

Illustrations of Scanlan's Rule(s)
(Aug. 27, 2008; rev. May 3, 2012)

[This is a PDF version of the [Scanlan's Rule](#) page of jpscanlan.com with notes as footnotes rather than endnotes.]

Prefatory notes:

1. This page was originally created in August 2008, after Bauld et al.¹ termed 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 "Scanlan's rule." In doing so, the authors noted that if governments fail to consider the rule, "they run the risk of guaranteeing failure, largely for conceptual and methodological reasons rather than social welfare reasons." The caution was not as ominous as it might sound, since the referenced failure solely involves the meeting of health disparities reduction goals set in relative terms. On the other hand, in the law as well as the social and medical sciences vast resources have been devoted to efforts to appraise the size of differences between outcome rates. And such efforts can yield little of value without consideration of the ways relative differences in experiencing and avoiding an outcome and other standard measures of differences between outcome rates are affected by the overall prevalence of an outcome. Subsequent to its creation, this page has been periodically updated and a variety of pages or sub-pages have been added to address related issues, some of which are treated in more abbreviated form below. An outline to this page and its subpages may be found [here](#) (though sub-pages are added frequently enough that the outline often may be out of date).

2. In the introduction below, the detailed descriptions of the patterns by which standard measures of difference tend to be affected by the overall prevalence of an outcome are based on hypothetical test score data, which approach allows examination of the patterns in their most essential form. Illustrations of the patterns based on actual data may be found in the sub-pages to this page styled [Framingham Illustrations](#), [Life Tables Illustrations](#), [NHANES Illustrations](#), and [Credit Score Illustrations](#), as well as in some of the tables on the [Mortality and Survival](#) page and the table and figures in "[Can We Actually Measure Health Disparities?](#)" (Chance 2006)

3. Most of the above-referenced illustrations and the discussion below involve measuring differences in the rates at which two groups experience in outcome. For example, suppose that in Year 1 the rates of experiencing some adverse outcome are 21.7% for a disadvantaged group (DG) and 10% for an advantaged group (AG) and in Year 2 those

¹ Bauld L, Day P, Judge K. Off target: A critical review of setting goals for reducing health inequalities in the United Kingdom. *Int J Health Serv* 2008;38(3):439-454. (Britannica online excerpt available at: <http://www.britannica.com/bps/additionalcontent/18/33140632/Off-Target-A-Critical-Review-of-Setting-Goals-for-Reducing-Health-Inequalities-in-the-United-Kingdom>)

rates are 12.7% and 5%. The question to be answered is whether the difference between the circumstances of the two groups regarding the outcome is greater in Year 1 or in Year 2. The question will be answered differently by those who rely on relative differences in experiencing an outcome and those who rely on relative differences in avoiding the outcome. But actually the differences are the same, each reflecting a situation where the difference between means of the underlying distributions is half a standard deviation.

The same figures could be analyzed in terms of whether the change for DG (from 21.7% to 12.5%) is larger than the change for AG (from 10% to 5%). Again, the question would be answered differently by those who rely on relative changes in rates of experiencing an outcome and those who rely on relative changes in rates of avoiding the outcome. And, again, in fact the changes are the same. Each reflects a shift in the underlying distribution of .36 standard deviations. This perspective may have more important practical implication than the perspective discussed in the prior paragraph, for it involves the interpretation of subgroup effects/differential effects/interaction/effect heterogeneity, which can influence decisions about how individual patients are to be treated. See the [Subgroup Effects](#) sub-page of this page.

4. An initial focus of this page and the related [Measuring Health Disparities](#) page (MHD) and the subsequently-created [Mortality and Survival](#) page involved the measurement of health and healthcare disparities and whether those disparities are changing over time or are otherwise larger in one setting than another. While vast resources are devoted to the study of such issues – and the overwhelming majority of such research is problematic for failing to recognize the statistical forces described here – for a time, such study was largely academic. That is, conclusions did not often affect policy. That changed when there arose a belief in the United States that pay-for-performance programs might exacerbate health inequalities and it was suggested that an institution's performance regarding health and healthcare disparities should be included in pay-for-performance programs. See the [Pay for Performance](#) and the [Between Group Variance](#) sub-pages of MHD.

5. As suggested in note 3 *supra*, another area where an understanding of these issues has important practical implications involves the interpretation of clinical outcomes with regard to identification of subgroup effects and, more important, estimating absolute risk reductions across a range of baseline rates based on an observed risk reduction as to one baseline rate. Thus, the most important treatment of the patterns described on this page may be found on the [Subgroup Effects](#) sub-page mentioned above. These [Illogical Premises](#), [Illogical Premises II](#), and [Inevitability of Interaction](#) sub-pages, which are related to the [Subgroup Effects](#) page, address why is it not merely incorrect to regard the rate ratio as a measure of association, but illogical to do so.

6. While most of the attention I have devoted to this subject in recent years has involved health or healthcare issues, the points discussed here have comparable significance in the law and the social sciences (as reflected in Section B *infra*.) Two matters recently in the news (as of the May 2, 2012 updating of this item) illustrate the importance of the

issues in non-medical contexts as well as how poorly the issues are understood by the United States Government in matters involving law enforcement. First, in December 2011, the Department of Justice announced a \$335 million settlement of a lending discrimination suit against Bank of America's Countrywide Financial unit. That matter fits into a pattern whereby since 1994, out of concern about large relative (racial) differences in adverse lending outcomes, the government has been encouraging lenders to reduce the frequency of such outcomes. In March 2012 the Department of Education released data showing large relative (racial) differences in public school discipline rates. Most observers attributed the disparities to stringent discipline policies in effect in recent decades, as the Department of Education and Department of Justice had done previously when, in July 2011, they jointly created the [Supportive School Discipline Initiative](#) to encourage public schools to relax disciplines. As should already be made clear from the discussion above, however, relaxing of lending criteria tends to increase, not reduce, relative differences in adverse lending outcomes and stringent discipline policies lead to smaller, not larger, relative differences in discipline rates. Thus, important federal law regulatory and law enforcement policies are based on an understanding of data that is the exact opposite of the reality. See the [Lending Disparities](#) and [Discipline Disparities](#) pages of this site. (Section A.6 *infra* discusses the misunderstandings of these issues by the government agencies involved with health policy.).

7. This page provides important information regarding the forces underlying the patterns of differences in rates that we observe in reality and the scope of the misinterpretations of such differences due to the near universal failure to understand these forces. The crucial point, however, is simply that all standard measures of differences between outcome rates are problematic for appraising the size of differences between outcome rates or otherwise appraising the strength of association and that the only theoretically sound measure of making such appraisals is the described on the [Solutions](#) sub-page of MHD and applied in the clinical setting on the [Subgroups Effects](#) sub-page to this page (and somewhat described in note 3 *supra*), or some refinement thereon.

8. The original version of this page was about 3200 words and the most recent version is about 14,000 words. The increase reflects the continuing revision of the item to expand on particular points, sometimes as a result of further understanding of a matter or its importance or of learning of the way a matter is treated in the literature. The manner in which this has been done has led to occasional redundancy as to particular points. The increasing length of the item has also caused me to put particularly important points in the prefatory notes, also leading to some redundancy (including within the prefatory section itself). I apologize for the redundancy. But, because sections are sometimes referenced in other materials, sections are never deleted.

Introduction

A great deal of material made accessible through this site (jpscanlan.com) involves patterns whereby measures of differences between rates of experiencing or avoiding an outcome tend to be correlated with the overall prevalence of the outcome. The

[Measuring Health Disparities](#) page (MHD) on this site contains links to something over 150 references concerning these patterns, dating back to 1987, including over 120 on-line comments on studies in medical and health policy journals (which can be separately accessed [here](#)). The most important published articles on these patterns and their implications regarding the interpretation of data on group differences in the law and the social and medical sciences may be found on MHD as items A1 ([Plain Dealer 1987](#)), A3 ([Public Interest 1991](#)), A4 ([Signs 1991](#)), A5 ([Chance 1994](#)), A10 ([Society 2000](#)), and A12 ([Chance 2006](#)). Particularly important conference presentations or on-line comments may be found on MHD as items B7 ([BSPS 2006](#)), B12([APHA 2007](#)), B13 ([ICHPS 2008](#)), B16 ([JSM 2008](#)), B19 ([JSM 2009](#)), D23 ([Comment on Trivedi NEJM 2005](#)), D43 ([Comment on Boström SJPB 2004](#)), D51 ([Comment on Casalino Health Affairs 2007](#)), D52 ([Comment on Morita Pediatrics 2008](#)), D53 ([Comment on Baicker Health Affairs 2004](#)), D72 ([Comment on Eikemo IJEF 2008](#)), D79 ([Comment on Huijts EJPH 2009](#)), D85 ([Comment on Bhopal Emerg Themes Epid 2009](#)), D89 ([Comment on Sun BMJ 2010](#)). A number of recent comments on some older *BMJ* articles (D121 [Comment on Chatellier BMJ 1996](#), D119 [Comment on Cook BMJ 1995](#)) are important because those articles underlie current guidance for adjusting an observed relative risk reduction for baseline risk. See the [Subgroup Effects](#) sub-page

The [Measuring Health Disparities](#) page and its subpages also contain narrative treatments of certain issues, which treatments overlap a good deal with the treatments below and in the sub-pages to this page. [Section E.7](#) of MHD, however, importantly summarizes the extent of scholarly agreement with the thinking in the references made available on that page. Various sub-pages of MHD, including [Reporting Heterogeneity](#) and [Relative Versus Absolute](#), could just as well be sub-pages to this page. Descriptions of the sub-pages of MHD and of this page may be found by means of these links: [MHD Subpages](#) and [Scanlan's Rule Sub-pages](#).

The [Mortality and Survival](#) page addresses the way researchers, particularly in cancer journals, often discuss disparities in mortality and survival interchangeably without recognizing, for example, that as cancer survival increases, relative differences in mortality tend to increase while relative differences in survival tend to decrease or that more survivable cancers will tend to show large relative differences in mortality but small relative differences in survival. The [Measures of Association](#) page simply serves as a reminder that the issues addressed on these pages about measuring differences between outcome rates involve all efforts to measure the strength of an association.

Almost all of this material addresses in some manner the pattern whereby when two groups differ in their susceptibility to an outcome, the rarer the outcome (a) the greater tends to be the relative difference in rates of experiencing it and (b) the smaller tends to be the relative difference in rates of avoiding it. In recent years, for ease of reference, I have sometimes termed this pattern “Heuristic Rule X (HRX)” or “Interpretive Rule 1 (IR1).” In a 2008 article in the *International Journal of Health Services* (discussed in [Section E.7](#) of MHD ([Abstract](#))), researchers from the United Kingdom referred to the pattern as “Scanlan’s rule” and, on this page I shall employ such usage as modified in the following respects.

Most of the referenced works principally concern relative differences between rates of experiencing an outcome, which have tended to be the most common measures of differences between the health status of demographic groups, and which, as suggested in the preceding paragraph, tend to yield contradictory conclusions concerning the comparative size of differences at different points in time depending on whether one examines rates of experiencing an outcome or rates of failing to experience the outcome (e.g., mortality rates or survival rates, mammography rates or rates of failing to receive mammography). The fact that relative differences in experiencing an outcome and relative differences in avoiding an outcome tend to change systematically in opposite directions as the prevalence of the outcome changes might be seen as reason to measure differences between rates in terms of the absolute difference between rates (sometimes termed the “risk difference”) or the odds ratio, since such measures are unaffected by whether one examines an outcome or its opposite. Thus, on occasion over the years, I have pointed out that these measures are also problematic for appraising changes over time, because, like relative differences, each tends to change solely as a result of changes in overall prevalence. See A3 ([Public Interest 1991](#)), A12 ([Chance 2006](#)), B2 ([ICHPR 2001](#)), C1 ([Unpublished 1992](#)). That is, in order for a measure to identify a change over time that reflects something other than the consequence of a change in overall prevalence akin to that resulting from the simple raising or lowering of a cutoff on a test, the measure must remain constant in the face of such changes in overall prevalence. Neither the absolute difference nor the odds ratio satisfies that criterion.

Recently, particularly with regard to the study of healthcare disparities, researchers have increasingly employed the absolute difference as a measure of the size of a disparity. Odds ratios are also increasingly employed in a variety of types of studies, often simply because they can be conveniently derived from logistic regression results.

Because of the increasing use of absolute differences and odds ratios, I have given them a good deal of attention over the last few years, pointing out that (a) like the two relative differences, absolute differences and differences measured by odds ratios tend to change systematically in opposite directions, and (b) they exhibit patterns of changes that are related to the relationship between the two relative differences. Specifically, I have pointed out in various places that, as an outcome changes in prevalence, absolute difference tend to change in the same direction as the smaller of (a) the ratio of advantaged group’s rate of experiencing the favorable outcome to the disadvantaged group’s rate of experiencing that outcome and (b) the ratio of the disadvantaged group’s rate of experiencing the adverse outcome to the advantaged group’s rate of experiencing that outcome (e.g., D23 ([Comment on Vaccarino](#)), B13 ([ICHPS 2008](#))) – a formulation that I shall modify somewhat below. Differences measured in odds ratios tend to change in the opposite direction of the absolute difference

Given the increasing importance of understanding all of these patterns, I shall on this page refer to the patterns concerning relative differences as Scanlan’s Rule 1 (SR1) and the patterns concerning absolute differences and odds ratios as Scanlan’s Rule 2 (SR2). Since the articulation of these rules raises certain semantic issues, these are treated under

the [Semantic Issues](#) sub-page on this page (though the modifications to SR2 discussed below diminish the significance of the material in Section A of that sub-page).

In many prior efforts to explain the ways various measures are correlated with the overall prevalence of an outcome, at least partly to facilitate the understanding of the illustrations, I have described these patterns in the context of changes in overall prevalence of an outcome. Ideally, however, one states a principle in its most encompassing form. That is, to state that the rarer an outcome the greater tends to be the relative difference in experiencing it encompasses both (a) the pattern whereby as an outcome becomes less common relative differences in experiencing it tend to increase and (b) the pattern whereby relative differences between rates of experiencing an outcome tend to be greater in populations where the outcome is rarer. It is easy enough to express SR1 in the more encompassing form, and I do so below.

Because of the greater complexity of SR2, it is more difficult to express that rule without reference to changes over time. So I continue with such approach for SR2 notwithstanding the formal inconsistency with the expression of SR1. Further, whereas SR1 holds regardless of the size of the difference between means, SR2 varies somewhat depending on the size of the difference between means. Therefore, rather than describing such variation simply in terms of large and small differences between means, I set out SR2 in terms of an illustration based on perfectly normal data and particular differences between means. While the discussion in terms of precise differences between means may give some readers pause (since in reality one usually is not dealing with perfectly normal data), an appreciation of the nuances that are related to particular differences between means in perfectly normal data is essential to an understanding of the forces underlying the patterns observed in reality.

Finally, I note that the discussion on this page regarding changes in differences between rates over time pertains to the common situation where the rates of each group change in the same direction – as typically occurs when either rate changes materially. If the two groups' rates change in different directions, there has obviously occurred a change reflecting something other than a change in prevalence. Similarly, in the comparison of two settings defined other than temporally, when one setting has both a higher rate for an outcome in the group with the higher rate of that outcome and a lower rate for that outcome in the group with the lower rate of that outcome than in another setting, the difference between rates, however defined, is larger in the first setting. More concretely, if in one setting the advantaged group's rate of experiencing a favorable outcome is 55% and the disadvantaged group's rate is 25%, and in another setting the advantaged group's rate is 50% and the disadvantaged group's rate is 30%, the disparity is plainly larger in the first setting. It is when the rates are, say, 60% and 40% in one setting and 40% and 23% in another that one has to consider the implications of SR1 and SR2.

Against the background set out above, SR1 and SR2 are set out below:

SR1:

When two groups differ in their susceptibility to an outcome, the rarer the outcome:

(a) the greater tends to be the relative difference in experiencing it, and

(b) the smaller tends to be the relative difference in rates of avoiding it.

SR2:

[Note to description of SR2: While the detailed description of SR2 in the next seven paragraphs may be interesting to some, the essence of that description may be distilled into the following rougher description: As relatively uncommon outcomes (less than 50% for both groups being compared) increase in overall prevalence absolute differences between rates tend to increase (at least until one group's rate reaches 50%); as relatively common outcomes (greater than 50% for both groups being compared) increase in overall prevalence absolute differences tend to decrease; when rates are above 50% for one group and below 50% for the other, the distributionally-driven patterns are difficult to determine. Difference measured by odds ratios tend to change in the opposite direction of absolute differences.]

In circumstances where two groups have normal distributions of risks of experiencing an outcome but the distributions have different means, and the two distributions have the same standard deviation, outside of a range of overall prevalence defined at one border by a one group's having a rate of approximately 50% for either outcome and at the other border by the other group's having a rate of approximately 50% for either outcome (hereafter "MR" for "midrange of prevalence"), the absolute difference between rates will tend to increase as overall prevalence moves toward MR and to decrease as overall prevalence moves away from MR. Viewed another way, prevalence is outside MR when the rates of two groups are both above or both below 50% for either outcome, in which case, again, changes in overall prevalence that move toward MR tend to increase absolute differences while those that move away from MR tend to reduce absolute differences.

Where the difference between means exceeds .49 standard deviations, within MR absolute differences will exhibit a pattern similar to an irregular convex arc (though not all absolute difference values within MR will necessarily exceed the absolute difference values at the borders of MR). The maximum absolute difference within the arc will be found at the intersection of the two relative differences for the opposite outcomes where the group whose rate is used as the numerator for the fraction used to derive the relative difference is different for each outcome (see Section A of the [Semantic Issues](#) sub-page). As is implicit in the above, the maximum absolute difference within the arc will also be the maximum absolute difference generally. The greater the difference between means, the greater the degree to which the absolute difference will exceed, in proportionate and absolute terms, the absolute differences at the borders of MR.

Where the difference between means is less than .50 standard deviations, within MR absolute differences will exhibit a pattern similar to an irregular concave arc. All

absolute differences within the arc will be smaller than the absolute differences at the borders of MR (which absolute differences at such borders, as qualified below, will be the maximum absolute difference generally). The smaller the difference between means, the larger (in proportionate terms) will be the difference between the minimum absolute difference within the arc and the maximum absolute difference.

I add here that, more precisely, where the difference between means is less than .5 standard deviations, the absolute differences reaches maximums both at (a) the point where one group's rates of experiencing an outcome is .49202 (or .51798 for the opposite outcome) and (b) the point where the other group's rate of experiencing the outcome is .51798 (or .49202 for experiencing the opposite outcome), except where the difference between means is .10, .20, .30, .40, in which cases the absolute differences reach maximums at the points either group's rate of either outcome is 50%. I note such fact, not because there is any likely utility in knowledge of it, but merely to qualify points in the prior paragraph, that, if not so qualified, would be incorrect.

The above points may be more broadly described in the following terms (with MR previously defined): Outside MR absolute differences will tend to increase as prevalence ranges move toward MR and decrease as prevalence ranges move away from MR. Within MR patterns will be difficult to predict and be affected by the size of the difference between underlying means.

Odds ratios will tend to change in the opposite directions of absolute differences.

The above discussion is generally consistent with my prior broader illustrations of patterns of absolute differences and odds ratios. However, my more precise expressions of patterns of absolute difference correlations with overall prevalence prior to March 2009 (as in B11-18, D23, D40, D41) were based on normal distributions with half a standard deviation difference between means. Having investigated the matter to some extent, I had assumed that the patterns would be approximately the same regardless of the difference between means. As indicated above, further investigation has revealed that I was mistaken in that assumption. The precise pattern I previously described does hold for all differences between means from .50 to 1.0 standard deviations, and, without yet investigating the matter, I am inclined to believe it holds for differences above 1.0 standard deviations. But it does not hold for differences between means ranging from .01 to .49.

A graphic illustration of these patterns for all four measures, as illustrated with a difference between means of .5 standard deviations (from Table 4 of B16), can be conveniently accessed as a PDF file by means of this link: [JSM 2008 Fig. 4](#). In light of the additional aspects of absolute differences described above, a version of the part of that figure involving absolute differences – and including, in addition to the illustration for a difference between means of .50 standard deviations, illustrations for differences between means of .85 and .15 standard deviations, is provided by means of this link: [Absolute Difference Illustrations](#).

Section B of this page lists various implications of SR1 and SR2 in the interpretation of group differences in the law and the social and medical sciences. Those who review these implications should be persuaded of the futility of evaluating the size of differences between rates, however measured, without understanding the tendencies these rules endeavor to capture. Initially, however, in Section A, I clarify a number of issues about SR1 and SR2.

A. Clarifying Points

1. *Merely Tendencies Though Powerful Ones.* It is important to recognize that SR1 and SR2 are merely tendencies or common patterns of correlation. Though they are pervasive and powerful tendencies, for a variety of reasons one will observe departures from them. In some cases, such departures may allow one to cautiously draw inferences about whether a difference is changing in some meaningful way over time or otherwise is larger in one setting than another (as discussed, for example, in items [A12](#), [D40](#), [D41](#), and [D43](#) on MHD). But the existence of departures by no means diminishes the importance of understanding these rules if one is to make sound judgments about the size of differences between rates. Rather, SR1 and SR2 provide a framework for recognizing the futility of attempting to appraise the size of differences between rates through standard measures of differences that are in some manner affected by overall prevalence. (See the [Case Study](#) sub-page of this page.) The [Solutions](#) sub-page of MHD then provides a theoretically grounded approach for appraising the size of differences between rates that is not affected by the overall prevalence of an outcome and the [Solutions Database](#) sub-page provides a database with which to implement that approach. As discussed in the introduction to the Solutions sub-page, a probit analysis achieves the same result. See the [Life Table Illustrations](#) sub-page regarding the predominance of the distributional tendencies in measures of demographic differences in mortality/survival at young ages compared with old ages.

2. *Theoretical Basis.* While I often illustrate these tendencies with actual data, it is important to understand that they have a theoretical basis derived from the shapes of distributions of factors associated with experiencing (or avoiding) some outcome. Further, while most of my illustrations of the rules involve underlying distributions that are close to normal, most distributions of factors associated with experiencing an outcome will in fact tend toward the normal. And even where distributions depart substantially from the normal, the measure of difference still will be in some manner affected by overall prevalence. Thus, the potential for the underlying distributions to depart from the normal by no means justifies the ignoring of overall prevalence in interpreting the meaning of differences between rates. Rather, such potential merely complicates the effort to interpret differences between rates while taking overall prevalence into account.

I add that, while I have not fully tested the proposition, SR1 ought to hold even when distributions depart substantially from the normal, so long as the distributions have no significant irregularities. That may not hold for SR2 (as discussed and illustrated in B11

([BSPS 2007](#)) and B13 ([ICHPS 2008](#)) on MHD). But the point of the preceding paragraph holds for both rules. See [Truncation Issues](#) sub-page of this page.

3. *Meaning of Changes.* Some read SR1 as suggesting that as adverse outcomes decline in prevalence, health disparities inevitably or almost inevitably increase. But the point is not that health disparities increase. Rather, the point is that, while as the overall prevalence of an adverse outcome declines relative differences in experiencing it tend to increase, such increase does not necessarily indicate an increase in disparity in any meaningful sense – as pointedly demonstrated when the relative difference in the opposite outcome is decreasing. Anyone disposed to debate that point ought at least to recognize both that changes in disparities that reflect something other than the standard consequence of a change in prevalence are of far greater consequence than those that simply reflect changes in prevalence and that there may occur meaningful changes in disparities in one direction notwithstanding that various measures indicate a change in the opposite direction.

4. *One Pattern Implied in the Other.* It may seem counterintuitive to some that, for example, a change in overall mortality over time could typically lead both to an increase in relative differences in mortality rates and to a decrease in relative differences in survival rates. That is, the patterns denoted (a) and (b) in SR1 may seem contradictory. In fact, however, (b) is implied in (a), if, indeed, (b) is not exactly the same thing as (a). For, if it is true that a decline in the overall prevalence of an outcome will tend to increase relative differences in experiencing the outcome, it follows that an increase in the overall prevalence of an outcome will tend to decrease relative differences in experiencing the outcome. And if one outcome is decreasing in overall prevalence (hence tending to increase the relative difference in experiencing it), the opposite outcome is necessarily increasing in overall prevalence (hence tending to decrease the relative difference in experiencing that outcome). Thus, to say the first part of the rule is also to say the second part of the rule.

I add here that there is a tendency for observers to think something meaningful has occurred when two groups do not experience the same proportionate decrease or increase in rates of experiencing an outcome. Yet anyone inclined to regard it as somehow natural that two different base rates should undergo the same proportionate change – say, for example, that an increase in HDL should achieve the same proportionate reduction in heart attack risk for men and women with similar profiles as to other risk factors – should recognize that it is mathematically impossible for groups with different rates of experiencing some outcome to undergo the same proportionate change in those rates while also undergoing the same proportionate change in the opposite outcome rates. That is, for example, when rates of 10% and 20% each are reduced by the same proportionate amount, the opposite outcome rates (90% and 80%) necessarily will undergo different proportionate increases. And since there is no more reason to expect two groups with different base rates of experiencing an outcome to undergo the same proportionate changes in rates of experiencing the outcome than there is to expect that they would undergo the same proportionate changes in the opposite outcome, there is no reason to regard it as natural that the groups should undergo the same proportionate change in

either outcome rate in the first place. See the [Illogical Premises](#) and [Subgroup Effects](#) sub-page of this page and [Comment on Sun BMJ 2010](#), [Comment on Chatellier BMJ 1996](#), and [Comment on Cook BMJ 1995](#).

(It should be obvious as well that groups with different base rates cannot undergo both the same proportionate changes and the same absolute change in their rates of experiencing (or avoiding) an outcome. But there is no need to give such point particular attention here.)

4a. *Scale Issues*

A first thought of some about the way relative differences in one outcome and relative differences in avoiding the outcome change systematically in opposite directions as the prevalence of the outcome changes is that the pattern may be simply a function of scale. A basis for such thought involves the fact that the same absolute change in each group's rate would necessarily cause a greater proportionate increase in one outcome rate for one group and a greater proportionate change in the other outcome rate for the other group. For example, where Group A's rate of experiencing an outcome is 10% (90% for the opposite outcome) and Group B's rate of experiencing the outcome is 20% (80% for the opposite outcome), and both group experience a 5 percentage point increase in the less common outcome, 5 would be the numerator of each fraction for calculating the proportionate changes in the outcomes and the prior rate would be the denominator. Thus, Group A would have the smaller denominator (10% versus 20%) for one outcome and Group B would have the smaller denominator the opposite outcome (80% versus 90%). Where two fractions have the same numerator, the fraction with the smaller denominator will be larger. That is, Group A shows the larger proportionate increase in one outcome (5/10 compared with 5/20), while Group B shows the larger proportionate decrease in the opposite outcome (5/80 compared with 5/90). As should be obvious from the example, the pattern whereby the group that has the smaller initial rate for one outcome will always have the larger initial rate for the opposite outcome.

But SR1 is not simply a function of scale. For there is no reason to expect that the absolute differences in two initial rates will show the same percentage point change. In fact, that would occur rarely (though, as discussed in the introduction, within MR, absolute difference changes often will be very similar).

4b. *Subgroup Effects*

In a number of places I have described the reasons to expect that, solely for statistical reasons, a factor that similarly affects two groups with different baseline rates of an outcome will tend to cause different proportionate changes in the outcome rates for each group. For example, a factor that decreases an outcome would be expected to cause a larger proportionate decrease in an outcome in the group with the lower base rate, while causing a larger proportionate increase in the opposite outcome for the other group. Moreover, since a factor cannot cause identical proportionate decreases in an outcome for two groups and at the same time cause identical proportionate increases in the opposite

outcome, it is unreasonable to expect identical proportionate changes in either outcome. This matter is treated in greater depth on the [Subgroup Effects](#) sub-page of this page. See also the [Illogical Premises](#) sub-page regarding why it is illogical to regard it as somehow normal that a factor that would cause groups with different base rates to experience equal proportionate changes in those rates.

5. *Representation among Population Experiencing and Failing to Experience an Outcome.* A corollary to SR1 is that as an outcome decreases in prevalence, the group that is more susceptible to the outcome will tend to comprise both (a') an increasing proportion of the population experiencing the outcome and (b') an increasing proportion of the population failing to experience the outcome. That may seem even more counterintuitive than SR1. But, as with SR1, (b') is implied in (a').

6. *The Value of Health Disparities Research.* Most of the post-1999 material on this site discussing SR1 and SR2 involve health or healthcare disparities, an issue to which, in 2003, the United States Government devoted \$2.6 billion, or approximately 9% of the budget of the National Institutes of Health. These funds and such other funds as may be devoted to health disparities issues by state and local governments, as well as by educational institutions and foundations, presumably include moneys directed to improving the health and healthcare of disadvantaged groups, as distinguished from studying disparities. The former moneys may be well spent. But, given the near universal failure to recognize the way binary measures of difference are correlated with the overall prevalence of an outcome, it is doubtful that moneys devoted to studying such questions as whether disparities are increasing or decreasing have yielded very much of value.

The extent to which scholars in the United States and abroad have accepted SR1 with regard to the measurement of health disparities is discussed in [Section E.7](#) of MHD. While partly covered there (and as discussed in various references on MHD), the health disparities measurement approaches of three agencies of the United States Government warrant brief mention here. In papers published in 2005, 2006, and 2009 that cited A5 ([Chance 1994](#)) and/or A10 ([Society 2000](#)) for showing that conclusions about the size of relative differences may turn on whether one examines the favorable or the adverse outcome – but without discussing the claim that the two differences tend to change systematically in opposite directions as the prevalence of an outcome changes (or the reasoning underlying such claim) – statisticians of the National Center for Health Statistics addressed the interpretative issue by simply recommending that all disparities be measured in terms of relative differences in adverse outcomes. The following of such recommendation will tend to ordain that improvements in both health and healthcare will be associated with perceived increases in disparities.

The Agency for Healthcare Research and Quality (AHRQ), which is responsible for the annual National Healthcare Disparities Report, primarily measures disparities in health and healthcare outcomes in terms of whichever relative difference (in the favorable or the adverse outcome) is larger. Since most of the things AHRQ measures involve situations where the adverse outcome is relatively uncommon – and thus where the relative

difference in the adverse outcome tends to be larger than the relative difference in the favorable outcome – AHRQ will usually measure disparities in the same manner as NCHS. But in circumstances where a favorable outcome is relatively uncommon – and where the relative difference in the favorable outcome tends to be larger than the relative difference in the adverse outcome – AHRQ may reach different conclusions from NCHS regarding changes over time. Further, where an outcome grows from being relatively rare to being quite common, the situation may change from one where the relative difference in the favorable outcome is larger to one in which the relative difference in the adverse outcome is larger. In such circumstances, assuming AHRQ follows its stated approach, solely for statistical reasons, the agency’s appraisal of the disparity may change from one where the disparity is decreasing to one where the disparity is increasing. See B12 ([APHA 2007](#) and [Addendum](#)) and D51 ([Comment on Morita](#)) on MHD.

Generally, researchers outside the federal government have adopted neither the approach of NCHS nor that of AHRQ and may not be fully aware of either approach. Research into healthcare disparities, much of it funded by AHRQ, typically relies on relative differences in receipt of beneficial procedures or appropriate levels of care (which reliance, given that rates of such receipt generally have been increasing, tend to result in findings that disparities are decreasing, as discussed, say, in D48 ([Comment on Escarce](#)) or D51 ([Comment on Morita](#)), or on absolute differences between rates (which reliance tends to result in findings that disparities in relative uncommon outcome are increasing and that disparities in relatively common outcomes are decreasing, as discussed, say, in D23 ([Comment on Vaccarino](#)) and B11 ([BSPS 2007](#))). And researchers often cite findings of other researchers for consistency or inconsistency with their own findings without evident recognition of the role of choice of measure in such consistency or inconsistency. Officials of AHRQ have several times cited a study that relied on absolute differences between rates of receiving appropriate care as indicating that improvements in care will tend to reduce disparities. But they have done so without recognizing that AHRQ’s own measurement methodology will tend more often to correlate improvements in care with increases in disparities. See items B12 ([APHA 2007](#)), D23 ([Comment on Vaccarino NEJM 2005](#)), D42 ([Comment on Aaron and Clancy JAMA 2003](#)).

In January 2011, the Centers for Disease Control and Prevention (of which NCHS is a part) issued its first Health Disparities and Inequalities Report. The Report principally relied on absolute differences between rates as a measure of disparity. Since absolute differences tend to change in the opposite direction of the larger of the two relative differences, such approach will tend to cause CDC to reach conclusions as to the directions of changes over time that are the opposite of those AHRQ would reach. It should not, however, be thought that the failure to understand these correlations or the implications of such failure are by any means limited to the study of health disparities. The same failure of understanding undermines efforts to evaluate the size of differences between rates in every area where the size of a difference between rates is a matter of consequence, as discussed in items A1 to A10 on MHD. Table 1 of item B16 ([JSM 2008](#)), which is used to illustrate problems of interpreting changes in healthcare outcomes, could as well be used to illustrate problems in determining whether hiring disparities are increasing or decreasing (as further developed on the [Case Study](#) sub-page

of this page and [Relative Versus Absolute](#) sub-pages MHD). See also the illustration in D58 of MHD ([Second Comment on Hetemaa JECH](#)) as to why it is impossible to appraise the size of a hiring disparity if one only knows a group's representation among hires and among applicants, a matter also addressed as Issue 3 on the Case Study sub-page), a matter further developed on the [Representational Disparities](#) sub-page of this page. A similar issues is treated on the [Gender differences in DADT](#) sub-page of the [Vignettes](#) page.

Apart from the general measurement problem in the National Healthcare Disparities Report, the report contains a number of technical problems (some related to that general problem and some not). These are addressed on the [NHDR Technical Issues](#) sub-page of MHD.

Note added in December 2011: While the mismeasure of health disparities involves a great deal of wasteful research, until recently such research seemed unlikely to cause specific harm. That is, even though much of that research was justified as an effort to learn how to reduce disparities, generally institutions continued to do the things that seemed to improve population health even though such measure might seem to increase disparities. For example, no one was likely to suggest that successful program like the Back-to-Sleep program should be abandoned because it led to increased relative differences in SIDS deaths. See [Comment on Pickett AJPH 2005](#). But recently there has arisen a perception (though, like other perceptions about health and healthcare disparities, one based on a misunderstanding of how to measure them) that improvements in care would tend to increase healthcare disparities. This led to recommendations for including effects on healthcare disparities as a performance criterion in pay-for-performance programs. To the extent that such recommendation are followed, institutions now will actually be paid according to unfounded perceptions about health disparities. See [Pay for Performance](#) sub-page of MHD. For reasons related to the failure to understand how measures of differences between rates are affected by the prevalence of an outcome (and some others as well), there is a basis to believe that a Massachusetts program tying pay-for-performance to healthcare disparities will increase healthcare disparities. See [Between Group Variance](#) sub-page of MHD.

7. *A Measure Unaffected by Prevalence.* Some of the references on MHD are skeptical of the existence of tools that can measure, for example, whether health or healthcare disparities are increasing or decreasing over time with sufficient reliability to justify the resources devoted to the study of such issues. Some relatively recent references proposing a plausible approach to measuring disparities (or anything else) that is not affected by the prevalence of an outcome are listed in Section E.6 of MHD and under the [Solutions](#) sub-page. A downloadable database to implement the approach is made available on the [Solutions Database](#) sub-page of MHD. As discussed in those references, this approach has certain weaknesses and the existence of such approach does not necessarily resolve the question of whether disparities can be measured with sufficient reliability to justify devoting substantial resources to health disparities research. But the approach is plainly superior to reliance on the standard measures of differences between

rates without regard to the way such measures are affected by the overall prevalence of an outcome. As mentioned in Item A.1 above, a probit analysis achieves the same result.

It should be recognized, however, that even assuming there exist tools for accurately measuring the size of health disparities in different settings, there may be reason to reconsider the level of resources committed to such study. One reason for the substantial commitment of resources to the study of health disparities is the perception that disparities in adverse outcomes have increased substantially. For example, between 1950 and 2000, the ratio of the black infant mortality rate to the white infant mortality rate increased from 1.64 (4.29%/2.68%) to 2.47 (1.41%/0.58%). In other words, a 64% greater black infant mortality rate increased to a 147% greater black infant mortality rate. On its face such increase seems to be a quite dramatic. Very likely the disturbing aspects of that change and others like it contribute to the support for health disparities research. Thus, should it be concluded that all or substantial parts of such increases are solely a statistical consequence of declining mortality, health disparities research may lose of some of its perceived urgency.

The above point is not to suggest that there in fact occurred no meaningful change in the racial disparity in infant mortality. According to the methodology discussed on the Solutions sub-page of MHD, the disparity increased from a 0.23 standard deviations difference between hypothesized means in 1950 to a 0.35 standard deviations difference in 2000. But it is in terms such as these that issues concerning changes in disparities should be appraised. But see also the [Irreducible Minimums](#) sub-page of MHD.

It should also be recognized, however, that health disparities research is lately much focused on healthcare. Although as discussed in item A.6 the approaches of both AHRQ and NCHS will tend to find healthcare disparities to increase as the overall quality of healthcare improves, the references on the [Solutions](#) sub-page of MHD seem to indicate that, to the extent changes over time can be effectively measured, the disparities are more often declining.²

8. *Meaning of “Overall Prevalence.”* Precise epidemiologists or statisticians may be troubled by the use of the phrase “overall prevalence” since overall prevalence is obviously a function of both the rates of each group in the population and the proportion each group comprises of the total population. It should be clear, however, that my use of such terms is intended to reflect a general pattern. As a rule, that general pattern can best be identified with reference to the situation of the advantaged group at least where the advantaged group is the majority of the population. But I think that, despite the imprecision, describing the general pattern in terms of prevalence or overall prevalence breeds less confusion than doing so in terms the rate of the advantaged group.

² As discussed in item D41 on MHD ([Comment on Trivedi JAMA](#)), the statistical analysis underlying the perception that improvements in care seem to reduce disparities in process but not disparities in clinical outcomes is flawed. But, as also noted, the reasoning offered to explain the perceived pattern is plausible enough and appropriate techniques for evaluating the size of disparities seem to support the conclusions.

9. *Settings Differentiated Temporally or Otherwise.* In various places I refer to different settings, often clarifying that one differentiator of such settings is time. And much of the material on MHD is devoted to discussion of changes over times. But I also make clear that the same principles apply to comparisons of disparities in settings that are differentiated other than temporally. Yet, while mathematically there is no difference as to the ways SR1 and SR2 apply to comparisons of differences between rates in different settings differentiated temporally and those differentiated otherwise (e.g., geographically, by subpopulation, or by condition), the following practical issue should be recognized.

As discussed most fully in items A12 ([Society 2000](#)) and B7 ([BSPS 2006](#)), in addition to being partly a function of overall prevalence, each measure of differences between rates is also a function of the size of the difference between means of underlying risk distributions (which is usually termed EES, for estimated effects size, in references discussed on the Solutions sub-page of MHD, or which could be more simply envisioned as the difference between mean scores, in terms of percentage of a standard deviation, in the case of two test score distributions). For any given level of an outcome for the advantaged groups in two settings, each of the four binary measures will be larger in the setting with the larger EES. Thus, as a rule, when the advantaged groups in the two different settings have different rates, any difference in the EES will tend to heighten the effect of overall prevalence as to two measures and counter the effect of overall prevalence as to the other two. And with regard to the latter two measures, the implications of the difference in averages may be either greater than or less than the prevalence effect.

Possibly to belabor an obvious point, I note that it should always be recognized that when one does not observe the standard patterns – as, for example, where as an outcome changes in prevalence relative differences in experiencing it and relative differences in avoiding it change in the same direction – that does not mean that the described statistical forces are absent. For the observed pattern may result from (indeed, probably results from) the fact that a meaningful change or difference is strong enough to overcome the statistical tendency. And, of course, where there occur only small changes in the rates of various groups, one may well be observing nothing but random variation.

With that background in mind, consider possible differences between comparisons of disparities in settings differentiated temporally and those differentiated otherwise. Given a particular difference between means of two identified populations at one point in time, such difference probably provides the best estimate as to what it will be at a later point in time (at least absent reason to believe that meaningful changes in the difference have taken place). And, when the points in time are quite close, one would tend to expect the differences to be rather similar at the two points in time. But in the case of populations defined geographically, there is no reason to believe that the difference between means in one population is the same as the difference in another population, save in the sense that we might expect that they would be broadly comparable, say, in two high-income countries with similar social policies, similar health care systems, and similar levels of income inequality. Put another way, at least absent some contrary information, there is little reason to expect differences between the means of the same populations to differ

very much at two points relatively close points in time. On the other hand, there is no particular reason to expect the differences between means of advantaged and disadvantaged populations in one county to be very similar to the difference between means of advantaged and disadvantage populations in another country.

For these reasons, we are likely to observe a more consistent inverse relationship between relative differences in experiencing and avoiding an outcome when examining inequalities in a particular country or geographic region over time than we are when comparing the size of inequalities across different countries or geographic regions.³

In addressing the strength of SR1 (as expressed in Race and Mortality) Houweling et al,⁴ apparently with the view that SR1 should theoretically lead to an inverse correlation between rankings of relative differences in one outcome and relative differences in the opposite outcome, examined such correlation and found mixed results. But it must be borne in mind that (a) whereas, all things being equal, the rarer an outcome the greater will tend to be the relative difference in experiencing and the smaller will be the relative difference in the opposite outcome (hence tending toward an inverse correlation between the two rankings), (b) it is also true that, all things being equal, the larger the relative difference in one outcome the larger will tend to be the relative difference in the opposite outcome (hence tending toward a direct correlation between the two rankings). The observed correlation will be a function of these contrasting tendencies. See D.72 [Comment on Eikemo](#), which also discusses Houweling but not as to the above point. See discussion of some competing forces as to absolute differences in D53 ([Comment on Baicker](#)).

³ The patterns by which each measure will vary from setting to setting (including settings defined temporally and otherwise) are functions of both (1) the above-described distributionally-driven forces and (2) the actual differences between the underlying means in each setting. The comparative influence of the two factors will be a function of the size of the difference in prevalence in the two settings and the size of difference between means in the two settings. One situation where the distributionally-driven patterns commonly predominate involves comparisons of effects of some factor on mortality across age groups. Almost invariably the relative effect of some factor on mortality (whether the factor be race, socioeconomic status, or some intervention) will be larger among the young (as will commonly be noted), while the relative effect on survival will be greater among the old (as will rarely be noted). For, while of course subject to where we draw the line distinguishing the young from the old, the difference in the prevalence of mortality among the young and the old usually is substantial. Hence, the described distributionally-driven patterns tend to be evident regardless of the size of the differences between the means of the distributions in the different settings. Similarly, the more substantial are temporal changes in overall prevalence the more likely it is that the distributionally-driven tendencies will predominate. But where the difference in prevalence from setting to setting is small, the effect of the distributionally-driven patterns will be small, and hence, for example, the setting with the larger relative difference in one outcome often may also have the larger relative difference in the opposite outcome.

⁴ Houweling TAJ, Kunst AE, Huisman M, Mackenbach JP. Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health. *International Journal for Equity in Health* 2007;6:15: <http://www.equityhealthj.com/content/6/1/15>

See the [Life Table Illustrations](#) sub-page regarding the predominance of the distributional tendencies in measures of demographic differences in mortality/survival at young ages compared with old ages.

10. *Effects of Lowering Cutoffs on Employment Tests.* In many places, in criticizing the way that researchers assume that increasing relative differences in adverse outcomes during periods of overall declines in the outcome indicate a worsening of the relative situation of disadvantaged groups, I have discussed employment tests on which one demographic group has lower average scores than another. I have pointed out that lowering a cutoff is universally regarded as reducing the disparate impact of such a test because it tends to reduce the relative difference in pass rates, even though lowering the cutoff tends to increase relative difference in failure rates. But such point leaves open the question of whether lowering a cutoff in fact reduces the disparate impact of a test in any meaningful sense. Such issue is addressed on the [Employment Tests](#) sub-page of this page, which, by means of the procedure described on the Solutions sub-page of MHD, shows that, in circumstances where selections among persons who pass a test are not correlated with test scores, lowering a cutoff does reduce the disparate impact of a test.

Like many things on this and related pages, the point demonstrated with some effort on the referenced sub-page is something that many will find quite obvious after giving the matter a little thought and some will find quite obvious even without giving the matter any thought.

11. *Absolute Differences.*

[The material that follows was written prior to the April 5, 2009 modifications to the Introduction. It warrants revision in light of the revised discussion of absolute differences in the introduction. It still tends to hold for situations where the difference between means is greater than .49 standard deviations.]

Determining the expected patterns of absolute differences in a particular setting is rather more complicated than determining the expected patterns of either relative difference. For purposes of considering expected patterns of absolute differences, it is useful to think of the area to the left of the intersection of the two ratios of rates as Zone A and that to the right as Zone B. Further increases in overall favorable outcomes where the ratios are in Zone A tend to increase absolute differences while further increases in favorable outcomes where the ratios are in Zone B tend to reduce absolute differences.

As noted in the Introduction, absolute differences are increasingly used as a measure of disparities in healthcare. The factors being measured include receipt of procedures, receipt of an appropriate level of care, and control of some adverse condition. Most procedure rates tend to be increasing. But the procedures examined include many relatively rare ones (like hip replacement) and many very common procedures (like mammography). Generally, when the total population of each group is used as the numerator, overall increases in the former tend to result in increasing absolute differences and overall increases in the latter tend to result in declining absolute differences. But, for

example, if a study of hip replacement rates is limited to a universe of persons with substantial hip impairment, hip replacement might be considered a high prevalence outcome where further increases would tend to reduce absolute differences. There may also be cases where the activity moves from Zone A to Zone B, resulting in an increase in absolute difference during part of the period examined and a decrease during part of the period examined.

Complexities increase when one examines things other than changes in disparities over time. Consider studies that have examined the correlation of quality of care (measured in terms of receipt of procedures or other favorable outcomes) with the size of disparities. For outcomes where the rates were in Zone A better quality would tend to be associated with larger absolute differences and for outcomes where the rates were in Zone B better quality would be associated with smaller absolute differences. See [D40](#), [D41](#). And consider a study that examined both (a) the correlation between the overall rates in different areas and the absolute differences between black and white rates in those areas and (b) the correlation of the black rates in different areas with the absolute differences in different areas. To the extent that areas with high rates for some procedures tend to have high rates for other procedures, such areas would tend to show comparatively large absolute differences between black and white rates for uncommon procedures and comparatively small absolute differences for common procedures). To the extent that high black rates tended to be correlated with high overall prevalence, high black rates would similarly tend to be correlated with large absolute differences for uncommon procedures and small absolute differences for common procedures. See [D53 \(Comment on Baicker\)](#).⁵

When the outcome examined is an appropriate level of care, one would expect the overall prevalence to be fairly high in the United States – that is, as a rule, well above 50% for all groups (though adequate hemodialysis only recently achieved such level (as discussed in [B12 \(APHA 2007\)](#) and [D23 \(Comment on Vaccarino\)](#)). Thus, as a rule, further improvement in appropriate care rates will tend to reduce absolute differences between appropriate care rates.

The rates of control of certain adverse conditions tend to be examined within the universe of persons needing such control (blood pressure control among those deemed hypertensive, HbA1c control among diabetics). That tends to place the ratios in Zone A, even though the ratios relating to the achievement on the control point within the total population are well into Zone B. Thus, for example, bringing the systolic blood pressure of everyone at below 150 to below 140, while tending to reduce absolute differences in high blood pressure among blacks and whites in the population at large, will tend to increase absolute differences in control of high blood pressure among blacks and whites deemed hypertensive. See [B12 \(APHA 2007\)](#), [B14](#), [D40](#).

⁵ These points may be illustrated by data in [BSPS 2006 Table 1](#). Keeping in mind that we are concerned with the favorable (pass) outcome, consider Rows F70 and E80 as reflecting areas with high and low overall (and black) rates for an uncommon procedure and Rows K20 and J30 as reflecting the areas with high and low rates for a common procedure.

12. Odds Ratios.

[This material is subject to the same qualifications as expressed in the prior subsection with regard to the treatment of absolute differences.]

As noted in the Introduction, differences measured in odds ratios tend to be inversely correlated with absolute differences. Apart from generally raising issues similar to those raised by absolute differences, odds ratios may create additional anomalies when used to adjust other measures for confounders. For example, consider a situation where, during a period of increases in a favorable outcome, ratios of experiencing that outcome decline. Suppose logistic regression were employed to determine whether the observed decrease in relative difference may have been a result of confounding by other factors. Assuming the potential confounders in fact had no role whatever, the odds ratios yielded by the logistic regression would tend to show a pattern that is the same as the pattern of the risk ratio if the ratios are in Zone B but a reverse pattern if the ratios are in Zone A. In the case where absolute difference is the measure being examined, the adjustment by means of logistic regression would reverse the pattern of apparent change regardless of the zone in which the activity occurred.

For example, consider that in the table below which is used in both the [Case Study](#) and [Relative Versus Absolute](#) sub-pages, one set of changes over time is reflected in Rows A and B and another is reflected in Rows C and D. In these cases, reliance on the absolute difference would cause one to observe an increase or decrease in the disparity and adjustment for confounders by means of logistic regression would reverse the patterns.

Table 1 (from Case Study sub-page) – Illustration of Associations of Measures of Difference Between Rates with Prevalence of an Outcome						
Settings	AGFR	DGFR	RelDiffFav	RelDiffAdv	AbsDiff	OR
A	20.0%	9.0%	55.0%	13.8%	0.11	2.53
B	40.1%	22.7%	43.5%	29.2%	0.17	2.29
C	59.9%	40.5%	32.3%	48.2%	0.19	2.19
D	90.0%	78.2%	13.1%	117.1%	0.12	2.50

13. *Phi Coefficient.* The phi coefficient is a measure of association between two binary variables. It was originally described by Udney Yule in a 1912 article in the *Journal of the Royal Statistical Society*. According to this [link](#), the numerical value of the phi coefficient of correlation is identical to that obtained by the Pearson product-moment coefficient of correlation. The phi coefficient and its perceived advantaged and disadvantaged are discussed in Fleiss et al, *Statistical Methods for rates and Proportions* (Third Edition, 2003) at 98-99. It can be calculated for the values in the Values Table of the [Solutions Database](#) by developing values a (1-[AGFail]), b ([AGFail]), c (1-[DGfail]),

d ([DGFail]) in accordance with the parentheses and applying the formula $\phi = \frac{([a]*[d]) - ([b]*[c])}{\sqrt{([a]+[b])*([c]+[d])*([a]+[c])*([b]+[d])}}$. Doing so will reveal the way the values change as overall prevalence changes, and, hence, that, like that other measures discussed above, the phi coefficient is not a useful indicator of the size of differences between rates (as illustrated in Table A of this [item](#)). As overall prevalence changes, the measure behaves essentially like the absolute difference, yielding a fractional value that is somewhat larger than the absolute difference as a fraction when prevalences are high or low, but for the combinations of advantaged and disadvantaged rates that yield the maximum absolute difference for each EES value, such value will be identical to the phi coefficient.

13a. Cohen's Kappa Coefficient. Cohen's Kappa Coefficient is ordinarily used as a measure of agreement between two raters' ratings of something and is supposed to take into account the extent to which agreement would occur by chance. The values in this measure change as overall prevalence changes (as illustrated in Table B of this [item](#)). Hence, it cannot provide a useful indicator of the size of a difference between outcome rates.

14. *Longevity Differences*. A central theme of the discussion above is that binary measures are not useful for measuring health disparities because all binary measures are affected by the overall prevalence of an outcome. Only truly continuous measures are useful for appraising the comparative situation of two groups, and the approach described on the Solutions sub-page is an effort to derive continuous measures from binary measures. Longevity appears to be a continuous measure. But it is a function of mortality and longevity differences tend to change solely because of changes in overall mortality rates. Thus, changes in longevity differences, whether measured in relative or absolute terms, do not provide useful information on whether health disparities have changed in some meaningful sense (as discussed at page 6-7, and illustrated in Table 2, of B7 on MHD ([BSPS 2006](#))).

15. *Gini Coefficient etc.* There exist varied other measures of health inequalities, such as the Gini coefficient, concentration index, relative index of inequality, slope index of inequality, which attempt to take into account the way the sizes of different groups change over time as well as the changes between outcome rates of the groups. All such measures, however, are affected by changes in overall prevalence, as illustrated, for example, with regard to the Gini coefficient in Table 1 of D43 ([Comment on Boström](#)). Hence, like longevity differentials, none is a useful indicator of whether differences are changing over time or whether differences between rates are otherwise larger in one setting than another. See A.17 below.

16. *Irreducible Minimums*. See the [Irreducible Minimums](#) sub-page of MHD regarding situations where an outcome approaches a level beyond which it may not be possible to achieve further reductions.

17. *Concentration Index*. The [Concentration Index](#) sub-page of MHD shows how the concentration index changes as prevalence changes.

18. *Probit Analysis*. As discussed in the introduction to the [Solutions](#) sub-page of MHD and the to the February 23, 2009 update to the [Comment on Morita](#), a probit analysis yields the same results as the more mechanically derived EES described on the Solutions sub-page of MHD. Thus, probit results are not functions of the overall prevalence of an outcome.

19. *Case Control Studies*. Implicit in the arguments made on this site as to the need to know the actual rates in order to effectively appraise an effect size is an argument that case control studies are fundamentally flawed because, while they enable one to derive an odds ratio, they do not enable one to determine the underlying rates. See [Case Control Studies](#) sub-page.

20. *Meta-analyses*. Implicit in the arguments made on this site as to the flaws of rate ratios and odds ratios as measures of association is an argument that meta-analyses are fundamentally flawed. See [Meta-Analysis](#) sub-page.

B. Illustrations of SR1 and SR2

Set out below are a number of illustrations of the importance SR1 and SR2 in the interpretation of data on group differences in the law and the social and medical sciences. Some illustrations are redundant of points made in Section A above and many are redundant among themselves. I give only limited attention to situations where observers regard increases in relative differences in mortality or other adverse health outcome during time of overall declines in such outcomes as increasing health disparities, without recognizing that declines in an outcome tend usually to increase relative differences in experiencing it and without consideration of whether the disparity in the opposite outcome has declined. But there exist hundreds of such studies and many are discussed in the on-line comments collection in Section D of MHD as well as in various items in Sections A-C of MHD. None of these studies offers useful insight into whether the disparities are increasing or decreasing in a meaningful sense because none considers the implications of changes in overall prevalence. Of course, the same holds for studies that find declining relative differences in outcome that are increasing – whether such outcome be beneficial health procedures or adverse outcomes like obesity and diabetes.

In many cases where studies seem to indicate that disparities have changed over time or are otherwise larger in one setting than another, researchers offer explanations as to the reasons for the differences or suggest implications of such differences. But in such cases, while the reasoning of the researchers may be plausible enough, as a rule the premise for such reasoning is flawed due to the failure of the researchers to correctly determine whether one disparity is in fact larger than another. Several such examples are presented among the illustrations below.

1. *Feminization of Poverty I*. In the late 1970s it was observed that the female-headed families were comprising an increasing proportion of the poor. This pattern was termed

the “feminization of poverty” and universally regarded as a very negative trend. In 1980 a presidential advisory panel lamented that, if current trends continued, by the year 2000 the poverty population would be entirely comprised of female-headed families. The pattern was in part a function of the fact that female-headed family members were becoming an increasing proportion of the total population. But it was also in substantial part a function of the fact that between 1959 and the mid-1970s there occurred dramatic declines in poverty, including the poverty of female-headed families. This went overlooked as did the fact that at about the time the feminization of poverty was discovered, poverty ceased to decline and, correspondingly, ceased to become more feminized. By the year 2000, poverty was no more feminized that it was when the feminization of poverty was discovered. Nevertheless, there continues to be a perception that poverty is still becoming increasingly feminized. A September 6, 2009 Google search for “feminization of poverty” yielded 828,000 results. The phrase “feminisation of poverty” [sic] also yielded 570,000 results, reflecting the way the concept has spread to the United Kingdom and other places that use British spelling. Virtually none of the discussions underlying these hits recognizes the connection between the feminization of poverty and declining poverty.

An interesting issue to ponder is the near universal failure to recognize that a natural consequence of reductions in some adverse outcome is that it would become increasingly concentrated in groups that are disproportionately susceptible to such outcome, as in the case of more poverty-prone groups, like single parent families, or that, were we to verge on the total elimination of poverty, the poor would be entirely comprised of groups that are disproportionately poverty-prone. See A1, A3, A4, A5, A10. Only on occasion will observers note with concern that poverty has spread to groups that are not ordinarily very poverty-prone. Such spread is properly recognized as a cause for concern. But it nevertheless goes unrecognized that the opposite trend is not necessarily a negative one.

Given the resources devoted to health disparities research in recent decades, or even the resources devoted to exploring health disparities measurement issues, the near universal failure to recognize SR1 within the health disparities research communities here and abroad must be considered at least somewhat remarkable. That is the more so when one considers that persons with statistical backgrounds would be expected to understand that a difference between the means of two groups that yields modest relative differences in rates of falling above or below either mean yields much larger relative differences with regard to falling above or below more extreme values. And to know such tendency is to know SR1. But at least with regard to things like relative differences in mortality rates, where the underlying distributions are not directly observable (even if distributions of closely correlated factors are observable, as with the systolic blood pressure levels underlying Figure 8 of B11 ([BSPS 2007](#)) or the folate levels discussed in D62 ([Comment on Dowd](#))), the seeming invisibility of the risk distributions undoubtedly contributes to the failure to recognize SR1 or its implications. Things are different in the case of the feminization of poverty and other areas where demographic differences in poverty are studied with a near universal failure to recognize that declines in poverty will almost always cause poorer groups to comprise a larger proportion of the poor and increase relative differences in poverty rates (as well as cause poorer groups to make up a larger

proportion of the non-poor and reduce relative differences in rates of avoiding poverty.) For data illustrating the relevant patterns are widely published, in particular the data on populations falling below various ratios of the poverty line that underlie the discussion of the feminization of poverty in MHD items A1 ([Plain Dealer 1987](#)), A3 ([Public Interest 1991](#)), A4 ([Signs 1991](#)), A5 ([Chance 1994](#)), A10 ([Society 2000](#)) and that are replicated in Table 1 of item A12 ([Chance 2006](#)). And, of course, the reader may note that everything on MHD flows more or less directly from A1, the 1987 *Plain Dealer* article styled “‘Feminization of Poverty’ is Misunderstood.” In any case, it is difficult to understand how researchers can look at data like that found in Table 1 of A12 and fail at least to recognize the likely effects of increases or decreases in poverty on relative differences in poverty rates.

2. *Feminization of Poverty 2.* The feminization of poverty is composed of two elements: (1) an increase in the proportion female-headed families comprise of the total population; and (2) an increases in the relative difference in poverty rates female-headed families and other part of the population. The latter can result from (a) declines in poverty or (b) factors that reflect something other than the consequences of a decline in poverty. It should be evident that it makes little sense to study a subject like the feminization of poverty without separate consideration of (1) and (2), since they are quite different phenomena and, moreover, are phenomena warranting quite different responses. Yet only a few studies have separately examined these issues. And researchers who have separately examined changes in the relative susceptibility to poverty of female-headed families (or women generally) and other parts of the population have done so without consideration of the extent to which such changes are a function of changes in overall poverty. See [A4](#) on MHD.

3. *Racial Differences in Infant Mortality.* At some point in the middle 1980s it was noted that in 1983 the racial difference in infant mortality had reached its highest point in history. It went unnoticed that in 1983 the infant mortality rates of blacks and white reached their lowest points in history or that the racial disparity in infant survival reached its lowest point in history. In the ensuing years, as infant mortality continued to decline generally, much attention was been given to the increasing relative difference in infant mortality, invariably without regard the fact that relative differences in experiencing an outcome tend to increase as an outcome declines in prevalence and without regard to whether relative differences in infant survival rates had declined. Some studies of increasing relative differences in infant mortality have casually referred to the pattern as one of an increase in the disparity in infant survival rates without recognizing that the survival disparity may in fact have declined.

4. *Racial Differences in Adverse Birth Outcomes among Advantaged Groups.* Studies have found that racial disparities in adverse birth outcomes like low birthweight and infant mortality among relatively advantaged groups (e.g., the well-educated, low-risk mothers) are as large as, or larger than, they are among relatively disadvantaged groups. A1992 study in the *New England Medical Journal* that found large relative differences in infant mortality where mothers had college educations received widespread publicity and led to suggestions that genetic differences may have had a role in higher black infant

mortality rates. It went unnoticed that large relative differences in adverse outcome rates, though small relative differences in favorable outcome rates, ought to be expected among relatively advantaged groups simply because of the low adverse outcome rates within such groups. See A1, A3, A10, A11, C1.

That racial disparities in birth outcomes tend to be larger among more advantaged groups has been suggested as indicating that social class is not the main component of racial differences in mortality. See D59. Similarly, that racial differences in mortgage rejection rates tend to be largest among the highest income categories has been interpreted as indicating that financial factors are not the principal reasons for observed racial disparities in mortgage rejection rates. See A10. Once one recognizes that large relative differences in adverse outcome rates (though small relative differences in favorable outcome rates) are to be expected within advantaged populations simply because adverse outcome rates tend to be low in such populations, arguments as to the meaning of large disparities in such populations lose all force.

5. *Whitehall Studies*. Research into health inequalities in the United Kingdom has been substantially driven by extensive studies of the health of cohorts of British civil servants (Whitehall studies). The studies found that within this relatively affluent and homogenous population socioeconomic differences in mortality were larger than in the UK population at large. Such finding has been interpreted to suggest that mortality differentials in the general UK population are underestimated due to the imprecision of socioeconomic measures in the general population. The presence of such disparities in a population of civil servants where few suffer from material deprivation has also been read to suggest that psychosocial factors, such as stresses related to lack of control over one's job, may play as large a role in mortality disparities as differences in material well-being. These and other interpretations of the large relative differences in mortality observed in the Whitehall studies, however, have overlooked that relative differences in mortality among British civil servants would be expected to be large simply because mortality is low within that population. Presumably, the data would also show that relative differences in survival are small within this group.

In the Whitehall studies as elsewhere, it has been observed that relative differences in mortality tend to be greater among the young than the old. And as elsewhere it is overlooked that one would expect such a pattern simply because mortality is lower among the young than the old. That the mortality gradient declines with age in the Whitehall population has been suggested as refuting a contention that poor health causes one to be in lower graded rather than the reverse. While the contention is an unlikely one, that smaller relative difference in mortality among the younger members of the Whitehall cohort does nothing to refute it. That relative differences in mortality in the Whitehall cohorts decline after retirement has been deemed to result from the post-retirement removal from the psychosocial stresses of the hierarchical work environment. But again the premise for the interpretation – that the disparity declines after retirement – lacks a foundation. See A13, B7, D17, D30, D34, D35, D37, D59.

6. *Nordic Health Disparities*. In 1997 a landmark study in *The Lancet* surprised observers by finding that the relatively egalitarian countries of Norway and Sweden showed comparatively large socioeconomic disparities in mortality. The finding continues to be much cited, and observers offer various explanation for the pattern. It goes overlooked, however, that there is reason to expect large relative differences in mortality in Norway and Sweden simply because mortality is low in those countries.

There exists a theory that the greater social cohesion of egalitarian societies improves the health of all members of such societies. To the extent the theory is sound (and it is not implausible) the more egalitarian a society the better would be its overall health and hence the larger would be the perceived disparity in mortality (though one must, of course keep in mind that, in a real sense, more egalitarian societies will likely have narrower differences between health of advantaged and disadvantaged groups, which would offset to some degree the tendency for relative differences in adverse health outcome to be larger because such outcomes are rarer, as discussed in Section B9 *infra*). See A13, B6, B7, D16, D17, D19, D50, D54, D59. See also D43 both with respect to this issue generally and with respect to the applying of the procedure described on the Solutions tab of MHD to mortality disparities in the countries in the *Lancet* study. And see D70 ([Comment on Mackenbach NEJM 2008](#)) regarding the apparent recognition (but subsequent ignoring) by the authors of the *Lancet* study of the implications of overall prevalence with regard to cross-country comparisons.

7. *NCAA's Proposition 48*. At various times, citing the fact that (even allowing for high black representation among collegiate athletes) blacks comprise an extremely high proportion of athletes disqualified from participation in intercollegiate athletics by Proposition 48 or other NCAA rules, observers have maintained that the eligibility requirements are too stringent. But the less stringent the standards the larger will be the black representation among those disqualified. See A2 ([National Law Journal 1989](#)), A3 ([Public Interest 1991](#)). See also the material in the [Employment Tests](#) sub-page on this page.

8. *Terminations from Employment*. In 1992, a U.S. court of appeals decision upheld a challenge to a performance standard on the grounds that the standard was too high and disproportionately disqualified older workers. The court relied on a ten-to-one disqualification disparity for older workers without recognizing that a lower standard would tend to increase the disparity in disqualification rates. See A6 ([Legal Times 1993](#)).

The discipline practices of the Internal Revenue Service were once subjected to intense scrutiny because of widely disparate rates at which blacks and whites were disciplined for workplace infractions. An extensive report was produced attributing the disparity largely to race-neutral factors, and recommending largely race-neutral approaches to address the situations causing the discipline problems. If such approaches are effective, they should reduce the racial disparity in avoiding discipline problems. But they may well increase further the racial disparities in discipline rates that attracted attention to the situation in the first place.

A study of racially disparate termination rates among Postal Service workers explored whether such disparities would be as great at this quasi-federal agency with an excellent reputation as a fair employer as they were in the private sector. But the authors failed to consider that the greater protections afforded public sector workers, by reducing overall termination rates, and hence reducing disparities in keeping one's job, tend to lead to greater disparities in losing one's job. See [A6](#).

9. *Racial Impact of the Three-Strikes Law*. California's three-strikes law has been viewed as having a stark impact on blacks because of the very disproportionately black representation among persons charged under the law. Limitations on prosecutor discretion have been seen as potential means of reducing the impact. Of course anything that reduces actual bias will reduce the impact, however measured. But to the extent that limitation of prosecutor discretion leads generally to a lower rate of applying the law, the tendency would be to further increase the black representation among those charged. The same holds for allowing courts to exercise discretion not to apply the law. See [A8](#).

10. *Studies of Healthcare Disparities Relying on Absolute Differences*. An August 2005 issue of the *New England Journal of Medicine* carried two articles on changes in healthcare disparities both of which relied on absolute differences between rates and both of which examined outcomes that were increasing. One study found that for the most part the disparities had increased; the other found that for the most part disparities had declined. The reconciliation of the two studies lay in the fact that the former study examined relatively rare outcomes (where overall increases tend to increase absolute differences) while the latter examined relatively common outcomes (where overall increases tend to reduce absolute differences). While the latter study found disparities to decrease for procedures, it found no change or increases for favorable clinical outcomes, like control of hypertension among persons diagnosed as hypertensive. And it has been cited in a number of places as showing that healthcare improvements tend to reduce disparities in processes but not in clinical outcomes. But rates of control of adverse health factors within a population suffering from a condition like hypertension tend to be low. Thus, the findings regarding control were to be expected because of the prevalence range for the favorable outcome. See Section A.11 above and Note i.

11. *NHANES Data*. National Health and Nutrition Survey (NHANES) data allow one to examine the consequences of health improvements for a variety of problems. Such data illustrate that measures that generally reduce adverse outcomes like high blood pressure, obesity, diabetes, low folate level in ways that cause all people between point A and point B with regard to some measure – as, for example, enabling everyone with systolic blood pressure (SBP) between 140 and 150 to reduce their SBP to 140 – tend to cause the relative differences in experiencing the adverse outcomes to increase and the relative difference in avoiding those outcomes to decline. Most of the activity for the adverse outcomes like those just discussed occurs in Zone B. See Section B11 above. Thus, the referenced improvements would tend to reduce absolute differences. But if the universe examined was limited to persons deemed to have a condition, with the focus on controlling some aspect of the conditions, overall improvements would tend to increase

absolute differences. See D23, 40, 41 on MHD. See also discussion in item D62 ([Comment on Dowd](#)) of this issue with regard to a study of the way measures to improve folate levels affected racial differences in rates of low folate levels.

12. *Pay-for-Performance Programs*. Pay-for-performance (P4P) is increasingly proposed as a means of generally improving health care. But there have been concerns about the impact of P4P on health disparities. In the United States, this concern largely arose from a study that examined the way a cardiac care report card tended to increase absolute differences in rates of receiving certain procedures. But, given that these were fairly rare procedures, and overall rates increased during the period, the increase in absolute differences does not indicate a meaningful worsening of disparities. The concern arising out of the study has prompted a movement to make effects on healthcare disparities part of a P4P program. The only P4P program that has implemented a measure of effects on health care disparities relies on absolute differences between rates of receiving certain procedures or certain levels of care. Assuming the program achieves its goals or increasing levels of appropriate care, the measurement approach should tend to find higher levels of care associated with larger disparities for relatively rare procedures/outcomes and with smaller disparities for relatively common procedures/outcomes. See the [Pay for Performance](#) sub-page of MHD.

13. *Age Patterns*. As discussed in various places above, observed patterns of relative differences in experiencing or avoiding an outcome are invariably a function of both the described statistical forces and the size of the difference between means. The [Life Table Illustrations](#) sub-page shows how the distributional forces will tend to predominate where prevalence differences are very large, is in the case of comparisons of demographic disparities in mortality among comparatively young age groups and much older age groups.