%	83.17%	4.16%	4.67%	1.12	0.51%
580	97.61%	89.18%	4.10%	4.43%	1.08	0.34%
560	98.72%	93.67%	4.05%	4.25%	1.05	0.20%
540	99.37%	96.52%	4.03%	4.14%	1.03	0.11%
300	100.00%	100.00%	4.00%	4.00%	1.00	0.00%

Notice that reducing the cutoff from 700 to 680 caused the absolute difference to decline while the relative difference continued to increase. I am uncertain whether this is a function of the fact that the population examined, being comprised of persons with credit scores above 300 (and who were offered and accepted loans), is a truncated population. See Truncation Issues and Credit Score Illustrations subpages of the Scanlan's Rule of jpscanlan.com.