Measuring Health and Healthcare Disparities

2013 Research Conference of the Federal Committee on Statistical Methodology

James P. Scanlan
Attorney at Law
Washington, DC
jps@jpscanlan.com

Table of Contents

Table of Figures and Tables

Abstract

Researchers employ a number of methods to measure health and healthcare disparities. The most common of these measures are relative differences in either a favorable outcome or the corresponding adverse outcome, absolute differences between rates, and odds ratio. Research relying on these measures has been problematic, however, because of a failure to recognize the patterns by which each measure tends to be affected by the prevalence of an outcome. The rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. Thus, as mortality generally declines, relative differences in mortality tend to increase while relative differences in survival tend to decrease; as rates of appropriate healthcare generally increase, relative differences in rates of failing to receive appropriate care tend to increase while relative differences in receipt of appropriate care tend to decrease. Absolute differences and odds ratios tend also to be affected by the prevalence of an outcome though in a more complicated way than the two relative differences. Broadly, as uncommon outcomes become more common absolute differences tend to increase; as already common outcome outcomes become even more common absolute differences tend to decrease. Further, as the prevalence of an outcome changes, absolute differences tend to change in the same direction as the smaller relative difference. Differences measured by odds ratios tend to change in the opposite direction of absolute differences as the prevalence of an outcome changes. Although these patterns are apparent in a wide range of data, they are little known among individuals and institutions attempting to determine whether health disparities are increasing or decreasing over time or are otherwise larger in one setting than another. This paper illustrates these patterns and describes a method of appraising differences in the circumstances of two groups reflected by a pair of outcome rates that is unaffected by the prevalence of an outcome.

Introduction

In recent decades many billions of dollars have been devoted of research into health and healthcare disparities. Very little of that research had been sound, however, as a result of the failure of those conducting it to recognize the patterns by which, for reasons related to the shapes of the underlying distributions of factors associated with experiencing an outcome, measures of differences between outcome rates of advantaged and disadvantaged groups tend to be systematically affected by the prevalence of an outcome. Section A describes the patterns themselves and some of their implications with respect to the interpretation of data on group differences in outcome rates in the context of health and healthcare disparities. Section B illustrates that there can be only one correct answer regarding whether the forces causing health and healthcare outcome rates of advantaged and disadvantaged groups to differ are stronger in one setting than another and explains a method for divining that answer and quantifying those forces in a way that is unaffected by the prevalence of an outcome. Section C describes the disarray of health and healthcare disparities research as a result of the failure of individuals and institutions to recognize the ways standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome and the particular implications of such failure with respect to the inclusion of disparities elements in pay-for-performance programs.

A. Patterns by Which Relative Differences in Favorable and Adverse Outcomes Tend to be Systematically Affected by the Overall Prevalence of an Outcome and Implications of those Patterns with Regard to Health and Healthcare Disparities Research

1. The Principal Patterns

This section describes the patterns by which, for reasons related to the shapes of the underlying distributions, standard measures of differences between outcome rates of advantaged and disadvantaged groups tend be systematically affected by the prevalence of an outcome and, hence, why none of those measures can effectively quantify the strength of the forces causing outcome rates of such groups to differ without taking the prevalence of an outcome into account.

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

Income data show that the lower the income level, the greater is the relative differences between the rates at which advantaged and disadvantaged group fail to reach the level while the smaller is the relative difference between the rates at which such groups reach the level. More concretely, such data show that reducing poverty tends to increase relative differences in poverty rates while

reducing relative differences in rates of avoiding poverty.[1] National Health and Nutrition Survey data show that increasing folate levels, hence reducing rates of low folate, tends to increase relative difference in low folate while reducing relative difference in adequate folate and that lowering blood pressure tends to increase relative differences in hypertension while reducing relative differences in rates of avoiding hypertension.[2] Life Tables show that the lower the age, the greater tend to be relative differences in failing to reach it while the smaller tend to be relative differences in reaching it.[3]

But in order to illustrate the patterns in their most essential form, and to provide the framework for a sound method of appraising the difference between the circumstances of two groups reflected by a pair of outcome rates, I base the illustrations below on normally distributed test score data.

Figure 1 is based on a situation where two groups have normal test score distributions with means that differ by half a standard deviation (and where the standard deviations of the distributions are equal).[4] The numbers at the bottom of the figure are the fail rates of the advantaged group, which are used as benchmarks for overall prevalence of test failure. The blue line with the diamond marker tracks the ratio of the fail rate of the disadvantaged (*i.e.,* lower-scoring) group (DG) to the fail rate of the advantaged (*i.e.*, higher-scoring) group (AG) and the red line with the square marker tracks the ratio of the pass rate of the AG to the pass rate of DG at each benchmark.[5] From left to right, the lines illustrate the effects on the two ratios of serially lowering the test cutoff from a point where almost everyone fails to a point where almost everyone passes, in each instance enabling all persons with scores above each new cutoff now to pass the test. And we see that as test failure becomes less common and test passage becomes more common, the relative difference in failure rates increases while the relative difference in pass rates decreases. Thus, we observe the common pattern whereby as the prevalence of an outcome changes relative differences in experiencing it and relative differences in avoiding it tend to change in opposite directions.

---

[1] See the table and figures in Scanlan 2006a. See also Scanlan 1987, 1991, 1994, 2000.

[2] See Scanlan 2008 and the NHANES Illustrations subpage of the Scanlan's Rule page of jpscanlan.com.

[3] See the Life Tables Illustrations subpage of the Scanlan's Rule page of jpscanlan.com.

[4] The data underlying Figures 1 through 3 may be found in Table 1 of Scanlan 2006b.

[5] The ratio is commonly termed the "rate ratio, "risk ratio," or "relative risk" (RR). The relative difference between rates is RR minus 1 where RR is greater than 1 (in which case the larger the RR the larger the relative difference) and 1 minus RR where RR is less than 1 (in which case the smaller the RR the larger the relative difference). In recent years I have generally used the larger figure as the numerator of the RR for both favorable and adverse outcomes. Thus, as to both outcomes, the larger the RR the larger the relative difference. Whether one uses the larger or smaller figure as the numerator in RR can affect the size of a relative difference. For example, in a case where rates are 30 percent and 40 percent, the former could be deemed 25 percent less than the latter or the latter could be deemed 33 percent greater than the former. But choice of numerator is irrelevant to issues about the comparative sizes of relative differences addressed here. Determinations as to which is the larger relative difference reflected by two pairs of rates of experiencing an outcome will always hold regardless of which figure is used as the numerator of the ratio.
.

Figure 1. Ratios of (1) DG Fail Rate to AG Fail Rate and (2) AG Pass Rate to DG Pass Rate at Various Cutoff Points Defined by AG Fail Rate
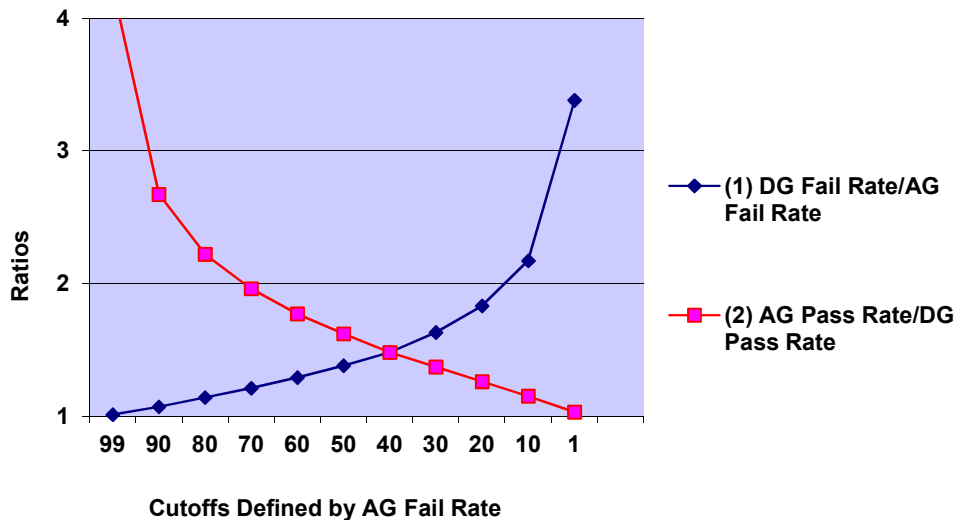
The pattern by which a change in the prevalence of an outcome tends to increase one relative difference while decreasing the other may seem counterintuitive at first. But one element of the pattern is implied in the other, if, indeed, the two elements are not exactly the same thing. That is, for example, if reducing the prevalence of an outcome increases the relative difference in experiencing the outcome, it follows that increasing the prevalence of an outcome decreases the relative difference in experiencing it. And if one outcome decreases in prevalence (hence, increasing the relative difference in experiencing it), the opposite outcome will necessarily increase in prevalence (hence, reducing the relative difference in that outcome).

There are two important corollaries to the pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. First, as an outcome changes in overall prevalence, groups with lower baseline rates for the outcome will tend to experience larger proportionate changes in those rates than groups with higher baseline rates for the outcome, while groups with higher baseline rates for the outcome will tend to experience larger proportionate changes in the opposite outcome. That is, in the case of the lowering of cutoffs reflected in Figure 1, for example, said lowering would cause larger proportionate decreases in failure rates for AG than DG, while causing larger proportionate increases in pass rates for DG than AG. Second, the rarer an outcome, the larger will tend to be the proportion groups most susceptible to the outcome comprise of both (a) the population experiencing the outcome and (b) the population failing to experience the outcome. That is, for example, the lowering of cutoffs reflected in Figure 1 would cause DG to comprise a

larger proportion of those who fail and a larger proportion of those who pass than it did before the cutoff was lowered.[6]

Appraisals of the comparative size of differences between outcome rates measured in absolute (percentage point[7]) terms or in terms of odds ratios are unaffected by which outcome one examines.[8] But in order for a measure to effectively quantify the difference between the circumstances of two groups reflected by a pair of outcome rates (or, put another way, to quantify the forces causing the rates to differ) a measure must remain constant when there occurs a general change in the prevalence of an outcome akin to that effected by the lowering of a test cutoff. And, like the two relative differences, the absolute difference and the difference measured by the odds ratios tend to change systematically as the prevalence of an outcome changes. They do so, however, in a more complicated way than the two relative differences.

Roughly, as uncommon outcomes (less than 50 percent for both groups being compared) become more common, absolute differences between rates tend to increase; as common outcomes (greater than 50 percent for both groups being compared) become even more common, absolute differences tend to decrease. In cases where the outcome is either common or uncommon, the pattern of direction of changes in absolute differences as the prevalence of an outcome changes

---

[6] The third and fourth last columns of Table 1 of Scanlan 2006a illustrate the pattern in circumstances where poverty is the outcome being reduced in prevalence. The table thus shows, for example, that reducing poverty will tend to cause blacks to comprise both a larger proportion of the combined black and white poor populations, and a larger proportion of the combined black and white non-poor populations, than they did previously.

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

[8] It would be correct to say both (a) that the size of the absolute difference is unaffected by which outcome one examines and (b) that the comparative size of differences measured in absolute terms is unaffected by which outcome one examines. While it is sometimes said that the size of the odds ratio or the difference measured by the odds ratio is unaffected by the prevalence of an outcome, neither is precisely correct. A group's odds of experiencing an outcome is the group's rate of experiencing the outcome divided by its rate of failing to experience the outcome. There are four possible ways to calculate the odds ratio. These include (1) AG's odds of experiencing the favorable outcome to DG's odds of experiencing the favorable outcome; (2) DG's odds of experiencing the adverse outcome to AG's odds of experiencing that outcome; (3) AG's odds of experiencing the adverse outcome to DG's odds of experiencing that outcome; and (4) DG's odds of experiencing the favorable outcome to AG's odds of experiencing that outcome. Methods (1) and (2) reach the same result as each other; Methods (3) and (4) also reach the same result as each other, and which are the reciprocals of the odds ratios calculated with Methods (1) and (2). For example, where AG's favorable outcome rate is 80 percent and DG's favorable outcome rate is 63 percent, Methods (1) and (2) would yield an odds ratio of 2.35 and Methods (3) and (4) would yield an odds ratio of .43, which is the reciprocal of 2.35. It is because an odds ratio calculated by any of the methods will be the same as, or the reciprocal of, the odds ratios calculated by the other methods that observers commonly say that the odds ratio is unaffected by which outcome one examines. But since whether the odds ratio is above 1 or below 1 is affected by which outcome is examined (and which group's odds is used as the numerator), it is not precisely correct to say that the difference measured by the odds ratio is unaffected by whether one examines the favorable or the adverse outcome, just as it would not be correct to say that the size of the relative difference is unaffected by which rate is used in the numerator of the rate ratio (see note 5 *supra*). For example, with respect to the 2.35 and .43 odds ratios just noted, in the first case one odds is 135 percent greater than the other while in the second case one odds is 57 percent less than the other (which figures are not reciprocals of one another). It is true, however, that the determinations as to the comparative size of differences between various pairs of rates as measured by the odds ratio will be unaffected by the method by which the odds ratio is calculated.

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

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

Figure 2. Absolute Differences between Rates of AG and DG Pass (or Fail) Rates at Various Cutoff Points Defined by AG Fail Rate
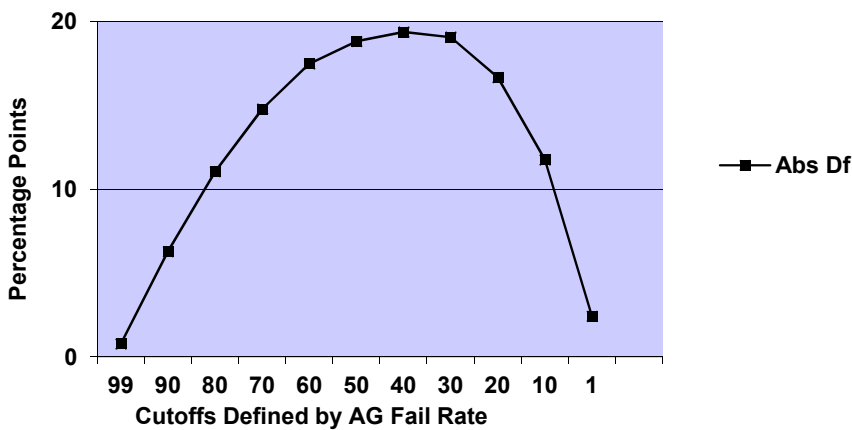
As the prevalence of an outcome changes, the difference measured by the odds ratio tends to change in the opposite direction of the change in the absolute difference. Figures 3 charts the odds ratio as it would be calculated based on either (a) AG's odds of passing to DG's odds of passing or (b) DG's odds of failing to AG's odds of failing (see note 8 *supra*) according to the specifications underlying Figures 1 and 2.

6

Figure 3. Ratio of DG Odds of Failure to AG Odds of Failure at Various Cutoff Points Defined by AG Fail Rate
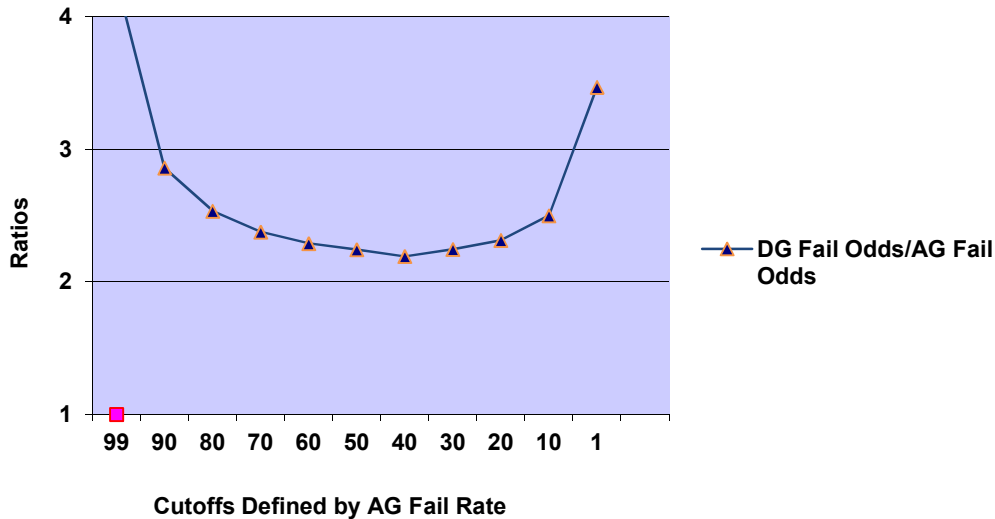
The described patterns of correlations between absolute differences and differences measured by odds ratios with the prevalence of an outcome have their corollaries as well. As with the patterns of correlations themselves, however, such corollaries do not lend themselves to succinct summary description. It suffices here to note that when uncommon outcomes become somewhat more common higher baseline rates tend to show larger percentage point increases than lower baseline rates; when common outcomes becomes even more common lower baselines rate tend to show larger percentage point increases than higher baseline rates.

2. Some Implications of the Patterns

The following are some of the implications of the above-described pattern of relative differences between the rates at which advantaged and disadvantaged groups experience favorable or adverse outcomes that are pertinent to the interpretation of data on health and healthcare disparities.

- When adverse health outcome decline, relative differences in rates of experiencing those outcomes tend to increase while relative differences in rates of experiencing the corresponding favorable outcomes tend to decrease. The opposite occurs when adverse health outcomes increase in overall prevalence.

- When overall rates of receiving beneficial health procedures or care (*e.g.*, mammography, immunization, prenatal care, adequate hemodialysis, coronary artery bypass grafting) increase, relative differences in rates of receiving such procedures or care tend to decrease while relative differences in rates in failing to receive them tend to increase.

- The more extreme the form of an adverse outcome, the greater will tend to be the relative difference in experiencing it and the smaller will tend to be the relative difference in avoiding it. Thus, for example, relative differences in poor health will tend to be greater than relative differences in health-less-than-good, while relative differences in health-better-than-poor will tend to be smaller than relative differences in health-good-or-better; relative difference in rates of falling below the cutoff for very low birthweight will tend to be greater than relative differences in falling below the cutoff for low birthweight, while relative differences in rates of falling above the cutoff for very low birthweight will tend to be smaller than relative differences in falling above the cutoff for low birthweight.

- The more survivable is a particular type of cancer, the smaller will tend to be relative differences in surviving it while the larger will tend to be relative differences in failing to survive it.

- Generally reducing blood pressure will tend to increase relative differences in hypertension while reducing relative differences in rates of avoiding hypertension; generally improving folate levels will tend to increase relative differences in low folate while reducing relative differences in adequate folate.

- Relative racial and gender differences in receipt of particular therapies will tend to be smaller, while relative differences in rates of failure to receive the therapies will tend to be larger, among subjects whose symptom/condition profiles call for generally higher rates of receipt of the therapies than among subjects whose symptom/condition profiles call for generally lower rates of receipt of the therapy.

- Relative differences in adverse outcome rates will tend to be large among comparatively advantaged subpopulations (where such outcomes are less common) – *e.g.*, racial differences in infant mortality where parents are highly educated compared with where parents are less educated; racial differences in low birth weight among low risk groups compared with high risk groups; racial, gender, and socioeconomic differences in mortality among the young compared with the old; occupational differences in mortality and morbidity among British civil servants compared with the United Kingdom population at large; racial and socioeconomic differences in failing to receive appropriate care among the insured compared with the uninsured – while relative differences in the opposite, favorable outcomes tend to be small among those subpopulations.

The following are some of the implications of the two corollaries to the described patterns of relative differences that are pertinent to the interpretation of data on health and healthcare disparities.

- Factors that exacerbate a condition will tend to cause larger proportionate increases in adverse outcome rates for groups with lower baseline rates, while causing larger proportionate decreases in the corresponding favorable outcome rates for other groups. For example, chronic conditions will tend to increase rates of health-less-than-good

proportionately more among higher socioeconomic groups while reducing rates of health-good-or-better proportionately more among lower socioeconomic groups.

- As adverse outcomes decline, those measuring disparities in terms of the concentration index applied to the adverse outcome will tend to find the disparity to have increased (*i.e..*, the adverse outcome will have become more concentrated in the disadvantaged group) while those measuring disparities in terms of the concentration index applied to the favorable outcome will tend to find the disparity to have decreased (*i.e.*, the favorable outcome will have become less concentrated in the advantaged group).

The following are some of the implications of the above described patterns of absolute differences between outcome rates of advantaged and disadvantaged group rates that are pertinent to the interpretation of data on health and healthcare disparities.

- As uncommon procedures (*e.g.*, coronary artery bypass grafting, knee replacement, certain types of immunization) increase, absolute differences tend to increase; as common procedures (*e.g.*, mammography, prenatal care, certain types of immunization) increase, absolute differences tend to decrease. Similarly, where appropriate healthcare rates are low, improvements will tend to increase absolute differences between rates; where rates are high, improvements will tend to reduce absolute differences.

- As rates of receipt of some type of care increase from very low levels to very high levels (as commonly occurs with respect to certain types of immunization and cancer screening, especially ones that have been recently developed or that have been recently recognized as important), absolute differences will tend to increase for a while and then decrease.

- As survival rates increase for cancers with generally low survival rates, absolute differences will tend to increase; as survival rates increase for cancers with generally high survival rates, absolute differences will tend to decrease.
  .
- For outcomes with generally low rates or in settings with generally low rates for an outcome, higher rates for the outcome will tend to be associated with larger absolute differences between rates; for outcomes with generally high rates or in settings with generally high rates for an outcome, higher rates will tend to be associated with lower absolute differences between rates. For example, with regard to types of outcomes for which favorable outcome rates tend to be low, higher-performing hospitals will tend to show larger absolute differences between rates than lower-performing hospitals; with regard to types of outcomes for which favorable outcome rates tend to be high, higher-performing hospitals will tend to show smaller absolute differences between rates than lower-performing hospitals.

3. Some Qualifying Considerations

The illustrations in Figures 1 through 3 are based on perfectly normal distributions. But the patterns whereby the two relative differences change in opposite directions as the prevalence of

an outcome changes, as illustrated in Figure 1, would hold so long as the distributions are not highly irregular. They would exist, for example, when distributions are uniform (rectangle-shaped).

The described patterns of absolute differences and odds ratios will not necessarily hold when distributions are not normal. With uniform distribution, for example, absolute differences would remain the same as the prevalence of an outcome changes up to the point where one group's rate for the outcome or its opposite reaches 0 percent/100 percent. Examples of the ways odds ratios change when the distributions are truncated parts of normal distributions may be found in Figures 8 and 10 (slides 13 and 15) of Scanlan 2008 and in Figure 3 accompanying the Credit Score Illustrations subpage of the Scanlan's Rule page of jpscanlan.com.

But, as reflected in the illustrations of these patterns mentioned at the outset, the underlying risk distributions for most outcomes do tend toward the normal. And, while the possibility that the underlying distributions may not be normal is an important consideration with respect to appraising the strength of an association reflected by a pair of outcome rates in a way unaffected by the prevalence of an outcome, that possibility does not detract from the importance of recognizing that standard measures of differences between outcome rates will almost invariably be affected by the prevalence of an outcome in some manner. For the fact that a measure of differences between outcome rates is in *any manner* affected by the prevalence of an outcome renders it an unsound measure of association, even when (or especially when) we do not know just how it tends to be affected.[9]

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

4. Failure of Researchers to Understand the Patterns

Few studying health and healthcare disparities are aware of any pattern by which a measure may be affected by the prevalence of an outcome. Indeed, researchers have commonly relied on a chosen measure without evidencing any recognition that other measure might yield contrary results, much less that they would tend systematically to do so. That occurs even when a measure that in fact would yield a contrary result in the situation examined may be the most commonly used measure. The extent of the misunderstanding in health disparities research is particularly evident in the discussion of disparities in cancer outcomes, where observers commonly refer to relative differences in survival and relative differences in mortality interchangeably, often purporting to examine one while in fact examining the other. They do so

---

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

without recognizing that the two tend to change in opposite directions as survival generally increases or that more survivable cancers tend to show larger relative differences in mortality but smaller relative differences in survival than less survivable cancers.[10]

Some researchers have lately given increased attention to the possibility that the relative difference they happen to be examining and the absolute difference may yield different conclusions about such things as whether disparities are increasing or decreasing over time and have stressed the importance of reporting both measures where that occurs. But they have generally done so without even recognizing that there even exist two relative differences, even though anytime a mentioned relative difference and the absolute difference have changed in opposite directions, the unmentioned relative difference will necessarily have changed in the opposite direction of the mentioned relative difference and the same direction as the absolute difference. See Scanlan 2012d, 2012e.

Commencing in 2004, there has been some recognition of the above-described patterns by which measures tend to be affected by the prevalence of an outcome, particularly of the pattern whereby relative differences in the favorable outcome and in the corresponding adverse outcome tend to change in opposite directions as the prevalence of an outcome changes. But such recognition has usually not shown a clear understanding of the forces at work and has yet to have a useful effect on health and healthcare disparities research.[11]

What was potentially the most notable of instance of a recognition of these patterns was by statisticians of the National Center for Health Statistics, who in a number of official or unofficial documents or articles between 2004 and 2009 recognized that determinations of the directions of changes in disparities over time would commonly turn on whether one examines relative differences in favorable outcomes or relative differences in adverse outcomes.[12] But rather than also recognizing that such pattern called into question the utility of either relative difference for quantifying the strength of the forces causing outcome rates to differ, and providing guidance either on how one might employ those or other measures while taking the effects of prevalence into account or on the availability of measures that are unaffected by prevalence, NCHS merely recommended that all health and healthcare disparities should be measured in terms of relative differences in adverse outcomes.

---

[10] See the Mortality and Survival Page of jpscanlan.com.

[11] Principal materials addressing aspects of the patterns by which measures tend to be affected by the prevalence of an outcome are discussed in Section E.7 of the Measuring Health Disparities page of jpscanlan.com. The treatments were occasioned by, or in response to, Scanlan 1991, 1994, 2000, 2006, 2006a.

[12] Published National Center for Health Statistics documents in some manner addressing issues concerning the ways relative differences in favorable and adverse outcomes lead to different conclusions about the directions of changes in disparities over time include Keppel et al. 2004, Keppel et al. 2005, Keppel and Pearcy 2005, Keppel and Pearcy 2006, and Keppel and Pearcy 2009. Unpublished NCHS materials addressing the issues are discussed in Section E.7 of the Measuring Health Disparities page of jpscanlan.com.

I will give further attention to NCHS in the course of describing the general disarray of health and healthcare disparities research in Section C.

B. A Theoretically Sound Method of Appraising the Differences in the Circumstances of Two Groups Reflected by a Pair of Outcome Rates

1. The Method Explained in Context of Refuting Claims That Choice of Health Disparities Measure Involves a Value Judgment

This section uses a hypothetical situation where employment bias at issue in order both to refute contentions that two measures yielding contrary results as to the comparative size of a disparity can both be in some way correct and that a value judgment is involved in choosing between them, and to demonstrate a method by which one may effectively appraise the differences between the circumstances of two groups reflected by a pair of outcomes rates/quantify the strength of the forces causing the rates to differ. The points of this section are implicit in Section A. Nevertheless, I believe that a hypothetical where most observers should easily recognize that there exists an underlying reality concerning the comparative strength of the forces causing the rates to differ, and that there can be only one such reality, contributes to the explication of the principal ideas of this paper. The illustration also suggests a theoretically sound method for appraising the differences in the circumstances of two groups reflected by a pair of outcome rates that is unaffected by the prevalence of the outcome.

Increasingly, health and healthcare disparities researchers discuss relative and absolute differences in circumstances where the examined relative difference provides a different interpretation as to the comparative size of a disparity from that provided by the absolute difference. Sometimes they do so simply to provide as complete a picture as possible. But sometimes researchers maintain that both measures provide valid information regarding a particular aspect or the matter and suggest that a value judgment is involved in the choice between measures. To the extent that such values are articulated, they generally involve an argument that absolute differences best reflect the extra burden of a disease on a disadvantaged group, while relative differences better reflect the degree of inequity. Harper and Lynch 2005, Harper et al. 2010. As discussed above, however, discussions about the choice between relative and absolute differences have yet to show a recognition that there exists two relative differences much less that, as a mathematical fact, anytime a mentioned relative difference yields a different conclusion as to the comparative size of two disparities – with respect, for example, to whether a disparity has increased or decreased over time – the unmentioned relative difference will yield a conclusion that is the opposite of that yielded by the mentioned relative differences and the same as the absolute difference. That fact aside, however, such discussions fail to reflect an understanding that the purpose of examining a pair of outcome rates is to understand the forces causing the rates to differ (including the components of those forces), whether the strength of those forces is increasing or decreasing over time, what causes them to increase or decrease over time or otherwise to be stronger in one setting than another, and what the comparative strength of the forces may suggest about related subjects.

Table 1 presents hypothetical hire rates of advantaged and disadvantaged groups applying for work at four employers, along with rate ratios for hire and rate ratios for rejection, as well as the absolute difference between rates and an odds ratio (reflecting the advantaged group's odds of selection over the disadvantaged group's odds of selection, see note 8). In a situation where it is

assumed that for each employer the qualifications of the applicants from the advantaged group do not differ from the qualifications of the disadvantaged group and all differences in outcome rates result from employer bias, the question to be addressed is how might the employers be ranked, from highest to lowest, according to level of bias. The numbers in parentheses for each measure reflect the ranking pursuant to that measure. I note in advance that I could make the same point more simply with two rows of data. But I use four rows to illustrate some of the issues concerning the ways the absolute differences and odds ratios alter their directions of change as overall prevalence changes. The use of four rows will also facilitate the discussion of certain issues about healthcare disparities and pay-for-performance programs in Section C.2.

Table 1. Hypothetical Patterns of Hiring Rates of Applicants from an Advantaged Group (AG) and a Disadvantaged Group (DG) at Four Employers, with Disparity Measures

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

There are four principal ways observers might rank the degree of bias of these employers. Those who rely on relative differences in favorable outcomes, such as might commonly occur in an employment discrimination case involving hiring or promotion, would rank them A,B,C,D. Those who rely on relative differences in adverse outcomes as the National Center for Health Statistics would do and as might also be done in an employment discrimination case where the favorable outcome is retention and the adverse outcome is termination, would rank them D,C,B,A, the opposite of the first approach.

A third approach would be to rank them according to the absolute difference between rates, such as researchers at the Health Care Policy Department of Harvard Medical School and the Centers for Disease Control and Prevention would commonly do and as Federal Reserve Board economists have done in some lending disparities studies. That ranking would be C,B,D,A. And those who rely on odds ratios to measure disparities, such as those who would attempt to evaluate the situation by means of logistic regression might do, would rank them A,D,B,C, the opposite of the ranking based on absolute differences.

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

The reasoning would hold as well if it was not known whether any of the employers was biased and the question to be answered involved the degree of difference in the qualifications of

applicants of the advantaged and disadvantaged groups that would be necessary to explain each difference in outcome rates as a result of something other than bias. The reasoning would also hold if, instead of representing the situations of four employers, the rows of data represented one employer at four points in time and the question to be answered was whether discrimination increased or decreased from each point in time to the next. It would hold as well if rows A to D reflected the hiring patterns of applicants with increasingly greater qualifications and the question to be answered involved what inferences about processes might be drawn from the comparative size of the disparities among applicants with stronger qualifications versus applicants with weaker qualifications.

Consider further how one might analyze these data (a) in order to determine whether it would be worthwhile to study the practices at the employers with larger disparities compared with the practices at employers with smaller disparities for purposes of identifying what types of practices are associated with larger and smaller disparities or (b) in a situation where the rows involved hiring patterns from year to year, in order to determine whether a manager hired specifically to reduce disparities should be deemed to have caused the situation to improve or caused the situation to worsen.

With respect to the implications of the differing patterns among applicants with varying levels of credentials, there exists a view, based on smaller relative differences in selection rates among applicants with stronger credentials than among applicants with weaker credentials, that employers tend to treat applicants of different demographic groups more equally when objective indicators of qualification are present, but will rely more heavily on stereotypes when objective indicators of qualifications are not present. See Kim 2013. Yet, those relying on relative differences in adverse outcomes like NCHS – and who would regard the disparities to increase as qualifications increased – would presumably reach exactly opposite interpretations of the implications of these patterns. Those relying on absolute differences, however, would base their interpretations on the view that bias tends to increase as qualifications increase until qualifications reach a certain point, at which time bias decreases as qualifications increase. Those relying on odds ratios, by contrast, would base their interpretations on the view that bias tends to decrease as qualifications increase until qualifications reach a certain point, at which time bias tends to increase. With respect to none of these issues can a value judgment aid one in appraising the strength of the forces causing outcome rates to differ in one setting compared with the strength of such forces in another setting.

What then can in fact be divined about the comparative degrees of bias reflected in the four rows? Each situation is based on the specifications underlying the figures in Section A where means of the underlying normal distributions differ by half a standard deviation. Thus, to the extent that we can measure the bias, we can only conclude that the degree of bias is exactly the same in each case. At any rate, there exists no rational argument that the degree of bias reflected in any of the rows differs from that in another. Hence, any measure that suggests that four situations involve different degrees of bias is an unsound measure.

Implicit in the illustration in Table 1, as well as in the reasoning throughout Section A, is that the only theoretically sound solution to the question of how to appraise a difference in circumstances reflected by a pair of rates – whether with regard to the degree or likelihood of bias in the employment setting, the comparative size of health or healthcare disparities in different settings, or any other matter as to which demographic differences in outcome rates are matters of concern

– involves deriving from a pair of rates the difference between the means of the hypothesized underlying distributions measured in terms of percentages of a standard deviation. I commonly term the value yielded by this approach "EES" for estimated effect size (and statisticians will recognize it as the result yielded by a probit analysis).[13]

Thus, just as we are able on the basis of a particular difference between means of the underlying distribution – half a standard deviation in each of the above illustrations – to determine the rate for DG corresponding to any rate for AG (or vice versa), the probit allows us to estimate from any pair of rates the difference between the means of the underlying distributions. For example, when favorable outcome rates for AG and DG are 30 percent and 10 percent we can estimate that the difference between the means is .757 standard deviations; when those rates are 15 percent and 5 percent we can estimate that the differences between the means is .608 standard deviations. To put these standard deviation figures in perspective, the .757 standard deviation difference reflects a situation where, in testing terms, approximately 22 percent of DG scores above the mean for AG; the .608 standard deviation difference reflects a situation where approximately 27 percent of DG scores above the mean for AG. Similar illustrations, by tenths of a standard deviation, based on situations at different levels of overall prevalence where the disadvantaged group's favorable outcome rate is four-fifths of the advantaged group's favorable outcome rate may be found in Table 1 of the Four-Fifths Rule subpage of the Disparate Impact page of jpscanlan.com.[14]

2. Illustrations of the Method in Context of Varying Interpretations of the Comparative Size of Health Disparities Based on Standard Measures

Tables 2 through 12 provide further perspective on the EES figure, as well as illustrate the problematic nature of standard measures of differences between outcome rates, in various contexts where the prevalence of an outcome changed over time or is larger with regard to one subpopulation or type of outcome than another. The tables also provide useful background for consideration of the disarray of health and healthcare disparities research that is the subject of Section C.

Each table shows the favorable outcome rates for the advantaged and disadvantaged group, along with the standard measures of differences between outcome rates that were shown in Figures 1 through 3 (RRFav = rate ratio for favorable outcome; RRAdv = rate ratio for the adverse outcome; AD = absolute difference; OR = odds ratio), as well as the EES.

Table 2 presents the white and Hispanic rates of receipt of mammography cited in Keppel 2005 et al. and Keppel and Pearcy 2005 when NCHS determined that for purposes of Healthy People 2010 these disparities would be deemed to be increasing on the basis of an increase in the relative difference in the adverse outcome rather than decreasing on the basis of the decrease in

---

[13] The method is more fully explained on the Solutions subpage of the Measuring Health Disparities page of jpscanlan.com.

[14] A principal purpose of the Four-Fifths Rule subpage is to show that the Four-Fifths Rule of the Uniform Guidelines on Employee Selection Procedures is an illogical measure of association. See Section A.3 of Scanlan 2013a. See also Scanlan 2013 and the Illogical Premises and Illogical Premises subpages of the Scanlan's Rule page of jpscanlan.com.

the relative difference in the favorable outcome (which, as discussed in Scanlan 2000, tended to be the common approach to measuring disparities in things like mammography). One remarkable thing about the NCHS action in this regard is that, so far as I am aware, no one found such a reversal jarring or questioned whether there might be an underlying reality as to whether the forces causing the rates to differ had grown smaller or larger. Presumably, anyone at NCHS who had theories as to the cause of a decrease in disparity would have to revise those views to accord with the revised determination as to how disparities should be measured. In any case, the EES did increase very slightly (from .19 to .21). But there is little reason to believe that the change reflected anything other than random variation or an irregularity in the distributions of factors associated with receiving a mammogram.

Table 2. Changes in Mammography Rates of Whites and Hispanics between 1990 and 2002, from Keppel et al. 2005, with Disparity Measures

Year	Wh Mam Rate	Hi Mam Rat	RRFav	RRAdv	AD	OR	EES
1990	52.70%	45.20%	1.17	1.16	0.075	1.35	0.19
1998	68.00%	60.20%	1.13	1.24	0.078	1.40	0.21

Table 3 presents information on mammography rates of the highest and lowest socioeconomic groups from Harper et al. 2009 (a study whose authors included the authors of the principal article concerning the importance of value judgments in choosing among health disparities measures and authors of the health disparities measurement guides discussed in Section C.1.d and C.1.f *infra*). The abstract for the study highlighted what it described as a 161 percent increase in relative area-socioeconomic differences in mammography. But the study in fact analyzed relative differences in failure to receive mammography. Such fact was made clear enough in the article, which cited Keppel et al. 2005 and Keppel and Pearcy 2005 to the effect that the size of relative differences can be affected by which outcome one examines. The study did not, however, note that both references also stated that the directions of changes over time are commonly affected by whether one examines the relative difference in the favorable outcome or the relative difference in the adverse outcome.[15]

Because of some presentation issues in Harper et al. 2009, I am not able to divine precisely how the authors derived the 161 percent increase for relative differences in failure to receive mammography. But Table 3 presents data from the study showing that the relative difference in failure to receive mammography increased by 236 percent (30 percent increased to 98 percent)

---

[15] Sections C.1.a and C.1.b *infra* address the confusion in the discussions of disparities issues arising from the pattern of stating indicators in terms of favorable outcomes while analyzing them in terms of adverse outcomes. With regard to healthcare outcomes this pattern is a largely a consequence of a conscious decision of the National Center for Health Statistics (apparently followed by the Agency for Healthcare Research and Quality) to continue to describe various healthcare disparities in favorable outcome terms while analyzing them in adverse outcome terms. The situation is rather different with respect to discussions of cancer outcomes where observers seem not to recognize the distinction between relative differences in mortality and relative differences in survival, or, at any rate, not to recognize the possibility that they might yield contrasting interpretations as to the comparative size of disparities, as discussed on the Mortality and Survival page of jpscanlan.com.

while the relative difference in receipt of mammography decreased by 64 percent (111 percent reduced to 40 percent). So the difference between the patterns of changes for the two relative differences was dramatic. The change in the EES figures from .60 to .62 standard deviation suggest that the strength of the forces causing outcome rates to differ was essentially unchanged.

Table 3. Changes in Mammography Rates of Highest and Lowest Socioeconomic Groups between 1987 and 2004, from Harper et al. 2009, with Disparity Measures

Year	High	Low	RRF	RRA	AD	OR	EES
1987	36.30%	17.20%	2.11	1.30	0.19	2.74	0.60
2004	77.40%	55.20%	1.40	1.98	0.22	2.78	0.62

One might regard the EES figures as also suggesting that the socioeconomic disparities reflected in Table 3 are much larger than the ethnic disparities reflected in Table 2. But even with a theoretically sound measure, such comparisons are problematic given that the size of socioeconomic disparities are functions of the proportions of the total populations that the analysts choose to include in the advantaged and disadvantaged groups whose rates are examined. That is, one will reach different conclusions based on comparisons of the highest and lowest tenths of the population from those based on comparisons of the highest and lowest quarters of the population.

Table 4 presents data on cervical cancer screening rates for the most and least deprived groups in the United Kingdom, from Baker and Middleton 2003, during a period when rates were generally increasing. This study relied on relative differences in the favorable outcome to find substantial decreases in disparities. The relative difference in screening rates decreased by 74 percent (115 percent reduced to 30 percent). By contrast, those relying on relative difference in failure to be screened, as in the study underlying Table 3 (and as NCHS would do), would have found a 470 percent increase (from 283 percent to 1614 percent).[16] The EES figures indicate that there was a fair sized increase in the strength of the forces causing the rates of the two groups to differ (.22 standard deviations) as well as the fact that the disparity was very large at both points in time. While in the Harper et al. 2009 study the .02 standard deviation increase in EES for mammography involved a small increase in the absolute difference, here the much larger increase in the cervical cancer screening EES involved a decrease in the absolute difference, a fact I merely note to illustrate the confusion that can be wrought by standard measures.[17]

---

[16] Not everyone would characterize the change in this manner. NCHS and AHRQ both measure changes in disparities in terms of percentage point changes in the relative difference, which would be a 1331 percentage point in the case of the increase here. While NCHS describes such changes in terms of "percentage point changes," AHRQ describes them as "percent changes," potentially resulting in considerable confusion. See the Percentage Points subpage of the Vignettes page of jpscanlan.com .

[17] There exist issues as to how one might characterize changes in EES where different baseline figures are involved, including whether one might regard one change to be larger than another. Those may or may not be resolvable objectively. But these issues are related to theoretically sound measures rather than to theoretically unsound measures.

Table 4. Changes in Cervical Cancer Screening Rates of Least and Most Deprived Groups in the United Kingdom between 1991 and 1999, from Baker and Middleton 2003, with Disparity Measures

Year	Least Dep	Most Dep	RRFav	RRAdv	AD	OR	EES
1991	84.09%	39.03%	2.15	3.83	0.45	8.26	1.28
1999	98.60%	76.00%	1.30	17.14	0.23	22.24	1.49

As I show below, the NCHS recommendation to measure disparities in terms of relative differences in adverse outcomes is not necessarily followed, or necessarily known, even in the United States. Naturally it will be less known and followed in the United Kingdom (or elsewhere). So there could be a tendency for the researchers in the two countries to systematically reach opposite conclusions on certain matters. More important, of course, neither group of conclusions would have a sound statistical foundation. See also discussion in connection with Table 9 regarding the way researchers in the United States and the United Kingdom have tended to reach opposite conclusions about the impact of pay-for-performance on healthcare disparities even when they use the same measure, but use it without regard to the implications of the overall prevalence of the healthcare outcome being examined.

This paper is principally focused on healthcare disparities issues in the United States. But I note the following issues regarding health disparities measurement abroad. With the exception of Carr-Hill and Chalmers-Dixon 2005,[18] measurement guides issued outside the United States have yet to show a recognition of the ways measures tend to change because the prevalence of an outcome changes.[19] In 2013, the World Health Organization (WHO) issued a *Handbook on Monitoring Health Inequalities* (WHO 2013). The handbook principally measures disparities in terms of relative differences in favorable outcome, while showing no awareness that measures tend to be affected by the prevalence of an outcome and, though citing Keppel et al. 2005, no awareness that the United States NCHS would measure disparities in terms of relative differences in non-receipt of care. Table 5 is based on the handbook's Table 3.8, which shows the ratios of the rates of births attended by skilled persons for the highest income quintile to the lowest income quintile for seven selected countries. Table 5 is limited to the two countries that WHO identified as showing the least and most inequality on the basis of those ratios. These happen also to the countries that NCHS would identify has having the lowest and highest inequality on the basis of ratios of failing to have skilled persons in attendance, but with an opposite ordering of lowest versus highest. The perspectives offered by the RRFav (WHO

---

[18] This measurement guide, along with other works addressing my arguments about ways measures tend to be affected by an outcome, is discussed in Section E.7 of the Measuring Health Disparities page of jpscanlan.com.

[19] See Scanlan 2013b regarding the understanding of these issues by the United Kingdom's National Health Service. I also note that, in contrast to cancer outcomes disparities research in the United States – which, as discussed, commonly measures outcome disparities in terms or relative differences in mortality (while terming them relative differences in survival) and tends to find general increases in survival to be associated with increasing racial disparities – research in the United Kingdom commonly measures cancer outcome disparities in terms of absolute differences between rates and hence tends to find general increases in survival to be associated with increasing socioeconomic disparities for less survivable cancers and decreasing disparities for more survivable cancers.

approach) and RRAdv (NCHS approach) columns differ dramatically. The EES figure indicates that, while inequalities are very large in both countries, they are considerably larger in Bangladesh.

Table 5. Rates of Births Attended by Skilled Measures for Highest and Lowest Quintiles in Columbia and Bangladesh, from WHO Handbook on Monitoring Health Inequalities 2013, with Disparity Measures

Country	Highest Q	Lowest QF	RRFav	RRAdv	AD	OR	EES
Columbia	99.40%	83.70%	1.19	27.17	0.157	32.26	1.34
Bangladesh	50.60%	4.90%	10.33	1.93	0.457	19.88	1.67

Table 6 presents 1989 and 1995 rates of pneumococcal and influenza vaccination among persons older than 65 for blacks and the population at large from an October 26, 1998 Health and Human Services (HHS) Report titled "Progress Review: Black Americans." On the basis of the declines in relative differences in vaccination rates reflected in these figures, HHS determined that there had been a narrowing of the disparities. Since 2004, however, NCHS would regard the figures to reflect increased disparities. The EES figures, however, indicate that the disparity for pneumococcal vaccine decreased substantially while the disparity for influenza vaccine increased modestly.   This is another situation where NCHS researchers exploring the reasons for progress in reducing disparities would instead have to search for reasons for increasing disparities.

Table 6. Changes in Pneumococcal and Influenza Vaccination Rates for the Total Elderly Population and the Black Elderly Population between 1989 and 1995, from Progress Review for Black Americans, with Disparities Measures

Type	Yr	Tot Rt	Bl Rt	RRFav	RRAdv	AD	OR	EES
Pneumococcal	1989	15.00%	6.00%	2.50	1.11	0.09	2.76	0.53
Pneumococcal	1995	34.00%	23.00%	1.48	1.17	0.11	1.72	0.33
Influenza	1989	33.00%	20.00%	1.65	1.19	0.13	1.97	0.42
Influenza	1995	58.00%	40.00%	1.45	1.43	0.18	2.07	0.47

Table 7 presents data on white and black Hepatitis B vaccination rates from Morita et al. 2008, a study that had received a Robert Wood Johnson Foundation award for addressing health disparities. The table presents data for fifth graders and ninth graders for the year before imposition of a school-entry Hepatitis B vaccination requirement and the two years following the imposition of the requirement. While not indicating an awareness of the manner in which NCHS would appraise the disparities, the authors relied on relative differences in vaccination rates as a measure of disparities, finding that the requirement, which dramatically increased overall vaccination rates, dramatically reduced racial and ethnic vaccination disparities. By contrast, NCHS would have found dramatic increases in disparities. In the case of the black-white differences in Table 7, those relying on absolute differences, as CDC would do in the circumstance, would reach different conclusions as to the different grades and different

conclusions as to Grade 5 with respect to the two years at issue (that is, would find an increase in disparity immediately after the requirement was imposed, then a decrease the following year). The EES indicates that, as one might expect to commonly occur when a mandatory requirement is imposed, the disparities decreased.

Table 7. Hepatitis B Vaccination Rates for Whites and Blacks In Grades 5 and 9 Before and After Implementation of School-Entry Vaccination Requirement, from Morita et al. 2008, with Disparity Measures

Grade	Year	Program	WhVacRt	BlVacRt	RRFav	RRAdv	AD	OR	EES
5	1996	Pre	8.00%	3.00%	2.67	1.05	0.05	2.81	0.47
5	1997	Post	46.00%	33.00%	1.39	1.24	0.13	1.73	0.34
5	1998	Post	50.00%	39.00%	1.28	1.22	0.11	1.56	0.29
9	1996	Pre	46.00%	32.00%	1.44	1.26	0.14	1.81	0.37
9	1997	Post	89.00%	84.00%	1.06	1.45	0.05	1.54	0.24
9	1998	Post	93.00%	89.00%	1.04	1.57	0.04	1.64	0.26

Table 8 presents data from a study by Hetemaa et al. 2003 of the effects of substantial increases in revascularization rates in Finland between 1988 and 1996 on demographic differences in revascularization rates among men and women hospitalized for cardiac conditions. The table provides figures on revascularization rates men and women in the highest and lowest income groups,[20] along with the same measures used in the tables above. All standard measures changed in the direction that they typically do in the circumstance, while the EES indicates a small decline in disparities for both men and women.

Table 8. Revascularization Rates of Finnish Men and Women in the Highest and Lowest Income Categories Hospitalized for Cardiac Conditions in 1988 and 1996, from Hetemaa et al. 2003, with Disparity Measures

Gender	Year	Highest	Lowest	RRFav	RRAdv	AD	OR	EES
M	1988	17.91%	8.27%	2.16	1.12	0.10	2.42	0.48
M	1996	41.27%	25.36%	1.63	1.27	0.16	2.07	0.44
F	1988	10.00%	3.70%	2.70	1.07	0.06	2.89	0.51
F	1996	30.75%	17.06%	1.80	1.20	0.14	2.16	0.45

---

[20] Rates are derived from figures in Hetemaa Table 1 with persons hospitalized for coronary heart disease used as the pool.

Table 9 presents data from Werner et al. 2005, a study examining (among other things) the way a coronary artery bypass graft (CABG) report card program, which was believed to cause general increases in CABG rates, affected racial and other differences in CABG rates among persons with acute myocardial infarction. In this instance, all standard measures changed in the way that they commonly do in the circumstances, though the EES figure indicates a modest reduction in disparity.[21]

Table 9. White and Black CABG Rates Before and After Implementation of a CABG Report Card, from Werner et al. 2005, with Disparities Measures

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

Had the authors examined relative differences in favorable outcome rates, as was probably the most common approach at the time, they would have would have found the disparity to have decreased. They instead relied on the absolute difference between rates and found that the disparity to have increased, concluding that incentive programs like the CABG report card program at issue are likely to increase healthcare disparities.

The authors made no mention of why they chose to use absolute difference as a measure of disparity, whether they regarded that as the usual measure in the circumstances, or whether they were aware that other measures would yield contrary conclusions. In any case, the Werner study was widely regarded as indicating that pay-for-performance (P4P) programs were likely to increase healthcare disparities. And, so far as is revealed in published literature, none of those regarding the study as indicating that incentive programs would tend to increase healthcare disparities gave any thought to the implications of the fact that the authors had relied on absolute differences and did so with respect to rate ranges where general increase tend to increase absolute differences.   Nor did such persons give thought to the fact that had the same sort of study been used with regard to an outcome where initial rates were quite high, the absolute differences between rates would have likely been found to decrease. Meanwhile, in the United Kingdom, researchers measuring disparities in terms of absolute differences between rates of receiving certain common types of care, and where P4P seemed to increase those rates, were concluding that P4P would likely reduce healthcare disparities.[22]

Table 10 is based on Albain et al. 2008, which studied racial differences in outcomes for various types of cancers with very different overall survival rates. While the article's title mentioned disparities in survival, the study in fact analyzed relative differences in mortality. As shown in Table 10, relative differences in mortality consistently increase, while relative differences in

---

[21] Because populations with acute myocardial infarction may be truncated parts of larger distributions, it is possible that the EES is a flawed measure in such circumstances. Such issue, however, ought not to detract from the illustration here.

[22] See the Pay for Performance subpage of the Measuring Health Disparities page of jpscanlan.com .

survival consistently decrease, as overall mortality declines.   The EES showed somewhat larger disparities for the more survivable cancers.[23]

Table 10. Survival Rates of White and Black Women for Various Types of Cancers, from Albains et al. 2008, with Disparity Measures

Type	W	B	RRFav	RRAdv	AD	OR	EES
premenopausal breast cancer	77.00%	68.00%	1.13	1.39	0.09	1.58	0.27
postmenopausal breast cancer	62.00%	52.00%	1.19	1.26	0.10	1.51	0.26
advanced ovarian cancer	17.00%	13.00%	1.31	1.05	0.04	1.37	0.18
for advanced prostate cancer	9.00%	6.00%	1.50	1.03	0.03	1.55	0.21

Table 11 is based on Robbins et al. 2007, which purported to examine differences between white and black men in survival from prostate cancer, but which in fact analyzed relative differences in mortality. The figures in the table pertain to the question examined in the study of whether racial differences were greater above or below the median age of diagnosis (age 68), which question the authors regarded as testing the hypothesis "that the survival disadvantage for Black men might be worse at younger ages." The authors concluded, apparently on the basis of the fact that the relative difference in mortality was not substantially greater among the under 68 group (142 percent versus 112 percent), that "the elevation in risk for Black men was quite similar above and below the median age at diagnosis." Whether or not the authors would have reached the same conclusion concerning similarity on the basis of relative differences in survival, the fact is that, while the mortality difference was larger in the under 68 group than the over 68 group, the survival difference was smaller in the under 68 group than the over 68 group (12 percent versus 17 percent), and the comparative patterns as to the two relative differences are what one commonly observes. That is, relative differences in mortality tend be larger, but relative differences in survival tend to be smaller, among the subpopulation where the adverse outcome is less common. The EES is exactly the same in the two age groups.

Table 11. Survival Rates of White and Black Men with Prostate Cancer below and above Age 68, from Robbins et al. 2007, with Disparity Measures

AgeGroup	W	B	RRFav	RRAdv	AD	OR	EES
68 and above	88.50%	75.60%	1.17	2.12	0.13	2.48	0.52
under 68	93.10%	83.30%	1.12	2.42	0.10	2.71	0.52

---

[23]   Other examples of patterns of measures of differences between rates according to survivability may be found in the Mortality/Survival Illustrations subpage of the Scanlan's Rule page of jpscanlan.com, which presents information on mortality and survival according to stage of cancer.

Table 12 is based on published life tables and presents rates at which white and black men survive to certain ages along with the measures employed in the earlier tables. The declining relative difference in mortality (RRAdv) is what underlies the common perception that demographic differences in mortality decrease with age (though relative differences in survival tend to increase with age). Throughout the table the EES indicates that for most of the life span the disparity increases with age, though there starts to develop a cross-over in later life.[24]

Table 12. White and Black Rates of Surviving to Various Ages, from Life Tables, with Disparities Measures

Age	W	B	RRFav	RRAdv	AD	OR	EES
20	98.70%	97.48%	1.01	1.94	1.22	1.96	0.28
30	97.37%	95.27%	1.02	1.80	2.10	1.84	0.27
40	95.85%	92.49%	1.04	1.81	3.36	1.88	0.3
50	92.66%	87.21%	1.06	1.74	5.45	1.85	0.32
60	86.04%	75.75%	1.14	1.74	10.30	1.97	0.39
70	72.97%	57.53%	1.27	1.57	15.44	1.99	0.44
80	48.07%	32.64%	1.47	1.30	15.43	1.91	0.41
90	15.32%	9.95%	1.54	1.06	5.37	1.64	0.26
100	0.80%	0.91%	0.89	1.00	-0.10	0.89	0.01

Many further illustrations of EES values associated with particular pairs of rates and the patterns by which the standard measures tend to be affected by the prevalence of an outcome may be found in the online comments collected in Section D of the Measuring Health Disparities page of jpscanlan.com and in the conference presentations and workshops collected in Section B of that page.

3. Some Shortcomings of the Method

The EES approach to measuring the differences in the circumstances of two groups reflected by a pair rates has a number of weaknesses. For example, it relies on an assumption that the underlying distributions of factors associated with experiencing an outcome are normal. Rarely can we be sure that the underlying distributions are normal and sometimes we will know that they are not normal, as, for example, when the distributions are truncated parts of normal

---

[24] See the Life Tables Illustrations subpage of Scanlan's Rule page of jpscanlan.com, along with the Life Table Information Document referenced on the subpage, for similar information on white men compared with white women, black men compared with black women, and white women compared with black women. The latter item also shows how interpretations may differ somewhat depending on whether one analyzes rates of surviving to a particular age or rates of surviving from one age to the next.

distributions. There also exists a range of more subtle issues.[25] But an approach of this nature (including one based on a sounder understanding of the actual shapes of the underlying distributions) is clearly superior to reliance on standard measures of differences between outcome rates without consideration of way the measure tends to be affected by the prevalence of the outcome at issue. For such an approach provides a benchmark for appraising the strength of the association reflected by any pair of rates and for comparing the strength of association reflected by two or more pairs of rates when standard measures would yield varying interpretations as the comparative size of differences between rates. And it can at least spare us from wrongly concluding, on the basis of one preferred standard measure or another, that there is reason to distinguish among the employers in Table 1 and then mistakenly devoting resources to exploring the reasons for the perceived differences, drawing inference based on the perceived differences, or making decisions of consequence based on the perceived differences. The same holds where the issue is the size of a health or healthcare disparity, such as those addressed in the studies underlying Tables 2 through 12, whether or not any part of that disparity is deemed to result from bias.

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

C. The Disarray of Health and Healthcare Disparities Research and the Consequences of the Misunderstanding of Measurement Issues in the Pay-for-Performance Context

The disarray of health and healthcare disparities research as a result of the failure to recognize the ways standard measures of differences between outcome rates tend to be affected by the prevalence of an outcome has already been suggested in the discussion at the end of Section A and in the discussion of the tables illustrating EES values in Section B. This section discusses the matter further with a focus on the activities of particular entities, including government

---

[25] For a fuller explanation of the varied problems with the described approach, see the Solutions, Irreducible Minimums, and Cohort Considerations subpages of the Measuring Health Disparities page and the Truncation Issues subpage of the Scanlan's Rule page of jpscanlan.com.

agencies that have an important role in health and health disparities research, as well as certain measurement guides. This section also discusses the way that a failure to understand these issues had led to an unfounded perception that pay-for-performance programs would tend to increase healthcare disparities and the way that the perception then led the Commonwealth of Massachusetts to unwisely include a health disparities measure in its Medicaid pay-for-performance program and to do so in a manner that is more likely to result in increased healthcare disparities than in reduced healthcare disparities.

While this section is strongly critical of the understanding of measurement issues among government agencies and private entities involved with health and healthcare disparities research, it should be recognized that the same failures of understanding discussed with respect to health and healthcare disparities measurement in this document are reflected in virtually all efforts to appraise group differences in outcome rates in the law and the social and medical sciences. By way of example based on matters lately receiving widespread media attention, out of concern about large relative differences in adverse lending or school discipline outcomes, federal regulators encourage lenders and public schools to reduce the frequency of such outcomes. But, unaware that reducing the frequency of such outcomes tends to increase relative differences in experiencing the outcomes, regulators continue to monitor the fairness of practices on the basis of relative differences in adverse outcomes. Thus, by complying with federal encouragements to reduce adverse outcomes, entities make it more likely that the government will find them guilty of discrimination.[26]

Even though income data in exactly the form such data are published make it clear that any reduction in poverty will tend to increase relative differences in poverty while any increase in poverty will tend to reduce relative differences in poverty (see Scanlan 2006a), for many decades observers have been purporting to find significance in changes in relative differences in poverty rates without consideration of the extent to which such changes are the functions of changes in overall poverty rates or consideration of the fact that examination of relative differences in rates of avoiding poverty would yield opposite conclusions about whether disparities in the circumstances of advantaged and disadvantaged groups are increasing or decreasing. Scanlan 1987, 1991, 1994, 2000. More recently observers have been analyzing changing poverty patterns in terms of absolute differences, with a tendency to reach opposite conclusions about changes in poverty disparities from those who rely on relative differences in poverty rates. But observers relying on absolute differences also overlook that the income data themselves show that reductions in poverty will tend to reduce absolute differences in poverty rates while increases in poverty will tend to increase absolute differences in poverty rates.

Given that discussion of changes in poverty rates are often couched in terms of the changes experienced by various groups rather that changes in disparities, it may be useful to restate the above points in terms of the corollaries discussed earlier. When poverty rates change for two groups (whether increasing or decreasing), the group with the lower baseline poverty rate tends to experience a larger proportionate change in its poverty rate while the other group tends to experience a larger proportionate change in its rate of avoiding poverty. Given the rate ranges that we commonly find for poverty, when poverty changes (again, whether increasing or

---

[26] See Scanlan 2012a, 2012b, 2012f, 2013a, 2014.   While the references discuss recent attention to the matters, the consequences of these misunderstandings are longstanding. See Scanlan 1993, 1996.

decreasing) the group with the higher baseline poverty rate tends to experience a larger percentage point change in its rate than the group with the lower baseline rate. But the matter is no better understood when addressed in terms of the comparisons of the sizes of changes in poverty rates than when addressed in terms of comparisons of changes in differences between poverty rates. The same holds in circumstances where observers appraise a situation in terms of the proportion a group comprises of those experiencing a favorable or adverse outcomes, as in the case of the failure to recognizes that decreases in poverty (including the poverty of female-headed families) tends to increase the feminization of poverty while increases in poverty (including the poverty of female-headed families) tend to decrease the feminization of poverty. [27]

Thus the misunderstandings of measurement issues in the study of health and health and healthcare disparities is no greater than that reflected in every other area where differences in outcome rates are deemed matters of concern (including even when the groups at issue are the treated subjects and the control subjects in a clinical trial).[28]

1. The Disarray of Health Disparities Research

a. National Center for Health Statistics

To date the National Center for Health Statistics (NCHS) is the only government agency to recognize in any way that standard measures of health or healthcare disparities may be affected by the prevalence of an outcome. As noted above, however, the agency failed to act on such recognition in a sensible manner. Further, NCHS has yet to suggest that it recognizes that there exists an underlying reality as to the strength of forces causing outcome rates to differ or how one might determine whether the strength of such forces has increased or decreased. In fact, so far as is reflected by publicly available materials, the agency's understanding of the issues has regressed in recent years, and in no public document since 2009 have agency representatives even suggested the possibility that one might reach different conclusions as to directions of changes over time depending on whether one examines relative differences in favorable outcomes or relative difference in adverse outcomes, much less that, as NCHS specifically indicated in earlier documents, that that would commonly occur. A 2010 NCHS presentation titled "Defining and measuring disparities, inequities, and inequalities in the Healthy People initiative" (Klein and Huang 2010) discusses the importance of value judgments in choosing between relative and absolute differences without showing any recognition that there exist two relative differences.

With respect to the measurement of health disparities for Healthy People 2010, a Technical Appendix (at A-8) in the *Healthy People 2010 Final Review* (NCHS 2012) states that, while objectives may be stated in terms of either favorable or adverse outcomes, in order to facilitate comparisons across different objectives, disparities are measured in terms of relative differences in adverse outcomes. Though the appendix cites three NCHS documents that discuss the fact that relative differences in favorable outcomes and relative differences in adverse outcomes commonly yield different conclusions about directions of changes in disparities over time, the

---

[27] See Scanlan 1987, 1991, 1992, 1994, 2000 and the Feminization of Poverty page of jpscanlan.com. See also with the Concentration Index subpage of the Measuring Health Disparities page of jpscanlan.com.

[28] See Scanlan 2013 and the Subgroup Effects subpage of the Scanlan's Rule page of jpscanlan.com.

appendix itself does not even suggest that it is possible that the two relative differences could yield contrary conclusions, and, as suggested above,[29] few readers would infer that on their own. Those who do not read the appendix would assume that discussion of changes in things like receipt of some beneficial procedure that are discussed in terms of favorable outcome are analyzed in such terms, and certainly few would understand that a statement that a disparity in some outcome has changed in a particular direction involves a situation where it was the opposite outcome that changed in the indicated direction and that the described outcome may have changed in the opposite direction from that stated.

In the case of things like immunization, which is the subject of Tables 6 and 7 above, it is not clear how to interpret the *Healthy People 2010 Final Review*. At one point, the document (at 14-6) states that "vaccination rates for preschool children in racial and ethnic groups with lower vaccination rates, however, have been increasing at a more rapid rate, significantly narrowing the gap." That when rates of receipt of such procedures increase they do so to proportionately greater degrees among groups with lower baseline rates does narrow relative differences between rates of receipt of such procedures. But, as explained above, and as exemplified in Tables 6 and 7,[30] typically difference in rates of non-receipt tend to increase. So it is not clear how the document is measuring the disparity in immunization rates.

More important, the quoted statement betrays a failure to recognize that anytime an outcome generally increases it tends to do so at proportionately greater rates among groups with lower baseline rates and that the principal purpose of health disparities research is to determine whether one is simply observing the consequence of changes in overall prevalence or something beyond those consequences.

Finally, while the above discussion goes mainly to healthcare outcomes, the same issues exist for analyses of disparities in health outcomes like mortality and survival, with respect to which NCHS has also recognized that interpretations as to changes in disparities over time would commonly depend on which relative difference is examined. Here, too, however, NCHS has failed to provide guidance on how one might determine whether observed changes in differences between outcome rates reflect anything other than changes in the prevalence of an outcome.

Presumably, NCHS will one day responsibly address the measurement issues outlined here. But its delay in doing so will ultimately be responsible for the expenditure of billions of dollars in research that is wasteful even when it is not misleading.

b. Agency for Healthcare Research and Quality

The Agency for Healthcare Research and Quality (AHRQ) has yet to show recognition, either in the yearly *National Healthcare Disparities Report* (NHDR) or in any other document it has produced, that measures of health and healthcare disparities may be affected by the prevalence of an outcome. In the NHDR, AHRQ describes the outcomes it measures in either favorable or adverse terms. And, for a time, the NHDR indicated that a disparity would be deemed important if the relative difference in either the favorable or the adverse outcome exceeded 10 percent

---

[29] The appendix cites Keppel 2005 and Keppel 2005a (references 18 and 19 in the document) and apparently intended to cite Keppel 2004 (reference 13, though the appendix, presumably erroneously, cited reference 12).

[30] See also the Immunization Disparities page of jpscanlan.com

(hence, for that purpose, relying on the larger of the two relative differences).[31] But in the report AHRQ has intended to measure disparities, consistent with Healthy People 2010/2020, in terms of relative differences in adverse outcomes. Apparently, however, the agency has not always successfully executed that intention. Among outcomes that the 2012 report highlight as involving the most rapid improvements in disparities are outcomes where the relative difference in adverse outcome in fact increased while the relative difference in the favorable outcome and the absolute difference decreased. Thus, NCHS would have found increasing disparities in these cases, as should AHRQ have according to the approach it believes it is implementing.[32]

 AHRQ also funds a great deal of health and healthcare disparities research. But it is virtually certain that AHRQ official making funding decisions, like those seeking the funding, do so without any understanding of the ways the measures to be employed in the study are likely to be affected by the prevalence of the outcomes at issue. Emblematic of the process is an AHRQ contract with the Institute for Medicine and Public Health of the Vanderbilt University Medical Center aimed at evaluating the effectiveness of quality improvement in reducing disparities in health and healthcare, The contract yielded a 475-page, peer reviewed report, issued in August 2012, that cites 4258 sources in providing a great deal of information on findings of studies as to the way improvements in healthcare affected health and healthcare disparities. But the report reflects no recognition whatever of the way the various measures employed in those studies may be affected by the prevalence of an outcome or even that it is possible that various measures could yield different conclusions as to directions of changes in disparities as healthcare improves. In discussing findings of various studies, the report does not identify the measures that were used.[33]

3. Centers for Disease Control and Prevention

Researchers from the Centers for Disease Control and Prevention (CDC) publish many studies on health and healthcare disparities issues, particularly with respect to immunization. The agency also itself issued in January 2011 and December 2013 *Health Disparities and Inequalities Reports* (CDC 2011, 2012). Most CDC research into health and health disparities appears to rely on absolute differences between rates as a measure of disparity, though some things in the 2011 and 2013 reports are measured in terms of relative differences in adverse outcomes. I am not aware of a publically available (non-NCHS) CDC document discussing the fact that different measure from that employed in a CDC study could or would yield a different conclusion as to the direction of changes over time or that any measure of differences between outcome rates tends to be affected by the prevalence of an outcome. Nor am I aware of a single CDC document indicating that NCHS, an arm of CDC, has chosen to measure disparities in

---

[31] The description of the outcomes in the NHDR and the discussion of relative differences in either the favorable or adverse outcome for purposes of determining important has led me to incorrectly describe the NHDR approach in a number of places. See, e. g., Scanlan 2007. Possibly that is also the reason for the Institute of Medicine's apparent misunderstanding discussed below.

[32] See the NHDR Measurement subpage of the Measuring Health Disparities page of jpscanlan.com. As of the finalizing of this paper, I had not yet determined how the agency in fact been measured disparities in the 2012 report.

[33] See the AHRQ's Vanderbilt Study subpage of the Measuring Health Disparities page.

terms of relative differences in adverse outcome or has concluded that determinations of changes in directions of disparities commonly will turn on whether one examines relative differences in favorable outcomes or relative differences in adverse outcome.

I have not systematically studied all CDC's activities relating to health disparities. But it is doubtful that any aspect of those activities reflects an understanding of the ways the measures employed tend to be systematically affected by the prevalence of an outcome. For example, the agency has a web page on Health Disparities in Cancer, which discusses, among other things, programs studying disparities in cancer survival. It is doubtful that CDC draws a distinction between disparities in cancer survival and cancer mortality or recognizes that relative differences in the two tend to change in opposite directions as survival generally increases.

d. National Cancer Institute

The National Cancer Institute (NCI) endeavors to provide guidance on the measure of disparities in cancer outcomes and cancer screening through a variety of means. Nothing it has produced of which I am aware reflects an understanding that measures tend to be affected by the prevalence of an outcome and or that relative differences in mortality and survival can (or typically do) yield different conclusions as to directions of changes in disparities over time. The agency continues to rely on a commissioned measurement guide (Harper and Lynch 2006), by authors who are among authors of the study discussed with regard to Table 3 and who also authored the University of Michigan Measuring Health Disparities course discussed below. While addressing some highly sophisticated topics, the document reflects no awareness that measures may be affected by the prevalence of an outcome or that relative differences in favorable outcomes and relative differences in adverse outcome could yield different results. The same holds for a 2007 follow-up document by the same authors (Harper 2007). Crucially, as with the other entities discussed above and below, nothing the agency has produced reflects an awareness that there exists an underlying reality with respect to whether the forces causing cancer (or screening) outcomes to differ by race have increased or decreased over time.

e. Institute of Medicine

The Institute of Medicine (IOM) of the National Academies of Sciences has been for a least a decade been providing guidance on health and healthcare disparities research issues, commencing with it 2003 treatise *Unequal Treatment: Confronting Racial and Ethnic Differences in Healthcare* (Smedley *et al.* 2003). That treatise, however, left measurement issues unaddressed, and IOM has yet to recognize that standard measures may be affected by the prevalence of an outcome.

In April 2010, IOM issued a 247-page document titled *Future Directions of the National Healthcare Quality and Disparities Report* (Institute of Medicine 2010). The report (at 82-83) discusses relative and absolute differences as disparities measures, presenting an example where the two change in opposite directions. The report mentions that the National Healthcare Disparities Report measures disparities in terms of relative difference in favorable or adverse outcomes (which mention suggests that IOM may not in fact understand the measure used in the report, see note 31 *supra*) but does not discuss any implications of there being two relative differences. The report cites Keppel et al. 2005 for the proposition that one should present both relative and absolute differences. But the report reflects no awareness that the two relative

differences can yield different conclusions as to the directions of changes over time, that they commonly or usually do so (as Keppel et al. 2005 had stated), and that they in fact did so in hypothetical the IOM report posited as an example of the way a relative difference and the absolute difference can yield different conclusions about changes over time. The report reflects no awareness that measures may be affected by the prevalence of an outcome and, expressing the view that normative judgments should inform choice of measure, reflects no awareness that there might be an underlying reality concerning the strength of the forces causing rates to differ or that health disparities research should be directed toward understanding that reality.

In September 2012 the IOM released a workshop report appraising progress in reducing health disparities since 2001 (Institute of Medicine 2012). The report reflected no understanding of the measurement issues discussed here. To date nothing IOM has issued reflects any understanding that the measures it discusses may be affected by the prevalence of an outcome.

f. University of Michigan Measuring Health Disparities Course

Since 2005, the Center for Social Epidemiology and Population Health of the University of Michigan has made available for free an online (or CD) *Measuring Health Disparities* course (Lynch and Harper 2005). The course is authored by authors of the NCI measurement guides mentioned above and the course is similar to the NCI guides in many respects. It reflects no understanding that a measure may be affected by the prevalence of an outcome. It contains much discussion of the choice between relative and absolute differences but without apparent recognition that anytime one relative difference disagrees with the absolute difference as to the comparative size of a disparity the other relative difference will agree with the absolute difference. The document does, however, discuss (at 57-58) the possibilities for different interpretations concerning the size of disparity depending on whether one examines relative differences in the favorable or the adverse outcome, which it describes as the positive and negative outcomes. While expressing no preference for one over the other, the document merely notes that a researcher should be consistent Such discussion, however, does not reflect an awareness of the possibility that conclusions as to directions of changes over time could differ depending on which outcome one examined much less that the such conclusions would tend usually to differ.

g. Harvard University and Massachusetts General Hospital.

Harvard University conducts a great deal of health and healthcare disparities research through the Harvard School of Public Health or the Health Care Disparities Group of Harvard Medical School. Whether or not Harvard produces the greatest volume of such research, simply because of the institution's reputation, its health and healthcare disparities research may be the most highly regarded in the world. Faculty of Harvard Medical School play a large role in the Health Disparities Solutions Center of Massachusetts General Hospital, which, among other things, conducts a Health Disparities Leadership program to train administrators in health disparities issues including how to measure them. No work emanating from any of these institutions has ever suggested that any healthcare disparities measure may be affected by the prevalence of an outcome and only very rarely has any research conducted by these institutions even shown an awareness that different measures of health or healthcare disparities in fact yielded different

results.[34] A June 2012 issue of *Health Services Research*, which purported to be devoted to the discussion of the measurement of health disparities with articles largely authored by Harvard researchers, makes no mention of the possibility that different measures might yield different conclusions as to changes over time or that any measure tends to be systematically affected by the prevalence of an outcome.

In approximately 2010, the Robert Wood Johnson Foundation and the National Quality Forum commissioned Harvard Medical School and the Disparities Solution Center of Massachusetts General Hospital to produce a healthcare disparities measurement guided titled *Commissioned Paper: Healthcare Disparities Measurement* (Weissman et al. 2011). Because of the prestige of the four involved entities, as well as the currency of the guide, it has potential to be the most important healthcare disparities measurement guide for years to come. The guide, which was released for comment in the summer of 2011, did indicate (though somewhat obscurely) that one might reach different conclusions as to directions of change over time depending on whether one examined relative differences in a favorable outcomes or relative differences in the corresponding adverse outcome. But it said nothing to suggest that this would commonly occur and it showed no recognition whatever that any measure it discussed tend to be affected by the prevalence of an outcome. That remained the case after comments were submitted.

My efforts to cause the authors and the responsible institutions to modify or withdraw the document are summarized or reflected in an October 26, 2012 letter to those institutions. The response of the Harvard Medical School and Massachusetts General Hospital is found in December 12, 2012 letter from the research integrity offices of those entities.[35] The response stated that issues I raised concerning the guide involved "a difference of scientific opinion" and not research misconduct, and that, absent the latter, and Harvard Medical School and Massachusetts General Hospital do not independently assess the merits of individual papers of their faculty members. Thus, the entities declined to withdraw the guide.

While few people recognize that standard measures of differences between outcome rates tend to be affected by the prevalence an outcome, no one has yet taken the position that they are not. So it is not clear that there exists anything that might be deemed a difference of scientific opinion. In any case, the document continues to bear the names of Harvard Medical School and Massachusetts General Hospital on its cover, and by the failure to mention that there exists issues (or a difference of scientific opinion) concerning the ways the measures it discusses may be affected by the prevalence of an outcome leads readers to believe that no such issues exist.

The above discussion of the disarray of health disparities research suggests that the named institutions, rather than providing sound research or guidance, are in fact contributing to the production of flawed research that will be wasteful even when it is not misleading. But the treatment of these institutions here should not be read as suggesting that their understanding of

---

[34]   A description of the health and healthcare disparities research at Harvard may be found at page 32-41 of my October 9, 2012 letter to the President of Harvard University, which was written in conjunction with an applied statistics workshop I gave at Harvard's Institute for Quantitative Social Science (Scanlan 2012c). The letter may be found on the Institutional Correspondence subpage of the Measuring Health Disparities page of jpscanlan.com.

[35] Both letters are available on the Institutional Correspondence subpage of the Measuring Health Disparities page of jpscanlan.com.

the measurement issues is any less sound than that found elsewhere. The same failure of understanding can be found at virtually every institution attempting to make a contribution to health disparities research or to ameliorate the differences in the health and healthcare of advantaged and disadvantaged groups reflected by their differing outcome rates.

2. Effects of the Misunderstanding of Disparities Measurement in the Pay-for-Performance Context

The problems with perceptions about the effects of pay-for-performance (P4P) programs on health and healthcare disparities and the potential harms to be caused by inclusion of disparities elements in P4P programs can be usefully illustrated with the information in Table 1 when viewed from the following perspective. Consider the favorable outcome rates in Rows A and B to be the rates at which advantaged and disadvantaged receive some relatively uncommon outcome during a period when such outcomes are generally increasing and Rows C and D to be the rates at which such groups receive some relatively common outcome during a period when those outcomes are generally increasing, and assume that all changes are attributed to incentive programs like P4P.

Observers who rely on relative differences in favorable outcomes would find that P4P tends to reduce disparities and those who rely on relative difference in adverse outcomes would find that P4P tends to increase disparities. And this holds with respect to both the A/B and C/D situations.

As it has happened, however, efforts to appraise the effects of incentive programs on healthcare disparities have tended to rely on absolute differences as a measure of disparity. As discussed with respect to Werner et al. 2005 and Table 9, those appraising such effects in the United States examined uncommon outcomes (the A/B situation) and concluded that incentive programs would tend to increase disparities. By contrast, those appraising such effects in the United Kingdom examined common outcomes (the C/D situation) and concluded that incentive programs would tend to reduce disparities. As stated or suggested earlier, no one drawing such conclusions or relying on them considered the implications of the ways absolute differences between rates tend to affected by the prevalence of an outcome.

When the research underlying Table 9 led to the perception in the United States that P4P would tend to increase disparities and hence that there was a need to include disparities elements in P4P programs, there were suggestion that such elements involve either appraisals of changes in disparities over time at individual institutions or appraisals of the comparative size of disparities across institution. Given that these elements would in fact be examined in terms of absolute differences between rates, I do not discuss here implications of reliance on either of the two relative differences, though the implications are implied in the above discussions.

Assuming that the disparities element were to focus on changes over time, the implications of reliance on absolute differences would tend to turn on the type of outcome examined, in the following way. If the favorable outcome examined is relatively uncommon (the A/B situation),[36]

---

[36] I couch the discussion in terms of favorable outcomes because that tends to be the focus in analysis of healthcare disparities. But to simplify conceptualization of the matter it is useful to think in terms of the frequency of the favorable outcome even when the institution is thinking in terms of the adverse outcome. To increase the favorable outcome, as the matter is put in the text above, is of course to decrease the adverse outcome.

the more an institution generally increases favorable outcome rates (reduces the corresponding adverse outcome rates), the more its disparities would appear to increase (at least until the point where one or more groups' rate or rates reached 50 percent). If the favorable outcome examined is relatively common (the C/D situation), the more an institution generally increases favorable outcome rates, the more its disparities will appear to decrease.

Assuming that the disparities element were to focus on comparisons across institutions, the implications of reliance on absolute difference would again also tend to turn on the type of outcome examined. But it would do so in the following way (considering for this purpose that Rows A and B reflect outcome rates for an uncommon outcome at different hospitals and Rows C and D reflect outcome rates for a common outcome at different hospitals). If the favorable outcome examined is relatively uncommon (the A/B situation), lower-performing hospitals would be deemed to have smaller disparities. If the favorable outcome examined is relatively common (the C/D situation), higher-performing hospitals would be deemed to have smaller disparities.

Massachusetts was the first to respond to calls for inclusion of disparities elements in a P4P program by including such an element in its Medicaid P4P program. But the program evaluated the size of disparities on the basis of a measure that was a function of absolute differences between rates, and it did so with regard to outcome rates that were generally quite high (above 80% for all types of care combined). Given the tendency for higher overall rates in such ranges to be associated with smaller absolute differences between rates, the program will tend to find healthcare disparities to be smaller at higher-performing hospitals than lower-performing hospitals and hence to reward higher-performing hospitals for reason unrelated to a useful indicator of the size of a disparity. Since higher-performing hospitals tend to have smaller minority representations among their patient populations than lower-performing hospitals, the inclusion of a disparities element in the Massachusetts P4P program – by diverting resources away from providers with large numbers of minority patients – is more likely to increase healthcare disparities than to reduce them.[37]

This is not say that there is no reason to expect P4P to increase healthcare disparities either simply because P4P programs generally favor higher-performing hospitals[38] or because such programs may cause hospitals to alter their patient mixes in ways adverse to disadvantaged groups. Hence, there may be reason to include an appraisal of healthcare disparities in such programs. But any such appraisal must involve a sound measure.

But regardless of whether the inclusion of a disparities element in a P4P program has the anomalous consequence observed in the Massachusetts Medicaid program, unless the program

---

[37] See Scanlan 2011 and 2012 and the Pay for Performance and Between Group Variance subpages of the Measuring Health Disparities page of jpscanlan.com.

[38] See Friedberg 2010 regarding the concern that P4P will generally increase disparities simply be channeling resources to higher-performing hospitals irrespective of any disparities element in the program. The particular anomaly of the Massachusetts program is that it is the disparities element of the program that contributes to the channeling of resources to higher-performing hospitals.

employs a sound measure of disparity, it will be allocating monetary incentive awards based on something that has nothing to do with the purpose of the award.

***

Many measurement issues related to those discussed above are treated in greater detail on the following pages of jpscanlan.com and their associated subpages: Measuring Health Disparities, Scanlan's Rule, Mortality and Survival, Measuring Association, Immunization Disparities, Lending Disparities, Discipline Disparities, Educational Disparities, Disparate Impact, and Feminization of Poverty.

References

Albain KS, Unger JM, Crowley JJ, et al. 2009. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *J Nat Cancer Inst,*101:988-992.
Baker E, Middleton D. 2003. Cervical screening and health inequality in England in the 1990s. *J Epidemiol Community Health*,57:417-423.

Carr-Hill R, Chalmers-Dixon P. 2005. The Public Health Observatory Handbook of Health Inequalities Measurement. Oxford: SEPHO. http://www.sepho.org.uk/extras/rch_handbook.aspx

Centers for Disease Control and Prevention. 2011. *CDC Health Disparities and Inequalities Report – United States, 2011*: http://www.cdc.gov/mmwr/pdf/other/su6001.pdf

Centers for Disease Control and Prevention. 2013. *CDC Health Disparities and Inequalities Report – United States, 2013.* http://www.cdc.gov/minorityhealth/CHDIReport.html#CHDIR

Department of Health and Humans Services. 1998. Progress review: Black Americans, Oct. 26. http://odphp.osophs.dhhs.gov/pubs/HP2000/PDF/prog_rvw/pr-black.pdf

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

Harper S, Lynch J. 2006. Methods for Measuring Cancer Disparities: A Review Using Data Relevant to Healthy People 2010 Cancer-Related Objectives. Washington DC: National Cancer Institute. http://seer.cancer.gov/archive/publications/disparities/

Harper S, Lynch J. 2007. Selected Comparisons of Measures of Health Disparities: A Review Using Databases Relevant to Healthy People 2010 Cancer-Related Objectives. NCI Cancer Surveillance Monograph Series, Number 7. National Cancer Institute. NIH Pub. No. 07-6281, Bethesda, MD. http://seer.cancer.gov/publications/disparities2/

Harper S, Lynch J, Meersman SC, et al. 2009. Trends in area-socioeconomic disparities in breast cancer screening, mortality, and survival among women ages 50 years and over (1987-2005). *Cancer Epidemiol Biomarkers Prev,*18(1):121-131.

Harper S, King NB, Meersman SC, et al. 2010. Implicit value judgments in the measurement of health disparities. *Milbank Quarterly,* Mar;88(1):4-29.
http://www.milbank.org/quarterly/8801feat.html

Hetemaa T, Keskimäki I, Manderbacka, et al. 2003. How did the recent increase in the supply or coronary operations in Finland affect socioeconomic and gender equity in their use? J Epidemiol Community Health, 57:178-185.

Institute of Medicine of the National Academies of Sciences. 2012. How far have we come in reducing health disparities? Progress since 2000. National Academies Press. Washington, D.C. http://www.iom.edu/Reports/2012/How-Far-Have-We-Come-in-Reducing-Health-Disparities.aspx

Keppel KG, Pearcy JN, Klein RJ. 2004. Measuring progress in Healthy People 2010. Healthy People statistical notes. No. 25. Hyattsville, Md.: National Center for Health Statistics. http://www.ncbi.nlm.nih.gov/pubmed/15446274

Keppel KG., Pamuk E., Lynch J., *et al.* 2005.   Methodological issues in measuring health disparities. *Vital Health Stat* 2005;2 (141): http://www.cdc.gov/nchs/data/series/sr_02/sr02_141.pdf

 Keppel KG, Pearcy JN. 2005. Measuring relative disparities in terms of adverse events. *J Public Health Manag Pract*, 11(6):479–483.

Keppel K.G., Pearcy J.N. 2006. Response to Scanlan concerning: measuring health disparities in terms of adverse events. *J Public Health Manag Pract,* 12(3):295. http://www.nursingcenter.com/library/JournalArticle.asp?Article_ID=641472

Keppel KG, Pearcy JN. 2009, Healthy People 2010: Measuring disparities in health. *Chance*, 22(1):6-9.

Kim S. 2013. Statistical Discrimination, Employer Learning, and Employment Differentials by Race, Gender, and Education. Population Association of America 2013 Annual Meeting, New Orleans, LA (Apr. 11-13). http://paa2013.princeton.edu/papers/130021

Klein R, Huang D. 2010. Defining and measuring disparities, inequities, and inequalities in the Healthy People initiative. http://www.cdc.gov/nchs/ppt/nchs2010/41_Klein.pdf

Lynch J, Harper S. 2005. Measuring health disparities. Center for Social Epidemiology. University of Michigan. http://sitemaker.umich.edu/mhd/home

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

National Center for Health Statistics. 2012 *Healthy People 2010 Final Review*. Hyattsville, MD. http://www.cdc.gov/nchs/data/hpdata2010/hp2010_final_review.pdf

35

Robbins AS, Yin D, Parikh-Patel A. 2007. Differences in prognostic factors and survival among white men and black me with prostate cancer, California, 1995-2004. *Am J Epidemiol ,*166:71-78.

Scanlan JP. 1987. The "feminization of poverty" is misunderstood. *Plain Dealer*, Nov. 11 (reprinted in *Current*, May 1988, and *Annual Editions: Social Problems* 1988/89, Dushkin 1988). http://www.jpscanlan.com/images/Poverty_and_Women.pdf

Scanlan JP. 1991. The perils of provocative statistics," *Public Interest*, 102 (Winter), 3-14 http://jpscanlan.com/images/The_Perils_of_Provocative_Stat.pdf

Scanlan JP. 1993. Getting it straight when statistics can lie. *Legal Times*, June 28. http://jpscanlan.com/images/Getting_it_Straight_When_Statistics_Can_Lie.pdf

Scanlan JP. 1994. Divining difference. *Chance*, 7(4):38-9,48: http://jpscanlan.com/images/Divining_Difference.pdf

Scanlan JP. 1996. When statistics lie. *Legal Times*, Jan. 1. http://jpscanlan.com/images/When_Statistics_Lie.pdf

Scanlan JP. 2000. Race and mortality. *Society*, 37(2):19-35 (reprinted in *Current* 2000 (Feb)): http://www.jpscanlan.com/images/Race_and_Mortality.pdf

Scanlan JP. 2001. The misunderstood relationship between declining mortality and increasing racial and socioeconomic disparities in mortality rates. "Making a Difference: Is the Health Gap Widening?," Norwegian National Institute of Public Health, Oslo Norway, May 14. http://www.jpscanlan.com/images/Oslo_presentation.ppt

Scanlan JP. 2006. Measuring health disparities. *J Public Health Manag Pract*,12(3):293-296. [letter]. http://www.nursingcenter.com/library/JournalArticle.asp?Article_ID=641470

Scanlan JP. 2006a. Can we actually measure health disparities? *Chance*, 19(2):47-51: http://www.jpscanlan.com/images/Can_We_Actually_Measure_Health_Disparities.pdf

Scanlan JP. 2006b. The misinterpretation of health inequalities in the United Kingdom. British Society for Population Studies Conference 2006. Southampton, England, Sept. 18-20. http://www.jpscanlan.com/images/BSPS_2006_Complete_Paper.pdf

Scanlan JP. 2007. Measurement problems in the National Healthcare Disparities Report. American Public Health Association 135th Annual Meeting & Exposition, Washington, DC, Nov. 3-7.
PowerPoint presentation. http://www.jpscanlan.com/images/APHA_2007_Presentation.ppt
Oral presentation (annotated). http://www.jpscanlan.com/images/ORAL_ANNOTATED.pdf
(Addendum to oral) (March 11, 2008): http://www.jpscanlan.com/images/Addendum.pdf

Scanlan JP. 2008. Can we actually measure health disparities? 7th International Conference on Health Policy Statistics, Philadelphia, PA, Jan. 17-18.
PowerPoint presentation. http://www.jpscanlan.com/images/2008_ICHPS.ppt
Oral presentation. http://www.jpscanlan.com/images/2008_ICHPS_Oral.pdf (oral)

Scanlan JP. 2011. Perverse perceptions of the impact of pay for performance on healthcare disparities. 9th International Conferences on Health Policy Statistics, Cleveland, Ohio, Oct. 5-7.
PowerPoint presentation. http://jpscanlan.com/images/Scanlan_ICHPS_2011.ppt
Oral Presentation. http://jpscanlan.com/images/ICHPS_2011_Oral.pdf

Scanlan JP. 2012. Flaws in tools for measuring healthcare disparities can exacerbate those disparities. Health Aff (Millwood), Jan. 30. [online response]
http://content.healthaffairs.org/content/30/6/1165/reply

Scanlan JP. 2012a. The lending industry's conundrum. National Law Journal, Apr. 2.

Scanlan JP. 2012b. "Disparate impact": regulators need a lesson in statistics. *American Banker*, June 4. http://www.americanbanker.com/bankthink/disparate-impact-regulators-need-a-lesson-in-statistics-1049886-1.html

Scanlan JP. 2012c. The mismeasure of group differences in the law and the social and medical sciences. Applied Statistics Workshop. Institute for Quantitative Social Science at Harvard University, Cambridge, MA, Oct, 17.
http://jpscanlan.com/images/Harvard_Applied_Statistic_Workshop.ppt

Scanlan JP. 2012d. Clarity in the reporting of health equity issues requires addressing measurement issues. *PLoS Medicine*, Nov. 3, 2012. [online response].
http://www.plosmedicine.org/annotation/listThread.action?inReplyTo=57571&root=57571

Scanlan JP. 2012e. Discussions of relative and absolute differences cannot ignore that there are two relative differences. *BMJ*, Nov. 8, 2012/ [online response].
http://www.BMJ.com/content/345/BMJ.e5774/rr/613496

Scanlan JP. 2012f. Misunderstanding of statistics leads to misguided law enforcement policies. *Amstat New*, Dec. http://magazine.amstat.org/blog/2012/12/01/misguided-law-enforcement/

Scanlan JP. 2013. Goodbye to the rate ratio. *BMJ,* Feb. 25. [online response]
http://www.BMJ.com/content/346/BMJ.e5793/rr/632884

Scanlan JP. 2013a. The mismeasure of discrimination. Faculty Workshop, University of Kansas School of Law, Lawrence, KA, Sept. 20.
http://jpscanlan.com/images/Univ_Kansas_School_of_Law_Faculty_Workshop_Paper.pdf

Scanlan JP. 2013b. The monitoring of health inequalities has never been sound. *BMJ*, Nov. 9. [online response]. http://www.bmj.com/content/347/bmj.f6576/rr/671152

Scanlan JP. 2014. Things government doesn't know about racial disparities," *The Hill*, Jan. 14. http://thehill.com/blogs/congress-blog/civil-rights/196543-things-the-legislative-and-executive-branches-dont-know

Smedley, BD, Sith AY, Nelson AR (eds). 2003. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Institute of Medicine of the National Academy of Sciences, Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Washington, DC. http://www.nap.edu/catalog.php?record_id=12875

Weissman JS, Betancourt JR, Green AR, et al. 2011. *Commissioned Paper: Healthcare Disparities Measurement*. Disparities Solution Center of Massachusetts General Hospital: http://www2.massgeneral.org/disparitiessolutions/z_files/Disparities%20Commissioned%20Paper.pdf

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

World Health Organization. 2013. Handbook on Monitoring Health Inequalities focused on low- and middle income countries. http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Fapps.who.int%2Firis%2Fbitstream%2F10665%2F85345%2F1%2F9789241548632_eng.pdf&ei=O476Ur6NBYfV0wG07ICgCg&usg=AFQjCNFzWcHv1r7f9-q2lws280nIL9nMLQ&sig2=jIwEYCybLPyYlRF34EbJbA&bvm=bv.61190604,d.dmQ