Methodological Issue in Comparing the Sizes of Differences Between Rates of Experiencing or Avoiding an Outcome in Different Settings

James P. Scanlan
Washington, DC
jps@jpscanlan.com

British Society for Population Studies 2007 Annual Conference
St. Andrews, Scotland, 11-13 Sep 2007

This presentation addresses certain issues in the use of four binary variables for measuring the size of differences between the rates at which two groups experience or avoid some outcome.[1]

[SLIDE 2]

These variables are:

1 Relative differences between rates of experiencing an outcome (Ratio 1)

2 Relative differences between rates of avoiding an outcome (Ratio 2)

3 Odds ratios

4 Absolute differences between rates

I'll first explain how Ratios 1 and 2 tend to changes systematically in opposite directions as an outcome changes in overall prevalence. Then I'll explain how odds ratios and absolute differences also tend to change systematically in opposite direction as an outcome changes in prevalence – though in a more complicated pattern.

I shall illustrate why these various measures will cause one to reach different conclusions about whether a difference between groups is larger in one setting or another. By different settings I mean at different points in time, in different geographical areas, or among different population subgroups.

My point does not involve which measure provides the most meaningful information. Rather my point is that none of these measures can, without more, provide accurate information about which difference is larger in a meaningful sense – that is, with respect to indicating whether the risk distributions of advantaged and disadvantaged groups are more similar in one setting than another

There may seem an unrealistic aspect to some of the points I'll be making. Many of you may think that numbers simply cannot be expected to behave according to the neat patterns I describe. And it is certainly true that few situations in reality will conform precisely to these patterns. But the underlying forces will have a sufficient role in almost every situation where we might try and compare the sizes of differences between rates in two or more settings that it makes no sense to do so while ignoring these tendencies. At the same time, the fact that we cannot predict the precise role of these tendencies in a particular setting will usually, and perhaps always, make it difficult or impossible to make reliable judgments while taking these tendencies into account.

[1] This presentation should be read in conjunction with PowerPoint presentation available at: http://www.jpscanlan.com/images/2007_BSPS_Presentation.ppt

This presentation is an extension of ideas presented at last year's BSPS conference in a presentation called "The Misinterpretation of Health Inequalities in the United Kingdom." There I tried to illustrate why the perceptions in the United Kingdom and virtually every other country studying health inequalities that such inequalities had been increasing does not have a sound foundation. Most notably, the perception is based on the fact that during periods of declining mortality relative differences in mortality rates between higher and lower social classes have been increasing. But such conclusions invariably have been reached without an appreciation of the way that, solely for reasons related to the characteristics of differing risk distributions, declines in mortality tend ordinarily to increase relative differences in mortality rates while reducing relative differences in survival rates.

Shortly after last year's conference I completed an extensive paper based on that presentation, which addresses a range of heath inequalities measurement issue with a particular focus on the UK. It is posted on my web site – jpscanlan.com – and is easy to find there.[2] While I will be going beyond the issues addressed in that paper, I mention it as a source for background to better understand the points I make today and as a source generally for understanding the problems affecting virtually all efforts to compare the size of health inequalities in different settings. Also, there is a Health Disparities Measurement tab on the site that contains a substantial body of material on these issues, including about 40 on-line comments – some quite extensive – questioning the reasoning of medical and health policy journal articles that fail to consider the extent to which the patterns they identify may be functions of the fact that an outcome is more common in one setting or another. I'll reference a few of those comments shortly.

So let's turn to Figure 1.

[SLIDE 3 - FIGURE 1]

Figure 1 is based on two normal distributions of factors related to experiencing some outcome, where the distributions have the same standard deviation and where the advantaged group (AG) has an average that is one half a standard deviation greater than the average for the disadvantaged group (DG). That means that about 30% of the group with the lower mean falls above the average for the other group. The numbers along the bottom, which are used as benchmarks for overall prevalence of some outcome, show the proportion of the advantaged group that falls below each point. Think of each point as representing a cutoff on a paper and pencil test and consider moving from left to right as reflecting the lowering of the cutoff on the test such as to serially enable the population between each point to pass the test. But the patterns I will show would also obtain if, instead of lowering the cutoff, we improved test performance to allow everyone between the two points to pass the test.

[2] The following is a direct link to the paper: http://www.jpscanlan.com/images/BSPS_2006_Complete_Paper.pdf.

The blue line with the diamond marker represents the ratio of DG's rate of falling below each point to AG's rate of falling below the point – Ratio 1 I've termed it. It reflects the relative differences between rates of experiencing the adverse outcome. And notice that that as we move from left to right and failure becomes less common, relative differences between failure rates tend to increase.

It is the failure to recognize this pattern – i.e., that the rarer an outcome the greater tends to be the relative difference between rates of experiencing it – that alone undermines so much research into health inequalities and every other area in which group differences are studied in terms of ratios of the rates of two groups.

[SLIDE 4 – FIGURE 2]

Next we examine the other side of the picture, the relative difference between rates of experiencing the opposite outcome – in this case the favorable outcome. Figure 2 adds to the first figure a red line with a box marker, which represents the ratio of AG's pass rate to DG's pass rate at each point (Ratio 2 I've termed it). And here we see that as we move from left to right and failure becomes less common – and success becomes more common – the relative difference in experiencing the favorable outcome declines.

Thus do we observe how relative differences in experiencing an outcome and relative differences in avoiding the outcome tend to move systematically in opposite directions as the prevalence of an outcome changes.

Note the point of intersection of these two ratios. For ease of reference, I'll refer to this as point Chi. Point Chi will always occur at a point where greater than 50% of the advantaged group experiences the favorable outcome. In the case of the half standard deviation difference in averages hypothesized here, the point occurs where approximately 58% of the advantaged group experiences the favorable outcome. So it shows up close to the point designated by the 40% failure rate for that group. To facilitate the subsequent discussion, I am going to call the area to the left of point Chi Zone A and that to the right of point Chi Zone B.

The next slide sets out a few of the implications of the pattern whereby relative differences between rates of experiencing an outcome and relative differences between rates of failing to experience an outcome tend to vary systematically in opposite directions as an outcome changes in prevalence.

[SLIDE 5 – IMPLICATIONS RE 1 AND 2]

First, and among the most important we have changes over time:

Whether an inequality is deemed to be increasing or decreasing will depend on which outcome we examine – mortality or survival, experiencing an illness or avoiding it.

Next, we have comparisons of inequalities in different populations:

British civil servants are deemed to have greater inequalities in mortality than the UK population at large. But because British civil servants have generally low mortality we should expect them to have large relative differences in mortality though small relative differences in survival.

Countries or regions with low mortality will tend to have high socioeconomic differences in mortality but low relative differences in survival,

In the US the black-white relative difference in infant mortality is greater among the well-educated than the poorly-educated. But that is to be expected simply because infant mortality is lower among the well educated. At the same time, the relative difference in infant survival is smaller among the well-educated.

In the US, healthcare disparities with respect to things like mammography have typically been measured in terms of relative differences in receipt of those procedures. Because these favorable outcome has been increasing, relative differences in receipt of them have been declining while relative differences in failure to receive them have been increasing.[3]

[SLIDE 6 – FIGURE 3]

Figure 3 adds a yellow line with a triangle maker. It represents the ratio of DG's odds of failure to AG's odds of failure. Some researchers favor odds ratios because differences measured by odds ratios remains the same whether one examines the favorable or the adverse outcome. Some also favor it because it can be easily derived from the standard output of a logistic regression.

The odd ratio starts out large when failure is rare, grows smaller as failure becomes more common, then grows large again when the outcome become rare. It is smallest near the intersection of Ratios 1 and 2.

[3] Because the Agency for Healthcare Research and Quality (AHRQ), the agency responsible for the US National Healthcare Disparities Report, has decided to measure healthcare disparities in terms of relative differences in experiencing adverse outcomes, it will tend to find a correlation between improvements in healthcare and increasing healthcare disparities, just the opposites of what the AHRQ expects. See:

Scanlan JP. Measurement Problems in the National Healthcare Disparities Report, presented at American Public Health Association 135th Annual Meeting & Exposition, Washington, DC, Nov. 3-7, 2007: http://www.jpscanlan.com/images/APHA_2007_Presentation.ppt; http://www.jpscanlan.com/images/ORAL_ANNOTATED.pdf

Scanlan JP. Recognizing the way correlations between improvements in healthcare and reductions in healthcare disparities tend to turn on the choice of disparities measure. *Journal Review* Nov. 9, 2007, responding to Kaytur FA, Clancy CM. Improving quality and reducing disparities. JAMA 2003;289:1033-34: http://www.journalreview.org/view_pubmed_article.php?pmid=12597759&specialty_id=22

[SLIDE 7 – FIGURE 4]

Figure 4 adds a chartreuse line with an x as a marker. It represents the absolute differences between rates. Some researchers favor absolute differences as measure of health inequalities because, like odds ratios, they are the same whether one examines the favorable or the unfavorable outcome. And some favor absolute differences over relative differences because they better reflect the proportion of the disadvantaged group that is harmed by its disadvantaged position. And we observed that absolute differences start out small when almost everyone experiences the adverse outcome, grow larger as the outcome becomes more common, then grow small again as the outcome becomes rare.

I have included the absolute difference on the same figure as the three ratios just discussed in order to show the relationship to the other measures, specifically that the absolute difference reaches a maximum at point Chi, the intersection of the Ratios 1 and 2 and the low point for the odds ratio. The absolute difference does not actually belong here, being a percentage point difference rather than a ratio, and it does not show up so well on the scale employed to show the patterns of changes in the other measures. So I have included the absolute difference alone in the next slide.

[SLIDE 8 - FIGURE 4A]

[SLIDE 9 - TWO CONTRASTING STUDIES]

Let us examine a couple of recent studies. The studies by Trivedi et al. and Jha et al. listed on the slide were among a group of three studies in a 2005 issue of the *New England Journal of Medicine* discussing changing racial inequalities in healthcare. The first study, that by Trivedi et al., examined a number of favorable healthcare outcomes – mainly process outcomes but a few clinical outcomes as well – where the overall rates were generally increasing over time. The study found that, for the most part, racial disparities – measured in terms of absolute differences between rates - were declining. Jha et al examined racial difference in rates of receiving certain procedures where the overall rates were also increasing. But the Jha study found that for the most part racial disparities – also measured in terms of absolute differences between rates – had been increasing.

So why the different findings?

Basically, Jha was examining relatively uncommon procedures like hip replacements where the rate ratios were generally far over in Zone A, where Ratio 2 was substantially greater than Ratio 1 and where overall increases tend generally to increase absolute differences. Trivedi, on the other hand, examined much more common outcomes that were usually in Zone B, and where further increases in prevalence tend usually to cause absolute differences to decline.[4]

[4] For further discussion of these and related studies, see the comments listed below:

[SLIDE 10 – TRIVEDI 2]

Trivedi et al. extended their work in a study appearing in 2006 in the *Journal of the American Medical Association.* The Trivedi *JAMA* study examined whether as plan quality improved (that is, plan quality measured in terms of overall rates of favorable outcomes), racial disparities (measured in terms of absolute differences between black and white rates) decline. The study found a very weak pattern to that effect. In this case the authors were looking at disparities entirely in clinical outcomes, and they opined that, consistent with patterns observed in the earlier study, apparently it was easier to reduce disparities in process outcomes than in clinical outcomes.

But in the 150 or so plans studied with respect to four types of clinical outcomes, overall, Ratio 1 was slightly higher than Ratio 2 – that is, just to the right of point Chi. This means that for some number of plans (those in Zone A) as quality improved absolute differences would be expected to increase, while in a majority of the plans (those in Zone B) increased quality would be expected to reduce absolute differences. Given that small absolute differences would be associated both with the lowest and the highest quality plans, the weak correlation of high plan quality and low absolute differences is pretty understandable. And as to the perception that improved quality reduces disparities in processes more so than for clinical outcomes, a good part of the reason for that difference

---

Scanlan JP. Effects of choice measure on determination of whether health care disparities are increasing or decreasing. *Journal Review* May 1, 2007, responding to 40. Understanding the ways improvements in quality affect different measures of disparities in healthcare outcomes regardless of meaningful changes in the relationships between two groups' distributions of factors associated with the outcome. *Journal Review* Aug. 30, 2007, Vaccarino V, Rathore SS, Wenger NK, et al. Sex and racial differences in the management of acute myocardial infarction, 1994 through 2002. N Engl J Med 2005;353:671-682 (and several other articles in the same issue):http://www.journalreview.org/view_pubmed_article.php?pmid=16107620&webenv=00P_2r_lHBKZPkExnEkCR_j5-u8waNcJ-87aLnoSJWxvN_ljFKstOR3CAx%402B600907661FF950_0034SID&qkey=1&rescnt=2&retstart=0&q=%22vaccarino+v%22+%22rathore+ss%22

Scanlan JP. Understanding the ways improvements in quality affect different measures of disparities in healthcare outcomes regardless of meaningful changes in the relationships between two groups' distributions of factors associated with the outcome. *Journal Review* Aug. 30, 2007, responding to Sequist TD, Adams AS, Zhang F, Ross-Degnan D, Ayanian JZ. The effect of quality improvement on racial disparities in diabetes care. Arch Intern Med. 2006;166:675-681: http://www.journalreview.org/view_pubmed_article.php?pmid=16567608&specialty_id=

Scanlan JP. Understanding patterns of correlations between plan quality and different measures of healthcare disparities. *Journal review* Aug. 30, 2007, responding to Trivedi AN, Zaslavsky AM, Schneider EC, Ayanian JZ. Relationship between quality of care and racial disparities in Medicare health plans. JAMA 2006;296:1998-2004: http://www.journalreview.org/view_pubmed_article.php?pmid=17062863&specialty_id=

Each of the above comments erroneously states that AHRQ (see note 3 *supra)* usually examines process outcomes in terms of relative differences in rates of experiencing the favorable outcome; but each was subsequently corrected in a later comment available at the same place.

is that satisfactory clinical outcome rates tend more often to be examined in zone a while process outcomes tend more often to be examined in Zone B. See D 41.

[SLIDE 11 – IMPLICATIONS RE 3 AND 4]

The discussion of these studies has largely illustrated the implications of measures 3 and 4 listed on slide 2 – odds ratios and absolute differences. But to summarize while focusing on absolute differences:

We should expect that when a favorable outcome is increasing in Zone A, the absolute difference increases; when it increase in Zone B, the absolute difference declines. When the changes occurs partly in Zone A then crosses over into Zone B, we would expect an initial increase in the absolute difference then a decline. Assuming that the favorable outcome is declining – that is, the pattern is moving from right to left – we should expect the opposite.

And for each thing that we observe for absolute differences, we would observe the opposite for odd ratios. So let us imagine that in an effort to make a comparison over time, a researcher decides it is important to make some adjustment for confounding factors. The most common means of doing so is by logistic regression. Since that usually will lead to the presentation of results in terms of odds ratios, there is reason to expect that the adjustment will lead to a conclusion about the direction of a change that is the opposite of that based on absolute differences.

[SLIDE 12 – FIGURE 5]

Now I'll quickly show how these patterns might appear with other than a perfectly normal distribution. Figure 5 is an illustration I typically use in the US in place of the one I have used here. It is based on black and white rates of having incomes below various ratios of the poverty line. And in the case of these near-normal distributions we see patterns very like those in the hypothetical data in Figure 4.

[SLIDE 13 – FIGURE 6]

Figure 6 is based on a truncated version of the distributions underlying Figure 4. It is based on the part of the total population falling below the indicator marked 30 in Figure 4. And we see that because this data is fairly regular, though no longer normal, we tend to observe the same patterns of changes in the relative differences in favorable and adverse outcome and in absolute differences observed in Figure 4. But notice that the odds ratio no longer behaves in the manner it did when the data were perfectly normal.

[SLIDE 14 – FIGURE 7]

Figure 7 is based on the systolic blood pressure (SBP) of black and white men between 55 and 64 in the US National Health and Nutrition Examination Survey. Again, the data are not close to normal. But we again see patterns for the relative differences in falling above and below each point – and for absolute differences – that we saw in Figure 4. But again, the pattern does not hold for odds ratios.

[SLIDE 15 – FIGURE 8]

Figure 8 is similar to Figure 7, but limited to a group that would meet the threshold for the systolic component of high blood pressure. This is akin to the universe at issue in the second Trivedi article where one of the outcomes was blood pressure control among persons diagnosed as hypertensive. The more limited universe is comprised of a very small sample, especially for blacks. So there is a great deal of variability. But again we see that the patterns tend to hold for relative differences and absolute differences, but not odds ratios.

And to add some concreteness to this discussion, let us suppose that different programs were able to achieve varying degrees of success in lowering blood pressure among hypertensive patients, such that the poorest programs enable only those with SPB not higher than 146 to reduce their SBP to 139 and better programs enable those with SBP levels much higher than 146 to reduce their SBP to 139. We would observe that, as programs grew more successful, relative differences in failing to control blood pressure would increase (blue line), relative differences in controlling it would decline (red line), and absolute differences would at first increase and then decline (chartreuse line).

[SLIDE 16 – WHICH MEASURE IS BEST?]

So, which measure is best?

None really tells the observer whether a change between rates is other than solely a consequence of changes in prevalence.

Further, each measure can change in one direction even when in fact there occurs a meaningful change in the opposite direction. For example, assume that improved test performance enabled 100% of one group but only 90% of the other group between two points to reach the higher cutoff. That would seem to be a meaningful change in the relative situation of the two groups. Yet in most cases each of the measures would still change in the same direction that we would observe in the case of simply lowering the cutoff.

[SLIDE 17 – HOW CAN WE MEASURE HEALTH INEQUALITIES?]

Next question: How can we measure meaningful changes over time? One way is to identify departures from the standard patterns, as, for example, when both relative differences change in the same direction (as discussed in the 2006 BSPS paper). But that only happens occasionally, and even then possible distributional irregularities can make the approach highly speculative.

[SLIDE 18 – OTHER POSSIBILITIES]

What are the other possibilities? It would seem that a meaningful change could be identified by comparing the effect size of the difference between averages in a true continuous variable. But many seemingly continuous variables are in fact functions of dichotomies.

As discussed in the 2006 BSPS paper, differences in longevity are affected by changes in mortality in ways such that changes in longevity differences cannot effectively identify changes in health inequality that are not solely reflections of overall changes in health. The same holds for measures like dichotomized self-rate health measures.

The next two measures – metabolic syndrome measures and cardio risk indexes – may seem like continuous variables where we might identify meaningful changes in the well-being of two groups by changes in the effect size of the difference between averages. But while each is seemingly continuous, in fact it is derived by attributing points to falling into categories and thereby implicating the same issues as longevity.

Allostatic load scores are based on points for falling in the quartile with the worst outcome. I think, but am not completely sure, that this could work as a means of appraising changing in relative health over time. The same holds for the next three measures. But, while these measures may not be subject to the problems with binary measures that I've described above, it is hard to know just how much they are going to tell us until one actually attempts to carry out analyses of inequalities through such approaches.

Further, smoking is one of the most important factors affecting health. Yet, as with other dichotomies, as smoking declines we should expect to observe difficult-to-interpret differential rates of decline in smoking – i.e., the group with the lower smoking rate will tend to show a larger proportionate decline in smoking rates while the other group will show a larger proportionate increase in rates of non-smoking. So suppose we observe some seemingly meaningful change in true continuous variable, but the change can then be traced to some uninterpretable differential change in smoking rates. In such circumstances, I am not sure just how we should regard the changes in the continuous variable.