

James P. Scanlan
Attorney at Law
1529 Wisconsin Avenue, NW
Washington, D.C. 20007
(202) 338-9224
jps@jpscanlan.com

October 9, 2012

Drew Gilpin Faust, Ph.D.
President
Harvard University
Cambridge, Massachusetts 02138

1. A table of contents to this letter may be found [here](#). 2. This letter, a follow-up letter pertaining to the withdrawal of the *Commissioned Paper: Healthcare Disparities Measurement* (the subject of Section E.2.c at pages 42-43) and the response to the follow-up letter by research integrity officers of Harvard Medical School and Massachusetts General Hospital are discussed in Section C.1.g, at pages 30-32, of my Federal Committee on Statistical Methodology 2013 Research Conference paper titled "[Measuring Health and Healthcare Disparities](#)" and at pages 16-17 of "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014). 3. The PowerPoint presentation for the applied statistics workshop at Harvard's Institute for Quantitative Social Science to which this letter is related is available [here](#).

Re: Measurement of Health Disparities and Other Demographic Differences by
Faculty and Researchers at Harvard Universityⁱ

Dear President Faust:

On occasion I write to institutions whose missions involve the interpretation of data on demographic differences in the law and the social and medical sciences, alerting those institutions to problems in their interpretations arising from the failure to recognize the ways that standard measures of differences between rates of experiencing favorable or adverse outcomes tend to be affected by the overall prevalence of an outcome. The problems I address are almost universal among institutions around the world involved in such things as teaching or research (including the publication of research) about differences in health or healthcare outcomes or other outcomes of public concern. Thus, the recipient institutions are chosen not because of any particular deficiency in their work compared with other work on similar issues, but on the basis of the institution's importance regarding matters where the measurement issues I raise are most pertinent.

I have for some time intended to write to Harvard University both because of its position as the leading university in the world and because of the volume and prestige of its health and healthcare disparities research. But this letter is immediately prompted by the scheduling of an October 17, 2012 Applied Statistics Workshop at Harvard's Institute for Quantitative Social Science where I will present a paper titled "[The Mismeasure of Group Differences in the Law and the Social and Medical Sciences](#)"ⁱⁱⁱ that will address many of the issues raised in this letter.

ⁱ To facilitate consideration of the issues raised in letters such as this I make available electronic copies of the letters on the Institutional Correspondence subpage of the Measuring Health Disparities page of jpscanlan.com. Underlinings in this letter reflect links to the underlined material in such a copy of the letter. If the letter is corrected after it is first posted on the website, such fact will be noted on the final page.

The timing of the letter is also influenced by Harvard's scheduling of an October 11, 2012 university-wide symposium titled "[Eliminating Health Disparities: Transdisciplinary Perspectives](#)." The discussion below will show that there exist grave problems with the research community's efforts to appraise the size of health and healthcare disparities, including efforts regarding the crucial issue of whether such disparities are increasing or decreasing over time. A central flaw in health disparities research involves the near universal failure to recognize the implications of patterns by which relative and absolute differences between outcome rates tend to be affected by the prevalence of an outcome. These patterns include that whereby relative differences in favorable outcomes and relative differences in the corresponding adverse outcomes tend to change systematically in opposite directions as outcome rates change generally. That is, for example, as mortality declines relative differences in mortality tend to increase while relative differences in survival tend to decrease. The patterns also include that whereby absolute differences between rates tend to increase as rare outcomes become more common and decrease as already common outcomes become even more common. Because the role of these patterns have never been considered in research appraising the difference between the circumstances of two groups reflected by a pair of outcome rates, the conclusions of such work have been invariably suspect and very often simply incorrect. And even when conclusions are broadly correct, such research is misleading by implying or representing that particular methods of quantifying differences in the circumstances of two groups are sound when in fact the methods are fundamentally flawed.

The policy implications of the failure to understand the patterns addressed here are substantial. In fact, as I will show below, misperceptions resulting from the failure to consider these patterns have led the Commonwealth of Massachusetts to unwisely include a health disparities measure in its Medicaid pay-for-performance program and to do so in a manner that is more likely to result in increased healthcare disparities than in reduced healthcare disparities. Thus, I had hoped to provide this letter sufficiently in advance of the October 11 symposium that the speakers at the symposium would have an opportunity to consider the bearing of the points it makes on the subjects they are to address. I regret that I was unable to provide it much earlier than I have been able to do.²

Irrespective of the symposium, however, the purpose of this letter is to cause Harvard University to comprehensively review the methods by which its various arms examine group differences in the law and the social and medical sciences and to ensure that the research and teaching of those arms with respect to such matters proceeds with a sound statistical foundation.

Section A addresses the patterns whereby relative differences in experiencing favorable and adverse outcomes tend to be systematically affected by the prevalence of an outcome and some of the implications of the universal or near universal failure to understand those patterns.

Section B addresses the patterns whereby absolute differences between outcome rates (and differences measured by odds ratio) tend to be systematically affected by the prevalence of an

² Because I would not be in a position to transmit the letter prior to October 9, 2012, on October 8, 2012, I sent the speakers an email advising them of the issues the letter would raise. Given that the letter is an open letter, I will provide the speakers links to the online copy of the letter as soon as the letter is mailed.

outcome and some of the implications of the universal or near universal failure to understand those patterns.

Section C addresses the implications of the failure to understand the patterns described in Sections A and B with respect to perceptions about the impact of pay-for-performance programs on healthcare disparities and with respect to the measurement of healthcare disparities in the Massachusetts Medicaid pay-for-performance program and any pay-for-performance programs in other jurisdictions that considers healthcare disparities issues among their performance criteria.

Section D addresses the fallacy of notions that two measures that yield contrasting interpretations as to the comparative strength of the forces causing an advantaged group's rate of experiencing an outcome to differ from a disadvantaged group's rate of experiencing the outcome, in one setting compared with another, at one point in time compared with another, or with respect to one outcome compared with another, can both be valid or that the choice between the measures involves a value judgment. The section also explains a method of appraising the differences in the circumstances of two groups reflected by a pair of rates that is theoretically unaffected by the prevalence of an outcome.

Section E addresses the disarray in health disparities research generally and at Harvard, showing the way that researchers commonly employ a chosen method either without recognizing or without acknowledging that different methods would yield contrary conclusions and, more important, without considering the patterns by which the chosen method tends to be systematically affected by the prevalence of an outcome and the implications of such patterns with respect to the validity of the measure. The section also discusses reasons why the recently issued [*Commissioned Paper: Healthcare Disparities Measurement*](#), a joint project of Harvard Medical School and Massachusetts General Hospital, should be withdrawn.ⁱⁱⁱ

Section F addresses reasons why it is important that Harvard address the issues in this letter in a timely manner and identifies persons who have some knowledge of those issues.

A. Patterns by Which Relative Differences in Favorable and Adverse Outcomes Tend to be Systematically Affected by the Overall Prevalence of an Outcome and Implications of the Failure to Recognize Those Patterns.

For about twenty-five years, I have been writing about the patterns by which, for reasons inherent in the distributions of factors associated with experiencing an outcome, standard measures of differences between outcome rates tend to be systematically affected by the overall prevalence of an outcome. Links to about 170 references explaining these patterns and the implications of the failure to understand them with respect to various issues in the law or the social and medical sciences may be found on the [Measuring Health Disparities](#) page (MHD) of [jpscanlan.com](#). The nuances of the patterns are described on the [Scanlan's Rule](#) page (SR) of the same site. Several score other pages and subpages on the site explore the failure to understand these patterns with respect to particular issues. For example, the [Mortality and Survival](#) page discusses the fact that in published studies in prestigious journals, particularly in the discussion of racial differences in cancer outcomes, researchers refer to relative differences in survival and

relative differences in mortality interchangeably, often stating they are analyzing one relative difference while in fact analyzing the other. They do so invariably without recognizing that the two relative differences tend to change in opposite directions as overall survival rates change or that more survivable cancers tend to show larger relative differences in mortality but smaller relative differences in survival than less survivable cancers. Extended discussion of like misunderstandings with regard to matters of substantial public importance may be found on the [Lending Disparities](#), [Discipline Disparities](#), [Educational Disparities](#), [Immunization Disparities](#), and [Feminization of Poverty](#) pages and the [Subgroup Effects](#) subpage of SR.

The more important published articles pertaining to these issues include “[Can We Actually Measure Health Disparities](#),” a guest editorial in the Spring 2006 issue of the American Statistical Association magazine *Chance*, as well as “[Race and Mortality](#)” (*Society*, Jan/Feb 2000, reprinted in *Current*, Feb. 2000), “[Divining Difference](#)” (*Chance*, Fall 1994), “[The Perils of Provocative Statistics](#)” (*Public Interest*, Winter 1991), and “[Feminization of Poverty’ is Misunderstood](#)” (*Plain Dealer*, Nov 11, 1987, reprinted in *Current*, May 1988, and *Annual Editions: Social Problems 1988/89*, 1988). Conference presentations containing particularly useful explanations concerning these issues include “[Can We Actually Measure Health Disparities?](#)” (7th International Conference on Health Policy Statistics 2008), “[Measurement Problems in the National Healthcare Disparities Report](#)” (American Public Health Association 2007), and “[The Misinterpretation of Health Inequalities in the United Kingdom](#),” British Society for Populations Studies 2006. Subsequent reference to these items will commonly employ some shorthand form.³

The online comments collected in [Section D](#) of MHD, which include a substantial number of comments on studies by Harvard researchers, will give an impression of how widespread is the misunderstanding of the issues I raise among health disparities researchers in the United States and abroad. [Section E.7](#) of MHD provides a summary of the extent of scholarly agreement with my interpretation of these issues. The most significant of the work addressing that interpretation can be described as specifically agreeing with it, or, while not necessarily agreeing with the interpretation as to the forces causing observed patterns, similarly concluding that it is not possible to conduct sound research on group differences without taking these patterns into account.

The most notable of the patterns by which standard measures of differences between outcome rates are affected by the prevalence of an outcome is that whereby the rarer an outcome, the greater tends to be the relative difference between the rates at which advantaged and disadvantaged groups experience it and the smaller tends to be the relative difference between the rates at which advantaged and disadvantaged groups avoid it. The pattern can be illustrated with virtually any data set where one can observe various point on a continuum of quantifiable factors associated with experiencing an outcome or simply observe the rates at which different groups experience or avoid an outcome as that outcome increases or decreases in prevalence.

³ Because many references in this letter are clipped from other materials, including many that use medical journal citation form, I have not endeavored to reconcile citation formats.

Such illustrations may be found on the [Framingham Illustrations](#), [NHANES Illustrations](#), [Life Tables Illustrations](#), [Income Illustrations](#), and [Credit Score Illustrations](#) subpages of SR. Illustrations may also be found in countless studies of health disparities or guides on the measurement of those disparities, even though the authors draw their conclusion based on only one of the outcomes and without consideration of that fact that examination of the opposite outcome would yield a contrary conclusion. For example, in the *Commissioned Paper* mentioned above, in each instance where the authors make some statement concerning the comparative size of a disparity based on a relative difference in a favorable or adverse outcome, the relative difference in the opposite outcome would support a statement to the opposite effect. The same holds with a number of the figures in the document that are presented simply to graphically show minority and white rates of experiencing some favorable or adverse outcome.

One will of course find many departures from this pattern and other patterns described here. 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 here. Society's interest is solely in (a). But only with a firm understanding of (b) can one discover (a).

An illustration showing the patterns by which relative differences between rates at which blacks and whites fall above or below various percentages of the poverty line change in opposite directions as that percentage increases or decreases may be found in Table 1 and Figures 2 and 3 of the 2006 *Chance* editorial.⁴ Compelling illustrations can also be drawn from published life tables, which allow one to show the patterns by which relative (racial and gender) differences in mortality decline with age while relative differences in survival increase with age (see [Life Tables Illustrations](#) subpage of SR and [Life Table Information Document](#)), and I will make recurring references to such patterns and the materials illustrating them.

But I believe the most useful illustration of patterns by which the two relative differences are affected by the prevalence of an outcome is based on hypothetical test score data, which show how lowering a test cutoff, thereby making test failure less common and test passage more common, increases relative differences in failure rates while reducing relative differences in pass rates.

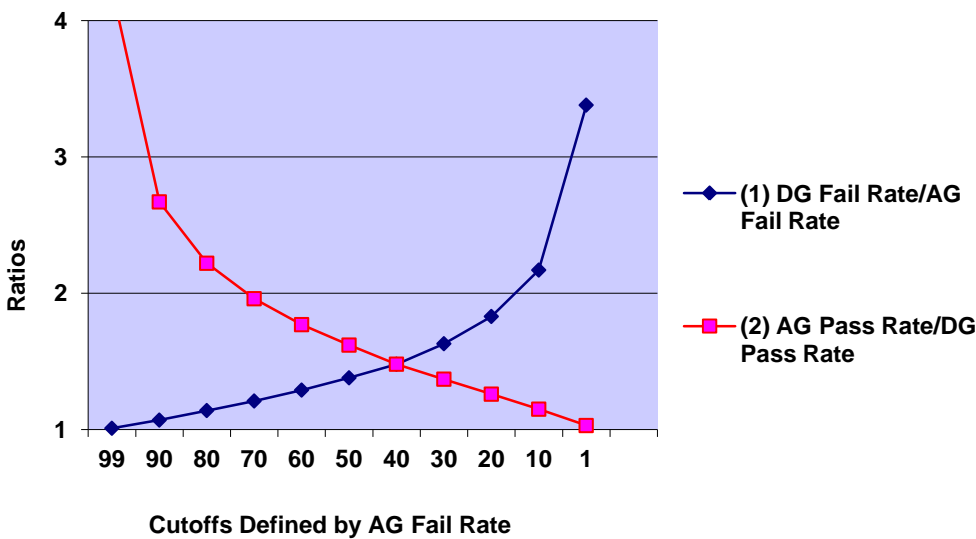
Figure 1 below is based on a situation where the two groups have normal test score distributions with means that differ by half a standard deviation (and where the standard deviations of the distributions are equal).⁵ A graphic illustration of the distributions themselves may be found in Figure 1 of the 1994 *Chance* article "[Divining Difference](#)." The numbers at the bottom of the figure are the fail rates of the advantaged group, which are used as benchmarks for overall prevalence of an outcome. The two lines represent the ratios of the fail rates of the disadvantaged

⁴ Useful illustrations of a similar nature may be found in Tables I through IV of my "[The Misunderstood Relationship Between Declining Mortality and Increasing Racial and Social Disparities in Mortality Rates](#)" Norwegian Institute of Public Health 2001.

⁵ The data underlying the figure may be found in Table 1 of my paper "[The Misinterpretation of Health Inequalities in the United Kingdom](#)," British Society for Population Studies 2006.

(*i.e.*, lower-scoring) group (DG) to the fail rate of the advantaged (*i.e.*, higher-scoring) group (AG) and the ratios of the pass rates of the advantaged group to the pass rates of the disadvantaged group.⁶ From left to right, the lines illustrate the effects on the two rate ratios of serially lowering the test cutoff from one point to another, in each instance enabling persons with scores falling between the two points now to pass the test. One thus observes the pattern whereby relative differences in the adverse outcome and relative differences in the favorable outcome change in opposite directions as the two outcome changes in overall prevalence.⁷

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



⁶ The relative difference is the rate ratio (RR) minus 1 where RR is above 1 and 1 minus RR where RR is below 1. It is the 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), as was done in the 2006 *Chance* editorial. But for reasons explained in the [Semantic Issues](#) subpage of SR and for ease of interpretation, I now prefer always to use the larger rate as the numerator, in which case all RRs will be above 1 and the larger is each RR the larger is the relative difference. Because choice of numerator affects the way a relative difference is characterized – *e.g.*, 80 is 20% less than 100 while 100 is 25% greater than 80 – choice of numerator can sometimes affect determination of which relative differences is larger or other matters. But the choice of numerator is of no consequence to any material point made in this letter.

⁷ Some readers may recognize that the entire situation is reflected in one line – *i.e.*, that one pattern is implicit in the other. See Section A. 4 of the [Scanlan's Rule](#) page. I think, however, that such readers will understand the need for both lines to illustrate the patterns to those to whom the redundancy of the lines may not be evident.

The same patterns will hold if, without any lowering of cutoffs, test performance is improved such as to allow all persons previously scoring between two points to reach the higher cutoff.

The following are some examples of the 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.

- When rates of achieving proficiency for various subjects in elementary and secondary schools generally increase, relative differences in rates of failing to achieve proficiency tend to increase while relative differences in proficiency rates tend to decrease.
- When poverty declines, relative differences in poverty rates tend to increase while relative differences in rates of avoiding poverty tend to decrease.
- When mortality declines, relative differences in mortality rates tend to increase while relative differences in survival rates tend to decrease.
- When adverse health outcomes increase in overall prevalence, relative differences in experiencing them tend to decrease while relative differences in rates of avoiding them tend to increase.
- The more survivable is a particular type of cancer, the smaller will tend to be relative differences in surviving it while the larger will tend to be relative differences in failing to survive it.
- When overall rates of receiving beneficial health procedures or care (*e.g.*, mammography, immunization, prenatal care, adequate hemodialysis, coronary artery bypass grafting) increase, relative differences in rates of receiving such procedures or care tend to decrease while relative differences in rates in failing to receive them tend to increase.
- Generally reducing blood pressure will tend to increase relative differences in hypertension while reducing relative differences in rates of avoiding hypertension; generally improving folate levels will tend to increase relative differences in low folate while reducing relative differences in adequate folate.
- Banks with more liberal lending policies will tend to have larger relative differences in mortgage rejection rates, but smaller relative differences in mortgage approval rates, than banks with less liberal lending policies.
- More lenient school discipline policies will tend to result in larger relative differences in discipline rates, though smaller relative differences in rates of avoiding discipline, than more stringent policies.

- Lower professional proficiency or employment performance standards will tend to result in larger relative differences in failing to meet them, though smaller relative differences in meeting them, than higher standards.
- At more selective universities, where graduation rates are generally higher than at less selective universities, relative differences in graduation rates will tend to be smaller, while relative differences in rates of failing to graduate will tend to be larger, than at less selective universities.
- Relative racial and gender differences in receipt of sophisticated therapies will tend to be smaller, while relative differences in rates of failure to receive the therapies will tend to be larger, among subjects whose symptom/condition profiles call for generally higher rates of receipt of the therapies than among subjects whose symptom/condition profiles call for generally lower rates of receipt of the therapy.
- Relative differences in adverse outcome rates tend to be large among comparatively advantaged subpopulations (where such outcomes are rare) – *e.g.*, racial differences in infant mortality where parents are highly educated; racial differences in low birth weight among low risk groups; racial differences in mortgage rejection rates among high income applicants; racial, gender, and socioeconomic differences in mortality among the young; occupational differences in mortality and morbidity among British civil servants; racial and socioeconomic differences in failing to receive appropriate care among the insured – while relative differences in the opposite, favorable outcomes tend to be small among those subpopulations.

To take a pair of examples from the *Commissioned Paper* that involve the types of patterns mentioned in the last two bulleted points, Figure 4 (at 33) shows that among appropriate renal transplantation candidates (where transplantation rates are generally higher than among inappropriate candidates), relative (racial) differences in rates of failing to receive a transplant are larger, while relative differences in rates of receipt of a transplant are smaller, than among inappropriate candidates. Figure 8 of the document (at 53) shows that within the highest income group (where rates of fair/poor health are much lower than within the lowest income group), the relative (racial) difference in fair/poor health is larger, and the relative difference in avoiding fair/poor health is smaller, than within the lowest income group.

While the focused discussion of the implications of the failure to understand these patterns on health and healthcare disparities research at Harvard and elsewhere will be presented later in this letter, I believe it would be useful at this point to summarize certain aspects of that matter. For simplicity I do so with respect to healthcare even though, as suggested above, issues concerning receipt or non-receipt of immunization, where the former is the favorable outcome and the latter is the adverse outcome, are equally implicated in discussions of survival and mortality. Also for simplicity, I cast the matter in terms of generally improving healthcare, which is the fairly

consistent pattern observed in the United States and most other high income countries. As suggested by Figure 1, as healthcare generally improves – which commonly means increased rates of receiving beneficial procedures like immunization and cancer screening as well as increased rates at which care is deemed to be generally appropriate – those who measure healthcare in terms of relative differences in the favorable outcomes (as used to be the predominant approach and is still a common approach) will tend to find decreasing disparities. Those who measures disparities in terms of relative differences in adverse outcomes (as has been the case at the National Center for Health Statistics (NCHS) since approximately 2004 and as will be the case for those who follow NCHS guidance) will tend to find increasing disparities. The Agency for Healthcare Research and Quality (AHRQ), in the *National Healthcare Disparities Report*, relies on the larger of the two relative difference, which, as reflected in Figure 1 will ordinarily be the relative difference in the favorable outcome when the favorable outcome is uncommon but will ordinarily be the relative difference in the adverse outcome when the favorable outcome has become quite common. Thus, as healthcare improves, AHRQ will tend to find decreasing disparities for uncommon procedures like knee replacement, coronary artery bypass grafting, and the less common types of immunization; increasing disparities for more common procedures like the more common types of immunization as well as generally appropriate care (which in a country like the United States usually is found in a substantial majority of cases); when a favorable outcome increases from being uncommon to being very common, AHRQ for a time will tend to find disparities to be decreasing but then will later tend to find them to be increasing.^{iv}

A description of the patterns of findings by those who measure disparities in terms of absolute differences between rates (as is the most common practice of researchers in the Health Care Policy Department of Harvard Medical School) is more complicated. So I will defer that discussion until after I have explained the patterns by which absolute differences between rates tend to be affected by the prevalence of an outcome.

I add at this juncture that the important point of this letter is not that various measures tend to yield different conclusions. And it is certainly not that the issues I raise can be addressed by presenting each of the various measures that yield different conclusions from one another. Indeed, as I show in Section D, any thought that the presentation of contrasting measures addresses the issues raised here betrays a fundamental misunderstanding of the purpose of an inquiry into the nature of the forces that cause the rates of an advantaged group and a disadvantaged group to differ. Rather, the important point is that none of the standard measures of differences between rates can provide useful information about the comparative circumstances of two groups without consideration of the way the measure tends to be affected by the overall prevalence of the outcome examined. That applies to health disparities research and to every other subject in the law and the social and medical sciences where the difference between the circumstances of demographic groups is an issue of consequence.

There are two important corollaries to the 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. First, as an outcome changes in overall prevalence, groups with lower baseline rates for the outcome will tend to experience larger proportionate changes in those rates than groups with higher baseline rates for the outcome, while groups with higher baseline rates for the outcome will tend to experience larger proportionate changes in the opposite outcome. Second, the rarer an outcome, the larger will tend to be the proportion groups most susceptible to the outcome comprise of both (a) the population experiencing the outcome and (b) the population failing to experience the outcome. See Table 1 of [Chance 1994](#), Table 1 of [Chance 2006](#), and Figure 3 of [American University Colloquium 2012](#).

Despite the fact that these patterns are apparent in a wide variety of readily available data, they are unrecognized in virtually every context where understanding them is critical to soundly interpreting data on group differences. Published income data, in precisely the form it is published (*i.e.*, in terms of numbers of persons and rates of falling below various percentages of the poverty line, broken down by race and family type) make it evident that when overall poverty rates change significantly advantaged groups will generally experience a larger proportionate change in their poverty rates than disadvantaged groups, while disadvantaged groups will generally experience a larger proportionate change in their rates of avoiding poverty than advantaged groups. See Tables I to IV of [Norwegian Institute of Public Health 2001](#).

As suggested in note 7 above, there is nothing contradictory between the patterns pertaining to poverty rates and those pertaining to rates of avoiding poverty. Rather, the latter is the same as the former, since disadvantaged groups have lower rates of avoiding poverty. But the contrasting interpretations based on the two outcome highlight that neither pattern is reflective of a meaningful change in the comparative circumstances of the two groups.

Nevertheless, when poverty declines observers commonly find significance in the fact that it decreased more among advantaged groups and when it increases they find it significant that it increased more among advantaged groups. Yet the only things that might be significant about such patterns, or the contrasting patterns regarding rates of avoiding poverty, are such things as might be divined by exploring the way observed patterns might differ from those one would predict on the basis of the pre-change income distributions. To my knowledge, however, that has never been done. Rather, among the hundreds or thousands of instances where either journalists or scholars have addressed the perceived significance of some change in rates during a period of general changes in some adverse or favorable outcome rate, I am unaware of single instance (save as might be reflected in the works discussed in [Section E.7](#) of MHD) where the journalist or scholar has considered the extent to which the observed pattern was a function of the general change. See the [Explanatory Theories](#) subpage of SR for discussion of some of the theories that have been devised to account for perceived patterns by which different groups do not share proportionately in some general improvement in health, but where no thought is given to the statistical forces described here.

The same holds for observations concerning changes in the proportion a group comprises of the part of the population experiencing an outcome. In 1978, following dramatic reductions in poverty (including the poverty of female-headed families), the term “feminization of poverty” was coined to describe the fact that female-headed families had come to comprise a much larger

proportion of the poor than they previously did. No thought was given to the extent that such pattern was a function of the decline in poverty or to the reasons to expect that any increases in poverty (including the poverty of female-headed families) would reduce the proportion female-headed families comprise of the poor. Increases in the proportion female-headed families comprised of the poor was also confused with a worsening of the circumstances of female-headed families, notwithstanding that increases in the proportion female-headed families comprised of the poor tend to be correlated with an improvement in the circumstances of female-headed families.⁸ Thirty-four years later, the same may be said about current discussions of this issue of continuing prominence. In fact, the feminization of poverty is commonly discussed as if it is an ongoing phenomenon, notwithstanding that – due in significant part to the unfortunate fact that poverty never declined substantially after the middle 1970s – the proportion female-headed families comprise of the poor differs little today from the proportion it was when the feminization of poverty was discovered. I note, however, that the pattern may indeed be ongoing in some developing countries where poverty, once nearly universally, is increasingly restricted to those most susceptible to it

Even works that address at some length the differences in interpretations as to the size of relative differences depending on whether one examines favorable or adverse outcomes commonly if not universally fail to recognize that the two relative differences tend to change in the opposite directions as the prevalence of the outcome changes. In one classic work, however, a Harvard Medical School professor came close to doing so when examining relative (gender) differences in mortality and survival among whites. In a 1958 Special Article in the *New England Journal of Medicine* styled “Shall We Count the Living or the Dead,”⁹ Mindel C. Sheps, while exploring issues concerning the size of mortality ratios and survival ratios, noted:

Such ratios have another feature: there is no predictable relation between relative mortality and relative survival. For example, among the three age groups in Table 2, the oldest group shows the greatest relative discrepancy by far in survival rates, but the smallest relative discrepancy in mortality rates.

But relative differences in mortality and survival do have a predictable relationship, as in fact is suggested by the coincidence of the lowest mortality rate ratio and the highest survival rate ratio

⁸ The feminization of poverty is also an unfortunate subject of study because it is influenced by two disparate phenomena – changes in relative differences between poverty rates of female-headed families and other units and changes in the proportion female-headed families comprise of the population. The same issue detracts from the utility of the concept for illustrating the statistical patterns described here. But, for reasons stated earlier, the reader should understand why decreases in poverty will tend to increase relative differences in poverty rates (but reduce relative differences in rates of avoiding poverty) of female-headed families and more advantaged demographic units, while increases in poverty will tend to have the opposite effects. See my “[The ‘Feminization of Poverty’ is Misunderstood](#)” (*Plain Dealer*, Nov 11, 1987); “[Comment on “McLanahan, Sorensen, and Watson's 'Sex Differences in Poverty, 1950-1980.'”](#)” (*Signs*, Spring 1991):409-13; “[The Perils of Provocative Statistics](#)” (*Public Interest*, Winter 1991): 3-14.

⁹ *N Engl J Med* 1958; 259:1210-14.

among the oldest age group noted by Professor Sheps, and as is consistently illustrated in the life table data mentioned earlier. The two relative differences tend to change in opposite directions as mortality and survival rates change generally.¹⁰ But for a certain irregularity in the data she examined, Professor Sheps would likely have discerned this herself, as she almost certainly would have done had she examined the same sort of data in the first section of either Table A or Table B of the [Life Table Information Document](#).^{11v} Professor Sheps would also have likely recognized such pattern had she examined the data with recognition that they presumably reflected underlying (probably more or less normal) male and female distributions of risks of mortality and survival.¹²

¹⁰ It would not, however, be accurate to say that the two rate ratios are inversely related. For any given level of prevalence, the larger is the relative difference in mortality the larger will be the relative difference in survival. That will hold for all measures.

¹¹ Professor Sheps examined the information set out in the three rows of Table 1 below, which show the female and male rates of surviving to age 40, surviving from age 40 to 60, and surviving from age 60 to 80, along with the male/female mortality ratios and the female/male survival ratios. In accordance with what I will term the standard pattern, the survival ratios increased as survival rates decreased. But contrary to the standard pattern whereby mortality ratios decrease as survival rates decrease, the mortality ratio first increased and then decreased. Had Professor Sheps examined solely the data in the first and third rows or solely the data in the second and third rows, she would have observed the standard pattern for both measures and might well have identified such pattern.

Table 1. White Male and Female Survival Percentages with Mortality Ratios and Survival Ratios (from Sheps NEJM 1958)

Age	Percent Male Surv	Percent Female Surv	M/F Mort Ratio	F/M Surv Ratio
Birth to Age 40	92.1%	95.1%	1.61	1.03
Age 40 to Age 60	81.8%	90.3%	1.88	1.10
Age 60 to Age 80	33.7%	49.5%	1.31	1.47

Whether or not there is something meaningful in the departure from the standard pattern in the mortality ratios of the first age group and the second age group in the data examined by Professor Sheps is hard to say. But it should be recognized that, while there may be variations in the strength of forces driving gender differences in mortality among persons under age 40 compared with persons over age 40 – and certainly the nature of the factors that might affect mortality of men and women differently under age 40 are likely to be different from those affecting men and women differently after age 40 – the comparatively few deaths under age 40 provide reason to expect more random variation under age 40 than over age 40. In any case, one finds the standard pattern rather consistently once the focus is restricted to the ages where the bulk of deaths occur. And rarely would one fail to observe the standard pattern when comparing, say, a 40 to 49 age group with a 60 to 69 age group, because the substantial differences in mortality between the two age ranges will cause the prevalence-related forces to predominate. See discussion in [Life Tables Illustrations](#) subpage of SR. The important matter, however, involves the way the patterns demonstrate the problematic nature of standard measures of differences in outcome rates and that the measure discussed in Section D infra is the only sound measure of the strength of the forces causing a pair of rates to differ.

¹² See the discussion in [Comment on Eikemo Int J Equity Health 2009](#) of the two studies discussed in [Section E.7](#) of MHD that, in responding to the descriptions in “Race and Mortality” or “Can We Actually Measure Health Disparities?” of patterns of correlations of overall prevalence with the size of standard measures of differences between rates, reached similar conclusions empirically. Failing to recognize the distributional forces underlying the patterns, however, both studies were unable to posit sound explanations for the patterns they observed.

As reflected by Figure 1, however, the pattern is inferable from an understanding of normal distributions in abstract terms – indeed from the widely recognized pattern whereby differences between means of distributions that yield modest relative differences in experiencing a common outcome yield very large relative differences in experiencing an outcome as the outcome is restricted toward the tail of the overall distribution.

Yet I am uncertain that a great many people know that it is even possible for the two relative differences to change in opposite direction, much less that it will tend to happen systematically. The simple fact that lowering test cutoffs or like standards tends to increase relative differences in failing to meet them seems almost universally unrecognized. And educational, professional, and employment standards have been questioned, challenged, or invalidated on the basis of perceptions that large relative differences in failing to satisfy a standard are functions of the fact that the standard is set too high, even though lower standards would yield larger relative differences in failing to meet them. Similarly, the high proportion a disadvantaged group comprises of those adversely affected by a disqualifying criterion is often regarded as evidence that the standard is set too high, when lowering the standard would increase that proportion.¹³

The scope of the misunderstanding of the relationship between the stringency of a standard and relative differences in failing to meet it is illustrated by two matters recently in the news where federal law enforcement policies are based on statistical perceptions that are the exact opposite of reality. One involves the settlements of lending discrimination cases against Countrywide Financial Corporation announced in December 2011 and against Wells Fargo Bank announced in July 2012, with recoveries totaling over half a billion dollars. For eighteen years, the federal government has been encouraging mortgage lenders to relax lending criteria that tend to cause minorities to have their mortgage loan applications rejected at higher rates than whites (as well as more generally to reduce all adverse lending outcomes that minorities experience more often than whites). Such encouragement accords with longstanding practice in the enforcement of equal employment opportunity laws where lowering test cutoffs has been universally regarded as reducing the disproportionate impact of employment tests on lower-scoring groups because, as shown in Figure 1, lowering cutoffs tends to reduce relative differences in pass rates. But, as also shown in Figure 1 and discussed above, lowering cutoffs tends to increase relative differences in fail rates.¹⁴ Apparently unaware that reducing the frequency of adverse outcomes tends to increase relative differences in rates of experiencing such outcomes, federal regulators continue to monitor lender practices on the basis of relative differences in adverse lending outcomes. Such approach – evident in the complaints underlying the Countrywide and Wells

¹³ See my “[An Issue of Numbers](#)” (*National Law Journal*, Mar. 5, 1990) and “[Getting it Straight When Statistics Can Lie](#)” (*Legal Times*, June 28, 1993).

¹⁴ Whether, and the extent to which, lowering a test cutoff in fact reduces the disproportionate impact of an employment test turns on whether, and the extent to which, lowering a cutoff increases the proportion members of the disadvantaged group make up of a finite number of selections from among test passers. Typically, lowering test cutoffs will increase that proportion though it does not have to do so. See the [Employment Tests](#) subpage of SR. When passing a test or meeting a criterion itself dictates receipt of the desired outcome, the cutoff would seem not to affect that impact of the test or device as that impact would reasonably be measured.

Fargo settlements – makes the lenders most responsive to encouragements to reduce adverse lending outcomes the most likely targets for litigation. See my “[‘Disparate Impact’: Regulators Need a Lesson in Statistics](#)” (*American Banker*, June 5, 2012), and “[The Lending Industry’s Conundrum](#)” (*National Law Journal*, Apr. 2, 2012), and “Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies” (*Amstat News*, December 2012, in press). See also the [Lending Disparities](#) page.¹⁵

The second matter recently in the news involves racial differences in public school discipline rates, which received substantial media attention after the Department of Education’s March 2012 release of data showing several-fold racial differences in rates of suspension or expulsion. Observers, including the Departments of Justice and Education, have attributed the size of these differences to zero tolerance discipline policies in effect in recent decades.¹⁶ Those agencies have even encouraged school systems to relax discipline standards in order to reduce the differences in discipline rates. In response to concerns about racial disparities, Colorado has already enacted legislation modifying discipline standards and many legislative or administrative actions to the same effect are under consideration in other jurisdictions. Yet, like low test cutoffs or relaxed lending criteria, lenient school discipline standards tend to result in larger, not smaller, relative differences in adverse outcomes than more stringent ones. And, as with the enforcement of fair lending laws, investigations of the extent to which racial or ethnic bias is responsible for observed disparities are likely to focus on the school systems that are most responsive to encouragements to relax discipline standards. See my “[Racial Differences in School Discipline Rates](#)” (*Recorder*, June 22, 2012), which, among other things, discusses the way racial disparities in discipline rates increased substantially after the Los Angeles Unified School District introduced a program to reduce discipline rates. See also Section A of the [Discipline Disparities](#) page, as well as the [Los Angeles SWPBS](#) subpage, which discusses the Los Angeles program, and the [Suburban Disparities](#) subpage, which discusses misperceptions concerning the fact that relative differences in discipline rates are substantially larger in suburbs than in cities.

See generally Sections B through D of the [Discipline Disparities](#) page and the [Less Discriminatory Alternative - Substantive](#) subpage of the [Disparate Impact](#) page regarding the fact that, in any circumstance where the racial impact of a policy or device is appraised in terms of relative differences in adverse outcomes, things that would commonly be regarded as less discriminatory alternatives to existing practices will tend to result in larger relative differences in adverse outcomes than the existing practices.

¹⁵ While settlements in lending discrimination cases of the scope of those in the cases against Countrywide and Wells Fargo are a recent phenomenon, the pattern whereby lenders most responsive to federal encouragements to relax lending criteria tend to be singled out for litigation is a longstanding one. See my “[When Statistics Lie](#)” (*Legal Times*, Jan. 1, 1996).

¹⁶ Such was the perception of a March 6, 2012 *New York Times* editorial styled “[The Wrong Approach to Discipline](#).” It urged the Department of Education to press “school systems with the worst records [*i.e.*, those with the largest relative differences in discipline rates] to develop fair and sensible strategies that involve working with troubled children and their families instead of reflexively showing them the door.”

***.

One situation where the failure to understand the above-described patterns affects a matter of special interest to highly selective educational institutions like Harvard involves the debate over affirmative action at such institutions. Proponents of affirmative action have found support for their position in the fact that relative differences in minority and white graduation rates tend to be smaller at more selective institutions than at less selective institutions, while opponents of such policies have found support for their position in the fact that relative differences in rates of failing to graduate tend to be larger at more selective institutions than at less selective institutions. Neither side has recognized that both patterns are to be expected simply because graduation rates tend to be high, and rates of failing to graduate tend to be low, at highly selective institutions.

Related to the misperceptions about the significance of the size of relative differences in favorable or adverse outcomes at highly selective institutions are common misperceptions about large relative differences in adverse outcomes among advantaged subpopulations. Observers commonly attribute significance to the fact that relative differences in adverse lending outcomes tend to be larger among higher income groups than lower income groups, invariably without recognition of the reason to expect such pattern, as well as the opposite pattern for favorable lending outcomes, simply because the former are less common and the latter are more common among higher income groups than among lower income groups. See the [Disparities – High Income](#) subpage of the [Lending Disparities](#) page. The same, of course, holds for patterns of racial disparities in discipline rates in suburbs compared with cities mentioned above.

I noted above that the charts in the *Commissioned Paper* illustrated some of the patterns whereby relative (racial) differences in adverse outcomes were greater, though relative differences in favorable outcomes were smaller, among advantaged subpopulations than among disadvantaged subpopulations. The *Commissioned Paper* did not comment on the former pattern, nor was there reason for it to, given that the paper was a measurement guide and not a commentary of the size of disparities. But commonly, and probably to an increasing degree until statistical analysis in health and healthcare disparities research is substantially reformed, commentaries on the size of disparities will emphasize patterns of larger relative differences in adverse outcome rates among advantaged subpopulations, as is done, for example, in the Commonwealth Fund's [Racial and Ethnic Disparities in U.S. Health Care: A Chartbook](#) (2008), but always without recognition that it would be unusual to observe anything else.

Ever since a 1992 *New England Journal of Medicine* study¹⁷ received widespread media attention concerning its report of large racial differences in infant mortality where parents were college-educated, large racial differences in infant mortality where parents are highly-educated has been provoking thought in analyses of health disparities. But in the 20 years of thought on the matter little attention has been given to the fact that large relative differences in adverse

¹⁷ Schoendorf KC, Hogue CJR, Kleinman JC, Rowley D. Mortality among infants of black as compared with white college-educated parents. *N Engl J Med* 1992; 326:1522

outcomes among advantaged populations, though small relative differences in the corresponding favorable outcomes, are to be expected simply because adverse outcomes are rarer among advantaged subpopulations. See “[Race and Mortality](#).”

For well more than 20 years in the United Kingdom researchers have been drawing provocative inferences from the fact that relative differences in mortality and other adverse health outcomes are larger among British civil servants than among the UK population at large. No thought, however, has been given to the fact that large relative differences in adverse health outcomes, though small relative differences in favorable health outcomes, are to be expected among British civil servants simply because they tend to be healthier than the UK population at large. See the [Whitehall Studies](#) subpage of MHD. The forces underlying these patterns are exactly the same as those that cause relative differences between mortality rates of racial or gender groups – or for that matter between Finnish owners and renters, persons of normal weight and persons who are obese, or any other set of groups that tend to have different mortality rates – to invariably be larger among persons in their forties (the advantaged subpopulation) than in their sixties (the disadvantaged subpopulation) while causing relative differences in survival to be larger among persons in their sixties than in their forties. See [Life Tables Illustrations](#) subpage of SR, [Life Table Information](#) document, Figure 6 of [Nordic Demographic Symposium 2008](#), [Table A](#) of the [Comment on Berrington de Gonzalez NEJM 2010](#), and Table 1 of the [Mortality and Survival](#) page.

A 1997 article in *The Lancet* found that, despite being comparatively egalitarian societies, Norway and Sweden had comparatively large socioeconomic differences in morbidity and mortality among western European countries. Precisely because Norway and Sweden are egalitarian societies, these findings have been a matter of great concern in those countries for the last 15 years. But efforts to investigate causes for the patterns have suffered from a failure to recognize the role in these patterns of the fact that the populations of Norway and Sweden are quite healthy. See [Comment on Hemmingsson Eur J Public Health 2005](#), [Comment on Wilkinson Lancet 2006](#), [Comment on Mackenbach Lancet 1997](#).

For many similar examples, see the 1991 *Public Interest* article “[The Perils of Provocative Statistics](#)” (which explains, among other things that one with a sound understanding of the patterns described here could easily predict, (a) that relative differences between poverty rates of female-headed and married-couple families are larger among whites (where poverty is less common) than among blacks, while relative differences between rates at which female-headed and married-couple families avoid poverty are smaller among whites than among blacks, as well as, correspondingly, (b) that relative differences between rates poverty rates of blacks and white are greater among married-couple families than among female-headed families, while relative differences between rates at which blacks and white avoid poverty are smaller among married-couple families than among female-headed families). See also examples in [Society 2000](#), [Chance 2006](#).¹⁸

¹⁸ The larger racial differences in fair/good health among the highest income category than the lowest income category in Figure 8 of the [Commissioned Paper](#), mentioned just after the bulleted points, is precisely the sort of pattern that underlies many perceptions that being of higher socioeconomic status benefits whites more than blacks.

Failure to understand these patterns makes it virtually certain that any effort to explore the interaction of race and socioeconomic status in health disparities research, as is recommended in the *Commissioned Paper* and many other places, will yield some misleading conclusion, while almost never providing an insight of value. See [Comment on Kawachi Health Affairs 2005](#) and [Comment on Thurston Am J Epidemiol 2005](#) (the subject studies of which are addressed further in Section E.2). But the same issues exist with respect to any effort to determine whether a health disparity has increased over time, which is merely an exploration of the interaction of race and time. Thus, the confusion in research regarding disparities among advantaged subpopulations merely highlights the confusion in research about every other aspect of health disparities.

Discussions of large relative differences in adverse outcomes among advantaged subpopulations – and the attendant suggestions of a need to examine interactions of some factor like socioeconomic status with group membership – invariably involve situations where the absolute difference between rates is smaller among the advantaged subpopulation than among the disadvantaged subpopulation. Thus, researchers who rely on absolute differences between rates would generally fail to find larger differences in adverse outcomes among advantaged subpopulations, reason to investigate possible explanations for such a pattern, or reason to examine interactive effects (unless with respect to opposite perceptions of interactive effects).

The reasons to expect different relative effects among groups with different baseline rates are also overlooked in circumstances where they have important clinical implications, as in settings where the disadvantaged groups are the controls and the advantaged groups are the treated subjects of clinical trials. The standard assumption in such trials is that a factor will ordinarily cause equal proportionate changes in baseline rates of different subgroups and that any departure from such pattern will be deemed a subgroup effect. Yet, such expectation is fundamentally illogical given the simple fact that a factor cannot cause equal proportionate changes to different baseline rates of experiencing an outcome while at the same time causing equal proportionate changes to the baseline rates for the opposite outcome. The only sound expectation is that set out as the first corollary to the pattern of relative differences illustrated in Figure 1 (which was also discussed with respect to changes in poverty rates). Such expectation is that a factor will tend to cause a larger proportionate change in the outcome for the group with the lower baseline rate while causing a larger proportionate change in the opposite outcome for the other group. The same holds with regard to the crucial estimation of an absolute risk reduction involving a baseline rate different from that in the clinical trial forming the basis of perceptions about the effects or an intervention. See the [Subgroup Effects](#) subpage of SR.¹⁹ These issues have special

Self-rated health data like that in Figure 8 underlie a host of misperceptions regarding a phenomenon termed “reporting heterogeneity.” See the [Reporting Heterogeneity](#) subpage of MHD.

¹⁹ Harvard researchers have provided important guidance respecting the need to distinguish apparent subgroup effects that may have occurred by chance from actual subgroup effects. See Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine – Reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007;357:2187-2194. But

salience with respect to efforts to comply with requirements of federal law that medical research supported by the National Institutes of Health attempt to determine whether the therapies or procedures being studied affect minorities and women differently from the way they affect whites and men. See [Comment on Thurston Am J Epidemiol 2005](#).

That the assumption of a constant rate ratio across different baseline rates is unfounded – an aspect of the fact that the rate ratio is simply an unsound measure of association (see the [Illogical Premises](#), [Illogical Premises II](#), and [Inevitability of Interaction](#) subpages of SR) – also undermines common approaches to meta-analysis. See [Meta-Analysis](#) subpage of SR.

Finally, there exists an important issue concerning the way misperceptions about patterns of relative differences influence judgments about matters beyond the particular relative differences examined. That subject is treated in Section E.2.a with respect to a recent study by researchers at the Harvard School of Public Health.

B. Patterns by Which Absolute Differences between Rates and Differences Measured by Odds Ratios Tend to be Systematically Affected by the Overall Prevalence of an Outcome and Implications of the Failure to Recognize Those Patterns.

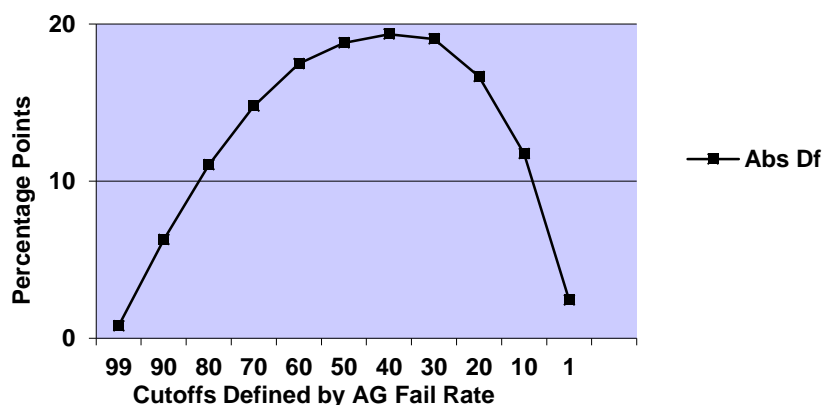
Absolute differences between rates and differences measured by odds ratios are the same whether one examines an adverse outcome or the opposite, favorable outcome. But for a measure to usefully indicate the degree of difference in the circumstances of two groups reflected by a pair of outcome rates the measure must remain constant when there occurs a change in overall prevalence akin to that effected by the lowering of a test cutoff.²⁰ And, like the two relative differences, absolute differences and differences measured by odds ratios tend to be affected by the overall prevalence of an outcome, though 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 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, as discussed in the introductory section of the [Scanlan's Rule](#) page. 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.

such guidance is set in the framework of the unjustified assumption that, absent a subgroup effect, one will observe the same relative effect across different baseline rates. See note 7 to the [Subgroup Effects](#) sub-page of SR.

²⁰ In many earlier discussions of these issues, I have referred to a measure of the difference between rates that is unaffected by the prevalence of an outcome. That characterization misses the point. Statements that 20% is 100 percent greater than 10% and that 2% is 100 percent greater than 1% are both obviously correct notwithstanding the difference in prevalence and notwithstanding that neither of the 2.0 rate ratios is a sound measure of association. The same holds for the 10 percentage point and 1 percentage point absolute differences. Such measures, however, simply fail to indicate the degree of association reflected by each pair of rates.

An illustration of patterns of changes in absolute differences based on the same specifications underlying Figure 1 above is set out in Figure 2 below.

Fig. 2. Absolute Differences Between Rates at Various Cutoff Points Defined by AG Fail Rate



The above illustration and discussion of absolute differences is based on the assumption that underlying distributions are normal, and I have explored few nuances of non-normal distributions. But I note that the patterns whereby relative differences in an outcome and its opposite tend to change in opposite directions as the prevalence of an outcome changes, as well as the two corollaries to the pattern, would hold even when the distributions are uniform (rectangular). By contrast, where the distributions are uniform, absolute differences would remain constant as overall prevalence changes. So the descriptions of absolute differences based on normal distributions would not hold where the distributions are uniform, but they would generally hold with respect to the normal or roughly normal distributions encountered in reality. Illustrations in [International Conferences in Health Policy Statistics 2008](#) show patterns similar to that in Figure 2 with data from other than precisely normal distributions, including distributions that are necessarily other than normal because they are truncated portions of larger normal and non-normal distributions.

Further, I note that the following relationships regarding patterns of relative and absolute differences hold regardless of the shapes of the underlying distributions. Anytime the two relative differences change in the same direction, the absolute difference will also change in that direction. But anytime the absolute difference changes in a direction that is different from the change in one relative difference, the other relative difference will necessarily change in the opposite direction of the first relative difference and hence in the same direction as the absolute difference. Thus, anytime an observer notes that a relative difference and the absolute difference have changed in different directions over time, the unmentioned relative difference will have changed in the opposite direction of the mentioned relative difference and in the same direction as the absolute difference.

The following are some examples of the above-described patterns of absolute differences between outcome rates and the overall prevalence of the outcome.

- As uncommon procedures (*e.g.*, coronary artery bypass grafting, knee replacement, certain types of immunization) increase, absolute differences tend to increase; as common procedures (*e.g.*, mammography, prenatal care, certain types of immunization) increase, absolute differences tend to decrease.
- As procedures go from being uncommon to being very common absolute differences will tend to increase then decrease.
- As survival rates increase for cancers with generally low survival rates, absolute differences will tend to increase; as survival rates increase for cancers with generally high survival rates, absolute differences will tend to decrease.
- Increases in student proficiency rates in more difficult subjects (where rates may be very low) will tend to increase absolute differences between rates, while increases in student proficiency rates in easier subjects will tend to reduce absolute differences between rates.
- For outcomes with generally low rates or in settings with generally low rates for an outcome, higher rates for the outcome will tend to be associated with larger absolute differences between rates; for outcomes with generally high rates or in settings with generally high rates for an outcome, higher rates will tend to be associated with lower absolute differences between rates.
- Since the Agency for Healthcare Research and Quality (AHRQ) measures disparities in terms of the larger relative difference (in either the favorable or the adverse outcome), observers who measure disparities in terms of absolute differences between rates will tend to reach opposite conclusions from those of AHRQ as to the comparative size of disparities in different settings or at different points in time.
- Since the National Center for Health Statistics (NCHS) relies on relative differences in adverse outcomes to measure all disparities, observers who rely on absolute differences between rates will tend to reach the same conclusions as NCHS where the adverse outcome is common (*e.g.*, failure to receive some uncommon procedure like coronary artery bypass grafting (CABG)) and different conclusions from NCHS where the adverse outcome is uncommon (mortality, failure to receive mammography).

None of these patterns is recognized in the increasing body of research and commentary that discusses things such as healthcare disparities or disparities in elementary and secondary school proficiency rates in terms of absolute differences between rates.

Finally, because of the discussion above concerning whether distributions are normal, I note that while the potential for distributions to vary from normal will affect the extent to which

distributionally-driven/prevalence-related forces will conform to those described above, such potential by no means provides a basis to rely on absolute differences without attempting to take the effects of overall prevalence into account. Indeed, even if the relationship between absolute differences (or any difference) and the prevalence of an outcome were commonly substantially different from the patterns described here, so long as a measure is in some manner affected by overall prevalence, the measure cannot be usefully employed without considering the implications of overall prevalence of the outcome.

Differences measured by odds ratios tend to change in the opposite direction of absolute differences and hence (subject to qualifications previously noted with respect to absolute differences when a rate is above 50% for one group and below 50% for the other) in the same direction as the larger of the two relative differences. An illustration of this pattern may be found in [Figure 5](#) of [Royal Statistical Society 2009](#), which sets out illustrations for all four measures together. Given the common use of logistic regression, understanding patterns by which odds ratios change as prevalence changes can be important (as suggested, for example, in item 1 of the [Adjustment Issues](#) subpage of the [Vignettes](#) page). But since the odds ratio is not the standard measure used by the governmental or Harvard entities discussed in various parts of this letter, I will give odds ratios only limited attention.²¹

C. Implications of the Patterns by Which Relative and Absolute Differences Tend to be Affected by the Prevalence of an Outcome with Respect to Pay-for-Performance Programs in Massachusetts and Elsewhere.

A particularly noteworthy consequence of the use of absolute differences as a measure of healthcare disparities without regard to the patterns described here is that (a) the study of absolute differences between rates at which advantaged and disadvantaged groups received an uncommon procedure that incentive programs seemed to increase generally (and where increases tend to be associated with increased absolute differences) has led to the perception in the United States that pay-for-performance (P4P) will tend to increase healthcare disparities, while (b) the study of absolute differences between rates at which advantaged and disadvantaged groups experience certain common healthcare outcomes that incentive programs seemed to increase generally (and where increases tend to be associated with decreased absolute differences) has led to the perception in the United Kingdom that P4P will tend to decrease healthcare disparities. Both patterns, however, are to be expected given the rate ranges at issue.

The study principally responsible for the perception in the United States involved (among other things) an examination of the way a coronary artery bypass graft (CABG) report card program, which was believed to cause general increases in CABG rates, affected racial and other

²¹ Some observers have asserted that odds ratios are constant, or close to constant, across different baseline rates. But they are not sufficiently close to constant to provide a superior measure to that described in Section D. As reflected in Tables 3 and 4 of the [Subgroup Effects](#) subpage of SR, however, the odds ratio provides a method of applying information derived in a clinical trial to estimate the absolute risk reduction that an intervention is likely to achieve for various baseline rates that is closer to the soundest estimate than either of the two observed relative risk changes provides.

differences in CABG rates.²² The key figures underlying the conclusion that incentive programs seemed to increase racial disparities in CABG rates are set out in Table 2, along with the absolute difference relied upon by the authors of the study and the rate ratios for the favorable and adverse outcomes, the odds ratio,²³ and the “EES” (as I will explain in Section D, a measure of the difference in circumstances reflected by a pair of rates that is theoretically unaffected by the prevalence of an outcome).

Table 2. White and Black CABG Rates Before and After Implementation of a CABG Report Card with Measures of Differences (from Werner et al. *Circulation* 2005)

Period	W Rt	B Rate	W/B Receipt Ratio	B/W non-Rec Ratio	Abs Df	Odds Ratio	EES
1	3.6%	0.9%	4.00	1.03	.027	4.11	0.58
2	8.0%	3.0%	2.67	1.05	.050	2.81	0.48

As shown in the table, observers who rely on relative differences in favorable outcome rates (or, as in the case of AHRQ, the larger relative difference), would have found the racial disparity to have decreased. So, too, would observers who analyze the matter using logistic regression with resultant odds ratio, which, as noted, tend to track the direction of the larger relative difference as the prevalence of an outcome changes. The EES figure suggests that, to the extent that the change in the disparity can be effectively measured, the disparity decreased.²⁴ As noted, however, the authors, relying on absolute differences between rates as a measure of disparity, found an increase in the racial disparity, as NCHS would also have done, but by relying on relative differences in failure to receive CABG.

Observers relied on the authors’ determination that disparities had increased and thereby came to believe that incentive programs like P4P would lead to increase healthcare disparities, which in turn led some of those observers to recommend that P4P programs include criteria for evaluating provider performance on the basis of the size of, or changes in the size of, healthcare disparities. Massachusetts was the first to respond to that call by including a healthcare disparities criterion in its Medicaid P4P program. But the program evaluated the size of disparities on the basis of a measure that was a function of absolute differences between rates, and it did so with regard to outcome rates that were generally quite high (above 80% for all types of care combined). Given the tendency for higher overall rates in such ranges to be associated with smaller absolute differences between rates, the program will tend to find healthcare disparities to be smaller at

²² Werner, RM, Asch DA, Polsky D. Racial profiling: The unintended consequences of coronary artery bypass graft report cards. *Circulation* 2005;111:1257–63.

²³ There are four ways to calculate the odds ratio, two of which yield one figure and two of which yield a figure that is the reciprocal of the first figure. The odds ratio shown, being above 1, could be either the ratio of the white odds of receipt of CABG to the black odds of receipt of CABG or the ratio of the black odds of failing to receive CABG to the white odds of failing to receive CABG.

²⁴ See [Second Comment on Werner Circulation 2005](#) regarding why the measure discussed in Section D might not be effective in a setting like that at issue in the CABG study.

higher-performing hospitals than lower-performing hospitals and hence to reward those higher-performing hospitals for reason unrelated to a useful indicator of the size of a disparity. Since higher-performing hospitals tend to have smaller minority representations among their patient populations than lower-performing hospitals, the inclusion of a disparities criterion in the Massachusetts P4P program – by diverting resources away from providers with large numbers of minority patients – is more likely to increase healthcare disparities than to reduce them. See the [Pay for Performance](#) and [Between Group Variance](#) subpages of MHD and my [Comment on Blustein Health Affairs 2011](#). See also “[Perverse Perceptions of the Impact of Pay for Performance on Healthcare Disparities](#)” (9th International Conferences on Health Policy Statistics 2011).

When I first created the [Pay for Performance](#) subpage questioning the wisdom of including disparities measures in pay-for-performance programs given the measurement issues I have raised in various places, I viewed the matter simply in terms of the misguided allocation of resources based on factors that were not in fact related to a sound measure of disparities. I did not consider the potential impact of P4P on healthcare disparities simply as a result of the diverting of resources to better-performing hospitals as a general matter (*i.e.*, without regard to the inclusion of any disparities measure), as has been suggested by Friedberg *et al.*²⁵ Having considered the healthcare disparities implications of simply diverting resources to higher-performing hospitals, I note that measuring disparities in terms of relative differences in favorable outcome rates (which tend to decline as favorable outcomes increase) will tend to favor higher-performing hospitals in the same way that the Massachusetts program does, while measuring disparities in terms of relative differences in adverse outcomes will tend to have the opposite effect.

I also did not originally consider the implications of the fact that, as reflected in the Massachusetts program, P4P programs that include disparities measures typically will do so with respect to indicators of receipt of generally appropriate care (rather than receipt of some uncommon procedure). Appropriate care rates usually will be well above 50% for all groups. Hence, as happened in Massachusetts, reliance on absolute differences between rates as a measure of healthcare disparity may commonly cause higher-performing hospitals to be perceived as having smaller disparities than lower-performing hospitals, with consequent diversion of resources to the former for reasons not necessarily having anything to do with the size of disparities as they might reasonably be measured.

²⁵ See Friedberg MW, Safran DG, Coltin K et al. Paying for performance in primary care: Potential impact on practices and disparities. *Health Aff* (Millwood) 2010;29:926-932.

D. The Fallacy of Notions That Choice of Disparities Measure Involves a Value Judgment or That Two Measure Yielding Different Conclusions About the Strength of the Forces Causing a Disparity are Both Valid (and Discussion of a Theoretically Sound Method of Appraising the Strength of Those Forces Reflected by a Pair of Rates).

The points made in this section are pertinent to an explanation provided by researchers from the Health Care Policy Department (HCPD) of Harvard Medical School in response to a letter to the editor concerning an article in a 2005 issue of the *New England Journal of Medicine*. The explanation, however, will be addressed in the discussion in Section E.2.a concerning the approach to healthcare disparities measurement of HCPD.

Increasingly, health and healthcare disparities researchers discuss relative and absolute differences in circumstances where the examined relative difference provides a different interpretation as to the comparative size of a disparity from that provided by the absolute difference. Sometimes they do so simply to provide as complete a picture as possible. But sometimes researchers maintain that both measures provide valid information regarding a particular aspect or the matter or suggest that a value judgment is involved in the choice between measures.²⁶ Commonly or invariably, when relative differences and absolute differences are presented in such circumstances, there is no mention of the other relative difference (which, as noted previously, will necessarily yield a conclusion that is the opposite of that drawn from the presented relative difference and the same as that drawn from the absolute difference). Rarely do researchers discuss choices between the two relative differences, even when mentioning the NCHS recommendation that all health and healthcare disparities be measured in terms of relative differences in adverse outcomes. But one situation where there is at least a suggestion that choice between the two relative differences might involve some sort of value judgment may be found in the *Commissioned Paper's* recommendation that the user should make the choice based on "context of the report." See Section E.2.c *infra*.

The above may be an inexact summary of published discussions of these issues. But, in any case, the points that follow should dispel any notion that there can be other than one correct answer in any effort to determine whether the strength of the forces causing two groups' rates of experiencing an outcome has changed over time or is greater in one setting than another.

Table 3 below shows hypothetical hire rates of advantaged and disadvantaged groups applying for work at four employers, along with the rate ratios for hire and the rate ratios for rejection, as well as the absolute difference between rates and an odds ratio. In a situation where it is assumed that for each employer the qualifications of the applicants from the advantaged group do not differ from the qualifications of the disadvantaged group and all differences in rates result from employer bias, the question posed is how might the employers be ranked, from highest to lowest,

²⁶ See the University of Michigan CD-ROM course [Measuring Health Disparities](#); Harper S, King NB, Meersman SC, et al. [Implicit value judgments in the measurement of health disparities.](#) *Milbank Quarterly* 2010.

according to level of bias. The numbers in parentheses for each measure reflect the ranking pursuant to that measure. I note in advance that I could make the same point more simply with two rows of data. But I use four rows to illustrate some of the issues concerning the ways the absolute differences and odds ratios alter their directions of change as overall prevalence changes.

Table 3. Hypothetical Patterns of Hire Rates of Advantaged and Disadvantage Groups at Four Employers and Measure of Differences between Rates of Hire or Rejection

Employer	AG Hire Rate	DG Hire Rate	AG/DG Hire Ratio	DG/AG Rej Ratio	Abs Df	Odds Ratio
A	20.1%	9.0%	2.22 (1)	1.14 (4)	0.11 (4)	2.53 (1)
B	40.1%	22.7%	1.77 (2)	1.29 (3)	0.17 (2)	2.29 (3)
C	59.9%	40.5%	1.48 (3)	1.48 (2)	0.19 (1)	2.19 (4)
D	90.0%	78.2%	1.15 (4)	2.17 (1)	0.12 (3)	2.50 (2)

There are four principal ways observers might rank the degree of bias of these employers.²⁷ Those who rely on relative differences in favorable outcomes, such as might commonly occur in an employment discrimination case involving hiring or promotion, would rank them A,B,C,D. Those who rely on relative differences in adverse outcomes as the National Center for Health Statistics would do 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.

A third approach would be to rank them according to the absolute difference between rates, such as researchers at the Health Care Policy Department of Harvard Medical School would commonly do, which would be C,B,D,A. And those who rely on the odds ratio, such as those who would attempt to evaluate the situation by means of logistic regression might do, would rank them A,D,B,C, the opposite of the ranking based on absolute differences.

I suggest, however, that it would be absurd to assert that one employer is more biased than another as to selection while another is more biased as to rejection. It would be similarly absurd to say that contrasting interpretations as to the degree of bias based on either of the two relative differences and the absolute difference (or odds ratio) would both be sound or that determining which employers are the most biased involves a value judgment. Rather there can only be one correct interpretation as to the comparative bias of the employers reflected in the data.

The same reasoning would hold if, instead of representing the situations of four employers, the rows of data represented one employer at four points in time and the question to be answered was whether discrimination increased or decreased from each point in time to the next. The reasoning would hold as well if it was not known whether any of the employers was biased and the question to be answered involved the degree of difference in the qualifications of applicants

²⁷ I omit the AHRQ approach of relying on the larger of the two relative differences partly for issues mentioned in note 6 and partly in order not to inordinately complicate the matter. But the approach would roughly approximate that shown for the odds ratio.

of the advantaged and disadvantaged groups that would be necessary to explain each difference in outcome rates as a result of something other than bias.

The same reasoning holds in any setting. The only value in exploring the meaning of a situation reflected by two unequal rates of experiencing an outcome involves determining what the rates signify about the strength of the forces causing the rates to differ. If two measures provide different answers as to the comparative strength of such forces, both cannot be correct. That hardly suggests, however, that one is correct.

What then can in fact be divined about the comparative degrees of bias reflected in the four rows? It is the same in each case. Each situation is based on the specifications underlying Figures 1 and 2 where means of the underlying normal distributions differ by half a standard deviation. Moreover, not only is the degree of bias the same in all cases, there exists no rational argument that the degree of bias reflected in any of the rows differs from that in another. Similarly, there is no rational argument that any measure that says the rows reflect different degrees of bias is a valid measure.

Table 3 is also a useful illustration because it implies the only solution to the question of how to appraise a difference in circumstances reflected by a pair of rates. The only theoretically sound method of conducting such an appraisal involves deriving from a pair of rates the differences between the means of the hypothesized underlying distributions measured in terms of percentages of a standard deviation. The method is explained on the [Solutions](#) subpage of MHD and illustrations of the values it would yield (which I commonly have termed “EES” for estimated effect size and which statisticians will understand as the value yielded by a probit analysis) for various pairs of rates (compared with rate ratios for those rates) may be found in Table 5 of [American University Colloquium 2012](#) (as well as Table 1 supra) and in many of the post-November 2007 comments collected in [Section D](#) of MHD.²⁸ The Solutions subpage also discusses various limitations of the method, which are not insubstantial. But at least it is a theoretically sound method. And it is, for example, the only method that would tell us that the degree of bias of each of the four employers in the table is the same. All other measures would falsely tell us that the degrees of bias are different.^{vi}

This would seem also a useful place to note that all thoughtful research about such things as health and healthcare disparities involves a search for the reasons why a disparity occurs. Absent the knowledge of the patterns explained here, and in a sense encapsulated in the method described in the prior paragraph, the information in Table 3 would leave observers with a basis for investigating the differences in the nature of employer practices that might account for the perceived differences in size of the four differences in outcomes rates. Or, in the situation where the rows reflect changes over time, observers without knowledge of the patterns described here might explore whether the perceived pattern of decreasing bias as to selection might have resulted from the replacement of a more biased with a less biased hiring official or whether the perceived pattern of increasing bias as to rejection might have resulted from the replacement of a

²⁸ Usually such figures are derived by means of the database described on the Solutions subpage of MHD, which involves a sort of virtual slide rule. As discussed on the Solutions subpage, an online calculator yields more exact result.

less biased with a more biased hiring official. It should be evident that there is nothing to be gained from such investigations and great risk that they would yield false conclusions. The same holds with respect to the investigation of any pattern of differences between rates appraised without regard to the patterns described here.

Three other matters relating to the patterns in Table 3, and the referenced method for appraising differences reflected by a pair of rates, warrant mention. The first, which is implicit in the varying interpretations standard measures would yield regarding the ranking of the bias reflected in the four rows of data in the table, is that standard approaches to measuring discrimination in employment (or any other discrimination issue) – even when the outcome rates are known and regardless of whether the approaches involve examining relative differences in favorable outcomes or relative differences in adverse outcomes – are unsound. The only sound method is that described above and which, as noted, would find the four situations to reflect the same amount of bias.

The second and third matters involve the fact that one must know the rates at which two groups experience an outcome to employ that method. Commonly, employment discrimination cases are not based on actual rates of experiencing an outcome but on the proportion a group comprises of those eligible to experience an outcome (*e.g.*, applicants, relevant labor market, promotable employees, all incumbent employees) and the proportion the group comprises of those experiencing the outcome. From such information it is possible to derive the rate ratios for experiencing the outcome (though not the rate ratios for avoiding the outcome).²⁹ But as already explained, one needs the actual rates in order to soundly appraise the strength of the forces causing the rates to differ. Thus, efforts to appraise the strength of such forces in many employment settings, and various other settings where one does not have the actual rates at which each group experiences the outcome (*e.g.*, where one knows only know the proportion a group comprises of drivers and of persons stopped by police, but not the actual rates at which drivers of various groups are stopped) are problematic. See the [Representational Disparities](#) subpage of SR.

That does not mean that such analyses can never prove discrimination. For they do involve a difference in outcome rates that may reflect a difference in treatment. But without the actual rates for each group one cannot draw valid inferences about the comparative strength of the forces causing the differences in different settings. Nor can one abstractly appraise the strength of the force causing the difference, save in the sense of knowing that a particular relative difference reflects a smaller force for a rare outcome than for a common outcome. And that appraisal is important for divining whether the likelihood that an observed difference is a function of failure to adequately adjust for nondiscriminatory factors. See the [Disparate](#)

²⁹ For example, where the disadvantaged group comprises 20% of applicants and 40% of hires, the advantaged group's hire rate is 2.67 times the hire rate of the disadvantaged group's rate (*i.e.*, $(80/60)/(20/40)$).

[Treatment](#) subpage of the [Discipline Disparities](#) page and the [Underadjustment Issues](#) subpage of [Lending Disparities](#) page.³⁰

The final matter involves case control studies. Such studies enable one to determine differences between the odds at which two groups defined by a condition happen to have a particular factor, which difference is also the difference between the odds at which persons with and without the factor experience the condition. And where the outcome is rare, the odds ratio approximates the ratio of the rate at which those with the factor experience the condition to the rate at which those without the factor experience the condition. But case control studies do not enable one to determine the actual rates at which those with and without the factor experiences the condition. As noted, such rates are necessary in order to derive an accurate estimate of a factor's effect. Thus, case control studies cannot provide such an estimate even when they provide fair estimates of relative risks. See the [Case Control Studies](#) subpage of SR. I do not know the practical implications of such fact. But appraisals of the utility of case control studies must recognize that the degree to which the odds ratio approximates the rate ratio does not resolve the matter,

E. The Disarray in Health Disparities Research Generally and at Harvard and the Questionable Value of All Such Research Undertaken without Regard to the Patterns by Which Standard Measures of Differences Between Outcome Rates Tend to be Affected by the Prevalence of an Outcome.

1. Health Disparities Measurement Generally

As previously explained, the crucial point of everything said above about the way standard measures of differences between rates tend to be associated with the prevalence of an outcome is not simply that researchers will commonly reach different conclusions about the comparative size of a difference between outcome rates depending on the measure employed, but that each of the measures is unsound. Nevertheless, the incongruities resulting from use of differing measures by various government agencies and arms of Harvard University, commonly without any mention that other approaches might yield different results, and the resulting disarray in

³⁰ In "[Illusions of Job Segregation](#)" (*Public Interest* Fall 1988) I discussed the impossibility of proving that a group was discriminatorily excluded from a better job on the basis of its higher representation in a poorer job. In doing so, I noted that a valid analysis must examine the subject group's representation among persons seeking the better job and its representation among persons securing the better job. I no doubt said many similar things in the articles on the same issues collected in Section A of the [Employment Discrimination](#) page. Whether or not such statements might be deemed technically correct, it is clear that a sound effort to prove exclusion from a job must examine the rates at which the putatively disfavored and the putatively favored groups are selected or rejected for the better job, from which can be derived the EES. That failing of the article, however, does not undermine the reasoning as to the various issues it addressed, including the illustration in Table IV that weaker qualifications of certain groups among persons selected for a job does not suggest that there was no discrimination against the group. See also "[Both Sides Misuse Data in the Credit Discrimination Debate](#)" (*American Banker* July 22, 1998). But the absence of the actual selection rates would undermine any interpretive point one might make on the basis of the differences in rate ratios for selection for persons with and without prior experience. See related discussion at the close of Section E.2.b. See the [Sears Case Illustration](#) subpage of SR regarding the pertinence of patterns reflected in Table IV to the principal issues addressed in this letter.

health disparities research warrant some discussion. Though I will not stress the point in each instance where it is pertinent, it should be recognized that, as suggested with respect to the employers in Table 3 at the close of the last section, the incongruity of the contrasting interpretations as to the comparative size of disparities is highlighted when observers posit explanation for observed patterns that are perceived to reflect differences in the size of disparities, including when disparities are believed to change over time. For researchers who reach different conclusions as a result of employing other methods would presumably have to posit quite different explanations.

To put the current disarray in health disparities research in context, one must look prior to the principal studies by Harvard researchers that I mention below. As of the end of the 1990s disparities in racial and other demographic difference in health status (as distinguished from healthcare) tended generally to be studied in terms of relative differences in adverse outcomes, such as relative differences in mortality. Since most adverse health outcomes were declining, and correspondingly relative differences in experiencing them tended to be increasing, health disparities were usually deemed to be increasing. No thought was given to the fact that, solely for reasons inherent in the underlying distributions, declines in adverse outcome rates would commonly be associated with increasing relative differences in adverse outcome rate irrespective of any meaningful change in the comparative situation of the groups. Indeed, it was commonly observed that relative differences in mortality had increased “despite declining mortality.” Nor was attention given to whether relative differences in favorable outcome were decreasing. Such matters were similarly ignored when researchers attempted to discern reasons either for what were perceived to be increasing disparities in adverse outcomes that were declining or for what were perceived to be decreasing disparities in adverse outcomes in the occasional situation where an adverse outcome was increasing. The same holds with respect to inquiries concerning large relative differences in adverse outcomes among advantaged subpopulations, as I believe was adequately addressed in Section A.

Meanwhile, solely as a matter of convention, disparities in receipt of beneficial health procedures tended to be measured in terms of relative differences in favorable outcomes. Since such procedures were generally increasing, differences tended to be regarded as decreasing. See Table 7 of [American University Colloquium 2012](#) for an example where, in an October 26, 1998 [Progress Review: Black Americans](#), the Department of Health and Human Services highlighted decreases in the racial difference in certain immunization rates on the basis of decreases in relative differences in the receipt of immunization.

Thus, as of the late 1990s, according to the methods commonly employed to appraise such matters, racial disparities in health status seemed to be increasing while racial disparities in healthcare seemed to be decreasing. But no health disparities research was exploring whether observed patterns were any different from those one would expect given the overall changes in rates. And several years into the implementation of the Race and Health Initiative of the Clinton Administration, there was little prospect that the achievements of that initiative would be reasonably appraised. Further, there existed the potential for great waste of resources in the federally-mandated inquiries into the way therapies might affect minorities and women differently from whites and men if such inquiries were undertaken without consideration of

reasons to expect interventions to cause different proportionate changes to different baseline rates.

Such at any rate was how I characterized the matter in the 2000 *Society* article titled “[Race and Mortality](#).” As also discussed in “Race and Mortality,” NCHS had acknowledged that it had not previously considered the measurement issues the article addressed (which I had brought to the attention of the officials directing the Race and Health Initiative in 1998.) Interactions with NCHS statisticians Kenneth Keppel and Jeffrey Percy concerning these issues over the ensuing year led NCHS to address in some manner the fact that relative differences in favorable outcomes and relative differences in adverse outcomes would tend to yield different conclusions as to the direction of the change in a disparity over time. In did so in documents issued in 2004³¹ and 2005. The more important of these was 2005 document “[Methodological Issues in Measuring Health Disparities](#)” (authored by Keppel, Percy, and seven other experienced health disparities researchers, including Joel S. Weissman of Harvard Medical School, who is also the principal author of the *Commissioned Paper* mentioned above).³²

For instant purposes, it merely warrants note that the document did not address the implications of the patterns by which relative differences in favorable outcomes and relative differences in adverse outcome tend to be affected by the prevalence of an outcome with respect to the utility of either measure for appraising the strength of the forces causing two rates to differ without consideration of those patterns. Rather, simply observing that determinations as to the directions of changes over time will turn on whether one examines relative differences in the favorable outcome or relative differences in adverse outcomes (thereby rather overstating the matter), NCHS recommended that henceforth all disparities (including both health and healthcare disparities) should be appraised in terms of relative differences in adverse outcomes.

The consequences of the recommendation, which underlies Health and Human Services’ appraisals of achievement of the health disparities reduction goals in Healthy People 2020, as it did also for Health People 2010, are substantial. In the case of healthcare disparities, as reflected in the above-mentioned Table 7 of [American University Colloquium 2012](#), decreases in immunizations that had been highlighted by HHS in the 1998 progress report now became substantial increases in disparities, and any NCHS researchers that might have been investigating the reasons for the decreases in the disparities now would need instead to investigate reasons for increases in the disparities. More generally a great many healthcare disparities that might previously have been deemed to be decreasing would now be deemed to be increasing and further improvements in healthcare would tend to be associated with increasing healthcare disparities.

It is important to note, however, that the NCHS failure to more directly address the measurement issues I had brought to its attention goes not merely to healthcare disparities, but to disparities in

³¹ Keppel KG, Percy JN, Klein RJ. [Measuring Progress in Healthy People 2010](#) (2004).

³² More direct responses by Keppel and Percy to “Race and Mortality” and the Fall 1994 *Chance* article “Divining Difference” are discussed in [Section E.7](#) of MHD.

things like mortality and morbidity, where conclusion about patterns of changes had no sounder a foundation than had conclusions about healthcare disparities.

Yet it is not clear the extent to which the recommendation regarding healthcare disparities is being followed even within the government. As previously noted, the Agency for Healthcare Research and Quality in the *National Healthcare Disparities Report* measures disparities in terms of whichever relative difference is larger. That approach will tend to yield the same conclusion as the approach of NCHS in situations where the adverse outcome is rare (and hence where the relative differences in the adverse outcome will tend to be larger than the relative difference in the favorable outcome, as reflected in Figure 1) and different conclusions from the approach of NCHS in situations where the adverse outcome is common (and hence where the relative difference in the adverse outcome will tend to be smaller than the relative difference in the favorable outcome, as reflected in Figure 1). The Centers for Disease Control and Prevention (CDC), while employing a number of measures in its 2011 [Health Disparities and Inequalities Report](#), measured immunization disparities in terms of absolute differences between rates. As discussed on my [Immunization Disparities](#) page, CDC researchers have also independently conducted studies of immunization disparities relying on absolute differences between rates. That approach will tend to yield the same conclusion as NCHS where immunization rates are low (as reflected in Figures 1 and 2) and a different conclusion from NCHS where immunization rate are high (as also reflected in Figures 1 and 2).³³ The disarray within the federal government is also illustrated in the way that officials of AHRQ have argued that improvements in healthcare will tend to reduce healthcare disparities while relying on a study that used a measurement approach that would tend to yield conclusions about the correlation between quality improvement and healthcare disparities that are the opposite of those AHRQ would tend to reach. See [Comment on Aaron and Clancy JAMA 2003](#) and “[Measurement Problems in the National Healthcare Disparities Report](#),” American Public Health Association 2007. But it is not uncommon for health disparities researchers to find their work to be consistent or inconsistent with other work without regard to whether the consistency or inconsistency is solely a function of the measure chosen.

Some private researchers have been measuring healthcare disparities in terms of relative differences in adverse outcomes, with or without reference to the NCHS recommendation. But the fact that the recommendation is not universally followed, and may not even be widely known, is reflected in a study by Morita et al. appearing in *Pediatrics* in 2008 that had received a Robert Wood Johnson Foundation award for addressing health disparities.³⁴ Relying on relative differences in vaccination rates as a measure of disparity, the authors found that a school-entry

³³ The Immunization page also describes study in which authors analyzed both relative differences in favorable outcomes (with respect to receipt of full immunization) and relative differences in adverse outcomes (with respect to failure to receive any immunization). While not necessarily reflected in the data in that particular study, the patterns described here suggest that increases in immunization rates will commonly decrease one disparity while increasing the other.

³⁴ Morita JY, Ramirez E, Trick WE. Effect of school-entry vaccination requirements on racial and ethnic disparities in Hepatitis B immunization coverage among public high school students. *Pediatrics* 2008;121:e547.

Hepatitis B vaccination requirement that dramatically increased overall vaccination rates also dramatically reduced racial and ethnic vaccination disparities. By contrast, NCHS would have found dramatic increases in disparities. See [Comment on Morita Pediatrics 2008](#) and Table 8 of [American University Colloquium 2012](#).

It is doubtful that federal funding of studies is in any way influenced by whether researchers are using methods that would tend to yield interpretations of patterns in disparities that are consistent with or inconsistent with the way that the funding agency would interpret such patterns. Even higher officials in many agencies do not know that such measurement issues exist and it is certainly doubtful that administrators involved in approving funding applications are any more aware of the ways different measures tend to yield different results than researchers are.

2. Health Disparities Measurement at Harvard

a. Health Care Policy Department of Harvard Medical School

At Harvard, since 2000 a vast volume of health and healthcare disparities research has been conducted wholly without regard to these issues, both before and after NCHS issued its recommendation concerning reliance on relative differences in adverse outcomes. Parts of this work analyze differences in things like healthcare funding or other health or healthcare issues involving continuous measures, matters to which the issues I raise would not appear to pertain save when a continuous measure is a function of some dichotomy. See [Comment on Chandola BMJ 2007](#) and [Comment on McWilliams Ann Int Med 2009](#).

In the case of the outcome rates to which the issues do pertain, the Health Care Policy Department (HCPD) of Harvard Medical School seems usually to measure healthcare disparities in terms of absolute differences between rates. Yet the most useful illustration of the issues addressed in this letter involves a situation where, though the corresponding author was from HCPD, the study relied on relative differences in favorable healthcare outcome rates.

In a study published in *American Journal of Public Health* in 2004, Escarce and McGuire,³⁵ examined changes in disparities in relatively uncommon medical procedures and diagnostics among elderly persons. The authors concluded that the disparities generally decreased during a period between 1986 and 1997 when rates were generally increasing.

Table 4, which is based on the [table](#) accompanying my [Comment on Escarce and McGuire APHA 2004](#), shows counts of the directions of changes for (a) the relative difference in the favorable outcome (the measure employed by the authors), (b) the relative difference in the adverse outcome (the measure NCHS would employ), (c) the absolute difference between rates (the measure HCPD would ordinarily employ) and (d) the EES figures described at the end of Section D, for the twelve situations involving the common pattern where rates generally

³⁵ Escarce JJ, McGuire TG. Changes in racial differences in use of medical procedures and diagnostic tests among elderly persons: 1986-1997. *Am J Public Health* 2004;94:1795-1799.

increased and there existed a difference between the black and white rates both at the beginning of the period and at the end of the period.

Table 4. Counts of Directions of Changes in Disparities in Healthcare Outcome Rates in Escarce and McGuire APHA 2004

Fav RR Dir	Adv RR Dir	Abs Df	EES	Total
Decrease	Increase	Increase	Decrease	9
Increase	Increase	Increase	Increase	3

The table shows that in most (9) of the 12 cases the relative difference in receipt of such procedures decreased (as commonly occurs when procedure rates generally increase), while in all 12 cases relative difference in failing to receive the procedures increased (as often occurs when procedure rates generally increase) and in all 12 cases the absolute differences between rates increased (as often occurs when procedure rates in the rate ranges at issue increase).³⁶

The authors posited a number of reasons for the perceived decreases in disparities. None of these reasons was implausible and, as indicated by the counts at the end (which include the EES), the authors would seem to have been correct with respect to the 9 cases where they identified a decrease in disparity. But the authors showed no recognition that, for reasons related to the underlying distributions, relative differences typically decline in these circumstances or that other measure would yield different conclusion.

Further, those who rely on relative difference in adverse outcomes or absolute differences between rates would presumably have to seek other reasons to explain the perceived increases in disparities in the 9 cases where Escarce and McGuire found decreases in disparity. It also warrants note that the patterns of changes were broadly similar to those Table 2, which, due to reliance on the absolute difference between rates as a measure of disparity, found an increase in the disparity examined, which finding then led to the perception in the United States that increases in procedure rates resulting from pay-for-performance programs will tend to increase healthcare disparities. Compare the [table](#) accompanying the comment with Table 2 above.

The five other studies of the HCPD on which I wrote comments – see [Comment on Schneider JAMA 2001](#), [Comment on Trivedi NEJM 2005](#), [Comment on Trivedi JAMA 2006](#), [Comment on Sequist Arch Int Med 2006](#), [Comment on McWilliams Ann Int Med 2009](#) (the last of which involved a study that used both continuous and dichotomous measures)³⁷ – all relied on absolute

³⁶ I use the word “often” in the latter two cases rather than “commonly” because, while the distributional forces I describe tend to drive the patterns of changes in those measures in the direction indicated, there is little reason to expect that it will typically be the case that such pattern will be observed as to every one of a substantial number of observations. Other forces are invariably involved. And in many situations, the diminishing in the strength of the forces causing rates to differ reflected by the decrease of the EES in nine cases in the Escarce and McGuire study could be substantial enough to cause the relative difference in the adverse outcome or the absolute difference to change in a direction contrary to that in which the distributional forces would tend to drive it.

³⁷ These responded respectively to (1) Schneider EC, Cleary PD, Zaslavsky AM, Epstein AM. Racial disparity in influenza vaccination: Does managed care narrow the gap between blacks and whites? JAMA 2001;286:1455-

differences between rates to interpret various patterns of differences between outcome rates. Some of the comments are rather involved because they attempt to interpret the data in the studies by taking prevalence issues into account (something that could be done far more efficiently by means of the method described at the end of Section D, but which I had not discovered as of the time of the comments). But all explain succinctly enough that the authors' efforts to draw conclusions of the types suggested in the study titles on the basis of absolute differences between rates were undermined by a failure to recognize the patterns described here. For the studies involved situations where the rates at issue were changing over time, or differed from setting to setting or with respect to the outcomes being examined, and the authors appraised the size of the disparities without any consideration of the way absolute differences tend to be affected by the rate ranges at issue in each situation being examined. Though the authors may have been broadly correct in many cases, as suggested previously, their quantification of differences in the circumstances reflected by the outcome rates examined could not have been and the implied representations as to the soundness of absolute differences between rates to quantify those differences in circumstances were misleading.

One of the articles warrants further discussion because it was part of a prominent series of articles that together reflect the broader disarray in health disparities research and because an exchange following it revealed some of the thinking underlying the measurement approaches of healthcare disparities researchers at HCPD. The 2005 article by Trivedi et al. was one of a series of three special articles on healthcare disparities in an August 18, 2005 issue of the *New England Journal of Medicine (NEJM)*, which also contained a commentary on the three articles. A study by Vaccarino *et al.*³⁸ relied on relative differences in favorable healthcare outcomes (though relative differences in adverse outcomes for health status issues) with regard to outcome rates that were not changing much in overall prevalence during the period examined; and, as commonly happens when overall prevalence does not change much, the study found little to remark on with respect to changes over time. A study by Jha *et al.*³⁹ relied on absolute differences between rate in examining disparities in rates of receiving certain fairly uncommon procedures that were generally increasing in overall prevalence; and, as commonly happens when outcome rates in the ranges at issue are generally increasing, the authors usually found increasing disparities. The study by Trivedi *et al.* relied on absolute differences between rates in examining adequacy of care (which included both treatment and control of conditions) where

1460; (2) Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Trends in the quality of care and racial disparities in Medicare managed care. *N Engl J Med* 2005;353:692-700; (3) Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Relationship between quality of care and racial disparities in Medicare health plans. *JAMA* 2006;296:1998-2004; (4) Sequist TD, Adams AS, Zhang F, Ross-Degnan D, Ayanian JZ. The effect of quality improvement on racial disparities in diabetes care. *Arch Intern Med* 2006;166:675-681; (5) McWilliams JM, Meara E., Zaslavsky AM, Ayanian JZ. Differences in control of cardiovascular disease and diabetes by race, ethnicity, and education, U.S. trends from 1999 to 2006 and effects of Medicare coverage. *Ann Int Med* 2009;150:505-515.

³⁸ Vaccarino V, Rathore SS, Wenger NK, et al. Sex and racial differences in the management of acute myocardial infarction, 1994 through 2002. *N Engl J Med* 2005;353:671-682.

³⁹ Jha AK, Fisher ES, Li Z, Orav EJ, Epstein AM. Racial trends in the use of major procedures among the elderly. *N Engl J Med* 2005;353:683-691.

adequacy of care rates (especially as to treatment) were at generally high levels and increasing; and, as commonly happens in such circumstances, the authors found absolute difference between rates usually to be decreasing (especially as to treatment).⁴⁰ A commentary⁴¹ discussed the various findings and their perceived implications and stressed the need for more health disparities research and action to reduce such disparities. As was common in 2005, as it is now, neither the commentary nor any of the articles mentioned anything about the way different measures might yield different conclusions or the way any measure might be affected by general changes in the outcome being examined.

A number of letters were published in response to the series.⁴² Only Keppel et al.⁴³ directly addressed measurement issues. The letter, by three authors of the NCHS 2005 guide recommending that all disparities be measured in terms of relative differences in adverse outcomes, presented an elaborate table to show, *inter alia*, that in four of the cases where Trivedi et al. found decreasing disparities, measurement in terms of relative differences in adverse outcomes would have shown increasing disparities. It urged greater consistency in the reporting of disparities. The letter did not discuss measurement in terms of relative differences in favorable healthcare outcomes, the approach used by Vaccarino et al. and still a common approach in 2005 notwithstanding the NCHS recommendation of a year earlier. Such an approach would have yielded a conclusion consistent with that yielded by the absolute difference in all four cases. As discussed above, when one relative difference shows a direction of change contrary to that shown by the absolute difference, the other relative difference will necessarily show a direction of change that is consistent with that shown by the absolute difference.

Three authors of the Trivedi *NEJM* article responded to the Keppel letter.⁴⁴ While noting that relative and absolute differences might yield different results, they stated that they “would not recommend a single approach to reporting disparities for all measures of healthcare disparities and utilization.” The Trivedi *NEJM* responders defended reliance on absolute differences in their study on the basis (1) that absolute differences showed the “percentage of eligible health-plan enrollees from the underserved groups who would benefit from the elimination of a disparity” and that such information would be more useful to clinicians and managers establishing healthcare priorities than information on relative differences and (2) that the

⁴⁰ See [Comment on Trivedi JAMA 2006](#) regarding the authors’ later effort to explain different patterns as to treatment and control, making very reasonable points in doing so, but without consideration of the generally lower rates of control compared with treatment.

⁴¹ Lurie N. Health disparities – Less talk. more action. *N Engl J Med* 2005;353:727-729.

⁴² Trends in racial disparities in care. *N Engl J Med* 2005;353:2081-2084.

⁴³ Keppel KG, Percy JN, Weissman JS. Untitled. *N Engl J Med* 2005;353:2082-2083.

⁴⁴ Trivedi AN, Zaslavsky AM, Ayanian JZ. Untitled letter. *N Engl J Med* 2005;353:2085.

absolute differences would yield a consistent result regardless of whether the favorable or adverse outcome was examined while the relative difference would not.⁴⁵

The Trivedi NEJM response also noted that changes in absolute differences and relative differences could have “sharply divergent policy implications depending on the proportions being compared,” pointing out that that “relative disparities are greatly magnified when adherence (or nonadherence) approaches 0 or 100 percent,” and presented the following example:

Consider, for example, hypothetical decreases in absolute nonadherence from 90 percent to 2 percent among blacks and from 60 percent to 1 percent among whites. Such changes would represent a substantial decrease in absolute disparity from 30 percent to 1 percent but a paradoxical increase in the corresponding relative disparity from 1.5 to 2.0.

The Trivedi NEJM response concluded:

For these reasons, rather than adopt a uniform method of reporting, the decision to report absolute disparities, relative disparities, or both should depend on the purpose of the analysis and the reasoned judgment of the investigators.

The practical importance of the absolute difference cited in the Trivedi response is a matter of genuine relevance and administrators might well focus on eliminating disparities in outcomes involving the largest absolute differences. But efforts to appraise success in correcting aspects of healthcare systems that cause a disparity can rarely if ever be of value if undertaken without consideration of the way absolute differences between rates are affected by changes in overall prevalence of the outcome in the rate ranges at issue. Otherwise, among many other things of comparable incongruity, higher authorities might sanction (or terminate) administrators who are successful at increasing rates of uncommon favorable healthcare outcomes, including such things as survival for those cancers where survival remains the exception (with corresponding increases in absolute differences) and reward administrators who are successful in increasing rates of common favorable outcomes such as generally appropriate care (with corresponding decreases in absolute differences), even though a sound measure might find that disparities actually decreased under the former administrator and increased under the latter administrator. And I have already shown in Section C the way that the failure to consider the effects of overall prevalence on absolute differences led Massachusetts to unwisely include a health disparities measure in its P4P program and to do so in a manner that is more likely to increase healthcare disparities than decrease them.

Yet, while the reference to magnified relative differences when a favorable or adverse outcome is rare possibly suggests an understanding of the way the two relative differences will tend to

⁴⁵ The responders noted that one of the instances where Keppel et al. found an increased disparity based on the relative difference in the adverse outcome there was a decrease in the relative difference in the favorable outcome. It is unfortunate that the responders did not point out that such was the case as to all four instances cited by Keppel et al. For that might have provoked some thought among readers.

change in opposite directions as prevalence changes, nothing in the response suggests a recognition that absolute differences will change as prevalence changes.⁴⁶ Nor is there any recognition that, if a measure is changing because prevalence changes, the change in the measure cannot accurately capture changes in the forces underlying the difference.

In any case, all of the referenced studies by HCPD relying on absolute differences were aimed at identifying the strength of forces causing rates to differ, at different points in time, in different settings, or with regard to particular kinds of outcomes. That includes the Trivedi *JAMA* 2006 study's efforts to explain differences in patterns in Trivedi *NEJM* 2005 as to treatment measures and control measures. See [Comment on Trivedi JAMA 2006](#). Such efforts can only be useful when undertaken with a recognition of the way absolute differences tend to be affected by the overall prevalence of an outcome in the rate ranges at issue, and those studies fail even to show a recognition that absolute differences might in some manner be affected by the rate ranges at issue.

Finally, even though each of the of the referenced studies in fact sought to determine the comparative size of forces causing disparities (though unarmed with recognition of the correlations of absolute differences with rate levels), the Trivedi *NEJM* response seems to reflect that same failure to recognize that there can be only one underlying reality that was addressed in Section D. The concluding point of the Trivedi response no doubt reflects the thinking of many health and healthcare disparities researchers, which thinking is much in evidence in the *Commissioned Paper* discussed in Section E.2.c. But as shown in Section D, there can be no reasoned judgment of health disparities investigators without recognition both of the fact that there exists only one underlying reality and of the necessity of considering the way a particular measure tends to be affected by the prevalence of an outcome when employing the measure to identify that reality.

b. Harvard School of Public Health

Researchers at the Harvard School of Public Health (HSPH) principally rely on relative differences between rates – in adverse outcome for health status, and, so far as I can tell, in favorable outcomes for receipt of healthcare. But, they do so without regard to the way such measures tend to be affected by the prevalence of an outcome and, so far as I am aware, without discussion of the possibility to reach different conclusions depending on the measure employed.

I have already made reference to the [Comment on Kawachi Health Affairs 2005](#) and [Comment on Thurston Am J Epidemiol 2005](#) with respect to issues involving perceptions about the interaction of factors like socioeconomic status with group memberships regarding some type of

⁴⁶ The quoted example might imply that as favorable outcomes increase (absolute differences decline) absolute differences will get smaller. But that is only the way absolute differences tend to change as the favorable outcome increases when the favorable outcome starts at a high level (unfavorable at a low level) level. The example presents a situation where the march from the situation of uncommon favorable (common adverse) outcome to a situation of very common favorable (very uncommon adverse) outcome typically would involve first an increase in the absolute difference and then a decrease in the absolute difference.

adverse outcome disparity. The study that is the subject of the first comment⁴⁷ pertained to the importance of considering both race and class in examination of health disparities issues. That would seem indeed important. But when the subject turns to appraisal of the implications of interactions and why a relative difference is greater within one subpopulation than another, the patterns described here become implicated and sound appraisals of that nature are impossible without consideration of those patterns. After pointing out a situation where the relative differences in adverse infant health outcome increase with parental education, the authors observe that the pattern is “incompletely understood.” It is fairer to say that it is not understood at all and will not be understood without recognition that among groups where adverse outcomes are rare, relative differences between rates of experiencing those outcomes tend generally to be large, while relative differences in favorable outcomes tend generally to be small.

The study by Thurston et al⁴⁸ attempted to determine whether the socioeconomic gradient in coronary heart disease is stronger in women than men. For reasons explained above and in the comment on the study, such an inquiry is futile without consideration of the different underlying rates and the reason to believe that the socioeconomic gradient in the adverse outcome would be greater among women (where the favorable outcome is rarer) while the socioeconomic gradient in the favorable outcome would be greater among men (where the favorable outcome is rarer). I merely note that the 2007 comment was written before I created the [Subgroup Effects](#) subpage treating the issue at greater depth (or the [Illogical Premises](#) and [Illogical Premises II](#) subpages of SR that suggest that, irrespective of the distributional forces I describe, any expectation of equal gradients as to one outcome for groups with different baseline rate of experiencing that outcome is illogical as well as incorrect). See also Table 1 of the [Framingham Illustrations](#) subpage of SR, which shows how increases in high-density lipoprotein reduces heart attack risk more among women than men while increasing the chance of avoiding health attack more among men than among women. Similarly, the [Life Table Information Document](#) shows how aging from age 40 to 60, for example, causes a larger proportionate increase in mortality for white women than white men, but causes a larger proportionate decreases in survival for white men than white women.

The [Comment on Baicker Health Affairs 2004](#) involves a study⁴⁹ where the principal author was from HSPH, and where the study mainly relied on relative differences in favorable outcomes in measuring healthcare disparities. The study did so in the context of appraising the comparative size of healthcare disparities from place to place. But it also endeavored to measure correlations between white and black rates in an area with absolute differences between black and white rates in the area. The comment is lengthy and probably does as good a job of explaining the issues as this letter. But in simplified summary, the effort to measure disparities from place to place was

⁴⁷ Kawachi I, Daniels N, Robinson DE. Health disparities by race and class: Why both matter. *Health Affairs* 2005;24(2): 343-352.

⁴⁸ Thurston RC, Kubzansky LD, Kawachi I, Berkman LF. Is the association between socioeconomic position and coronary heart disease stronger in women than in men. *Am j Epidemiol* 2005;162:57-64.

⁴⁹ Baicker K, Chandra A, Skinner JS, Wennberg JE. Who you are and where you live: how race and geography affect the treatment of Medicare beneficiaries. *Health Affairs* 2004:Var-33-Var-44.

problematic for failure to recognize the way that relative differences in favorable outcomes would tend to be smaller where the outcomes are more common. For reasons suggested above, NCHS would commonly reach contrary conclusions, or, at least, the statistical forces I describe here are of a nature to cause NCHS commonly to reach different results from those of the study.

The effort to appraise the correlations of black and white rates with absolute differences between rates – as with the works of the HCPD – was problematic for failure to recognize that for procedures that were relatively uncommon, higher rates for blacks and whites would tend to be associated with larger absolute differences between black and white rates, while for procedures that were relatively common higher rates for blacks and whites would tend to be associated with smaller absolute differences between black and white rates. The comment also explains that, as is inferable from the discussion above, black and white favorable outcome rate rates would tend to be inversely correlated with relative differences in favorable outcomes (*i.e.*, higher black and white favorable outcome rate would tend to be associated with smaller relative differences in favorable outcomes) while black and white favorable outcome rates would tend to be directly correlated with relative differences in adverse outcomes (*i.e.*, higher black and white rate favorable outcome rates would tend to be associated with larger relative differences in adverse outcomes).

One recent study by researchers at HSPH warrants mention because it illustrates the way mistaken perceptions about the comparative size of relative differences can influence interpretations of matters beyond those specifically involved in the data underlying the perceptions. I have already mentioned that observers have found significance in the fact that mortgage rejection rate disparities are commonly larger among higher income applicants than lower income applicants. They have done so not to suggest that rejection rate disparities among higher income borrowers warrant special attention but rather in order to refute contentions pertaining to all income levels that unaccounted for income differences might explain observed rejection rate disparities.⁵⁰ I suspect that similar conclusions about a range of things are at times based on perceptions about effects of socioeconomic status on different racial groups, as suggested in “Race and Mortality” and “The Perils of Provocative Statistics.”

In any case, an HSPH study in a September 2012 issue of the *New England Journal of Medicine* explored the effects of expanded Medicaid in a number of states on mortality rates in those states, finding that the expansion appeared to reduce mortality among adults under 65.⁵¹ I have not sufficiently examined the study to form a view as to the validity of its general methodology or findings. I did notice, however, that as a basis for maintaining that observed mortality reductions among the populations under 65 were unlikely to be a solely a reflection of general declines in mortality in the expansion states, the authors noted that reductions in mortality were greater among adults aged 20 to 64 than among those over 65. But there is reason to expect that

⁵⁰ As discussed on the [Disparities – High Income](#) of my [Lending Disparities](#) page the argument would not be sound even if views about the size of disparities were correct. But such fact is not germane to the instant point.

⁵¹ Sommers BD, Baicker K, Epstein E. Mortality and access to care among adults after state Medicaid expansions. *N Engl J Med* 2012;367:1025-34.

general reductions in mortality will commonly result in larger proportionate reductions in mortality among younger adults than older adults, along with larger proportionate increases in survival among older adults than younger adults, simply because baseline mortality rates are lower among younger adults than older adults. The pattern cited by the authors is the same as that underlying common observations to the effect that some demographic disparity is greater among the young than the old – which merely means that the factor defining advantaged status (whether the factor involves race, socioeconomic status, body mass index, or home ownership) reduces mortality proportionately more among the young than the old – but without consideration of the distributional forces driving such pattern or recognition that the factor tends to increase survival proportionately more among the old than the young. Thus, the observed pattern cannot support the authors' general finding unless evaluated with regard to the statistical patterns described here.

An illuminating example of the way a perception about relative differences has been deemed to support a broad interpretation of a subject, but only when considered without an understanding of the patterns described here, may be found in the commentary on two 1991 *New England Journal of Medicine* studies on gender differences in receipt of various cardiac diagnostics or treatments, which studies were the subject of considerable media attention.⁵² The authors of both studies left open the question of whether gender differences they found were the results of gender bias. But the journal published with the studies a commentary by the director of the National Institutes of Health (NIH) that, focusing on relative differences in favorable outcomes, emphasized that gender differences in rates of receipt of a particular type of care were smaller among persons who had stronger symptoms of heart disease.⁵³ One finds similar reasoning where employment discrimination is at issue and observers point out that relative difference in selection rates are smaller among applicants with experience or credentials than applicants without experience or credentials (such as would be observed in the data in Table IV of the article discussed in note 30). Such reasoning is that biased decision-makers will more commonly make objective decisions when there exist objective indicators of the comparative situation of two groups than when such indicators are absent, in which case the decision-maker is more likely to rely on (accurate or inaccurate) perceptions about differences in average characteristics of the groups, and hence that the existence of a pattern of smaller differences in the former situation than the latter suggest the likelihood that bias exists, though it may operate with greater force in the latter than the former situation. Such reasoning is plausible in these and other contexts assuming that the measure is accurately comparing the strength of the forces causing the observed differences in the two situations.⁵⁴ But those basing their reasoning on the smaller

⁵² The studies were (1) Ayanian JZ, Epstein AM. Differences in the use of procedures between men and women hospitalized for coronary artery disease. *N Engl J Med* 1991; 325:221-5 and (2) Steingart RM, Packer M, Hamm P, et al. Sex differences in the management of coronary artery disease. *N Engl J Med* 1991; 325:226-30. .

⁵³ Healy B. The yentl syndrome. *N Engl J Med* 1991; 325:274-75.

⁵⁴ The implications, however, differ by context. It is unlawful for an employer to base judgments on perceptions about group averages. A physician, however, would be negligent in failing to consider certain gender differences in susceptibilities (such as those that are incorporated into online Framingham calculators) in making treatment judgments respecting heart disease. More difficult issues arise with respect to such things as the consideration of group characteristics concerning adherence to a particular therapeutic regimen, where most would regard

relative difference in some favorable outcome among categories of persons that have generally higher rates of experiencing the outcome are overlooking that, for reasons unrelated to any difference in the strengths of those forces, one commonly finds smaller relative differences in experiencing an outcome where the outcome is common, but also larger relative differences in failing to experience the outcome in those circumstances.

With respect to the manner in which NIH director drew certain flawed inferences on the basis of perceptions concerning relative differences in favorable outcome rates, it warrants note, that under the NCHS approach of recent years, one would be interpreting the comparative size of the relative differences in exactly the opposite way. That is, focusing on relative differences in failure to receive a type of care, one would find that the gender disparity is larger among persons with the stronger symptoms. And observers might be prompted to make a provocative point to the effect that, among patients most in need of a particular type of care, the disparities are even larger than they are among those who need the care less – just as points along the same line are commonly made about large rejection rates disparities among high income mortgage applicants or about large relative differences in adverse outcome rates regarding a variety of health and healthcare issues among any subpopulation that has generally low rates of experiencing such outcomes.

Thus, while perceptions about the comparative size of relative differences in favorable outcome rates in various settings or the comparative size of relative differences in adverse outcome rates in various settings underlie many common perceptions about broad issues that are frequently debated, it is doubtful that a statistically sound statement has ever been made about such a matter. The same, of course, holds when the interpretation concerning the comparative size of relative differences is itself the principal issue.

Save for the recent *New England Journal of Medicine* article by researchers at HSPH that prompted the discussion in the immediately preceding paragraphs, the Harvard studies mentioned above simply happen to be those on which I posted comments between 2007 and 2009. There is otherwise nothing special about them, save that some of them were quite prominent and continue to be frequently cited. They are discussed here merely to show the inconsistencies of the methods by which various arms at Harvard analyze differences in outcome rates pertaining to health and healthcare disparities, and, more important, that none of those methods is sound.

consideration of socioeconomic status or race to be improper, and where such consideration could well be unlawful when laws proscribe consideration of certain demographic characteristics with respect to allocation of medical care. In any context, however, a common difficulty in the interpretation of statistical disparities is that the more sound is the basis for a perception about differences in group averages as to pertinent characteristics, the more likely it is both (a) that decision-makers will rely on such a perception and (b) that difference in group characteristics may explain an observed difference. That is, for example, a situation where physicians are influenced by socioeconomic status in making decisions about adherence and a situation where physicians make detailed inquiries into the objective factors associated with adherence may result in the same socioeconomic differences in treatment recommendations, even though rates of correct decisions will be higher in the latter situation than in the former.

c. The Commissioned Paper: Healthcare Disparities Measurement

I mentioned at the outset the [Commissioned Paper: Healthcare Disparities Measurement](#), a joint project of Harvard Medical School and Massachusetts General Hospital, which was sponsored by the Robert Wood Johnson Foundation and the National Quality Forum. It was released in final form in November 2011. The document provides the reader no basis to believe that any measure of health disparities might be affected by the prevalence of an outcome in the manner explained in this letter or in any other way.

The document first came to my attention in draft form in early September 2011, a few days after the close of the period for public comment. I nevertheless submitted comments by email directly to the authors and they included the comments among an online [collection](#) of timely submitted comments. The comments in the form of the email I sent, including the active links it contained to referenced materials, but also with the authors' responses to the comments, may be found [here](#).

The comments, which are essentially a more succinct version of this letter and which can largely speak for themselves, urged the authors not publish the document without making the substantial revisions necessary to address the way the measures it discussed tended to systematically be affected by the prevalence of an outcome. The authors responded by adding the eight words I have italicized in the quotation below to the introductory sentence of a Section 4.c, which section is styled "Absolute Versus Relative Differences and Favorable Versus Adverse Outcomes."

As modified that sentence read:

While calculations of disparities can be straightforward, comparisons of disparities among entities or over time can be sensitive to the calculations chosen, *especially when the prevalence of an outcome changes.*

Whatever the added language might suggest to the reader, it obviously does not alert the reader to the fact that the key measures discussed in the paper tend to change in certain ways when prevalence changes, much less to the issue of the ways such fact calls into question the validity of each measure commonly used. Rather the document continues to give advice on the measurement of health disparities of the same type that has caused virtually all attempts in health and healthcare disparities research to appraise the size of the difference in circumstances reflected by a pair of rates to be fundamentally flawed.

Thus, assuming there is validity to the points raised in this letter, the document, as amended and then published, will mislead the public, policy makers and other researchers, while causing the waste of substantial resources. The document is also likely to cause the production of incorrect or misleading findings, including in circumstances where such findings, as in the Massachusetts pay-for-performance program, may lead to increased healthcare disparities. And the document will do these things to a greater degree than other documents of the same nature that have been issued in the past. For it is the most current extended explication of disparities measurement

issues and it bears the name of Harvard Medical School, along with three other prestigious entities. Thus, I suggest, the responsible course for the University is, after consultation with those entities, to withdraw the document.

Further, the University should do that immediately. In a just a few months, such a document may influence scores of readers with regard to how to undertake health or healthcare disparities research. A guidance document of this nature ought not to be allowed to influence any readers unless that guidance can be defended in every material respect.

If the points I raise here are unfounded, that is quite another matter. But those points are not matters of nuance. If they are correct, the document is certain to cause flawed and misleading research and the University should withdraw it. Assuming those points are correct, the University should also thoroughly review all ongoing health and healthcare disparities research that has been undertaken without consideration of the issues raised here, including that in the process of publication. For publication of research that implicates the issues addressed here, but that fails to address those issues, has potential for the same long term consequences as the several of the studies discussed above that continue to be frequently cited for conclusions that may be unfounded and that may be commonly emulated with respect to methods that are unsound.

F. Generally Reviewing the Measurement of Demographic Differences at Harvard

I am somewhat familiar with the work at Harvard on health and healthcare disparities that involves interpretation of differences between outcome rates. I am also somewhat familiar with the work at Harvard concerning subgroup effects as well as with the general nature of research commonly conducted aimed at identifying subgroup effects and with standard guidance on methods for applying a risk reduction observed in a clinical trial to estimate absolute risk reductions in circumstances involving baseline rates other than baseline rates at issue in the trial. I am largely unfamiliar with Harvard's work in other areas where statistical issues addressed here are implicated. But, while there may be exceptions, I assume that in general – save as may have been influenced by a September 20, 2005 [discussion](#) of “Race and Mortality” on the Social Science Statistics Blog or individual contacts with faculty – teaching and research on demographic differences at Harvard is much as one finds in other universities. Thus, I assume that, as seems to be the case around the world, teaching about the “feminization of poverty” or changes in relative poverty of demographic groups is done without consideration of the extent to which observed patterns are functions of changes in overall poverty. I also assume that at the Law School, as at the Departments of Justice and Education, it is not recognized that relaxing lending standards tends to increase relative differences in mortgage rejection rates or that stringent public school discipline standards tend to cause smaller, not larger, racial differences in discipline rates than more lenient policies, and that the misperceptions that generally undermine efforts to appraise demographic differences in the enforcement of civil rights laws and that confound the interpretation of other legal issues involving quantitative reasoning are as common at the Law School as they are elsewhere. In the event that I am substantially mistaken regarding any of these assumptions, that merely means that there exists a body of thought within the University that should be supportive of the sort of review I propose.

In any case, I suggest that Harvard's review of its teaching and research on the measurement of demographic differences should be a university-wide project. Moreover, I suggest, it should be a project of great priority. As already suggested with regard to health and healthcare disparities issues, research conducted without consideration of the statistical patterns described here, not only wastes resources, but contributes to a body of work that is misleading both as to its substance and as to the validity of the indicators employed. Further, each day students make academic and career decisions based on perceptions about the nature of issues discussed in standard research and about the validity of the measures commonly employed in such research. Such decisions necessarily will be ill-informed if reached without an understanding both of the problems with standard measurement tools and of how much research to date has been undertaken without recognition of the shortcoming of those tools. So the harms of deferring the consideration of the issues I raise, or actions based on such consideration, are substantial.

In considering the issues raised in this letter, the University may find it useful to consult with the following faculty members who have some familiarity with those issues.

Christopher Winship, Diker-Tishman Professor of Sociology, and member of the faculty of the Harvard Kennedy School of Government, is the person responsible for my invitation to speak at the Applied Statistics Workshop. It was as a result of exchanges with Professor Winship that my 2000 article "Race and Mortality," mentioned a number of times above, was discussed on the Social Science Statistics Blog by then Ph.D. student Felix Elwert. While I can say that Professor Winship is generally interested in my work on issues addressed in this letter, I will not attempt to describe his particular views.

Alan Zaslavsky, Professor Health Care Policy (Statistics) of the Department of Health Care Policy (HCPD) at the Harvard Medical School was the discussant at a [session](#) I organized to discuss health disparities measurement issues at the 7th International Conference on Health Policy Statistics in 2008. He is generally familiar with the issues I raise. He is also the co-author or a number of the articles by researchers at HCPD that are subjects of my comments mentioned in Section E.2.a and one of the authors of the 2005 response in the *New England Journal of Medicine* articulating HCPD view on the measurement of healthcare disparities that was discussed in that section.

Joel S. Weissman, Associate Professor of Health Care Policy at Harvard Medical School, co-authored the 2005 NCHS measurement guide mentioned above in which NCHS adopted the policy whereby all health and healthcare disparities would be measured in terms of relative differences in adverse outcomes and one of the authors of the letter to the *New England Journal of Medicine* to which the item co-authored by Professor Zaslavsky responded. Professor Weissman is also the principal author of the *Commissioned Paper: Healthcare Disparities Measurement* that, in Section E.2.c above, I suggested the University withdraw, as well as co-author of the subject of the [Comment on Blustein Health Affairs 2011](#) mentioned in connection with the Massachusetts Medicaid pay-for-performance program. Thus, Professor Wiseman should be well acquainted with the issues raised in this letter.

David A. Wise, John F. Stambaugh Professor Political Economy at the Harvard Kennedy School of Government, also warrants special comment. Professor Wise testified as a defense expert

witness in *EEOC v. Sears, Roebuck and Co.*, a nationwide gender discrimination in employment case tried in 1984 and 1985, which has been the subject of considerable scholarly commentary. As counsel for the plaintiff, I examined Professor Wise at length both in deposition and at trial concerning his testimony. As a result of my familiarity with his testimony and the reasoning underlying it, I have some confidence that, on giving thought to the statistical issues addressed in this letter, Professor Wise would agree with the reasoning in the letter in most or all material respects.

Other Harvard faculty members with whom I have corresponded regarding issues addressed in this letter, and who have responded indicating some recognition of the issue raised, include: Stephan Thernstrom, Winthrop Research Professor of History; Norman Daniels, Mary B. Saltonstall Professor of Population Ethics and Professor of Ethics and Population Health, James H. Ware, Frederick Mosteller Professor of Biostatistics, and Rui Wang, Associate Professor of Medicine, at the Harvard School of Public Health; Thomas G. McGuire, Professor of Health Economics, and J. Michael McWilliams, Assistant Professor of Health Care Policy, at the Harvard Medical School; Benjamin Le Cook, Assistant Professor of Psychiatry, Cambridge Health Alliance. While some responses I received from these faculty members may have involved simple courtesy, irrespective of the communications, each of the mentioned persons should be in a position to comment on the issues raised here.

Outside of Harvard, but still in Cambridge, Christopher Zegras, Associate Professor of Planning and Urban Studies at Massachusetts, teaches a course in quantitative reasoning and statistical methods for urban planners that has for some years required the reading of the 1991 *Public Interest* article "The Perils of Provocative Statistics," an article mentioned several times above that addresses key issues raised in this letter. Since the course syllabus describes that article as a strong indictment of conventional social indicators, I assume that Professor Zegras would at least generally agree with much said in this letter.

Other scholars with views on the measurement issues addressed here are discussed in [Section E.7](#) of MHD, which section was mentioned several times above. These include some of the principal European authorities on the measurement of health disparities. As discussed earlier, the most significant of the work addressing interpretations of the type stated in this letter either specifically agrees with it, or, while not necessarily agreeing with the interpretation as to the forces causing observed patterns, similarly concludes that it is not possible to conduct sound research on group differences without taking these patterns into account.

Should there be any question about matters addressed in the letter, the October 17 Applied Statistics Workshop may provide an opportunity to resolve them. I am otherwise available to respond to questions at any time as well.

I do not, however, believe that there is a plausible basis for questioning that standard measures of differences between outcome rates are problematic for appraising the differences in the circumstances of two groups, that reliance on such measures in research without consideration of the ways the measures are affected by overall prevalence is misleading to the public, to policymakers and to other researchers, or that teaching about such measures, or the substantive topics where important research has relied on the methods, without addressing the problematic

Drew Gilpin Faust, Ph.D., President
Harvard University
October 9, 2012
Page 46

nature of the methods, does a signal disservice to students seeking the exceptional education Harvard is believed to provide. And I expect that members of the Harvard faculty who carefully read this letter are more likely to find the points it raises self-evident than to find them incorrect.

Thus, while a review of statistical methods at Harvard of the nature proposed in this letter unquestionably would be a substantial undertaking, such review offers the University an opportunity to assume the leadership in this area that it has so often assumed in others.

Sincerely,

/s/ James P. Scanlan

James P. Scanlan

Correction Record:

Minor corrections were made to this letter on October 10, 2012. Only two warrant mention:

(1) Third last paragraph of Section E.1, line 9: “tend to be larger” changed to “tend to the smaller.”

(2) Second row of Table 4, EES column: “Decrease” changed to “Increase.”]

Minor corrections were made to this letter on October 14, 2012. Only three warrant mention:

(1) Sixth paragraph of Section E.2.b, last line: “though I have not explored that matter” changed to “as suggested in ‘Race and Mortality’ and ‘The Perils of Provocative Statistics’”

(2) Last line of Section E.2.c: the following words added: “and that may be commonly emulated with respect to methods that are unsound.”

(3) Eighth paragraph of Section F, first sentence: “health disparities measurement issues or subgroup effects” changed to “issues addressed in this letter”

Minor corrections were made to this letter on October 15, 2012. None warrants mention.

Minor corrections were made to this letter on November 6, 2012. None warrants mention. The following significant correction made that date warrants mention.

The sentence beginning with “Or” in the last paragraph beginning on page 26 was changed in various ways to correct a misreading of patterns of increasing and decreasing rate ratios in Table 3.

The sentence originally read:

Or, in the situation where the rows reflect changes over time, observers without knowledge of the patterns described here might explore whether the perceived pattern of increasing bias as to selection might have resulted from the replacement of a less biased with a more biased hiring official or whether the perceived pattern of decreasing bias as to rejection might have resulted from the replacement of a more biased with a less biased hiring official.

It was changed to read:

Or, in the situation where the rows reflect changes over time, observers without knowledge of the patterns described here might explore whether the perceived pattern of decreasing bias as to selection might have resulted from the replacement of a more biased with a less biased hiring official or whether the perceived pattern of increasing bias as to rejection might have resulted from the replacement of a less biased with a more biased hiring official.

On November 22, 2012, were added on page 11 and in the title of Table 1 in note 10 at page 12 to clarify the gender comparisons in the Sheps article involved whites men and women.

Minor corrections were made to this letter on December 8, 2012. None warrants mention.

On April 19, 2014, in the fifth line of the first full paragraph on page 32, the date “August 18, 1995” was changed to “August 18, 2005.”

ⁱ As discussed on the home page of jpscanlan.com, this letter, commonly referred to as the Harvard University Measurement Letter, contains the most comprehensive discussion in a single document of the ideas addressed on the [Measuring Health Disparities](#), [Scanlan’s Rule](#), [Mortality and Survival](#), [Measures of Association](#), [Lending Disparities](#), [Immunization Disparities](#), [Educational Disparities](#), [Discipline Disparities](#), [Feminization of Poverty](#), and [Disparate Impact](#) pages (and their several score subpages) of jpscanlan.com. In order that the letter may continue to serve as a somewhat comprehensive guide to the issues addressed on those pages, including discussion of such subsequent developments as further correspondence to Harvard University (which, as of November 17, 2012 is comprised of the October 26, 2012 [Harvard et al. Commissioned Paper Letter](#), and which likely will eventually also include at least a letter regarding the subject of the [Subgroup Effects](#) sub-page of the [Scanlan’s Rule](#) page), the letter will be periodically annotated by means of endnotes.

ⁱⁱ The PowerPoint presentation used at the workshop is available [here](#).

ⁱⁱⁱ By [letter](#) of October 26, 2012 to the presidents of Harvard University, National Quality Forum, Robert Wood Johnson Foundation, and Massachusetts General Hospital and the Dean of the Harvard Medical School, I addressed with all entities responsible for the *Commissioned Paper* the reasons why the document should be withdrawn. The response from the Research Integrity Officers of Harvard Medical School and Massachusetts General Hospital is discussed in prefatory note 9 to the [Scanlan’s Rule](#) page and, as discussed in the text box on the first page, in in Section C.1.g, at pages 30-32 of my Federal Committee on Statistical Methodology 2013 Research Conference paper titled “[Measuring Health and Healthcare Disparities](#).”

^{iv} The statement in the above paragraph that in the NHDR AHRQ measures disparities in terms of the larger relative difference is incorrect. While it relied on the larger relative difference to determine whether a disparity is important, it has intended, consistent with Health People 2010/2010, to measure disparities in terms of relative differences in adverse outcomes. See my March 17, 2014 BMJ comment titled “[The need for new thinking about how to measure disparities II](#).” (This note was added on April 19, 2014.)

^v Rows 1 to 3 of Table EN1 below replicate the data in Table 1 of the letter (at 12 n.10) and rows 4 to 6 present the same type of information based on 2006 life tables underlying the [Life Table Information Document](#). The more recent data do not contain the anomaly found in the third data column of row 2 (highlighted) where the male/female mortality ratio decreased rather, rather than increased, as mortality became less common.

Table EN1. White Male and Female Survival Percentages with Mortality Ratios and Survival Ratios (from Sheps NEJM 1958 and 2006 Life Tables)

Row	Source	Age	Percent Male Surv	Percent Female Surv	M/F Mort Ratio	F/M Survival Ratio
1	Sheps	Birth to Age 40	92.10%	95.10%	1.61	1.03
2	Sheps	Age 40 to Age 60	81.80%	90.30%	1.88	1.10
3	Sheps	Age 60 to Age 80	33.70%	49.50%	1.31	1.47
4	2006 Life Tables	Birth to Age 40	95.85%	97.81%	1.89	1.02

5	2006 Life Tables	Age 40 to Age 60	89.77%	93.87%	1.67	1.05
6	2006 Life Tables	Age 60 to Age 80	55.87%	67.63%	1.36	1.21

^{vi} The following potential problem with this approach in the employment context, and possibly certain other contexts, warrants mention. The problem may best be illustrated with respect to Employer A in Table 3, the employer with the lowest selection rates, which involves a situation where the issue is probably more likely to be of consequence than in the case of any of the other three employers. Assume that the selection rates in the table are based on an examination of the employer's applications. But suppose that, given the high number of applications for the limited number of positions, the employer in fact only examined half the applications (and that the decisions as to which applications to examine were entirely random.) That would mean that among applications to which the employer gave actual attention, the selection rates 40.2% for AG and 18.0% for DG. These are the pertinent figures for determining the strength of the forces causing rates to differ (regardless of the extent to which those forces may be comprised of bias or differences in qualification). For these selection rates, the approach described in Section D would yield an EES of .67 standard deviations rather than the .50 standard deviations underlying the illustration. This issue is akin to that addressed in the [Irreducible Minimums](#) subpage of the [Measuring Health Disparities](#) page with respect to the measurement of health disparities.

The issue would seem unlikely to have any bearing on analyses of lending disparities where all applicants are examined. But it may have considerable implications for tester studies (depending on the natures of the study), some of which are explored in my "[Measuring Hiring Discrimination](#)" (*Labor Law Journal*, July 1993), though at a time when my thinking on measurement issues was less developed than it is today.

This endnote was originally added on added on May 15, 2013. In a September 20, 2013 University of Kansas Law School Faculty Workshop Paper titled "The Mismeasure of Discrimination" (at 21), I further developed the issue concerning unreviewed applications. (This note was amended on April 20, 2014.)