Comment on Baicker et al. Health Affairs 2004

The material below was originally published on journalreview.org on March 28, 2008. That site was closed down at an unknown data prior to August 30, 2012. References below from that site have been replaced. This item was written before discovery of certain nuances regarding the relationship between absolute differences between rates and the prevalence of the outcome that are discussed in the introduction to the [Scanlan's Rule](#) page. But the implications of those nuances are limited.

Comparisons of the sizes of differences between black and white rates for different procedures are not informative without consideration of the overall levels for each procedure

Baicker et al.[1] discuss various patterns of racial disparities in receipt of 10 healthcare procedures among Medicare beneficiaries in 306 hospital referral regions (HRRs). In the text of the article, the authors discuss the disparities in terms of relative differences between rates of receiving the procedures (*i.e.*, 1 minus the black-white ratio of rates of receiving the procedure, which ratio is shown in the last column of Baicker's Exhibit 1). In the case of certain analyses of correlations between such things as the sizes of disparities for different procedures within individual HRRs, however, the authors measure the disparity in terms of absolute differences between rates (as clarified in the Appendix Exhibits).

The various conclusions of the authors about the patterns they observe and the implications of such patterns warrant some rethinking, however, because such conclusions were reached without consideration of the ways measures of differences between rates tend to be affected by the overall prevalence of an outcome. To begin with, the more common an outcome, the smaller tends to be the relative difference between rates of experiencing the outcome (in this case, receipt of the procedure) and the greater tends to be the relative difference between rates of failing to experience the outcome (failing to receive the procedure.) Thus, comparisons of the sizes of various disparities will tend to yield opposite results depending on which measure is chosen. With regard to the main textual discussion, since the authors measure disparities in terms of relative differences between rates of receiving procedures, there will be a tendency to observe smaller disparities for the more common procedures.[1a-3a]

Confronted with illustrations of patterns whereby relative differences in experiencing an outcome and relative differences in failing to experience the outcome tend to yield opposite results regarding the comparative size of disparities, the National Center for Health Statistics (NCHS) (while not discussing whether such patterns tend to be systematic), has recommended that, for reasons of consistency, all disparities, including disparities in receipt of beneficial healthcare procedures, be measured in terms of relative differences between rates of adverse outcomes (that is, in the case of healthcare procedures, failing to receive the procedure).[4-6] Such approach underlies the measurement of disparities in Healthy People 2010, the Midcourse Review of which was summarized in a recent Health Affairs article.[7] Thus, NCHS would use an approach to measuring each of the disparities shown in Exhibit 1 that is different from the approach of Baicker et al., and the implications of the difference in approaches will be discussed further below.

(I should qualify the statement regarding how NCHS would measure such disparities in the following regard. Apart from consistency considerations, NCHS, focusing on rare adverse outcomes, has also supported reliance on relative differences in adverse outcomes on the basis that in such circumstances the relative difference in the favorable outcome would seem trivial.[5,6]. Thus, it is not completely clear that NCHS would rely on the relative differences in the adverse outcome, where, as with many of the procedures examined by Baicker et al., that difference might seem trivial.)

The Agency for Healthcare Research and Quality (AHRQ), which is responsible for the annual National Healthcare Disparities Report, measures disparities in terms of whichever relative difference (in the favorable or the adverse outcome) is larger.[8,9] Since for most of the procedures in the Baicker study the former difference is larger than the latter, AHRQ would usually measure the disparity in the same manner as Baicker et al. But in two cases (HbA1c testing and eye test), AHRQ would do so in the manner of NCHS.

In addressing the way NCHS would measure the disparities, I do not mean to suggest that it is a superior way of measuring disparities to that employed by Baicker et al. And I note that the Baicker article was published before the NCHS issued the key 2005 report.[5] But the point is not that one measure is superior to another, since neither distinguishes between differences that are functions of overall prevalence from those that reflect something more meaningful. Rather, the point is that neither measure – nor absolute differences and odds ratios, as discussed below – can alone provide meaningful information about the comparative size of disparities. But I discuss the NCHS and AHRQ approaches to illustrate the general confusion in the area.

Figures that clarify the above discussion, along with figures underlying the discussion that follows, are set out in Table A to this comment, which may be accessed by this link: http://www.jpscanlan.com/images/Table_A_to_Baicker_Comment.pdf

While absolute differences are the same whether one examines the favorable or the adverse outcome, as with relative differences, absolute differences also tend to change as the overall prevalence of an outcome changes. Absolute differences tend to be small when an outcome is rare, grow larger as the outcome becomes more common, and then grow small again when the outcome becomes nearly universal.[3] Absolute differences tend to reach a high at approximately the point where the ratio of the advantage group's rate of experiencing the favorable outcome to that of the disadvantaged group's rate of experiencing that outcome (Ratio 1) equals the ratio of the disadvantaged group's rate of experiencing the adverse outcome to the advantaged group's rate of experiencing that outcome (Ratio 2).[3,10-12]

(As discussed in the references just listed, odds ratios tend to change in the opposite direction of the absolute difference. While such fact is important with regard to notions that odds ratios might not be affected by overall prevalence the way relative and absolute differences are, as has been recently suggested,[13,14] there is little point in encumbering this complex comment with discussions of odds ratios.)

As discussed in reference 9, Ratio 1 uses a different numerator from that typically used to calculate relative differences in experiencing a favorable outcome (and used by Baicker et al. here). While the difference in approach is relevant to determining which relative difference is larger (an issue of consequence to AHRQ) and the point of intersection of the two relative differences (an issue of consequence to certain interpretive issues addressed here), the difference in approach does not involve a substantive issue. In any case, consistent with ways I have commonly illustrated the matter, Table A presents Ratios 1 and 2, as well as the relative differences in favorable (and adverse) outcomes as such differences are typically calculated. To facilitate further discussion, areas where Ratio 1 exceeds Ratio 2, and where higher rates tend to be correlated with larger absolute differences, will be referred to as Zone A; areas where Ratio 2 exceeds Ratio 1, and where higher rates tend to be correlated with smaller absolute differences, will be referred to as Zone B, as in the figures in reference 12.

Some may be troubled by a seemingly artificial aspect to the zone concept, particularly since comparisons may be made across zones. I would add as well that even in normal data absolute difference may exhibit very little change (and exhibit some inconsistency in the patterns of such changes) across a fairly substantial spectrum in the portions of each zone bordering the intersection of the two ratios, as illustrated in Figure 5 of reference 12. Such factors can complicate the effort to interpret patterns of absolute differences, especially when Ratios 1 and 2 are close to equal in size, as discussed in reference 15. This may even be relevant to the mammography and diabetic procedures examined by Baicker et al. For the weighted averages in Exhibit 1 that underlie the calculations of Ratios 1 and 2 in my Table A may well encompass black and white rates in particular HRRs where the comparative size of the ratios (i.e., with respect to which of the two ratios is larger) may differ from that based on the weighted average. Nevertheless, I think the zone designations facilitate the discussion here, particularly with regard to the situations that are solidly in Zone A.

The tendencies described above, of course, are by no means the only factors underlying observed patterns of differences between rates, however measured. There also commonly exist meaningful differences between the comparative situations of two groups with regard to different outcomes or different settings. Such meaningful differences may be conceptualized as differences in the sizes of the differences between means of underlying distributions of factors associated with an outcome (which matter will be discussed further below). They may also be conceptualized in terms of the simple fact that for any given white rate the black rates may vary in different settings. And it should be obvious that the greater the meaningful difference between the situation of blacks and whites, the greater will be each of the measures of differences between rates for any given white rate, with such meaningful difference sometimes driving a pattern in the same direction as the statistical tendencies and sometimes driving a pattern in the opposite direction of the statistical tendencies.[3a] In sum, the statistical tendencies rarely will constitute the whole story with regard to observed patterns of varying sizes of differences between rates. But they usually will play a substantial role in that story and often play a predominating role.

Thus, the existence of these tendencies raises questions about the possibility of making meaningful comparisons about the size of disparities with regard to different outcomes or different settings. One approach to doing so is described in reference 17 (where it is described as Approach 2, which term will also be used as a shorthand reference here) and employed with respect to a variety of data in references 9 and 18-20. Based on each pair of rates, the approach derives an estimate of the difference between means (in terms of standard deviations or percentages thereof) of hypothesized underlying distributions of factors associated with the likelihood of experiencing the outcome. The approach is necessarily speculative in the typical situations involved in the study of health and healthcare disparities where the underlying distributions cannot be directly observed. Nevertheless, I think the approach is superior to anything else currently in use and certainly superior to the near universal approach of relying on one or another standard measure of differences between rates without regard to the extent to which the size of such difference is a function of the overall prevalence of an outcome. I discuss it here, however, not particularly to defend it, but merely to inform the discussion of the various analyses conducted by Baicker et al. For whatever its strengths or weaknesses, the approach provides a framework within which to evaluate the plausibility of conclusions based on other methods. Thus, Table A also includes the estimates just described, identified as EES (for estimated effect size).

Against that background, I now examine certain points made by Baicker et al. regarding varying sizes of disparities.

A. Disparities Discussed or Illustrated in Terms of Relative Differences in Favorable Outcomes

Baicker et al. note as an example of variations in disparities by type of procedure that whites get three time as many carotid endarterectomies as blacks but only 30% more angiograms. But, given the higher overall rates for angiograms than carotid endarterectomies, a disparity in this direction would be expected solely for statistical reasons. Similarly, solely for statistical reasons, we generally would observe a larger relative difference in rates of failing to receive angiograms than carotid endarterectomies (which in fact we find in Table A), as well as, given that the ratios of rates are in Zone A, a larger absolute differences for angiograms (which in fact is shown in Exhibit 1 of Baicker et al.). Approach 2, I note, in fact shows a larger difference for cardiac endarterectomies than angiograms (.34 versus .13 standard deviations). But that should not be read as a validation of reliance on relative differences without regard to the considerations discussed above.

Based on the differing sizes of relative differences, the authors also cite particular instances of regional variation in the size of disparities for certain procedures. One of these involves a 14% lower black than white rate of HbA1c testing in Washington, DC (where the black rate is 59% and the white rate is, by my calculation, 68.6%) compared with a 4% lower black rate in the Bronx (where the black rate is 53% and the white rate is, again by my calculation, 55.2%). The other involves a 74% lower black than white rate for hip replacement in Manhattan (where the black rate is 0.08% and the white rate is 0.31%) compared with a 22% lower black than white rate in Raleigh (where the black rate is 0.18% and the white rate 0.23%). (Baicker et al. show the Manhattan white hip replacement rate as 0.21% in the text, at Var-38, but the referenced 74% relative difference suggests that in fact the white rate is 0.31%).

It would seem likely that, in a meaningful sense, the Hba1c testing disparity is larger in Washington than the Bronx, among other reasons, simply because, contrary to the standard pattern, the larger relative difference is found where the procedure is more common. In the case of hip replacement, assuming I am correct about the Manhattan white rate, the disparity would seem obviously larger there, inasmuch as the black rate is lower in Manhattan than in Raleigh, while the white rate is higher in Manhattan than in Raleigh. But, again, reliable conclusions about disparity size cannot be based on relative differences without regard to other considerations.

Similarly, any effort to appraise the comparative size of the disparities shown in Exhibit 1, based either on the relative or absolute differences shown in that exhibit, will be fruitless unless undertaken with an appreciation of the role of overall prevalences (and perhaps even then).[3,3a] The very small relative differences in the diabetic procedures compared with the other procedures are hardly surprising given the high rates for those procedures. And the much larger relative differences in rates of failing to receive those procedures – the measure that would be used by NCHS and, for these procedures, by AHRQ – are shown in Table A. It might also be noted that, contrary to the tendency described above, HbA1c monitoring, the more common of the two diabetic procedures, shows a larger relative difference than eye exams, the less common procedure. Such fact might be cautiously regarded as evidence that the disparity in HbA1c monitoring is larger than the disparity in eye exams in a meaningful sense, as would the larger EES shown in Table A.

One may observe varying patterns of consistency or inconsistency with the described tendencies with regard to the specific differences shown for the three procedures that are neither diabetic-related nor cardiac-related (mammograms, hip replacement, back surgery). I merely note that while Baicker Exhibit 1 shows a much smaller relative difference for mammography (the far more common procedure) than the two others, Approach 2 suggests that the disparity in fact is substantially larger for mammography.

With respect to the five cardiac procedures (which the text and Appendix Exhibit A1 indicate include catheterization, though a line divides it from the other cardiac procedures in Exhibit 1), Table A shows that, apart from some inconsistence with regard to angiography and catheterization, the patterns of relative differences in favorable and adverse outcomes are basically what one would expect given the overall frequency of each procedure. That is not to say there are no meaningful differences between the disparities – and the EES figures in Table A reflect an attempt to quantify such differences – but merely that the standard measures are inadequate for identifying those differences.

Finally, I note that Exhibit 1 itself shows for the cardiac procedures an almost complete reversal of the rankings of the size of disparities depending on whether one relies on relative or absolute differences, which is the sort of pattern the statistical tendencies would drive. And I note that many researchers, such as those to which references 10, 15, and 16 mainly respond, would rely on absolute differences to compare the size of disparities. Again, however, I note such fact merely to illustrate the anomalies in this area of research rather than to suggest that one measure is better than another or even is useful at all.

B. Analyses of Correlations

Baicker et al. conducted analyses of correlations between the sizes of disparities for different procedures within HRRs and between black rates and sizes of disparities for different procedures across HRRs, both of which analyses warrant comment. As noted above, these analyses rely on absolute differences between rates as measures of disparities.

1. Correlations between Disparities for Different Procedures within HRRs

With respect to the between-procedure correlations within HRRs, the authors find the absolute differences between black and white rates in cardio procedures to be strongly correlated with one another but not to be particularly correlated with other procedures. They also find the sizes of the disparities in the other procedures not to be much correlated with one another. A problem with deriving meaning from this analysis is that, both where correlations are found and where they are not, it is not possible to know the extent to which the results are driven by similar (or different) disparities in the white and black distributions of factors related to receiving the various procedure or by similar (or different) overall levels of the procedures. In the case of the strong correlations between pairs of cardio procedures, I suspect that the pattern is substantially driven by the fact that HRRs with high (or low) levels of one procedure also have high (or low) levels of the other procedures, with corresponding high (or low) absolute differences for most of the procedures. If relative differences in receiving the procedure were used as the measure of disparity, there is no reason to expect the results to be markedly different. That is, whereas an area that showed high rates for each cardio procedure would tend to show large absolute differences, but small relative differences, the degree of correlation between the sizes of the differences for each procedure within HRRs ought not to vary systematically depending on which measure is used.

That point, however, holds only as to cardiac procedures, for which all ratios of rates are well into Zone A (i.e., Ratio 1 is far larger than Ratio 2). Exploring correlations between the cardiac procedures and the diabetic procedures adds an additional wrinkle since the ratios of rates for those procedures are in Zone B. Thus, for example, if HRRs that have high (or low) levels of cardiac procedures have similar levels of diabetic procedures – and it would not be surprising to find some correlation between those levels within HRRs – the statistical factors would tend toward showing comparatively large absolute differences for the former but comparatively small absolute differences for the latter. By contrast, however, since relative differences tend to be smaller where overall levels are higher regardless of the zone, the statistical tendency would drive those differences in the same direction for both sets of procedures.

2. Correlations between Black Rates and the Size of Disparities across HRRs

With regard to the correlations between black rates and the size of disparities, the authors make the following point.

"Racial disparities in utilization of some surgical procedures seem to be driven by above-average white rates, rather than by below-average black rates, with positive correlation coefficients between disparities and black rates for CABG ($\rho = 0.20$, $p = .07$) and carotid endarterectomy ($\rho = 0.43$, $p < .01$). This seemingly paradoxical result is driven by the fact that black rates for these procedures tend to be somewhat higher in regions where white rates are very high."

It first warrants note that, in fact, black rates tend usually be higher where white rates are higher. But the statistical factors associated with the high overall rates in such circumstances will tend toward causing a comparatively narrow relative gap, but toward causing a comparatively large absolute gap in Zone A and a comparatively small absolute gap in Zone B. The correlation of large black rates with large disparities thus should not be surprising for CABG and cardiac endarterectomy, for which the ratios of rates are well into Zone A, even though there exists the countervailing fact that for any given white rate a lower black rate will reflect a larger disparity (hence tending to correlate lower black rates with larger absolute (and relative) differences). Even less surprising should be the fact that, as shown in Exhibit A2, the strongest negative correlations are for eye exams and HgA1c monitoring. The ratios of rates for these procedures are in Zone B. Thus, the negative correlation will be driven both by the tendency for higher black and white rates to be correlated with lower absolute differences and by the fact that, for any given white rate, the higher the black rate the lower the absolute difference.

Were relative differences between rates of receiving the procedure the measure of the disparity in this analysis, it is likely that most or all procedures would show negative correlations, and possibly quite strong ones, between the disparity and the black rate. For, as in the case of absolute differences in Zone B, both forces tend in that direction.

Finally, I note that some may (very reasonably) question my usage of terms like "common outcome" and "overall prevalence (or level)," since the overall prevalence will be a function of the black and white rates as well as the proportion blacks comprise of the total population. And in fact, while commonly using the phrase without explanation, I generally proxy it by means of the rate of the advantaged group (assuming, as discussed in the second endnote to reference 15, that my meaning is clear enough). It warrants note, however, that correlations of the disparities with the white rate would likely be substantially or entirely driven by the statistical tendencies, since the effects of any meaningful differences between disparities, such as would be reflected by differing black rates for a given white rate, would generally be random. That is not to say there could be no correlation (positive or negative) between true disparities and overall white rates. There may be such correlations, though they likely will vary with the character of the locale. But it is not necessary to explore such issues to make the several points I want to make about measurement issues raised in the article by Baicker et al.

References:

1. Baicker K, Chandra A, Skinner JS, Wennberg JE. Who you are and where you live: how race and geography affect the treatment of Medicare beneficiaries. *Health Affairs* 2004:Var-33-Var-44.

1a. Scanlan JP. Divining difference. *Chance* 1994;7(4):38-9,48: http://jpscanlan.com/images/Divining_Difference.pdf

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

3. 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

3a. 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

4. Keppel KG, Pearcy JN, Klein RJ. Measuring progress in Healthy People 2010. Healthy People statistical notes. No. 25. Hyattsville, Md.: National Center for Health Statistics: http://www.cdc.gov/nchs/data/statnt/statnt25.pdf

5. Keppel KG, Pamuk E, Lynch J, et al. Methodological issues in measuring health disparities. Vital and health statistics. Series 2. No. 141. Washington, D.C.: Government Printing Office, 2005.(DHHS publication no. (PHS) 2005-1341.): http://www.cdc.gov/nchs/data/series/sr_02/sr02_141.pdf

6. Keppel, KG, Pearcy JN. Measuring health disparities in terms of adverse outcomes. J Public Health Management Practice. 2005:11(6): 479-83

7. Keppel KG, Bilheimer L, Gurley L. Improving population health and reducing health disparities. Health Affairs 2007;26(5):1281-1292.

8. Agency for Healthcare Research and Quality, 2006 National Healthcare Disparities Report: http://www.ahrq.gov/QUAL/nhqr06/nhqr06.htm

9. Scanlan JP. Addendum (Mar. 11, 2008) to 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/Addendum.pdf

10. Scanlan JP. Effects of choice measure on determination of whether health care disparities are increasing or decreasing. *Journal Review* May 1, 2007, responding to 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://jpscanlan.com/images/Vaccarino_NEJM_2005.pdf

11. 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.

PowerPoint Presentation: http://www.jpscanlan.com/images/APHA_2007_Presentation.ppt
Oral Presentation: http://www.jpscanlan.com/images/ORAL_ANNOTATED.pdf

12. Scanlan JP. Can We Actually Measure Health Disparities, presented at the 7th International Conference on Health Policy Statistics, Philadelphia, PA, Jan. 17-18, 2008 (invited session).

PowerPoint Presentation: http://www.jpscanlan.com/images/2008_ICHPS.ppt
Oral Presentation: http://www.jpscanlan.com/images/2008_ICHPS_Oral.pdf

13. Houweling TAJ, Kunst AE, Huisman M, Mackenbach JP. Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health. *International Journal for Equity in Health* 2007;6:15: http://www.equityhealthj.com/content/6/1/15 (accessed 22 Feb 2008)

14. Scanlan JP. Reconsidering a landmark study. *Lancet* Feb. 25, 2008 Mackenbach, JP, Kunst, AE, Cavelaars, et al. Socioeconomic inequalities in morbidity and mortality in western Europe, *Lancet* 1997; 349: 1655-59: http://jpscanlan.com/images/Mackenbach_Lancet_1997.pdf

15. 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://jpscanlan.com/images/Trivedi_JAMA_2006.pdf

16. 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://jpscanlan.com/images/Sequist_Archives_Int_Med_2006.pdf

17. Scanlan JP. Comparing the size of inequalities in dichotomous measures in light of the standard correlations between such measures and the prevalence of an outcome. *Journal Review* Jan. 14, 2008, responding to Boström G, Rosén M. Measuring social inequalities in health – politics or science? *Scan J Public Health* 2003;31:211-215: http://www.jpscanlan.com/images/Bostrom_and_Rosen_Comment.pdf

18. Scanlan JP. Comparing health inequalities across time and place with an understanding of the usual correlations between various measures of difference and overall prevalences. *Journal Review* Jan. 30, 2008, responding to Moser K, Frost C, Leon D. Comparing health inequalities across time and place—rate ratios and rate differences lead to different conclusions: analysis of cross-sectional data from 22 countries 1991–200. *Int J Epidemiol* 2007;36:1285-1291: http://jpscanlan.com/images/Moser_Int_J_Epidemiol_2007.pdf
-

19. Scanlan JP. Pay-for-performance implications of the failure to recognize the way changes in prevalence of an outcome affect measures of racial disparities in experiencing the outcome. *Journal Review* Feb. 8, 2008 (responding to Werner, RM, Asch DA, Polsky D. Racial profiling: The unintended consequences of coronary artery bypass graft report cards. *Circulation 2005;*111:1257–63): http://jpscanlan.com/images/Werner_Circulation_2005.pdf

20. Scanlan JP. Perceptions of changes in healthcare disparities among the elderly dependant on choice of measure. *Journal Review* Feb. 12, 2008 (responding to Escarce JJ, McGuire TG. Changes in racial differences in use of medical procedures and diagnostic tests among elderly persons: 1986-1997. Am J Public Health 2004;94:1795-1799): http://jpscanlan.com/images/Escarce_McGuire_2004.pdf