Exploring Methods to Measure Health Inequalities That are Unaffected by Changes in the Prevalence of an Outcome[1]

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

Social, Cultural and Economic Determinants of Health: International Perspective for Global Action (1st Conference of the Journal Public Health, Journal of the Royal Institute of Public Health), Lisbon Portugal,
May 9-11, 2007

For several decades researchers around the world have been studying changes in the size of health disparities and comparing the size of health disparities in different settings. They have generally found, at least with respect to mortality, that disparities have been increasing. They have also found unexpectedly large inequalities within relatively advantaged subgroups of the population (e.g., larger social inequalities among British civil servants than found in the UK at large, larger racial inequalities in infant mortality rates among the well-educated than found in the US population at large).

The main findings of this research, however, are suspect for failure to consider the ways the sizes of differences in experiencing or avoiding an outcome are affected by the prevalence of the outcome. Most notably, the research has failed to recognize the statistical tendency whereby the rarer an outcome, the greater the relative difference in experiencing it and the smaller the relative difference in avoiding it. Thus, studies have failed to appreciate that, solely for statistical reasons, declining mortality would be expected to result in increasing relative differences in mortality or that relative differences in mortality will tend to be high in certain settings simply because mortality is

[1] To be read in conjunction with PowerPoint presentation available at: http://www.jpscanlan.com/images/J_Public_Health_Conf_PowerPoint_Presentation.ppt

low in those settings. Further, all measures of health inequalities based on binary variables change solely as a consequence of changes in the prevalence of mortality and other outcomes and hence cannot offer means of identifying changes in the relative well-being of two groups that are not solely results of changes in prevalence of an outcome.

[SLIDE 2]

I'll be addressing two matters. The first involves demonstrating the problems with measuring health disparities using binary variables. It is important to understand that my point in this regard is not that one may get different results or impressions depending on the choice of a measure of health disparities – though that is certainly true enough and a matter of some consequence. But my point is that none of the measures that are currently used to appraise the size health disparities is useful for that purpose. Further, it is not clear that there exist tools for appraising changes in disparities in things like mortality. That is, there seem to be no tools that can distinguish between changes in rate differences that are solely the result of changes in overall prevalence of an outcome and those that reflect something of greater consequence.

The second matter involves exploring whether there are ways to evaluate changes in disparities over time – or otherwise to compare the sizes of disparities in different settings – that are not subject to the same problems as binary variables.

[SLIDE 3 – IR1]

The next slide sets out what I call Interpretative Rule 1 (or IR1), a principle stated a moment ago but that warrants special emphasis.

> When two groups differ in their susceptibility to an outcome, the rarer the outcome, the greater tends to be the relative difference between rates of

experiencing it and the smaller tends to be the relative difference between rates of avoiding it.

Now this is merely a tendency that can be enhanced or mitigated by other factors. But it is a powerful tendency. And while it may not be the whole story as to every analysis of the size of health disparities, it is almost always an important part of that story.

I will demonstrate the tendency with several illustrations. And to the extent that I leave you puzzled or unpersuaded, I refer you to several papers on my web site for amplification. They are easy enough to find. The articles from Chance[2] and Society[3] under the first publications tab and the UK paper under the first presentations tab. There is also a page that separately compiles a large volume of materials addressing health disparities measurement issue in assorted contexts.

[SLIDE 4 – REFERENCES]

The tendency I have called IR1 can be illustrated with virtually any set of data that allows one to examine the rates at which two groups fall above or below various points on continuum of factors associated with some outcome. Here I use hypothetical test score data. In other places I have used actual income data But we can expect the same patterns to exist whenever two groups differ with respect to their distributions of

---

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

[3] Scanlan JP. Race and mortality. Society 2000;37(2):19-35: http://www.jpscanlan.com/images/Race_and_Mortality.pdf.

3. Scanlan JP. 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.

risk of experiencing some outcome, even when, as in the case of mortality and other adverse health outcomes, we cannot directly observe the distributions.

[SLIDE 5 – FIGURE 1 – FAILURE RATIOS]

The figures that follow are based on normal distribution of test scores of two groups – the higher-scoring group termed AG, for advantaged group, and the lower-scoring group termed DG, for the disadvantaged group. The distributions have the same standard deviation and the means differ by one half a standard deviation. In Figure 1, the numbers along the bottom reflect various cutoffs based on descending failure rates of AG, the higher-scoring group, which we use for reference. The Blue line with the diamond marker reflects the ratio of DG's failure rate to AG's failure rate at each cutoff – in other words, the relative difference in experiencing the adverse outcome. As we move from left to right, we observe what would occur if the cutoff is serially lowered from a point where almost no one passes the test to a point where virtually everyone passes the test. That is, we simulate the situation where an adverse outcome becomes increasingly rare.

And we observe that as we lower the cutoff, the ratio of the failure rate of the lower-scoring group to that of the higher-scoring group increases. For example, at the point where 50 percent of AG fails the test, the fail ratio is 1.4; at the point where 3 percent of AG fails the test, the ratio is 2.8. That is what I mean when I say the rarer an outcome, the greater tends to be the relative difference in experiencing it. Be mindful, moreover, that we would observe the same pattern if instead of lowering the cutoff, we improved test performance such that everyone scoring between two points was enabled to achieve the higher cutoff score.

[SLIDE 6 – FIGURE 2 – SUCCESS RATIOS]

Now let us examine the other side of the picture – the relative differences in pass rates. Figure 2 adds to figure 1 a red line with a square marker, which shows for each cutoff the ratio of AG's pass rate to DG's pass rate. That ratio declines – i.e., the difference grows smaller – as the overall failure rate declines. Thus, we observe that the size of relative differences in experiencing an outcome and in avoiding the outcome tend to move systematically in opposite directions as the prevalence of the outcome changes.

This has an important implication with respect to the evaluation of changes in the size of disparities. Some might be inclined to maintain that an increase in the difference between rates of experiencing an adverse outcome reflects some true worsening of the relative status of the disadvantaged group, even when the increase results solely from a general decline in the frequency of the outcome. Even allowing the plausibility of the point for a moment, one would have to regard such change as a matter of far less significance than a change that went beyond the usual consequences of an overall decline in an outcome. But any plausibility of a claim that an increase in the relative difference in adverse outcomes that flows solely from a decrease in prevalence of the outcome somehow reflects a true worsening of the relative situation of the disadvantaged group disappears when one recognizes that, if one appraises the same matter in terms of the favorable outcome, one has to conclude that the disparity has declined.

As it happens, relative differences in many indicators have traditionally been measured in terms of the favorable outcome. In the United States, the disproportionate impact of employment tests has long been evaluated in terms of relative differences in pass rates. Beneficial health procedures (e.g., prenatal care, immunization, mammography) have traditionally been evaluated in terms of relative differences in rates of receiving the procedure. Thus, the increased availability of such procedures has led to a perception that healthcare inequalities are declining, even as that same increased availability, by reducing certain types of mortality, has led to the perception that racial differences in those types of mortality are increasing.

[SLIDE 7 – FIGURE 3 – ODDS RATIOS]

Now let consider the odds ratio. The odds ratio is the same regardless of whether one examines the favorable or the adverse outcome. What we need in order to identify meaningful changes in disparities over time, however, is a measure that does not change solely as a result of a change in prevalence – in order that when it does change, we know the change really means something. But we see from the yellow line with the triangle marker in Figure 3 that odds ratios also change simply because of changes in prevalence of an outcome. They are large where the adverse outcome is nearly universal, grow smaller as the outcome becomes less common, then grow large again as the outcome becomes quite rare. So changes in odds ratios cannot be used to identify changes in health disparities that are not solely the result of changes in prevalence of an outcome.

[SLIDE 8 – FIGURE 4 – ABSOLUTE DIFFERENCES]

Absolute differences between rates are also the same regardless of whether one examines the favorable or the adverse outcome. For that and other reasons some favor absolute differences over relative differences in measuring disparities. But we see in Figure 4 that absolute differences also change as the prevalence of an outcome changes. They are small where failure is nearly universal, grow larger where failure becomes less common, and then grow small again as failure becomes rare. So, like the measure already discussed, absolute differences do not provide a ready means of identifying changes in the relative situations of two groups that are not solely the consequence of changes in overall prevalence of an outcome. So far as I can tell, at least with respect to binary outcomes, there is no measure that serves such a purpose and therefore no measure that we can use to measure health disparities without consideration of the statistical tendencies I just described.

Now, in some cases, by taking the above tendencies into account we might identify meaningful changes in health disparities, most obviously, where a mortality rate for one group increases while that for another declines. But in the more usual case, particularly when changes are substantial, the rates of advantaged and disadvantaged groups tend to move in the same direction. In such cases, drawing inferences will involve a good deal of guesswork about the characteristics of risk distributions that we cannot directly observe.

One objection some might make to the above illustrations is that distributions of risks of adverse outcomes like mortality will not always be perfectly normal. But the distributions are likely to be close enough to normal sufficiently often that it makes no sense to draw conclusions about changing health disparities without consideration of the tendencies I describe above. Actually the larger significance of the possibility that risk distribution will depart substantially from the normal is that such possibility further complicates efforts to draw meaningful inferences about changes in health disparities while attempting to take the tendencies I described into account.

[SLIDE 9 – POSSIBLE SOLUTIONS]

All that said, let us consider possibilities for measuring changes in disparities that are not subject to the problems discussed above. One such possibility was the subject of Dr. Bird's presentation[4] – allostatic load. Further, at least with respect to the US, the NHANES survey data used by Dr. Bird provides a broad sample of information not merely for allostatic load (which is a composite) but for the continuous variables that underlie the allostatic load scores. I know that there is also some similar data in the Whitehall studies, which I discuss in the latter part of the UK paper I mentioned.

In monitoring health disparities over time – or otherwise comparing them from one setting to another – I think that what we are interested in is the difference in averages on a continuous indicator of health status, measured probably in terms of the effect size (i.e., the difference in average over the pooled standard deviation). And I tentatively think that even in circumstances where distributions are changing for both advantaged and disadvantaged groups – which is what typically underlies the changes in overall prevalence of mortality and other adverse outcomes that we observe – we may be able to divine whether the distributions are becoming more or less alike. Now I recognize that there may be odd distributions where one might say that the effect size is not such a good measure of difference. But at least we can observe the distributions and figure those things out.

Data of that nature can also be used to test the premises underlying the main points I have made here. But whether or not they bear out my belief that the risk distributions will be more or less normal most of the time, I don't think there can be anything in such data that will tell us we can sensibly continue to measure differences in

[4] Bird C, Finch B, Escarce J, et al. Race/ethnicity, neighbourhood socioeconomic status and allostatic load.

rates of mortality and other adverse – or favorable – health or healthcare outcome without regard to the way such differences may change solely because of changes in the overall prevalence of the outcome.