Perverse Perceptions of the Impact of Pay for Performance on Healthcare Disparities

9th International Conference on Health Policy Statistics
Cleveland, Ohio, October 5-7, 2011

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

Oral presentation accompany PowerPointPresentation

Good Afternoon. I am James Scanlan, a lawyer from Washington, DC.

My presentation is titled “Perverse Perceptions of the Impact of Pay for Performance on Healthcare Disparities.”

[SLIDE 2]

I have four main points to make here:

- Standard measures of differences between outcome rates (proportions) are problematic for appraising the size of health and healthcare disparities because each is affected by the overall prevalence of an outcome.
- Healthcare disparities research is in disarray because of observers’ reliance on various measures without recognition of the way each measure is affected by the overall prevalence of an outcome.
- There exists only one answer to whether a disparity has increased or decreased over time or is otherwise larger in one setting than another.
- That answer can be divined, albeit imperfectly, by deriving from each pair of outcome rates the difference between means of the underlying risk distributions.

[SLIDE 3]

Since I may fail to persuade you of these points completely in this short presentation, I will post a copy on jpscanlan.com, which will have links to key references, which I have listed here.

The Measuring Health Disparities page (MHD) has about 150 references (going back to 1987) explaining these points in particular settings in the law and the social and medical sciences.

The Pay for Performance sub-page of MHD provides quite a few references pertaining to the instant topic.

The Relative Versus Absolute sub-page of MHD refutes contentions that choice of measure involves some sort of value judgment.

The Solutions sub-page of MHD explains the approach to measuring disparities that is unaffected by the overall prevalence of an outcome.

Section E.7 of MHD shows the extent of scholarly agreement with my thinking.

The Scanlan's Rule page explains the nuances of the patterns I describe here and the Immunization Disparities of the Scanlan's Rule page provides a number of illustrations that are particularly pertinent to the instant topic.

The other references can speak for themselves. But I hope together they will answer any questions you have after I've finished today.

[SLIDE 4]

I first illustrate the way four standard measures of differences tend to be affected by the overall prevalence of an outcome. Since this presentation is focused on healthcare disparities – and rates of appropriate healthcare tend to be increasing – I cast the matter in terms of an increasing outcome. As that happens:

- Relative differences in experiencing the outcome tend to decrease.
- Relative differences in failing to experience the outcome tend to increase.
- Absolute differences between rates tend to increase to the point where the first group's rate reaches 50%; behave inconsistently until the second group's rate reaches 50%; then decline. Absolute differences tend also to move in the same direction as the smaller relative difference.
- Differences measured by odds ratios tend to change in the opposite direction of absolute differences.

[SLIDE 5]

Figure 1 is based on a situation of normal tests score distribution of an advantaged and disadvantaged group where the means differ by half a standard deviation. Think in terms of the effects on each measure of serially lowering a test cutoff to a point defined by certain pass rates for the advantaged group. From left to right we observe the implications of going from a point where pass rates are very low to a point where pass rates are very high.

These are the patterns I just described, including that relative differences in success rates and relative differences in failure rates tend to move systematically in opposite directions as the prevalence of test passage changes. It also shows the more complicated patterns for absolute differences and odds ratios. But in this short presentation, I can't spend much of time on this illustration, which I have used many times before. Again, the presentation is available on my website.

[SLIDE 6]

Figure 2 presents a clearer picture of the pattern of absolute differences. But even with perfectly normal data this pattern can be irregular in the mid ranges, as explained on the Scanlan's rule page.

[SLIDE 7]

While I have here used hypothetical, perfectly normal data, there are many publicly available data sets illustrating the same patterns, as shown in tables in the linked materials.

These data show, for example,

(1) how lowering poverty tends to increase relative differences in poverty rates while reducing relative differences in rates of avoiding poverty (see the table and figure in Can we actually measure health disparities?)

(2) how increasing folate levels tends to increase relative difference in low folate while reducing relative difference in adequate folate or how lowering blood pressure tends to increase relative differences in hypertension while reducing relative differences in rates of avoiding hypertension (see NHANES Illustrations sub-page of Scanlan's Rule page); and

(3) how relative differences in mortality tend to be greater among the young while relative differences in survival tend to be greater among the old (see Life Tables Illustrations sub-page of Scanlan's Rule page).

But similar patterns can be found in any data set that allows one to observe points on a continuum of factors associated with experiencing or avoiding an outcome.

[SLIDE 8]

I add here three reminders to prevent anyone from going off on the wrong track:

First, it does not matter that one observes departures from the described prevalence-related (or distributionally-driven) patterns. Of course one will. For actual patterns are functions of both (a) the prevalence-related forces and (b) the differences between the underlying distributions in the settings being compared.

[SLIDE 9]

Second, that the prevalence-related forces may depart from those I describe (e.g., distributions may be irregular) may complicate efforts to appraise the size of disparities using the method I describe below. But such possibility cannot justify reliance on standard measures of differences between outcome rates without consideration of the prevalence-related forces.

[SLIDE 10]

The third reminder is more of a caution. It would be mistake to find these points "interesting" then go on to do research using standard measures. If the points made here are valid, interpretations of patterns of changes using standard measures of differences between rates are invalid. They do not provide satisfactory results; they provide misleading results.

[SLIDE 11 - GOVERNMENT APPROACHES]

This slide lists three approaches of key federal government agencies to measuring health and healthcare disparities.

- NCHS always relies on relative differences in adverse outcomes.

- AHRQ relies on whichever relative difference (favorable or adverse) is larger.

- CDC relies on absolute differences between rates.

It should be evident from what I have already said that these approaches will commonly lead to different conclusions. The next slides will illustrate that fact as well.

[SLIDE 12]

Table 1 is based on a 2005 article in *Circulation* by Werner et al. that probably is the main reason that there is some interest in making reduction of disparities a performance criterion in pay for performance programs. The authors examined the effects of a coronary artery bypass graft (CABG) report card program on racial disparities in CABG rates. Relying on absolute differences between rates (Column 6), they found that the disparities increased as CABG rates increased.

But the initial rates were in ranges where absolute differences commonly increase solely for reasons relating to the underlying risk distributions. One will see in Columns 4 through 7 that each of the standard measures behaves in the way in which the distributional forces typically drive such measures.

The final column, termed EES for estimated effect size, involves deriving from each pair of rates the difference between means of the hypothesized underlying distributions. The procedure has apparently been around for quite some time in the form of the probit analysis. The Solutions sub-page of MHD mentioned earlier discusses some of its shortcomings. But it at least has a rational basis. And for all its shortcomings, the approach remains vastly superior to reliance on standard measures of differences without regard to the way such measures are affected by the overall prevalence of an outcome.

And the EES shows that contrary to the finding that cause such a stir, the racial disparity in CABG rates declined following implementation of the report card program.

The referenced comment discusses the study in somewhat greater detail.

[SLIDE 13]

Table 2 uses data from another study where the procedure rates at issue, like those in the Werner *Circulation* study, were in ranges where further overall increases tend to increase absolute differences between rates (as in fact happened as shown in column 7). But here the authors relied on relative differences in the favorable outcomes and found decreasing disparities, as reflected in the highlighted column 5. Of course, the relative difference in adverse outcome increased (Column 7). So NCHS would have found increased disparities. The final columns shows that, to the extent the disparity can be rationally measured, the disparities in fact decreased.

SLIDE 14

Table 3 is based on an award winning study of the effects on immunization disparities of a school entry hepatitis B vaccination requirement. Relying on relative differences in immunization rates, the authors found that the requirement, which dramatically increased overall vaccination rates, dramatically reduced disparities. The Favorable Ratio Column (COLUMN 6) shows the basis for that view. The Adverse Ratio column (Column 7) shows that NCHS would have found the disparities to be dramatically increased. These two same columns also suggest the varying views AHRQ would have as to different points in times. The absolute difference column shows the various perceptions of those like CDC who rely on the absolute difference.

The final column shows that in the main the disparities decreased.

[SLIDE 15

Table 4 is based on a 2003 study where, measuring disparities in terms of absolute differences between rates, Sehgal found that during a period of substantial increases in overall rates of adequate hemodialysis, the black white disparity decreased.

The article is interesting for a couple of reasons. First, it is commonly cited as evidence that improvements in healthcare tend to reduce disparities. Those so citing it include AHRQ, which measures disparities in terms of the larger relative difference. That approach tends systematically to reach conclusions that are the opposite of those one would reach based on the absolute difference.

The article is also interesting because it involves a situation where the rates initially were in ranges in which overall increases tend to increase absolute differences then moved into ranges where further increases tend to reduce absolute differences. But the figures in the table also show a common reality of standard measures tending to more or less move in accordance with the underlying distributional forces, but also interacting with meaningful changes and also showing the random variation that will commonly be observed from one year to the next. The first and last years, 1993 and 2000 provide the best information for drawing meaningful conclusions. But those conclusions cannot be based on standard measures, at least not without consideration of the way such measures are affected by the overall prevalence of an outcome.

[SLIDE 16]

Table 6, which is from the Relative Versus Absolute sub-page of MHD. It reflects an effort to demonstrate that choice of measure is not simply a matter of different ways of looking at things but that there can be only one reality as to the comparative size of disparities. The table shows the hypothetical hiring patterns of four employers in a situation where the qualification of the advantaged and disadvantage groups are exactly the same, and all differences in selection rates are due to employer bias.

So which employer is the most biased. The parenthetical figures show the ranking from most to least biased according to the four standard measures. These figures show, for example, exact opposite results for relative risks of selection and relative risks of rejection.

I use this example because I hope it will be obvious that it makes no sense to say that one employer is most biased as to selection while another is most biased as to rejection. It likewise makes no sense to say that one employer is more biased in relative terms and another is more biased in absolute terms. Only one employer can be the most averse to hiring the disadvantaged group. As some may surmise, the hypothetical is based on the same data as the earlier figures. And in fact each setting reflects exactly the same amount of bias – half a standard deviation between the means.

SLIDE 16 –CONCLUSION

- Researchers and governmental bodies need generally to rethink the way they measures health and healthcare disparities.
- Certainly we do not want to start paying providers on the basis of perceived effects on healthcare disparities until measurement issues are resolved.