Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities?

These rapidly proliferating programs do not appear to be devoting much attention to the possible impact on disparities in health care.

by Lawrence P. Casalino and Arthur Elster

ABSTRACT:

Pay-for-performance (P4P) and public quality-reporting programs can increase the quality of health care for the services being measured. However, unless carefully designed, these programs may have the unintended consequence of increasing racial and ethnic disparities. This paper describes ways in which P4P and public reporting programs may increase disparities and suggests ways in which programs might be designed that will make them likely to reduce, or at least not increase, disparities. [Health Affairs 26, no. 3 (2007): w405-w414 (published online 10 April 2007; 10.1377/hlthaff.26.3.w405)]

The quality of health care in the United States is variable and not what it could be, given the country's technical resources. During the past decade, public and private purchasers of health care have begun implementing "pay-for-performance" (P4P) and public quality-reporting programs to induce physicians and hospitals to invest in improving quality. During the same period, it has become evident that the quality of care for racial and ethnic minority patients is worse than that for whites, and increasing attempts have been made to develop interventions to help physicians and hospitals reduce this disparity.

These two movements--P4P and public reporting, on the other hand, and attempts to reduce disparities, on the other--have rarely intersected. This is unfortunate, because P4P and public reporting can have serious unintended consequences, and one of these consequences may be to increase health care disparities. During the next few years, as the number and size of programs expands--including, perhaps, the introduction of P4P for physicians by the Centers for Medicare and Medicaid Services (CMS)--there will be an important opportunity to design programs to minimize unintended consequences and to reduce disparities.

The purpose of this paper is to explore two questions: First, what consequences are P4P and public reporting programs, as currently implemented, likely to have on disparities? Second, how can these programs be designed to reduce disparities instead of increasing them? Three caveats regarding the scope of this paper: First, the paper focuses on physicians, but analogous points can be made for hospitals and for health plans. Second, generally speaking, the unintended consequences of P4P and public reporting programs are likely to be similar. They should both be understood as external incentives—that is, as incentives in addition to physicians' professional desire to provide better care—to induce physicians to invest in improving quality. We use the term "external incentives" to refer to both P4P and public reporting programs, and we refer specifically to P4P or public reporting when the point we make is specific to that type of program only. Third, we use the term "minority patients" to refer to patients who are members of ethnic/racial minority groups. "Minority patients" vary tremendously in income, education, acculturation to the United States, and English language ability; the unintended consequences discussed are likely to be
more severe for the more disadvantaged range of this spectrum. Although low-income, poorly educated white patients are not usually included in the term "minority patients," it is important to note that similar unintended consequences of external incentive programs may increase health care disparities between affluent and poor whites as well.

How Might P4P And Public Quality Reporting Increase Disparities In Health Care Delivery?

Empirical data on the effects of P4P and public quality reporting are still relatively scarce; the evidence to date may be summarized as suggesting that external incentives lead to positive but generally modest improvements in the areas of quality that are measured, at least in the short run. There are very few data available on the effects of external incentives on disparities in quality; they may be summarized by stating that such effects do in fact occur. These data (which are noted where relevant in the remainder of this section), surveys of physicians, experience in other industries, and theoretical considerations based on agency theory suggest at least five ways in which external incentives for quality may have the unintended consequence of increasing health care disparities.

Reduction in income for physicians in poor minority communities. Both P4P and public reporting may adversely affect the income of physicians practicing in minority communities--particularly poor minority communities--thereby potentially reducing both the number of physicians who work in such communities and their ability to invest in processes to improve quality. Physicians in these areas are doubly disadvantaged in trying to achieve high quality scores: First, their "payer mix" is likely to include a high proportion of uninsured and Medicaid patients, so there will be less revenue for them to invest in information systems, staff, and the development of organized processes to improve quality. Second, patients in these areas might be less likely to adhere to treatment recommendations. They might, for example, be less likely to obtain preventive care such as mammograms and Pap smears and less likely to return for follow-up of abnormal results, because of problems with transportation and child care and because of difficulty comprehending the recommendation. If compared directly to physicians in wealthier areas, physicians in poor minority communities might be less likely to receive P4P incentive pay and more likely to be listed in public report cards as poor-quality physicians. If, as is sometimes proposed, health plans require patients to make higher copayments for seeing these "poor-quality" physicians, poor patients who live in these areas and lack the means to easily visit physicians in other areas could be harmed. Only one research study provides data directly relevant to this "poor may get poorer" problem: In the first evaluation of a P4P program recently introduced for primary care physicians by the British National Health Service (NHS), practices that served lower-income populations had lower quality scores.

"Color-blind" quality improvement programs. P4P and public reporting could increase disparities if they induce medical groups to create organized processes to improve quality--such as programs to help patients with chronic illnesses learn "self-management" of their illness--and if minority patients are less likely to benefit from these programs. This would be the case, for example, if education--written or oral--is given in a language or cultural style or at a literacy level that minority patients understand poorly or feel is culturally insensitive.

"Teaching to the test" might disproportionately affect minorities. P4P and public reporting could induce physicians to focus their time and attention (consciously or unconsciously) on types of care that are being measured, to the detriment of nonmeasured areas that could be equally or more important. This "teaching to the test," which has been found in other industries, could disproportionately affect minorities. For example, with a relatively uneducated diabetic patient who speaks poor English, the physician might focus on making sure the patient has a hemoglobin A1c test (because this is measured) but not on the time-consuming task of explaining to the patient how to control his or her diabetes and blood pressure (because control of these things is
not measured). The same physician might not "teach to the test" to the same degree with a better educated, English-speaking patient, because he or she is more comfortable with that patient or believes that the patient is more capable of adhering to a therapeutic regimen, or because the patient is more assertive in demanding time and explanation from the physician.

Avoiding patients perceived as likely to lower quality scores. P4P and public reporting might induce individual physicians and medical groups to avoid patients whom they perceive as being likely to lower their quality scores, particularly if quality measures are not adequately adjusted for the patients' overall health status and perhaps for racial or socioeconomic characteristics as well. There is ample evidence that physicians perceive minority patients as less likely to comply with their recommendations for treatment and preventive services and more likely to have bad outcomes. There is also evidence that physicians subject to external incentives will try to avoid minority patients because they perceive them as more likely to have poor outcomes from treatments, even when this is not the case. After New York State initiated its report card measuring death rates from coronary artery bypass graft (CABG) surgery for individual surgeons and hospitals, the gap between CABG rates for whites and blacks increased. The report card appears to have made surgeons more reluctant to operate on black patients.

It is widely recognized that outcome measures of quality should be adjusted for patients' health status, and some attempt to do so is often made. However, racial or socioeconomic factors, or both, may also be associated with worse outcomes. For example, hospitalized children whose parents did not speak English well had a twofold increased risk for a serious medical event. The use of process measures, rather than outcome measures, is often proposed as a solution to the problem that patients' characteristics, in addition to physicians' actions, affect quality scores. It is often stated, without evidence being cited, that patients' characteristics do not affect physicians' scores on process measures and therefore that process measures, unlike outcome measures, need not be adjusted for patients' health status or socioeconomic status. To practicing physicians, this might seem implausible. For example, it should be easier for physicians who practice in affluent areas such as Marin County, California, than for those who practice in poor areas of Oakland to achieve higher rates of screening mammography: Their patients are more likely to be wealthy, well-educated, well-insured women who are aware of the benefits of mammograms, are more likely to request them, and are more able to travel to obtain them.

Many studies have indicated that scores on process measures are likely to be affected by the characteristics of patients, including their health status, primary language, and socioeconomic status (SES), although one important recent study showed little or no effect. Patients with lower SES have been found to be less likely to obtain Pap smears, mammograms, and diabetic eye exams, holding physician practice constant. Health plan scores on Health Plan Employer Data and Information Set (HEDIS) measures have been shown to be affected by patients' race and SES. Fully insured women with diabetes have been found to be one-third less likely than women without diabetes to have a screening mammogram. Patients with chronic illnesses, which are more common among minorities, may find it more difficult to travel to obtain these tests and may be less interested in having them than is the case for healthy patients, and their physicians may be so focused on one of the patients' illnesses that preventive care and the care of other illnesses may slip through the cracks.

Patients' ability to use public quality reports. Public reporting might benefit affluent or educated patients more than it benefits poor or uneducated patients. The affluent and educated might be more likely than their poor or uneducated peers to come into contact with physician report cards and more likely to understand them. They might also be more able to act on their understanding, either because the physicians or hospitals that are listed as high quality are close to their home or workplace, or because they have the income and work schedule flexibility to travel to more distant physicians. The results of a recent study indicated that black patients are 60 percent less likely than whites to travel outside their local area to be admitted to a hospital.
public reporting disproportionately benefits white and affluent patients, it could increase disparities even as it led to an overall improvement in quality. On the other hand, if public reporting provides information for minorities that they would not otherwise receive but that the more affluent and educated already receive informally through other channels, public reporting could reduce disparities by leveling the knowledge playing field.

**Designing P4P And Public Quality-Reporting Programs With Disparities In Mind**

Creating programs likely to reduce, or at least not to increase, disparities will not be easy, technically or politically. The discussion above suggests six design elements that, if implemented, would be likely to help.

**Reward both absolute quality scores and improvement over time.** To the extent that physicians practicing in poor minority areas start with lower quality scores, rewarding improvement could at least mitigate the potential "rich get richer, poor get poorer" effect of P4P. However, the same cannot be said of public reporting, where an improving physician group will still be ranked lower than a physician group with a high absolute quality score.

**Use risk adjustment or stratified analyses.** To the extent that minority patients--particularly poor minority patients--have worse health status, risk adjustment could be used both to reduce the "poor get poorer" problem and to minimize physicians' incentive to avoid minority patients. Risk adjustment should be used for process as well as outcome measures, except where data show that adjustment is not necessary. Risk adjustment for health status alone would not be enough, however. Adjustment for patients' race/ethnicity or SES, or both, would be necessary to level the playing field.

Stratified analyses could be used instead of or in addition to risk adjustment. For example, physicians practicing in poor minority areas could be compared--for purposes of P4P or public reporting, or both--with other physicians practicing in similar areas. Alternatively, organizations could be compared based on their care of minority patients in general, of African American patients, of low-SES patients, and so on. Public reporting programs would provide measures of a physician's, medical group's, hospital's, or health plan's care for these populations. P4P programs would provide rewards both for overall quality scores and for scores for minority populations, thus directly incentivizing the reduction of disparities.

If they can be done well, stratified analyses have three other advantages as well. First, they provide information that might be useful for directing quality improvement activities--for example, how physicians in a group are performing for specific categories of patients. Second, while risk adjustment at best reduces physicians' incentive to avoid certain patients, stratification gives a more tangible and positive incentive: "If you provide high-quality care for minority patients, we will pay you more." Third, compared to risk adjustment, stratified analyses are more transparent and may be more believable to physicians.

There are sizable technical barriers to using stratification or risk adjustment for minority and low-SES patients. The measures of race/ethnicity and SES would have to be defined, the data would have to be collected, and risk-adjustment methods would have to be refined. There is an additional, very important barrier to using stratification. Unlike risk adjustment, which can be done for each individual patient, reliable and valid stratified analyses of quality require that a physician or medical group have a large number of, for example, African American patients with diabetes. Most do not, so stratified analyses would be reserved for large medical groups, hospitals, and health plans--and even some of these organizations might not have enough minority patients to permit stratified analyses.
In addition, risk adjustment and stratification are two-edged swords: They would reduce physicians’ incentive to avoid patients perceived as likely to lower their quality scores but would also reduce their incentive to improve care for these patients. By rewarding physicians even when their quality scores for minority patients are lower, P4P programs would avoid penalizing physicians for caring for these patients but would also risk rewarding them for continuing to provide mediocre care. This problem probably cannot be entirely eliminated but would be reduced by basing rewards partly on absolute quality scores and partly on improvement over time. When sufficient numbers of patients are involved to permit reliable measurements, P4P programs might reward both absolute quality scores and improvement for an organization's patient population as a whole and--based on stratified analyses--for the organization's minority patients. Public reporting programs could report both overall quality scores and stratified scores for minority patients.

An alternative approach, used by the British NHS P4P program, is to permit physicians to exclude patients from quality measurements by designating them as "exceptions," based on rather broad criteria. Overall, physicians designated only 6 percent of patients as exceptions, but the range of patients so designated varied from zero to 86 percent. One percent of practices excluded more than 15 percent of their patients.

**Reward both overall quality and reduction in disparities.** As just noted, external incentives could be given both for quality of care for all of an organization's patients and for one or more subgroups of patients--specifically, minority patients. Alternatively, incentives could be given both for quality for all of an organization's patients and for reducing quality disparities between white and minority patients. Either of these alternatives would be technically challenging and politically controversial. Massachusetts, however, has recently announced a plan to include reduction of racial and ethnic disparities as a goal for its Medicaid P4P program for hospitals.

**Use a variety of methods to minimize the "teaching to the test" problem.** Rotating the measures used or adding new measures at relatively frequent intervals would reduce the "teaching to the test" problem. Including data from patient satisfaction surveys could counterbalance incentives to focus on narrow quality measures at the expense of communicating with patients and coordinating patient care. More ambitiously, in a model that comes from agency theory, some or all high-scoring physicians (or medical groups) would be evaluated on a second set of quality measures that would not be announced in advance. If they score well on these measures, they would receive substantial additional reward, but if they score poorly, they would lose some of their reward for scoring well on the initial set of measures. Payments would be designed so that physicians would, on balance, expect to gain from being scored on the second set of measures and would want to be scored on them. Since the second set of measures would involve a limited number of physicians or medical groups, they could (although need not) be directed at types of quality that are important but relatively costly to measure—for example, measures that require review of medical charts.

**Only use P4P and public reporting when statistically reliable and valid measurement can be done.** Because of the relatively small number of patients involved, this might be particularly difficult for quality measures for individual physicians and small physician practices. External incentive programs based on small numbers of patients will intensify physicians’ incentives to avoid patients who the physicians believe are likely to lower their scores.

**Include attention to the effects of the P4P or public reporting program on disparities.** This design feature might seem obvious, but to date it has not typically been included in evaluations, perhaps because of the difficulty in gathering the necessary data or in including large enough numbers of minority patients for reliable and valid comparisons.

**Concluding Comments**
Awareness that P4P and public reporting may increase disparities does exist among the leaders of medical societies, of organizations that devise quality measures, of organizations that are creating public reporting and P4P programs, and among public policymakers. Nevertheless, it appears that most P4P and public reporting programs are not well designed to avoid increasing disparities, much less to reduce them. The next few years will be critical for the future of P4P and public reporting programs in the United States. These programs are proliferating rapidly but do not appear to be devoting a great deal of attention to their possible effects on health care disparities. Programs designed to minimize the unintended consequences of increasing disparities will be more costly and time-consuming to implement, maintain, and modify over time than those that ignore the possibility of such consequences. But they should have more staying power and should be less likely to generate a backlash from physicians and patients. Only well-designed programs are likely to lead both to improved quality for all and to the reduction of disparities in health care.

This paper was funded by the American Medical Association (AMA) on behalf of the Commission to End Health Care Disparities. The viewpoints expressed in the paper are those of the authors and do not necessarily reflect the views and policies of either the AMA or the organizations of the commission. The authors thank Andrew Eisenberg, Christopher Ervin, Evelyn L. Lewis, John Montgomery, and Diana Ramos for helpful comments.

NOTES


7. Werner and Asch, "Unintended Consequences."


13. T. Doran et al., "Pay-for-Performance Programs in Family Practices in the United Kingdom," *New England Journal of Medicine* 355, no. 4 (2006): 375-384. The difference in quality scores between practices serving low-income populations was statistically significant, although not large. However, the measures used resulted in nearly all practices having high scores, so variation among practices was not large.


21. See, for example, MedPAC, *Report to the Congress*; and American College of Physicians, *The Use of Performance Measurements to Improve Physician Quality of Care* (Philadelphia:


30. Ibid.

31. See, for example, P.A. Braveman et al., "Socioeconomic Status in Health Research: One Size Does Not Fit All," Journal of the American Medical Association 294, no. 22 (2005): 2879-2888.


33. Doran et al., "Pay-for-Performance Programs."


40. A recent survey by Casalino and colleagues shows that general internists are generally in favor of financial incentives for quality but are very concerned about unintended consequences...
and overwhelmingly do not trust health plans or government programs to design appropriate measures. Casalino et al., "General Internists’ Views."

Larry Casalino (casalino@health.bsd.uchicago.edu) is an assistant professor, Health Studies, at the University of Chicago in Illinois. Arthur Elster is director of the Division of Medicine and Public Health, American Medical Association, also in Chicago.

DOI: 10.1377/hlthaff.26.3.w405
©2007 Project HOPE—The People-to-People Health Foundation, Inc.