Slides 145 to 150 of the “The Mismeasure of Health Disparities in Massachusetts ar

The comment below was posted on journalreview.org on June 2,2007. In light the closing of that site, the comment is reproduced here.

A study with a variety of problems.

The article by Schulman et al.[1] was widely reported by the media immediately after its publications, with most reportage to the effect that blacks and women were 40% less likely to be referred for cardiac catheterization than whites and men and that black women were 60% less likely to be referred than white men.

Table 4 of the study, however, provided the following unadjusted referral rates for whites and blacks and for men and women.

White 90.6% Blacks 84.7%
Men 90.6% Women 84.7%

This information, along with information in a table indicating that rates for white men and black men were the same and the rates for white men and white women were the same, allowed the astute reader to divine the rates for each racial/gender group. Schwartz et al.[2] did so, and presented the following figures in a Sounding Board comment.

White men 90.6%
White women 90.6%
Black men 90.6%
Black women 78.8% (actually 78.9%)

Schwartz et al. explained that the figures reported in the media were actually odds ratios, which, for a common outcome, tended to substantially overstate the relative risk. They pointed out that the .6 odds ratio for blacks versus whites and women versus men reflected risk ratios of .93 and the odds ratio of .4 for black women versus white men reflected a risk ratio of .87. That the figures reported in the media were odds ratio was evident from the Schulman article, though the authors had not corrected the media on that point. Schwartz et al. also criticized the article for reporting a racial difference and a gender difference when only black women had lower referral rates than the other racial/gender groups. They suggested that the exaggeration of the differences by use of odds ratios and the presenting of findings as if all blacks and women were referred at lower rates (rather than only black women) would tend to undermine patient trust. They also raised a question about the high level of referral rates in general for a procedure that is not always the appropriate course.

Schulman and two co-authors responded,[3] agreeing that odds ratios had the potential to mislead that in retrospect they should have converted odds ratios to risk ratios and underscored the absolute risks. But they defended their reporting of results for blacks generally and women generally as being in accord with fundamental statistical principles, since the study hypothesis had specified racial and gender comparisons. And they disputed that the study would undermine trust between physicians and patients,

maintaining that the study would foster more honest dialogue and encourage members of the medical profession to seek ways to eliminate unconscious bias. They also included a table showing the referral rates, relative risks, and odds ratios for each race/gender/age group, which showed, inter alia, that the 78.9% black female rate that Schwartz et al. had extracted from the original article was the result of a 73.3% referral rate for the 70 year-old black actress and an 84.4% rate for the 55 year-old black actress.

White 55 year-old man 91.1% White 70 year-old man 90.0%
Black 55 year-old man 91.1% Black 70 year-old man 90.0%
White 55 year-old woman 92.2% White 70 year-old woman 88.9%
Black 55 year-old woman 84.4% Black 70 year-old woman 73.3%

Schulman et al. defended the failure to include the table in the original article on the grounds that the grant proposal did not include a hypothesis based on three-way interaction of race, sex, and age, and that a joint test of the three-way interaction was not significant.

The Journal's editors also replied to Schwartz et al.[4], taking responsibility for what they termed "the media's overinterpretation of the article," and stating that they (the editors) should not have allowed the use of the odds ratio in the abstract. They also noted that it was unfortunate that the table provided by Schulman in response had been removed from the original article (apparently in response to a reviewer's concern about subgroup analyses). Noting that the table showed that the study's findings were largely due to the physicians' recommendations concerning the 70-year-old black actress, and to a lesser extent the 55 year-old black actress, the editors observed that "this important point should have been made more explicit in the article."

Notwithstanding the Schwartz criticism, the editors' acknowledgment, and a modest amount of reportage of the issues raised by Schwartz and her colleagues, the Schulman article continued to exert considerable influence, being cited in debates leading to the passage of the Minority Health and Health Disparities Research and Education Act of 2000. One commentator has observed that "publication of the Schulman study did more than any other single event to put the matter of racial disparities in health and medical care on the American public policy agenda – and to frame political discussion of the topic."[5] And the article is still regarded as probative of racial and gender differences in cardiac care, currently with little attention paid to the early criticism. The technique employed in the study – having physicians review videotapes of black and white actors reading identical scripts and review identical sets of test results – has also been cited as an example of a technique that can better identify bias by eliminating the effects of potentially confounding factors that call many other studies into question.

The criticism by Schwartz et al. was sound. But there were a number of additional problems with the Schulman study that were unaddressed or inadequately addressed by Schwartz et al. or other critical commentary. These involve both strict data issues and conceptual issues. Because of the continuing significance of the Schulman article, including that it may serve as model for further research, these problems warrant detailed

discussion

Four issues are addressed below. The first involves the way decisions the authors made about the conduct of the study, including the decisions to craft symptom scenarios that would call for catheterization in a very high proportion of cases, essentially foreordained that the study would yield small relative risk differences but large odds ratio differences. That issue also involves the fact that the logistic regression, and hence the deriving of an odds ratio, was unnecessary.

The second issue involves the fact that, assuming it made sense to present overall gender differences when only black women had rates different from any other racial/gender group, the gender effect, whether in terms of relative risk or odds ratio, was substantially overstated because of a failure to adjust for the fact that white women vastly outnumber black women in the population at large.

The third issue involves the presentation of overall race and overall gender effects when only black women were referred at a lower rate than the other groups; the authors' obscuring of the fact that white men, black men, and white women were all referred at the same rate; and the author's failure to scrutinize their anomalous results. That issue also involves the need for special scrutiny of anomalous results in circumstances where such results do not involve decisions made about many individuals, but decisions made about only one or two individuals.

The fourth issue involves a general questioning of whether discrimination can be identified in a simulated setting of this nature as well as the authors' reasoning in attempting to adjust for physicians' perceptions of the subjects' personal characteristics.

A. Reasons That the Study Design Would be Expected to Yield Both Small Relative Risk Differences and Large Odds Ratio Differences

One defense that has been offered of the way Schulman et al. presented the referral recommendation differences in terms of odds ratios rather than relative risks was that, had they presented the results in terms of the relative risks of failing to be referred for catheterization, the differences would have been just as striking as the odds ratios. That is, the black non-referral rate would have been 1.7 times the white rate; the female non-referral rate would have been 1.7 times the male rate; and the black female non-referral rate would have been 2.4 times the rate of each of the other groups. This is of course true and in fact a press release issued by Georgetown University Medical Center in conjunction with the publication of the study described the differences in terms of relative differences in rates of not being referred. It also warrants note that, according to recommendations of the National Center for Health Statistics (NCHS) that all health disparities, including disparities in receipt of beneficial procedures, be measured in terms of relative differences in adverse outcomes (i.e., in the case of beneficial procedures, rates of failing to receive the procedures)[6-9], that is exactly how the differences should have been presented.

The NCHS recommendations are unsound, however, among other reasons, for the failure to recognize that relative differences in a favorable outcome and relative differences in the contrary (adverse) outcome tend to vary systematically in opposite directions as the prevalence of the outcomes changes.[10-12] But that tendency, and other ways measures of differences between rates vary depending on the prevalence of an outcome are also important to appraising certain decisions made in the Schulman study. While Schwartz et al. maintained that the high overall referral rates were suggestive of unjustified referrals, it must be recognized that the situation did not involve symptoms of some group of actual patients where such high referral rates might indeed reflect unjustified referrals by physicians (depending of course on the subject universe). Rather, the fictional symptoms of the subjects were crafted by the designers of the study, presumably with the recognition that they were of nature that would generally result in catheterization recommendations in close to 90% of the cases. Indeed, given that there were 18 symptom scenarios, it seems likely that a high proportion of them resulted in referral recommendations in 100% of cases (and the non-referral recommendations may have been concentrated in several categories). Further, one would expect these symptom scenarios overall to yield even higher referral rates in a real-world setting where malpractice concerns would likely tend to elevate referral rates. The decision to fabricate circumstances where referral would almost invariably be the norm, rather than one involving a substantial amount of physician discretion, seems an odds one. But it presumably was deliberately made and it would have been useful for the authors to have explained the rationale for such a course.

In any event, there are reasons to expect certain implications of that decision with respect to the types of referral disparities that would likely result (assuming that there should be any tendency toward racial or gender differences in referral rates). When there is some difference in group susceptibility to an outcome, the more common the outcome, the smaller will tend to be relative differences in rates of experiencing the outcome and the greater will tend to be relative differences in rates of failing to experience the outcome.[10-15] For example, as shown in Table 1 of reference 12, where distributions are normal, and these is one half a standard deviation difference between averages (and the distributions have the same standard deviation), at the point where the group more likely to experience an outcome experiences it 90 percent of the time, the other group will experience it 78.2 percent of the time, a relative risk of .87 for the less susceptible group (though that group's relative risk of failure to experience the outcome will be 2.2); by contrast, where only 50 percent of the more susceptible group experiences the outcome, the less susceptible group will experience it 30.9 percent of the time, a relative risk of .62 (though a relative risk of failure to experience the outcome of only 1.4). Thus, by choosing to craft symptom scenarios that would yield referral recommendations in the overwhelming majority of cases, the Schulman group was tending to guarantee that any modest race or gender effect (whether the result of discrimination or of characteristics of the actors) would result in only a small relative difference in rates of referral. At the time of the study, it warrants note, studies of healthcare disparities tended generally to examine rates of receipt of procedures rather than non-receipt of procedures, an approach which seems still the norm notwithstanding the recent NCHS recommendations to the contrary.

On the other hand, however, the same context would tend to a yield a very large odds ratio difference, which is typically the case when the overall rate is very high.[10,12]. At least the initial deriving of an odds ratio appeared to be justified (even to Schwartz et al.) on the basis that the odds ratio is the output of the logistic regression Schulman et al. employed for their analysis (though Schwartz et al. thought each odds ratio should then have been converted to a relative risk). But that raises the question of why Schulman et al. needed a logistic regression at all. Logistic regression is a useful procedure for analyzing the role of various factors in situations where subjects differ in a variety of respects. Yet the distinctive feature of the study – indeed, the one must emphasized by the authors – was that the actors were made to appear exactly alike with respect to the factors affecting suitability for catheterization. Thus, a particular value of the study design lay in that there was no need to use logistic regression or any other procedure to adjust for differences among the groups that would tend to affect referral rates.

The study did include the physicians' estimates of the probability of heart disease for each subject in the primary analysis (a factor that was not self-adjusting), and the authors regarded this as a strength of the study. But those estimates did not alter the results, and, indeed, the odds ratios yielded by the logistic regression that took those estimates into account and that were reported in the study's Table 5 (.6 for blacks compared with whites, .6 for women compared with men, and .4 for black women compared with white men) were rounded equivalents of the odds ratios one would derive from the unadjusted referral rates (.61, .61, and .39), while the odds ratios for black men and black women compared with white men, at 1.0, were exactly what one would derive from the unadjusted referral rates. Although it might be said that the regression did usefully show that the physicians' evaluations of the likelihood of heart disease had no effect on the results, once that fact was revealed, the results could just as well have been reported in terms of the relative risks of the unadjusted referral rates.

That is not to say that the authors understood and thought through the implications of their approach. Indeed, the failure to appreciate the way various measures of difference between rates will tend to yield predictable results in particular settings is a serious problem with the great majority of health disparities research.[10-15] But it still warrants note that the disparity between the impressions conveyed by the seemingly high odds ratio difference and the seemingly small relative risk difference was the result of the authors' decision to use a logistic regression when it was probably unnecessary. And it also warrants note that the both the differences measured by odds ratios and the differences measured by relative risks were of a general size that should be expected as a consequence of the decision to craft symptom scenarios that would yield such high rates of referral.

B. The Calculation of the Gender Effect

By conducting the same number of analyses of black and white subjects the study in effect substantially over-sampled blacks. This is a perfectly legitimate way of enhancing the power of the study to detect differences. But the broad purpose of this study, and its

implications as they would be reported in the media, did not involve the disparities in the treatment of the 8 subjects, but disparities in the larger populations that the 8 subjects were intended to represent. In circumstances such as here where the differences were driven solely by the results of a subgroup, that fact has important implications for the calculation of the differences in the larger population.

That is, assuming that it makes sense to present differences between men and women generally when only black women differ from any other group, one needs still to take into account that in the United States at large, there are approximately 6.6 white women for every black woman. Thus, an appropriately weighted female referral rate would have been 89.1 percent, hardly distinguishable from the aggregated male figure. The same reasoning applies to the male rate. But since the black male rate was the same as the white male rate, a weighted average referral rate for men would be the same as the unweighted average. Thus, properly calculated, the female-male relative risk would have been .98 and the odds ratio would have been .80.

This reasoning also applies to the overall racial differences, which should be adjusted for differences in the gender composition of the two racial groups. Since women comprise a slightly higher proportion of American blacks than of American whites, an appropriate adjustment would slightly increase the black-white difference. But the increase is not substantial enough to warrant addressing the matter in any detail.

C. The Reporting of Overall Racial and Gender Differences and Obscuring of the Results Concerning White Men, Black Men, and White Women, and the Apparent Inattention to the Dominating Role of the Physicians' Perceptions of a Single Actress.

But does it make sense to report an overall racial difference or overall gender black difference when only black women had different rates from any other group? A related question is why physicians who treated white women the same as white men and black men the same as white men would treat black women differently from white men, as well as black men and white women?

The article did not address such questions and in fact obscured the results in ways that would avoid the questions. As noted, the fact that white men, black men, and white women all had the same referral rate could be derived from tables in the article. But nowhere in the text of the article was such fact noted. The statements in the conclusion of the abstract that the "findings suggest that the race and sex of a patient independently influence how physicians manage chest pain" seems plainly incorrect given that neither race nor gender had an effect alone. Even if the language could be somehow construed as technically correct because of the findings of a racial effect and a gender effect, the language would certainly lead one away from an understanding that only black women had a different referral rate from the other racial/gender groups. While the statement that "black women were the only patients who were significantly less likely to be referred for cardiac catheterization than white men" is consistent with the findings of equal referral rates for white men, black men, and white women, it nevertheless leads the reader away

from appreciating those findings. For the reader would typically regard the phrase as meaning merely that any differences among the other racial/gender groups did not rise to the level of statistical significance.

Indeed, assuming that the study provided useful evidence of anything (as the authors presumably believed, though see Section D infra), its most striking results are those indicating that, contrary to the findings of many studies referenced by the authors, white women and black men were referred at the same rate as white men. In terms of numbers of people potentially affected with respect to knowing whether they have reason to suspect that they might be discriminated against in these circumstances, the most numerous are white women and the second most numerous are black men. While one must be cautious about inferring from a study's failure to find a difference that the study offers evidence that there is no difference, reporting of findings of identical referral rates for white men, white women, and black men would have given white women and black men reason to believe that they did not have to fear that they would be less likely to receive appropriate care than white men (or at least that this study gave them no reason to have such fears). But the study, as written, and as reported, gave them reason to believe just the opposite. In such circumstances, the authors' later claim that they believed their study would promote honest dialogue seems baseless. The claim that it would encourage physicians to seek ways to eliminate unconscious bias is likely correct, though it would cause most physicians to look for biases that the study suggested did not exist. In any case, the authors could not have failed to appreciate that the way they reported their findings would mislead some readers, if not the overwhelming majority of readers. Thus, even if there were more merit to the authors' claim that they reported the overall race and gender results because of the way the study hypotheses were framed, there can be no excuse for failing, both in the study itself and in the post-publication interaction with the press, to make it absolutely clear that only black women had lower referral rates.

The article did imply that white women were not referred at lower rates than white men when, in discussing potential explanations for observed differences, it noted the possibility that physicians were influenced by data suggesting that women had worse outcomes than men after bypass surgery and angioplasty, and added that it was unclear why the physicians' familiarity with such data would affect recommendations for black women and not white women. That is as close to as the article came to posing or attempting to answer the question of why physicians who apparently did not consider race when making referrals for white and black men and did not consider gender when making referrals for white men and women, would consider race and gender when it came to making referral for black women. Were one to ask that question, or to otherwise ponder any curious results in the study, one of the first things to consider is the peculiar nature of the study.

This study was akin to the tester studies used to identify discrimination in housing and employment. Such studies seem to be sound enough in housing, particularly with respect to realtors, where the characteristics of the testers ought to have little role in whether they are shown a property. In some of the employment tester contexts involving interviews, however, the situation is rather different, since the way the testers come across in the

interview is of considerable importance. This can be a particularly serious issue when there are only a few testers, and in at least one study too little regard was given to the way overall results were influenced by the treatment of a single tester that was inconsistent with the patterns found among the other testers of the same racial group.[16]

In the Schulman study the anomalous results regarding the black female subjects were driven in substantial part by the even more anomalous results for the older black actress. In any tester situation where overall results are driven by seemingly anomalous results concerning one or two individuals, the first question one should ask about the seeming anomaly is whether there is some way that the subject tester from a particular group failed to appear comparable to testers from other groups. And here, of course, the authors were well-acquainted with the results for the older black woman, such results having been in a table that was originally intended for publication. Thus, one would expect the thoughtful researcher to carefully examine those results to try and determine why they so diverged from those of the other testers. But there is no indication that the authors gave the matter any thought. While before conducting the test the authors had cardiologists reach some level of concurrence on the scripts as reflecting definite angina, possible angina, or nonanginal pain, the authors apparently sought no similar appraisal on how the actors and actresses came across in describing their chest pain. Allowing that the failure to consider that issue at the outset might be excusable, it is hard to say the same of ignoring the issue once the results were reviewed.

A further point in this regard is that the authors noted that age was a significant predictor of referral, by which they meant, as shown in Table 4, that older workers were less likely to be referred, a result, like all other results shown in Table 4, driven largely by the referral rate of the single 70 year-old black actress Table 2, however, showed that the physicians (understandably) estimated a higher probability of coronary artery disease among the older subjects, a factor that presumably would tend toward higher rates of recommendations among older testers. Thus, that the referral recommendation rates were nevertheless lower for older subjects suggests that a perception of the general physical condition of the subjects, in terms of ability to deal with the stresses and potential problems associated with catheterization, may have had some role in the referral decision. That is a factor that could be present regardless of whether the setting was such that the way the tester came across could have any bearing on the judgment as to the nature of the chest pain.

Let us suppose, however, that the authors did scrutinize this matter and concluded that, inasmuch as the manner of presentation by the actor or actress could have had no role in the outcomes, the seeming anomaly could be disregarded. There still would seem no excuse – particularly in a setting where the authors presented so much information on the varied factors that proved to make no difference – for failing to alert the reader to the issue and provide sufficient information to enable the reader to appraise the reasonableness of such conclusion.

D. Whether a Study of this Nature Would be Likely to Reveal Such Physician Bias as Might Exist

The authors undertook to determine whether lower referral rates of blacks than whites and of women than men found in other studies were a result of physician bias. But they did not discuss whether such bias as may affect decisions in clinical settings would be expected also to be observed in the simulation they created. The authors noted that the physicians were not told that the purpose of the study was to determine the effect of patient race or gender on physician decision-making. But even putting aside that some physicians might nevertheless suspect such purpose (more so, probably, for race than gender), there is reason to question whether biased physicians would allow their biases to be reflected in their responses in this setting. Certainly, if the issue were employment, one would not expect biased personnel officers to exhibit the same level of bias when appraising hypothetical candidates in a monitored setting that they would in making actual decisions about whom to hire.

The employment situation is somewhat different from the healthcare setting, since the reasons personnel officers might discriminate are understood, while the reasons physicians would invidiously discriminate in healthcare remain something of a mystery. Also, the implications of the artificial setting might well vary depending on whether the bias is conscious or unconscious. But even assuming that physicians in fact are biased in a variety of ways that lead to lower referral rates for blacks and women, there is still a question as to whether the bias manifested with respect to the treatment of their patients would be reflected in this artificial setting, particularly an artificial setting that appears to have left little room for physician judgment. Thus, such refutation of the existence of generalized physician bias as one might otherwise find in the identical referral rates for white men, black men, and white women can hardly be found here. This point, however, is not a justification for researchers who believed their study was capable of identifying physician bias to fail to give attention to such salient findings.

Another conceptual shortcoming of the study involves its effort to determine whether factors such as the physicians' perception of the personal characteristics of the subjects might explain any observed differences. Apparently, the purpose of the analyses of physician perceptions of such things as whether the patient was likely to miss follow-up appointments or comply with the therapy was to determine whether differences in these perceptions with regard to the different subjects would provide nondiscriminatory explanations for the different referral rates. And in a setting where physician interaction with patients led legitimately to different conclusions about such factors among patients of different racial or gender groups, such conclusions might indeed provide nondiscriminatory explanations for observed disparities in healthcare decisions. But when there is no such interaction, and where the subjects say exactly the same things in recorded interviews, any physician perception of differences that correlated with race or gender would seem in fact to be based on race or gender. Thus, contrary to the authors' intended use of differential perceptions of these characteristics, such differential perceptions as might have been found ought not to have been regarded as nondiscriminatory reasons for differences in referral rates.

References:

1. Schulman: Schulman KA, Berlin JA, Harless, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. N Engl J Med 1999;340:618-26.

2. Schwartz: Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. N Engl J Med 1999;341:279-283.

3. Schulman KA, Berlin JA, Escarce JJ. Race, sex and physicians' referrals for cardiac catheterization. N Engl J Med 1999;341:286.

4. Curfman GD, Kassirer JP. Race, sex and physicians' referrals for cardiac catheterization. N Engl J Med 1999;341:287. 5. Bloche MG. Race and discretion in American medicine," Yale J Health, Pol Law & Ethics 2001;1:95-121. 6. 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

7. 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

8. Keppel KG, Pearcy JN, Weissman JS. Untitled letter. N Engl J Med 2005;353:2082-2083: http://content.nejm.org/cgi/content/full/353/19/2081.

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

10. 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

11. Scanlan JP. Measuring health disparities. J Public Health Manag Pract 2006;12(3):294 [Lttr]: http://www.nursingcenter.com/library/JournalArticle.asp?Article_ID=641470

12. Scanlan JP. The misinterpretation of health inequalities in the United Kingdom: Paper presented at: British Society for Population Studies Annual Conference 2006, Southampton, England, Sept. 18-20, 2006: http://www.jpscanlan.com/images/BSPS_2006_Complete_Paper.pdf

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

14. Scanlan JP. Divining difference. Chance 1994;7:38-39,48:
http://jpscanlan.com/images/Divining_Difference.pdf

15. Scanlan JP. Effects of choice of measure on determination of whether healthcare disparities are increasing or decreasing. Journal Review May 1, 2007:
http://jpscanlan.com/images/Vaccarino_NEJM_2005.pdf

16. Scanlan JP. Measuring hiring discrimination. Labor Law Journal 1993;44:387-394:
http://www.jpscanlan.com/images/Measuring%20Hiring%20Discrimination.pdf