Vol. XV, No. 4
May 2012

Changes in Health Care Financing & Organization (HCFO)

findings brief

key findings

- High- versus low-quality hospitals report lower rates of disparities, although minority patients are under-represented in high-quality hospitals and over-represented in low-quality hospitals.
- Simulations of different P4P programs resulted in modest reductions in disparities. However, the decrease was greater when hospitals were ranked on disparity scores as opposed to quality scores.

Pay-for-Performance Programs to Reduce Racial/Ethnic Disparities: Limitations of a One-Size-Fits-All Approach

Overview

Pay-for-performance (P4P) programs are broadly defined as performance-based payment arrangements that are designed to promote improvement in health care quality while reducing costs. Often absent in the equation is the issue of disparities. Racial and ethnic disparities in health care persist, and some believe that P4P programs have the potential to exacerbate such inequities in the quality of care received by minority patients. For example, P4P programs could incent hospitals to give preference to patients who are more likely to generate improved quality scores, thus reducing overall access to other patients. Moreover, when facing additional performance requirements, hospitals starting out with fewer resources may not perform as well as financially robust providers.

In an HCFO-funded study, Joel S. Weissman, Ph.D., of the Center for Surgery and Public Health at Brigham and Women's Hospital; Lisa Iezzoni, M.D., and Christine Vogeli, Ph.D., of Massachusetts General Hospital Institute for Health Policy; Romana Hasnain-Wynia and Raymond Kang of Northwestern University; Robin Weinick of RAND; and MaryBeth Landrum of Harvard Medical School examined the quality of hospital care, estimated the proportion of patients receiving recommended care, created new measures of patient care quality, and simulated the impact of several P4P scoring methods on hospital rankings. Findings from their analyses were published in the February 2012 issue of the *Journal of Health Care for the Poor and Underserved*.[1] The study team determined that standard P4P programs based on ranking providers by overall quality of care may not be the most effective way to target disparities. Rather, more tailored approaches could serve the dual purpose of improving quality and reducing disparities.

Robert Wood Johnson Foundation

Changes in Health Care Financing and Organization is a national program of the Robert Wood Johnson Foundation administered by AcademyHealth.

"Some policy experts have been talking about how P4P might increase or decrease disparities, but it's never really been tried on a large scale. We decided to use actual data from the [Centers for Medicare and Medicaid Services] CMS Inpatient Quality Reporting system to simulate what might happen to national-level disparities under different behavior change scenarios," says Weissman.

Methods and Analyses

The researchers surmised that the process of ranking hospitals by focusing exclusively on quality could fail to capture accurately hospitals' record on disparities. Accordingly, the researchers designed their study to respond to two questions: (1) Do hospitals with the highest quality also have the lowest disparities? and (2) How do U.S. hospitals with the highest quality or the lowest disparities treat minority patients compared with how they treat white patients? To examine these questions, the researchers used individual-level data for patients age 18 and older admitted to over 4,500 hospitals from 2005 through 2006 from the CMS Hospital Inpatient Quality Reporting System, the most comprehensive hospital data available. The data set includes information on acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PNE), and the individual-level data contain information on patient race and other demographic information. The researchers categorized race/ethnicity as black/African American, Hispanic, Asian, American Indian/Alaska Native, Native Hawaiian or other Pacific Islander, and white. The researchers constructed an "all-minority" group that included all race/ethnicity distinctions except white. In addition to the CMS data, the researchers turned to the 2006 American Hospital Association (AHA) annual survey to obtain hospital characteristics such as bed size and teaching status.

Drawing on the individual-level CMS performance data, the researchers were able to determine if patients received all applicable care processes for AMI, HF, and PNE. They then used the information to create quality scores at both the hospital and national levels. Next, they calculated disparity scores by subtracting the quality score for minority patients from that of white patients.

To model the results, the researchers estimated the effects of two types of P4P designs: (1) one that ranked hospitals by an overall quality-of-care score for all patients (quality rank) and (2) one that ranked hospitals by the disparity score (disparity rank). To simulate the effect of these P4P designs on quality and disparities, the researchers assumed the following:

- Sufficient P4P incentives could stimulate bottom performers.
- Scores in top performers would not improve.
- Results would be seen only in hospitals with a sufficient number of cases.
- Results would vary for incentives for overall quality or in response to disparities.

The analysis excluded hospitals if they did not have at least 30 minority cases and 30 white cases, thereby significantly reducing the sample; the exclusion is in line with how HQA measures are calculated. In general, hospitals in the South and teaching hospitals were most likely to meet the 30/30 threshold, whereas Midwest and public hospitals fell short.

As part of sensitivity testing, the researchers conducted additional simulations by using a combination of high-quality scores and low-disparity scores (combination rank) to examine the impact of the P4P program designs on quality and disparities.

Results

The number of hospitals included in the final analyses by conditions totaled 4,267 for AMI, 4,655 for HF, and 4,764 for PNE. The number of patients by conditions totaled 1,090,210 for AMI, 1,793,140 for HF, and 1,719,634 for PNE.

Based on quality measures, the highest-performing hospitals exhibited lower disparities than did lower-performing hospitals. High-quality hospitals had disparity scores between -1.4 and 1.9 percent for the three conditions (a negative disparity score means that minority patients received higher-quality care than did white patients), whereas low-quality hospitals had scores between 6.8 and 10.2 percent. However, the researchers also found that minorities were under-represented in high-quality hospitals and over-represented in low-quality hospitals.

In the simulations, the researchers found that disparity scores would potentially decline for all minority groups with the introduction of P4P programs. However, the reduction varied by condition and ranking method. For example, at baseline, the disparity scores for AMI were 3.9 percent for black/African Americans, 6.5 percent for Hispanics, and 1.9 percent for Asians. When the simulated P4P program using the quality ranking system was applied, those numbers decreased by approximately 2 percentage points. When a P4P program using the disparity ranking system was applied, national disparities decreased by 3 to 4 percentage points. The combined ranking produced reductions between these two results. Similarly, the introduction of all three models of P4P programs was associated with an increase in overall quality, though with similar variation.

Policy Discussion

The analyses generated several conclusions of interest to policymakers and practitioners involved in developing the most effective incentive programs for improving quality, lowering costs, and reducing disparities in hospital care. First, the researchers were pleasantly surprised to find that a large number of hospitals treat white and minority patients equitably. These hospitals could serve as models for other hospitals as P4P programs gain widespread implementation. Second, the findings were consistent with earlier work that suggested that minority patients could be disadvantaged in a P4P program. Despite an overall reduction in disparities, more minority patients were served at lower- versus higher-performing hospitals. Therefore, P4P could have the unintended consequences

of penalizing hospitals (and, by extension, their patients) that serve a disproportionate share of minority individuals. Third, the structure and design of P4P programs may affect overall quality and racial/ethnic minorities in different ways. Greater reductions in disparities may be achieved by providing incentives based directly on hospital-specific disparities.

The research team acknowledges that a national P4P program probably cannot be built solely on disparity measures in the absence of quality measures. The team therefore suggests that the best solution for P4P programs aimed at improving quality while reducing disparities is to rank hospitals based on a combination of quality and disparity measures. The researchers go on to suggest a potential two-step model for P4P programs that would first provide hospitals with financial incentives tied to quality measures and then provide additional incentives for a reduction in disparities. The researchers also suggest disincentives to hospitals that improve quality without reducing disparities.

"The time has come to move beyond mere documentation of the problem and instead focus on possible solutions. P4P can be a valuable instrument in a collection of strategies aimed at improving quality and reducing disparities, but only if designed in a thoughtful manner that recognizes the strengths and weaknesses of different providers," says Weissman.

Limitations

There are several limitations to the analyses. For example, the study measured quality by using a composite score based on a patient's receiving all recommended services for his or her condition. Other studies may measure quality differently and arrive at different conclusions. Similarly, the analyses used absolute rather than relative differences in quality scores when evaluating disparities. In addition, the analyses used measures that are typically high-performing areas for hospitals, as opposed to measures in areas that are traditionally challenging. The researchers looked at P4P programs that provided incentives for process of care, and assumed accurate reporting of race and ethnicity. Other measures in the Hospital Quality Alliance data set, such as hospital re-admissions, 30-day mortality, and patient experience surveys, are not available at the person level, making analysis by race/ethnicity impossible.

Finally, and perhaps most important, the analyses are based on a limited number of P4P scenarios that are all generally optimistic. Several other scenarios could alter the results by, for example, concentrating on hospitals that improve quality for only one race or focusing on easier-to-treat cases.

Conclusion

As policymakers and practitioners continue to develop incentive programs to improve quality and decrease costs in the U.S. health care system, it will be important to consider the potential for unintended consequences among certain vulnerable populations, especially minorities. With the appropriate design and focus, P4P programs offer the potential to produce a positive impact on care for all patients.

For More Information

Contact Joel Weissman at jweissman@partners.org.

About the Author

Christina Zimmerman is a research assistant for the HCFO program. She may be reached at 202-292-6736 or christina.zimmerman@academyhealth.org.

Endnotes

1. Weissman JS et al. Pay-for-performance programs to reduce racial/ethnic disparities: what might different designs achieve? J Healthcare Poor Underserved. February2012: 23 (1): pp. 144-160.