Can We Actually Measure Health Disparities?

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

7th International Conference on Health Policy Statistics, Philadelphia, PA, Jan. 17-18, 2008

[The PowerPoint presentation accompanying this oral presentation may be found here. Related materials may be found on the Measuring Health Disparities and Scanlan's Rule pages of jpscanlan.com.]

This presentation examines four binary measures 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

2 Relative differences between rates of avoiding an outcome

3 Odds ratios

4 Absolute differences between rates

I'll explain how the first two measures 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'll illustrate why these various measures can cause one to reach different conclusions about whether a difference between groups is larger in one setting than another. Most importantly, the settings are defined temporally – that is, with regard to the measure of change over time. But the settings can also be defined – among other ways – geographically, nationally, or by some characteristic that distinguishes between or among subpopulations.

It is important to understand that my point does not simply involve the fact that one will tend to reach different conclusions depending on the choice of measure (though that is certainly a matter of some consequence). Nor does it involve which of these measures may offer the most meaningful information. Rather, the point is that none of these measures can, without more, provide useful information about which disparity is larger in a meaningful sense – that is, with respect to whether the underlying risk distributions of advantaged and disadvantaged groups are more similar in one setting than another.

One will of course observe many departures from the patterns I describe. These can occur because some meaningful change in the relative situation of two groups is sufficient to outweigh the tendencies. And it is in these situations that we have the greatest opportunity to identify important changes in disparities. But one may also observe departures from the patterns because of irregularities in the underlying distributions of factors associated with some outcome.

Yet the underlying forces will have a sufficient role in almost every situation where we might attempt to compare the sizes of differences between rates in two or more settings that it makes no sense to do so while ignoring the tendencies. At the same time, the fact

[1] This presentation should be read in conjunction with the PowerPoint presentation found at:

that we cannot predict the precise role of these tendencies in a particular setting will greatly complicates efforts to appraise the size of disparities while taking these tendencies into account.

[SLIDE 2 – REFERENCES]

Within a few days, this presentation will by up on my web site – jpscanlan.com – and soon thereafter a more extensive elaboration of some of the points. On that site there is a "health disparities measuring" tab listing 70 or so efforts, going back to 1987, at explaining the implications of these tendencies in various contexts, including the law as well as the social and medical sciences. This slide lists a few of the more important references, with section numbers indicating where they appear on the web page.[2] I will also make a few references to other items on this page according to those section numbers. So it you remain puzzled by anything I say here, there ought to be ample clarification on that site.

[SLIDE 4 - 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). 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 test on which two groups differ in their average performance, and consider moving from left to right as reflecting the lowering of the cutoff such as to serially enable the population between each point to pass the test. But recognize that we would observe the patterns if, instead of lowering the cutoff, we improved test performance such as to allow everyone between the two points to pass the test at the higher cutoff.

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, hence failing the test. For ease of reference in the following discussion I call such ratio AOR – for "adverse outcome ratio."

---

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

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

The Misinterpretation of Health Inequalities in the United Kingdom, presented at the British Society for Populations Studies Conference 2006, Southampton, England, Sept. 18-20, 2006:http://www.jpscanlan.com/images/BSPS_2006_Complete_Paper.pdf.

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/ORAL_ANNOTATED.pdf

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 – that is, 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 disparities and every other area in which group differences are studied in terms of ratios of the adverse outcome rates of two groups. Almost universally, during times of declining mortality and other adverse health outcomes, increasing relative differences in mortality have been interpreted to mean that health inequality is increasing in some meaningful sense. And such interpretation has been made without regard to whether the observed increases in adverse outcome ratios are more than – or less than – what would be expected to occur solely due to the decrease in overall prevalence. It has also been made without regard to whether the disparity in the opposite outcome is decreasing.

[SLIDE 5 – FIGURE 2]

So let us now 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 – termed FOR for "favorable outcome ratio." 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.

Sometimes relative differences in favorable outcomes are presented with AG's rate as the denominator – the opposite of what I have here. That has a minor implication as to the size of the percentage difference; but it is not otherwise of consequence. I specifically use DG's rate as the denominator here to facilitate putting both ratios on one figure, and to illustrate an additional matter concerning the intersection of the two ratios. And in that regard note that I have identified the intersection of the two ratios as Point X; the area to the left of Point X as Zone A; and that to right as Zone B. I'll give further attention to these designations shortly.

At this juncture, I note that many disparities between groups tend to measured in terms of relative differences in favorable outcomes. That is how the discriminatory impact of tests has typically been measured; and the reducing the cutoff has generally been regarded as a means of reducing the racial impact of a test because it reduces the relative difference in pass rates (even though doing so tends to increase the relative disparity in failure rates). Until recently relative differences in healthcare – mammography, prenatal care, and immunization – were typically measured in terms of the favorable outcome. And as those beneficial outcomes were becoming more common, the disparities were usually deemed to be declining.

But since 2004 NCHS has recommended that all disparities in health and healthcare be measured in terms of relative differences in the adverse outcomes. I am not sure that recommendation has yet had much effect outside of government. But, as Dr. Keppel will discuss, it does inform the government's approach to health disparities measurement.

I have challenged this approach in various places, and have done so particularly with regard to the National Healthcare Disparities Report published yearly by the Agency for Healthcare Research and Quality. For, even though AHRQ seems to believe that improvement in quality will tend to reduce healthcare disparities, improvements in quality – like reducing cutoffs or improving test performance – while tending to reduce relative differences favorable outcome rates, will tend to increase relative differences in adverse outcome rates. Thus, under the usual measurement approach of AHRQ, improvements in healthcare will tend to be perceived as increasing healthcare disparities.

[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 remain the same whether one examines the favorable or the adverse outcome. Some also favor it because it is a function of 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 AOR and FOR, that is, Point X.

And here let me clarify a key purpose of these illustrations. In order for a measure to reliably identify a meaningful change in a disparity – that is, one that is not simply the consequence of an overall change in prevalence – the measure must remain constant when there occurs solely a change in prevalence such as that effected by the lowering of a cutoff. That way, when the measure does change, we can know it means something of consequence. The prior figures illustrated whey neither AOR nor FOR serves that purpose. And, notwithstanding that odds ratios are functions of both failure rates and success rates, Figure 3 shows that odds ratios do not serve that purpose either.

[SLIDE 7 – FIGURE 4]

Figure 4 adds the absolute difference between rates, indicated by a chartreuse line with an x as a marker. Because the absolute difference involves a different scale, I have broken the figure into two parts. Some researchers favor absolute differences as measures of disparities 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.

But we observe that, like the other measures, absolute differences tend also to change solely because of a change in prevalence. The absolute difference starts out small when almost everyone experiences the adverse outcome, grows larger as the outcome becomes more common, then grows small again as the outcome becomes rare. Thus, changes in absolute differences do not alone provide means of identifying changes that are other than a result of changes in overall prevalence.

I have presented the various measures together in order to illustrate the relationship to the absolute difference to the other measures other measures – specifically, that the absolute difference reaches a maximum at approximately the intersection of the increasing AOR and the declining FOR, and that the absolute difference exhibits a pattern that is the opposite of the odds ratio pattern.

[SLIDE 8 – FIGURE 5 (midsection of Fig.4)]

Figure 5 provides somewhat greater detail across a mid-range of values that encompass Point X. I present this illustration because I have written a good deal about the way absolute differences tend to change in the two zones and I don't want to create an unjustified impression that great exactness. So you see that, even in this normal distribution, there is a broad range where it is hard to know exactly to expect with regard to absolute differences and odds ratios. That will be even more so in distributions where there any modest irregularity.

There is increasing debate about relative and absolute differences as a measure of health and healthcare disparities. And many often emphasize the importance of presenting both measures, sometimes observing that both provide important information (even when they provide opposite interpretations of a pattern of change over time). I maintain, however, that in fact neither changes in relative differences (whether AOR or FOR) nor changes in absolute differences provide useful information about meaningful changes over time unless examined with an understanding of the ways such measures tend generally to change as the prevalence of an outcome changes.

[SLIDE 9 – FIGURE 6 (income data)]

Next I want to quickly illustrate the same patterns with some other than hypothetical data. Figure 6 is a counterpart to Figure 4, based on black and white income, with the reference points on the X-axis being various percentages of the poverty line and the adverse and favorable outcomes being having an income falling below or above each of these points. And we observe that the same general patterns of the four measures that we observed with Figure 4 with respect to patterns of falling above and below each point. To make the illustration a bit more concrete, we observe in the far right how decreasing poverty, such as, for example, to elevate from poverty everyone between the poverty line and 50 percent of the poverty line, will tend to increase relative differences in poverty (though decrease relative differences in rates of avoiding poverty). On the other hand, increasing poverty would have the opposite effect. Because these changes would occur in Zone B, the decline in poverty would reduce the absolute difference between rates, while an increase in poverty would increase the absolute difference.

[SLIDE 10 – FIGURE 7 (NHANES SBP)]

Figure 9 is based on actual data on black and white systolic blood pressure (SBP) levels (based on men age 45-64 in the 1999-2000 and 2001-2002 NHANES samples). Now here there is a great deal of irregularity in the data, for the black sample is small. But we generally observe the same patterns of differences in rates of falling above and below certain levels that one finds in normal data. Thus, we see, for example, that – using 140 as a cutoff for hypertension – a program that brought everyone with SBP above 150 down to 140 would increase relative differences in exceeding the adverse outcome, , reduce relative differences in the favorable outcome. And, since we are here dealing with Zone B, absolute differences would decline while odds ratios would increase. Serially bringing under control the SBP of those at even higher levels would continue to show these patterns.

[SLIDE 11 – SEHGAL]

I want to further illustrate some of the implications of these tendencies by reference to a couple of studies. The next slide presents data from a 2003 study in the *Journal of the American Medical Association*. The study found that, during a period of substantial increase in rates of adequate hemodialysis, absolute differences between black and white rates declined. It has been often cited as showing how improving healthcare reduces disparities, including by AHRQ officials responsible for the national healthcare disparities reports. The figures on the chart show the black and white rates at the beginning and end of the period along with the AOR and FOR, and the absolute differences. The absolute difference and the relative difference in receipt of adequate care both declined during this period, indeed to the point of being what some might deem negligible. But note that the relative difference in adverse outcome, which is what AHRQ would use to measure this disparity, has increased. This is discussed further in the references noted at the bottom.[3]

I have also added two figures at the bottom. These reflect the effect size of the difference between means on a hypothetical underlying, continuously-scaled distribution of factors associated with receiving adequate hemodialysis. I'll return to these figures later when I discuss possibilities for addressing the problems I've raised with standard approaches to measuring disparities.

[SLIDE 12 - TWO CONTRASTING STUDIES]

3

The two studies on this slide were among a group of three studies in a 2005 issue of the *New England Journal of Medicine* discussing changing racial disparities 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 to the left-hand side of Zone A, 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.

Now I want to address certain aspect of the patterns I describe with respect to other than normal settings and in doing so address certain perceptions about the way improvements in health care may affect disparities in different types of outcome.

[SLIDE 13 – FIGURE 8 (normal truncated at 30)]

Figure 8 is simply a replication of Figure 4, except that it is restricted to the population below the point defined by a fail rate of 30 percent for the advantaged group. Now this population, a truncated part of a normal distribution, is not itself normal. Nevertheless, that within this truncated population that, as cutoffs are further lowered to allow parts of this subpopulation also to pass the test, we observe the same patterns of changes relative difference and the absolute difference as in the overall population. But we do not observe the same pattern for the odds ratio.

Now I show this partly as an exploration of the way that certain of the patterns observed in normal distributions tend also to exist in non-normal distributions. But it is also intended to illustrate something else. For sometimes health disparities are examined within overall populations, and sometimes they are examined with populations defined by the need for special care, as for example, among the hypertensive. And though the same tendencies may be found in the truncated population as in the larger one, there still may be different perceptions about disparities depending on which population we examine. This has certain implications as to the comparative size of AIR and FOR. But I will limit my observations here to implications for absolute differences.

[SIDE 14 – FIGURE 9 (normal and truncated AD)

The top part of the figure show the pattern of the change of the absolute difference in the total population as the cutoff is serially lowered. The bottom part shows the impact within the truncated population, as, say, the cutoff is serially lowered beyond point 30. And we see that because we are well over in Zone B in the larger population, the lowering tends to continue reduce the absolute difference within that population. Thus, for example, when the cutoff is reduced from point 30 to point 15, the absolute difference in the overall population falls steeply. But within the restricted universe, the changes occur in Zone A; thus, the absolute difference between rates, as examined within the restricted universe, increases.

[SLIDE 15 – FIGURE 10 (NHANES SBP>139)

Now let's consider the same thing with actual data. Figure 10 is based similar to Figure 7, which was based on NHANES data on systolic blood pressure of black and white men 45 to 64. But Figure 10 is limited to a group that would meet the threshold for the systolic component of high blood pressure. And we see that within this population, while there is much irregularity, we see essentially the same patterns of changes that we saw with truncated normal data. That is, we observe generally the same patterns of changes for relative and absolute measures that we observed within the larger universe – but a somewhat different pattern for odds ratios. But note that a good part of the focus within the truncated population will be in Zone A. Whereas within the larger population, point X was at about 135, within this truncated population, point X is just above 150. So, let us imagine that we reduce below 140 the blood pressure of everyone with blood pressure initially below 150. Within the overall population we would observe absolute differences to decline. But within the restricted population we would observe absolute differences increasing. These patterns as to absolute differences are shown in Figure 11, which is a counterpart to Figure 9.

[SLIDE 16 – FIGURE 11 (NHANES and truncated AD)

I present this illustration because the pattern observed in the Trivedi study whereby improvements in health care reduced absolute differences between rates was much more pronounced for process outcomes than for clinical outcomes, like control of hypertension; and the study is increasingly cited as illustrating the way improvements in healthcare tend to reduce process outcomes but not clinical outcomes. But I think the closely examined, the differences between the two types of disparities are functions of the different universes that tend to get examined and the differences with respect to where point X falls within each universe.

[SLIDE 17 – Continuous possibilities)

So let me briefly turn now to prospects for measuring disparities notwithstanding the issues I have raised above. The next slide lists some measures that on their face seem to be continuous. But often measures that seem continuous are merely functions of changes in some dichotomy and hence are subject to the same problems as standard dichotomous measures. In any case, various such measures are listed here, and, as I say, this presentation will be available on my web site for further review. But there is a question whether the measures most likely to meet the criterion of not being a function of a dichotomy, like self-rated health on a continuous scale, are going to prove that interesting to study.

[SLIDE 18 – Approaches 1 and 2]

In a few places I have broadly discussed the possibilities for identifying meaningful changes based on departures from the typical patterns I describe. In reference D43, I attempt to apply such approaches to data on changes over time in certain European countries. Also, in D43, I attempt an approach alluded to earlier with regard to the data on adequate hemodialysis in the Sehgal study. That is, one can take two groups' rates of experiencing an outcome at two points in time and derive an effect size for the difference between means of a hypothetical, underlying normal distribution of continuously scaled factors associated with the outcome, In the case of the Sehgal study, the decline from a .26 to a .14 standard deviation differences in hypothesized means employs the same principle.

You will see in D43 repeated questioning of the reliability of either approach given that we are not sure what the underlying distributions actually look like. But whether or not either approach is reliable enough give one any confidence in the results, to my mind each is superior to studying health disparities the way it is currently done. For an approach that is unmindful of the implications of the way measures tend to change solely due to prevalence changes not only erroneously attribute significance to changes that may be nothing other than the standard result of a change in prevalence comparable to the lowering of a test cutoff. Such approach also fails to identify meaningful changes when they do occur.