

Health Affairs

At the Intersection of Health, Health Care and Policy

Cite this article as:

Jan Blustein, Joel S. Weissman, Andrew M. Ryan, Tim Doran and Romana
Hasnain-Wynia
Analysis Raises Questions On Whether Pay-For-Performance In Medicaid Can
Efficiently Reduce Racial And Ethnic Disparities
Health Affairs, 30, no.6 (2011):1165-1175

doi: 10.1377/hlthaff.2010.1022

The online version of this article, along with updated information and services, is
available at:

<http://content.healthaffairs.org/content/30/6/1165.full.html>

For Reprints, Links & Permissions:

http://healthaffairs.org/1340_reprints.php

E-mail Alerts : <http://content.healthaffairs.org/subscriptions/etoc.dtl>

To Subscribe: <http://content.healthaffairs.org/subscriptions/online.shtml>

Health Affairs is published monthly by Project HOPE at 7500 Old Georgetown Road, Suite 600, Bethesda, MD 20814-6133. Copyright © 2011 by Project HOPE - The People-to-People Health Foundation. As provided by United States copyright law (Title 17, U.S. Code), no part of *Health Affairs* may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by information storage or retrieval systems, without prior written permission from the Publisher. All rights reserved.

Not for commercial use or unauthorized distribution

By Jan Blustein, Joel S. Weissman, Andrew M. Ryan, Tim Doran, and Romana Hasnain-Wynia

Analysis Raises Questions On Whether Pay-For-Performance In Medicaid Can Efficiently Reduce Racial And Ethnic Disparities

DOI: 10.1377/hlthaff.2010.1022
HEALTH AFFAIRS 30,
NO. 6 (2011): 1165–1175
©2011 Project HOPE—
The People-to-People Health
Foundation, Inc.

ABSTRACT In 2006 Massachusetts took the novel approach of using pay-for-performance—a payment mechanism typically used to improve the quality of care—to specifically target racial and ethnic disparities in hospital care for Medicaid patients. We describe the challenges of implementing such an ambitious effort in a short time frame, with limited resources. The early years of the program have yielded little evidence of racial or ethnic disparity in hospital care in Massachusetts, and raise questions about whether pay-for-performance as it is now practiced is a suitable tool for addressing disparities in hospital care.

In addition to expanding health insurance coverage, Massachusetts' 2006 health care reforms included a large increase to the state's Medicaid payment rate for hospitals. Along with this increase came greater accountability in the form of a pay-for-performance program, under which hospitals that performed well on specified metrics would receive financial bonuses.

As of 2006 more than half of the nation's Medicaid programs had at least one pay-for-performance component, with the majority of those programs focused on managed care.¹ Massachusetts was one of four states choosing to target hospitals specifically. However, the Massachusetts legislature took the unusual step of allotting a portion of the program's incentive payments to the reduction of racial and ethnic disparities in health care. Addressing the task of reducing disparities by applying a tool generally used to improve quality was a novel approach. The legislation also set out an ambitious timeline, allocating just over a year for the program's development.

We begin this article by reviewing some current thinking about pay-for-performance and efforts to reduce disparities. We then discuss the Massachusetts experience as a case study, describing the environment in which the policy

was developed and identifying key decision points and choices that the state made. We also present data on hospital performance. Along the way, we point to the many challenges faced by administrators of the pay-for-performance program. We close with a discussion of what the program can teach other states, other payers, and policy makers with an interest in targeting "inequality in quality."²

The Massachusetts Executive Office of Health and Human Services was responsible for carrying out the rate increase and the accompanying pay-for-performance program. Within that agency, the Office of Medicaid had day-to-day responsibility for MassHealth, the Medicaid program in Massachusetts. In preparing this article, we reviewed public documents and met with staff at the Office of Medicaid. We also analyzed hospital performance data that the office provided to us and spoke with eleven hospital administrators and three staff members at the Massachusetts Hospital Association who agreed to share their experiences.

Using Pay-For-Performance To Reduce Racial And Ethnic Disparities

The Institute of Medicine has identified equity—the provision of care “that does not vary in qual-

Jan Blustein (jan.blustein@wagner.nyu.edu) is a professor of health policy and medicine at the Wagner School of Public Service and the School of Medicine, both at New York University, in New York City.

Joel S. Weissman is an associate professor of health policy at the Mongan Institute for Health Policy at Massachusetts General Hospital and Harvard Medical School, in Boston, Massachusetts.

Andrew M. Ryan is an assistant professor of public health at the Weill Cornell Medical College, in New York City.

Tim Doran is a clinical research fellow in public health at the School of Health Sciences, University of Manchester, in England.

Romana Hasnain-Wynia is the director of the Center for Healthcare Equity and an associate research professor at the Feinberg School of Medicine, Northwestern University, in Chicago, Illinois.

ity because of personal characteristics such as gender, ethnicity, geographic location, or socioeconomic status”—as one of six key aspects of health care in need of improvement.^{3(p6)} A growing number of initiatives by health care organizations have focused on reducing racial and ethnic disparities.^{4,5} However, no state-level pay-for-performance programs specifically targeted disparities before the effort in Massachusetts. Other pay-for-performance programs may have improved the care of minorities—as the saying has it, a rising tide lifts all boats—but gaps in access remain.^{6–8} Thus, there has been great interest in pay-for-performance programs that directly address disparities.^{3,9}

ESSENTIAL INGREDIENTS Implementing a pay-for-performance program is complicated.^{9–11} Adding the reduction of racial and ethnic disparities as a goal introduces additional tasks, such as collecting data on race and ethnicity at the patient level.^{12,13} Fortunately, Massachusetts had previously mandated that for all inpatient stays, hospitals had to report patients’ race and ethnicity, using standard categories.¹⁴

Choosing an aspect of performance to measure is also crucial. Measures could reflect hospital structure, clinical processes, patient outcomes, or patient experience. The National Quality Forum has identified several criteria for choosing performance measures when disparities are a concern: The measures should involve medical conditions that are highly prevalent in disadvantaged populations; there should be evidence that the targeted providers give different treatments to patients in minority and majority groups; and there should be research linking better performance to improved outcomes.¹⁵

A pay-for-performance program to reduce racial and ethnic disparities must translate differences in care among groups into a metric or statistic that indicates the extent of disparities. But the choice of statistics determines the extent and even the direction of the disparities that are found.^{16,17}

Consider a hospital that provided higher-quality care to majority patients than it did to patients from a disadvantaged racial or ethnic minority. Most people would consider this a disparity in care because the disadvantaged group received lower-quality care. But if a hospital provided higher-quality care to a disadvantaged minority than to whites, not everyone would agree that this should be considered a disparity. Program leaders must take a position on this issue when they choose a disparities statistic, as we show below. Carefully defining the disparities that are being targeted is essential in choosing the best statistic.^{18,19}

Finally, pay-for-performance programs need

to reliably identify providers that perform well or poorly in the area of interest.¹⁸ For a program that is designed to identify providers with low levels of disparity, this means that the disparities statistic should consistently identify the same providers as either high or low performers, if their performance is measured repeatedly. But small samples—often all that is available when measuring racial and ethnic disparities—yield low reliability. For instance, many hospitals have few minority inpatients.^{19–21} The smaller the numbers, the more likely that apparent disparities will reflect chance rather than true differences.

THE SETTING Massachusetts has a long history of progressive health policy innovation.^{22,23} However, the state is not particularly diverse, compared to the entire United States: Only 6.3 percent of Massachusetts’ residents are African American, as opposed to 12.9 percent of US residents,²⁴ with comparable figures for Latinos of 4.8 percent and 9.0 percent, respectively.²⁵

In addition, minorities are geographically clustered within Massachusetts. Two-fifths of the state’s African Americans live in one of its fourteen counties, Suffolk, which is home to just a tenth of the state’s population. Twenty-five percent of all Latinos in the state also live in Suffolk County.²⁶

Developing The Program

In designing the racial and ethnic disparities pay-for-performance program, the Massachusetts Medicaid office consulted with members of the hospital community, major payers in the state, and committees assembled under governmental auspices, including the Massachusetts Health Care Quality and Cost Council and the MassHealth Payment Policy Advisory Board.²⁷ The Massachusetts Medicaid Disparities Policy Roundtable—an independent body comprising representatives of hospitals, payer groups, and the local academic community—also made recommendations to the office.²⁸ This group suggested that the ethnic and racial disparities program focus on two kinds of quality measures: clinical performance measures reflecting hospitals’ success at minimizing disparities in the processes of delivering care; and structural measures reflecting hospitals’ efforts to reduce disparities, such as activities consistent with the culturally and linguistically appropriate services practice standards developed by the Department of Health and Human Services.²⁹

CLINICAL MEASURES The Massachusetts Medicaid office reviewed nationally endorsed clinical performance measures but found few that were suitable for the MassHealth popula-

tion, which is disproportionately composed of children and pregnant women. So the office worked with MassPro, the state's quality improvement organization, and other experts to develop suitable measures of maternity care. It adopted measures of newborn care developed by the Leapfrog Group.³⁰

The Medicaid office eventually decided to use five groups of clinical performance measures, relating to surgical infection protection and pneumonia (both from Medicare's Hospital Compare database); pediatric asthma (from the National Hospital Inpatient Quality Measures); and the maternity and newborn measures.³¹ All of the measures are based on documentation in patient charts.

STRUCTURAL MEASURES The Medicaid office found no nationally recognized structural measures to match the national culturally and linguistically appropriate services standards. However, Massachusetts had for some time required hospitals to report on a range of activities related to caring for minority patients, using a checklist known as the Cultural Competence Organizational Self-Assessment.

The checklist grew out of research by one of the Medicaid office's staff members.³² It contains twenty-eight items divided into four domains: governance, administration and management, service delivery, and customer relations. Exhibit 1 shows three representative items from each domain. (The full checklist is available on the Massachusetts Executive Office of Health and Human Services website.)³³ For each of the items, hospitals were required to indicate their current level of activity, ranging from "no plans" or in the "planning stage" to "in place for

>5 years."

Prior to the legislation introducing pay-for-performance in Massachusetts, hospitals had simply submitted completed checklists every year to the state Medicaid office. But if a hospital wanted to participate in the new program, it also had to submit a five-page summary of the activities it had undertaken to reduce racial and ethnic disparities, along with detailed supporting documentation.³⁴

This meant that the Medicaid office was responsible for determining whether checklist responses were "valid"—that is, adequately supported by the documentation. For example, the office had to determine whether a hospital's mission statement "articulates development of cultural diversity" and whether patient data had been sufficiently "analyzed by race, ethnicity, and languages spoken."³³ This was a difficult task. As described below, it raised a set of issues, ranging from uncertainty about what constituted sufficient documentation to broader questions such as whether better performance on the measures would decrease disparities.

Linking Performance To Payment

For the first year of the pay-for-performance program, Massachusetts' Rate Year 2008 (October 1, 2007, through September 30, 2008), \$4.5 million was available for incentives related to the structural measures; no funds were allocated for incentives related to the clinical measures. The \$4.5 million represented approximately 0.4 percent of the state's Medicaid hospital spending for the year.²⁷ At the hospital level, the potential incentive payments' portion of total revenues—

EXHIBIT 1

Examples Of Items Included In Massachusetts' Cultural Competence Organizational Self-Assessment

Domain	Sample item
Governance policy	Board of directors has adopted a mission statement that articulates development of cultural diversity Organizational strategic plans incorporate cultural competence goals and strategies Board of directors and senior management reflect the racial and ethnic mix of the actual population mix being served
Administration and management	Policies specify strategies to actively recruit racially/ethnically diverse medical/nursing/senior management staff Hospital provides diversity training/orientation programs for all clinical and nonclinical staff Hospital patient data are analyzed by race, ethnicity, and languages spoken
Service delivery	Policies exist to include racial/ethnic communities in the planning and design of health care services Patient education materials are translated into languages reflecting non-English-speaking groups served Hospital interpreters are members of a professional medical interpreter association
Customer relations	Patient satisfaction surveys are translated for non-English-speaking patients Interagency collaborative projects exist in racial/ethnic neighborhood communities in the service area Hospital makes provisions for accessing culturally and linguistically appropriate procedures to resolve service grievances

SOURCE Note 33 in text. **NOTES** As explained in the text, the Cultural Competence Organizational Self-Assessment is the instrument that was used in assigning scores on the structural measure. The items that appear above were used in Rate Year 2008, the period covered by our data. Items have been rephrased slightly from the original. The current version is online (see Note 41 in text).

in other words, money received from all payers—averaged 0.02 percent.³⁵

For Rate Year 2010 (October 1, 2009, through September 30, 2010), incentive payments were scheduled to rise to \$20 million for performance on the structural measures and \$12 million for reducing disparities according to the clinical measures.³⁶ If divided equally among the sixty-six acute care hospitals in the state, this 2010 allocation would have exceeded \$300,000 per hospital for the structural measures and \$180,000 per hospital for the clinical measures. These are large sums relative to those used in other pay-for-performance programs. For example, the Hospital Quality Incentive Demonstration Project—a joint project of the Centers for Medicare and Medicaid Services and Premier Healthcare Alliance, which provided bonus payments to hospitals that achieved high levels of quality in several clinical areas—paid incentives of approximately \$33,000 per hospital per year, from 2003 through 2006.³⁷

DEVELOPING AN INCENTIVE PAYMENT FORMULA

Hospital performance on the various ethnic and racial disparity measures was linked to incentive payments in a series of steps. First, performance data were validated by the state Medicaid office. For the clinical measures, this meant comparing reports with a sample of medical records from each hospital. For the structural measures, it meant reviewing the documentation, as described above. Hospitals whose data the Medicaid office validated were eligible to receive incentive payments in proportion to the number of eligible MassHealth patients they had discharged.

Although hospitals reported data on five clinical conditions, the small numbers of eligible Medicaid patients in each group in most hospitals precluded calculating meaningful scores stratified by race and ethnicity. The state Medicaid office therefore generated a single score for each racial and ethnic group in each hospital, combining data across all five clinical conditions. These scores were computed using the “opportunities” approach. Each score is a ratio: the sum of all instances in which the targeted care was provided, divided by the total number of opportunities to provide care to eligible patients.³⁸ These scores can be calculated and compared within hospitals or across all of the hospitals in the state. In the remainder of this article we refer to these as *opportunity scores*.

MEASURING DISPARITY In order to identify hospitals with low ethnic and racial disparities, the opportunity scores for each racial and ethnic group had to be translated into a metric reflecting the degree of within-hospital disparity. The state Medicaid office considered several ap-

proaches, focusing on two disparities statistics: the “absolute risk difference” and “between-group variance.”

The absolute risk difference is the difference in opportunity scores (each expressed as a proportion) between two racial or ethnic groups. For example, if the proportion of patients receiving the recommended care for a condition is 0.70 for white patients and 0.55 for black patients, then the white-black absolute risk difference is 0.15. A value of 0 means that there is no disparity between the groups, while +1 or -1 indicates the maximum disparity possible. For comparisons of multiple racial or ethnic groups, the measurement must be repeated for each pair of groups.

In contrast, between-group variance provides a single measure of the consistency of care provided. It is calculated by summing the variations between a hospital’s opportunity scores for each racial or ethnic group and the institution’s overall opportunity score. Values range from 0 to 0.25, with the former representing complete equality across groups and the latter representing care for one group delivered at 100 percent of the opportunities for it and care delivered at no opportunities for any other group. Further details and equations for both measures are available in Appendix Exhibit 1.³⁹

The Massachusetts Medicaid office chose to use the between group variance measure in awarding payments, for reasons described below. Office staff translated the values on this measure into “performance scores” relative to the previous years’ performance statewide, using the “benchmark, improvement, attainment” approach that has been proposed for Medicare’s value-based purchasing program.⁴⁰ That approach compares each hospital relative to statewide benchmarks in a prior year. Hospitals are eligible for payment based on either their absolute performance or their improvement over time.

Analysis Of Hospital Performance In The First Year

We now turn to the data on hospital performance that were provided to us by the state. We look at the extent of racial diversity within hospitals and examine disparities as they are captured by the two disparities statistics, with special attention to reliability. We also report on the extent to which hospitals were able to submit acceptable documentation on the structural measures.

Performance: Data And Methods

To assess performance on the clinical measures, we used 2008 hospital-level opportunity scores

for racial and ethnic groups—white, Latino, black, Asian, and other—provided by the state Medicaid office. Consistent with the program design, the data pertained only to eligible MassHealth patients who had conditions that fell into one of the five measure groups.

We began by examining the between-hospital distribution of patients by race or ethnicity. We calculated opportunity scores and then absolute risk difference statistics for each hospital, comparing whites and blacks, whites and Latinos, and whites and all minorities combined. To see if we could reliably differentiate among hospitals, we computed 95 percent confidence intervals for each hospital's absolute risk difference, by calculating standard errors from 1,000 resampling iterations using the bootstrap method.

For each racial and ethnic group, we also calculated an unweighted statewide mean absolute risk difference across all hospitals. We compared each hospital's confidence interval with the statewide mean. Then we repeated this procedure for the between-group variance statistic, calculating scores for each hospital and the unweighted statewide mean of these statistics across hospitals. We computed 95 percent confidence intervals for each hospital following the method described above and again compared those confidence intervals with the statewide mean.

For the structural measures, we calculated the percentage of hospitals whose data had been deemed valid by the state Medicaid office. And we examined the correlations between measures of hospitals' performance on the between-group variance statistic, structural measures, total opportunity scores, number of opportunities, and proportion of minority patients.

Performance: Results

CONCENTRATION OF MINORITY MEDICAID PATIENTS Massachusetts hospitals reported a total of 11,187 opportunities from the five measure groups for 2008. Data on racial or ethnic composition were missing from two hospitals. In the remaining sixty-four hospitals, 55.7 percent of the opportunity scores came from white patients; 14.1 percent from black patients; 23.0 percent from Latino patients; 3.0 percent from Asian patients; and 4.2 percent from patients of other races. This is a substantially higher proportion of minorities than is found in the state's population at large.

However, minority MassHealth patients received their care in a limited subset of hospitals. For black and Latino patients, approximately two-thirds of the opportunities came from ten hospitals, and 90 percent came from twenty hos-

pitals. In contrast, for white patients, 40 percent of opportunities came from ten hospitals, and 63 percent came from twenty hospitals. Eight of the state's hospitals reported zero opportunities for minority Medicaid patients; one hospital reported zero for white patients.

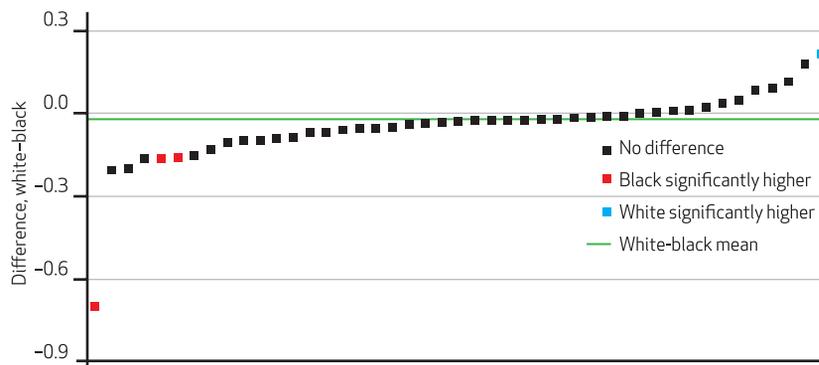
LITTLE EVIDENCE OF DISPARITIES IN CARE When we combined data from all of the hospitals, opportunity scores were slightly higher for black patients than for white patients, which in turn were slightly higher than for Latino patients. For blacks the statewide opportunity score was 91.4, for whites it was 89.6, and for Latinos it was 86.2. The corresponding statewide absolute risk differences were -1.8 for whites compared to blacks (95 percent confidence interval: $-4.0, 0.4$); 3.5 for whites compared to Latinos (95 percent confidence interval: $0.3, 6.8$); and 2.0 for whites compared to all minorities (95 percent confidence interval: $-0.8, 4.9$).

As shown in Exhibits 2 and 3, for most hospitals the within-hospital absolute risk difference scores of whites compared to blacks and whites compared to Latinos were small and could not be reliably estimated. The confidence intervals are typically wide and pass through 0, a result that is consistent with no disparity (see Supplemental Exhibits 2 and 3 in the online Appendix for confidence intervals).³⁹ This was true even when nonwhite patients were combined into a single minority category in each hospital (Exhibit 4; Supplemental Exhibit 4).³⁹

CONSISTENT CARE WITHIN EACH HOSPITAL Small sample sizes meant that the absolute risk difference measure would not reliably show disparities. The state Medicaid office turned to the

EXHIBIT 2

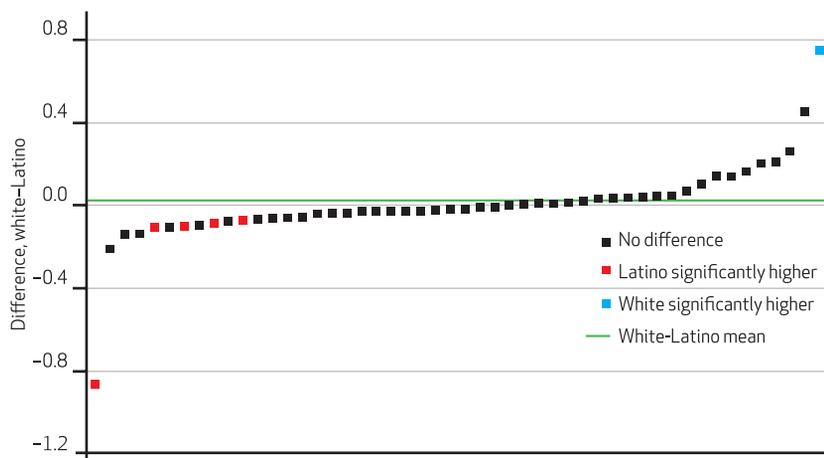
2008 Hospital-Level White-Black Differences In Quality Of Care In Massachusetts, According To The Absolute Risk Difference Measure



SOURCE Authors' analysis using 2008 hospital performance data from the Massachusetts Office of Medicaid. **NOTES** Only hospitals reporting at least one opportunity to provide care for both groups are shown ($n = 45$). "Significant" means statistically significant at the 0.05 level. A version of this exhibit showing 95 percent confidence intervals is available as Supplemental Exhibit 2 in the online Appendix; see Note 39 in text.

EXHIBIT 3

2008 Hospital-Level White-Latino Differences In Quality Of Care In Massachusetts, According To The Absolute Risk Difference Measure



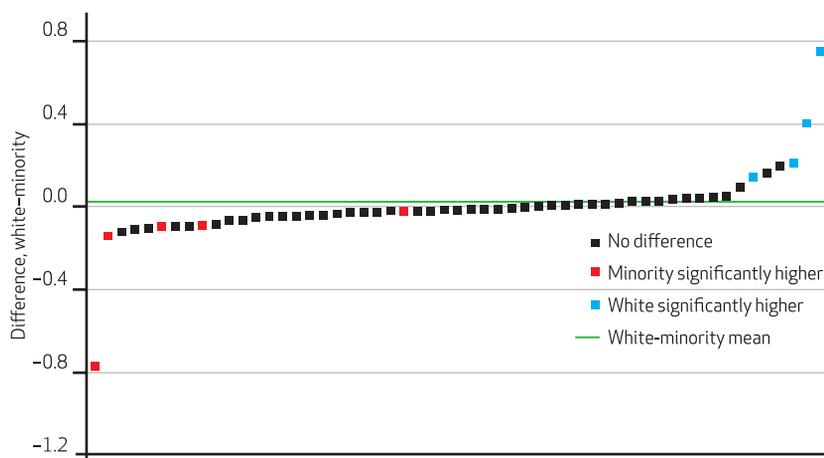
SOURCE Authors' analysis using 2008 hospital performance data from the Massachusetts Office of Medicaid. **NOTES** Only hospitals reporting at least one opportunity to provide care for both groups are shown ($n = 50$). "Significant" means statistically significant at the 0.05 level. A version of this exhibit showing 95 percent confidence intervals is available as Supplemental Exhibit 3 in the online Appendix; see Note 39 in text.

between-group variance measure, which it believed would allow it to more accurately identify low- and high-performing hospitals.

Our analysis of the data showed that for the fifty-five hospitals with eligible patients of more than one race, the statewide mean between group variance was very low (0.0058), consist-

EXHIBIT 4

2008 Hospital-Level White-Minority Differences In Quality Of Care In Massachusetts, According To The Absolute Risk Difference Measure



SOURCE Authors' analysis using 2008 hospital performance data from the Massachusetts Office of Medicaid. **NOTES** Only hospitals reporting at least one opportunity to provide care for both groups are shown ($n = 55$). "Minority" means all patients who are not white. "Significant" means statistically significant at the 0.05 level. A version of this exhibit showing 95 percent confidence intervals is available as Supplemental Exhibit 4 in the online Appendix; see Note 39 in text.

tent with uniform treatment of patients, regardless of their racial or ethnic groups.

Exhibit 5 shows values for each of the hospitals. There was minimal variation across hospitals, with values clustered around zero. Nevertheless, it was possible to identify eighteen hospitals whose performance was better (significantly lower) than the statewide mean, and two hospitals whose performance was worse (significantly higher).

However, further analysis showed that the apparent high degree of reliability, as indicated by the narrow confidence intervals for many hospitals, was an artifact of these hospitals' having small numbers of nonwhite patients, high clinical quality overall, or both. By definition, for hospitals with these characteristics, the resampling in bootstrapping iterations yielded between-group variance estimates that were essentially identical each time, resulting in small standard errors and narrow confidence intervals (see Supplemental Exhibits 2–5 in the Appendix).³⁹

DIFFICULTIES WITH THE STRUCTURAL MEASURES CHECKLIST In Rate Year 2008, the first year that the Massachusetts Medicaid office scored structural measures, twenty-one of the acute care hospitals (nearly a third) failed to provide adequate documentation for their responses, according to the office. Those hospitals received no incentive payments. The following year hospitals did better: Fifty-six submitted acceptable documentation.

LESS DISPARITY, HIGHER QUALITY OF CARE, LESS DIVERSITY Across hospitals, a lower (better) between-group variance was associated with better opportunity scores and better scores on the structural measures. However, it was also associated with having a lower proportion of minority patients (see Appendix Exhibit 2).³⁹

Responses From The Hospital Community

In our discussions with hospital administrators and staff from the state hospital association, we found support for the pay-for-performance program's goal of reducing racial and ethnic disparities in care. Our informants also noted that the program's reporting requirements—particularly those related to the structural measures—had led to discussions about disparities in their institutions. For example, hospitals analyzed the racial and ethnic compositions of their patient populations, inventoried their programs that targeted minority communities, and discussed whether their mission statements adequately addressed disparities in health care.

Even as they were immersed in these efforts, hospitals faced reporting requirements from

many other sources. People expressed frustration at having to adapt to yet another system. They were also skeptical about the choice of clinical measures. Some informants doubted that routine treatment at their institutions had anything to do with a patient's race or ethnicity. One administrator said that many in the hospital community believed that the program was "barking up the wrong tree."

Some of our informants—even those whose institutions had performed well—were quite negative about the structural measures. They found some of the items to be unclear, others to be highly subjective, and still others to be inappropriate to their institutional environment. For example, the standard that the hospital's mission statement "articulates cultural diversity as a core value"⁴¹ suggests that there is a straightforward way to interpret hospital mission statements. The requirement that patient data be "analyzed by race, ethnicity, and languages spoken" implies that there is sufficient diversity in the patient population to make that effort worthwhile.

Informants also reported receiving insufficient (and sometimes contradictory) advice about what would count as adequate documentation of their activities. These frustrations were compounded by delays of many months as the small staff of the state's Medicaid office worked to validate data.

But perhaps the greatest dissatisfaction came from what the hospitals saw as a focus on documentation at the expense of practical advice that might help reduce disparities. One hospital quality improvement officer described a "compliance mind-set rather than improvement mind-set." He contrasted the Massachusetts program's approach with the interactive one taken during accreditation visits by the Joint Commission, whose investigators reframe requests for documentation if the hospital's initial responses are not adequate.

Although our informants felt that instructions from the state Medicaid office were becoming clearer, the process was still irksome. Many people questioned whether it would improve care.

Lessons Learned

The state Medicaid office accomplished a great deal in a short time, following passage of the legislation that created the Massachusetts pay-for-performance program. In spite of limited resources, the office engaged stakeholders, developed a program, and heightened awareness of disparities within the hospital community.

Although the program is still in its early stages, some lessons are clear.

EXHIBIT 5

2008 Hospital-Level Uniformity In Quality Of Care In Massachusetts Across All Racial And Ethnic Groups, According To The Between-Group Variance (BGV) Measure



SOURCE Authors' analysis using 2008 hospital performance data from the Massachusetts Office of Medicaid. **NOTES** Only hospitals reporting at least one opportunity to provide care for white and nonwhite patients are shown ($n = 55$). "Significant" means statistically significant at the 0.05 level. A version of this exhibit showing 95 percent confidence intervals is available as Supplemental Exhibit 5 in the online Appendix; see Note 39 in text.

ADDRESS DISPARITIES IN HOSPITAL CARE EFFICIENTLY Massachusetts' racial and ethnic disparities legislation was based on the assumptions that there were racial and ethnic disparities in the treatment of patients within the state's hospitals and that every hospital's patient population was sufficiently diverse to make a statewide intervention sensible. Our analysis does not support either assumption.

This gap between legislative impulse and on-the-ground circumstance is striking. Perhaps if the program had used different measures, disparities would have been evident. But the problem of inequity may be at a level that is relatively untouched by pay-for-performance.

Recent evidence suggests that patients of color across the United States receive their care from a relatively small subset of providers, and that those providers may on average offer lower-quality care.^{19–21} To the extent that disparities in acute care arise more from between-hospital differences in quality than from within-hospital disparate treatment, a more effective way to reduce disparities may be to target hospitals serving minority populations. Moreover, to the extent that some disparities reflect structural barriers beyond hospital walls, it might not make sense to hold hospitals accountable.⁴²

IMPROVING CARE IS FOREMOST By introducing a program to reduce racial and ethnic disparities, the Massachusetts legislature presumably hoped to ensure that minority patients would receive

better care. However, given the short time available, the state's executive branch focused on measuring quality of care.

As members of the hospital community noted, it is not clear that the chosen measures were connected with disparities. Nor is there evidence that better performance on many of the structural measures would actually improve care for minority patients.

FIND THE TRUE TOP PERFORMERS Sample size is not just a statistical issue; it affects the fairness and plausibility of a program's results. Without sufficient numbers of patients, program administrators cannot reliably identify high-performing providers and therefore cannot make fair decisions about distributing incentive payments.

The state Medicaid office turned to the between-group variance measure to solve a technical problem. Our analyses confirm that for this sample, this approach measures performance more precisely than the absolute risk difference measure. However, given the difficulty with interpreting scores on the between-group variance statistic, a statistically lower between-group variance score did not mean that hospitals treated their patients more equally. Because there was little evidence that hospitals treated patients differently by race or ethnicity, it is unlikely that any statistic would have been helpful in identifying high-performing hospitals.

NOT ALL DISPARITIES STATISTICS ARE EQUAL Different statistics capture different dimensions of disparities. The between-group variance measure reflects uniformity of care. Yet as Appendix Exhibit 3 shows,³⁹ that measure has two disadvantages. First, it is directionless, producing the same score whether minorities or whites receive better care. Second, it is sensitive to racial or ethnic case-mix, which means that hospitals with more-diverse patient populations receive worse scores than less-diverse hospitals, even if they perform at a similar level. By definition, hospitals with few minority patients have low (good) scores. Penalizing diversity in this way would seem to be at odds with the goal of reducing disparities.

The legislation creating Massachusetts' pay-for-performance program was designed to reduce racial and ethnic disparities in health care and did not explicitly refer to the historically poorer treatment of minority patients or hospital diversity. As we noted above, there are differing conceptions of disparities. However, to the extent that reducing disparities involves redressing historical wrongs, all of these limitations to the between-group variance measure deserve attention going forward.

AVOID FRICTION IN REGULATING QUALITY As one of our informants noted, Massachusetts' approach differed from the Joint Commission's "back-and-forth" process, with which hospitals are more familiar. The Joint Commission assesses whether hospitals are working to meet standards and helps them develop a road map they can follow. The Massachusetts Medicaid office took a more regulatory approach, in part the result of resource constraints. Hospital managers found this troubling because they wanted to understand how to improve their performance and because large amounts of money were at stake.

Checklists may be weak catalysts for organizational change and poor tools for monitoring progress toward reducing disparities. To assist providers, both the National Quality Forum⁴³ and the Joint Commission⁴⁴ have offered detailed road maps and sets of preferred practices rather than simple lists of performance measures. However, using those approaches takes more time and resources than working with a checklist.

This case study—of a program implemented under legislative mandate in a very short time—provides a clear contrast with the more interactive and labor-intensive approach to quality improvement used by standard-setting bodies.

Epilogue

Massachusetts announced that as of the summer of 2010, it would no longer use structural measures in its pay-for-performance program to reduce racial and ethnic disparities, implicitly acknowledging that the measures needed to be revised. However, interest in cultural competence—a major issue targeted by the structural measures—is growing nationwide. The National Committee for Quality Assurance's new Multicultural Health Care Standards encompass the collection of data on race and ethnicity and attention to Culturally and Linguistically Appropriate Services practice standards. The national standards themselves are undergoing revision.⁴⁵

Clinical measures are still being used to monitor disparities in Massachusetts hospitals, and incentive payments continue to be awarded to hospitals that score well on the disparities measure. Although national bodies are making progress in developing measures suitable for tracking disparities in ambulatory care,^{46,47} there are still no nationally vetted measures available for use in the hospital setting. ■

A version of this paper was presented at the Veterans Affairs Disparities Conference, in Boston, Massachusetts, in September 2010. This work was supported in part by the Northwestern University Center for Healthcare Equity. Jan Blustein was supported by a New York University/Wagner Faculty Research Fund grant. Andrew Ryan was supported by a Career Development Award from the Agency for Healthcare Research and Quality (Grant No. 1 K01 HS018546-01A1). Tim Doran was supported by the Commonwealth Fund, as a Harkness Fellow in Health Care

Policy and Practice. The authors thank Terri Yannetti and Phyllis Peters of the Massachusetts Office of Medicaid for their help and guidance. Their willingness to share their experiences with a wider audience shows a dedication to public service, along with an awareness of the program's limitations. The authors are also grateful to Anuj Goel, of the Massachusetts Hospital Association, for his assistance. From 2008 to 2010 Joel Weissman served as senior health policy adviser to the secretary of the Massachusetts Executive Office of Health and Human

Services, JudyAnn Bigby. However, he had no particular role in the initiation or operation of the Massachusetts pay-for-performance program. The views expressed here are those of the authors, and no official endorsement by the Massachusetts Executive Office of Health and Human Services, the Massachusetts Office of Medicaid, the Massachusetts Hospital Association, Northwestern University, or the Commonwealth Fund is intended or should be inferred.

NOTES

- 1 Kuhmerker K, Hartman T. Pay-for-performance in state Medicaid programs: a survey of state Medicaid directors and programs. New York (NY): Commonwealth Fund; 2007 Apr.
- 2 Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA*. 2000;283(19):2579-84.
- 3 Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington (DC): National Academies Press; 2001.
- 4 National Committee for Quality Assurance. 2009 innovative practices in multicultural health care [Internet]. Washington (DC): NCQA; 2009 [cited 2011 May 4]. Available from: http://www.ncqa.org/Portals/0/HEDISQM/CLAS/CLAS_InnovPrac_09.pdf
- 5 Robert Wood Johnson Foundation. Finding answers: disparities research for change [Internet]. Princeton (NJ): RWJF; [cited 2011 May 4]. Available from: <http://www.solvingdisparities.org/>
- 6 Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D. Will pay-for-performance and quality reporting affect health care disparities? *Health Aff (Millwood)*. 2007; 26(3):w405-14. DOI: 10.1377/hlthaff.26.3.w405.
- 7 Chien AT, Chin MH, Davis AM, Casalino LP. Pay for performance, public reporting, and racial disparities in health care: how are programs being designed? *Med Care Res Rev*. 2007;64(5 Suppl):283S-304S.
- 8 Chien AT. The potential impact of performance incentive programs on racial disparities in healthcare. In: Williams RA, editor. *Eliminating healthcare disparities in America: beyond the IOM report*. Totowa (NJ): Humana Press; 2007. p. 237-56.
- 9 Rosenthal MB, Dudley RA. Pay-for-performance: will the latest trend improve care? *JAMA*. 2007;297:740-4.
- 10 Christianson J, Leatherman S, Sutherland K. Financial incentives, healthcare providers, and quality improvements: a review of the evidence [Internet]. London: The Health Foundation; 2009 [cited 2011 May 4]. Available from: <http://www.health.org.uk/public/cms/75/76/313/571/Financial%20incentives.pdf?realName=5GmFYW.pdf>
- 11 Doran T, Roland M. Lessons from major initiatives to improve primary care in the United Kingdom. *Health Aff (Millwood)*. 2010;29(5):1023-9.
- 12 Ver Ploeg M, Perrin E, editors. *Eliminating health disparities: measurement and data needs*. Washington (DC): National Academies Press; 2004.
- 13 Ulmer C, McFadden B, Nerenz DR, editors. *Race, ethnicity, and language data: standardization for health care quality improvement*. Washington (DC): National Academies Press; 2009.
- 14 Weinick RM, Caglia JM, Friedman E, Flaherty K. Measuring racial and ethnic health care disparities in Massachusetts. *Health Aff (Millwood)*. 2007;26(5):1293-302.
- 15 National Quality Forum. National voluntary consensus standards for ambulatory care: measuring health-care disparities [Internet]. Washington (DC): NQF; 2008 [cited 2011 May 4]. Available from: http://www.qualityforum.org/Publications/2008/03/National_Voluntary_Consensus_Standards_for_Ambulatory_Care%e2%80%94Measuring_Healthcare_Disparities.aspx
- 16 Harper S, King NB, Meersman SC, Reichman MA, Breen N, Lynch J. Implicit value judgments in the measurement of health inequalities. *Milbank Q*. 2010;88(1):4-29.
- 17 Hebert PL, Sisk JE, Howell EA. When does a difference become a disparity? Conceptualizing racial and ethnic disparities in health. *Health Aff (Millwood)*. 2008;27(2):374-83.
- 18 Normand SL, Wolf RE, Ayanian JZ, McNeil BJ. Assessing the accuracy of hospital clinical performance measures. *Med Decis Making*. 2007; 27(1):9-20.
- 19 Barnato AE, Lucas FL, Staiger S, Wennberg DE, Chandra A. Hospital-level racial disparities in acute myocardial infarction treatment and outcomes. *Med Care*. 2005;43(4):308-19.
- 20 Hasnain-Wynia R, Baker DW, Nerenz D, Feinglass J, Beal AC, Landrum MB, et al. Disparities in healthcare are driven by where minority patients seek care. *Arch Intern Med*. 2007;167(12):1233-9.
- 21 Jha AK, Orav EJ, Li Z, Epstein AM. Concentration and quality of hospitals that care for elderly black patients. *Arch Intern Med*. 2007; 167(11):1177-82.
- 22 Families USA. *Confronting disparities while reforming health care: a look at Massachusetts* [Internet]. Washington (DC): Families USA; 2008 Jan [cited 2011 May 4]. Available from: <http://www.familiesusa.org/assets/pdfs/ma-disparities-case-study.pdf>
- 23 Commission to End Racial and Ethnic Health Disparities. *Final report* [Internet]. Boston (MA): The Commission; 2007 Aug [cited 2011 May 4]. Available from: http://www.mass.gov/hdc/about/2006_report.pdf
- 24 Census Bureau. *The black population: 2000* [Internet]. Washington (DC): Census Bureau; 2001 Aug [cited 2011 May 4]. Available from: <http://www.census.gov/prod/2001pubs/c2kbr01-5.pdf>
- 25 Census Bureau. *The Hispanic population: 2000* [Internet]. Washington (DC): Census Bureau; 2001 May [cited 2011 May 4]. Available from: <http://www.census.gov/prod/2001pubs/c2kbr01-3.pdf>
- 26 Census Bureau. *Census 2000 redistricting data (Public Law 94-171), summary file, matrices PL1 and PL2* [Internet]. Washington (DC): Census Bureau; [cited 2011 Jun 2]. Available for download from: <http://factfinder.census.gov>

- 27 Yannetti T. MassHealth hospital pay for performance. Paper presented at: 10th Annual MassHealth Symposium; 2007 Oct 24; Worcester, MA.
- 28 Massachusetts Medicaid Policy Institute. Pay-for-performance to reduce racial and ethnic disparities in health care in the Massachusetts Medicaid program: recommendations of the Massachusetts Medicaid Disparities Policy Roundtable [Internet]. Boston (MA): The Institute; 2007 Jul [cited 2011 May 4]. Available from: <http://www.massmedicaid.org/~media/MMPI/Files/Pay%20for%20Performance%20to%20Reduce%20Racial%20and%20Ethnic%20Disparities.pdf>
- 29 Department of Health and Human Services, Office of Minority Health. National standards for Culturally and Linguistically Appropriate Services in health care: final report [Internet]. Washington (DC): HHS; 2001 Mar [cited 2011 May 4]. Available from: <http://www.omhrc.gov/assets/pdf/checked/finalreport.pdf>
- 30 Leapfrog Group [home page on the Internet]. Washington (DC): Leapfrog Group; [cited 2011 May 4]. Available from: <http://www.leapfroggroup.org/>
- 31 Massachusetts Executive Office of Health and Human Services. Rate Year 2010 EOHHS technical specifications manual for MassHealth acute hospital quality measures (version 3.0) [Internet]. Boston (MA): The Office; 2009 Aug 31 [cited 2010 Aug 30]. Available from: http://www.mass.gov/Eeohhs2/docs/mashealth/acutehosp/ry2010_technical-specs-manual-3.0.pdf
- 32 Garcia-Caban IA. Improving systems of care for racial and ethnic minority populations: exploring cultural competence in Massachusetts acute care hospitals [dissertation]. Waltham (MA): Brandeis University; 2002.
- 33 Massachusetts Executive Office of Health and Human Services. Acute Hospital RFA Appendix G 2008: Cultural Competence Organizational Self-Assessment (CCOSA) checklist [Internet]. Boston (MA): The Office; 2008 [cited 2011 May 4]. Available from: http://www.mass.gov/Eeohhs2/docs/mashealth/acutehosp/2008-appxG_1.0.xls
- 34 Massachusetts Executive Office of Health and Human Services. Technical specifications manual for MassHealth hospital P4P measures, appendix G 2008 reporting (version 1.3) [Internet]. Boston (MA): The Office; 2007 Dec 20 [cited 2011 May 4]. Available from: http://www.mass.gov/Eeohhs2/docs/mashealth/acutehosp/2008-techspec_manual-1.3.txt
- 35 Although 0.02 percent was the mean potential incentive payment, this varied across hospitals. For the lowest quartile (the twenty-fifth percentile), potential payments were 0.01 percent of revenues; for the fiftieth percentile, 0.02 percent; for the seventy-fifth percentile, 0.03 percent; and for the ninety-ninth percentile, 0.08 percent, according to the authors' analysis of data provided by the Massachusetts Medicaid office.
- 36 Massachusetts Executive Office of Health and Human Services. MassHealth hospital P4P program update: health disparities measures. Boston (MA): The Office; 2009 Dec 9.
- 37 Premier. About the Hospital Quality Incentive Demonstration (HQID) [Internet]. Charlotte (NC): Premier; 2008 [cited 2011 May 4]. Available from: http://www.premierinc.com/p4p/hqi/year-3-results/HQID-FactonlyPartic%20_2_.pdf
- 38 Landrum MB, Bronskill SE, Normand ST. Analytic methods for constructing cross-sectional profiles of health care providers. *Health Serv Outcomes Res Methodol*. 2000; 1(1):23-47.
- 39 To access the Appendix, click on the Appendix link in the box to the right of the article online.
- 40 Centers for Medicare and Medicaid Services. Report to Congress: plan to implement a Medicare hospital value-based purchasing program [Internet]. Baltimore (MD): CMS; 2007 Nov 21 [cited 2011 May 4]. Available from: <http://www.cms.hhs.gov/AcuteInpatientPPS/downloads/HospitalVBPPlanRTCFINALSUBMITTED2007.pdf>
- 41 Massachusetts Executive Office of Health and Human Services. Revised Cultural Competence Organizational Self-Assessment (CCOSA) checklist [Internet]. Boston (MA): The Office; 2010 [cited 2011 May 4]. Available from: http://www.mass.gov/Eeohhs2/docs/mashealth/acutehosp/ccosa_2010-revised-form.pdf
- 42 Blustein J. Who is accountable for racial equity in health care? *JAMA*. 2008;299(7):814-6.
- 43 National Quality Forum. A comprehensive framework and preferred practices for measuring and reporting cultural competency [Internet]. Washington (DC): NQF; 2009 Apr [cited 2011 May 4]. Available from: http://www.qualityforum.org/Publications/2009/04/A_Comprehensive_Framework_and_PREFERRED_Practices_for_Measuring_and_Reporting_Cultural_Competency.aspx
- 44 Joint Commission. Advancing effective communication, cultural competence, and patient- and family-centered care: a roadmap for hospitals [Internet]. Oakbrook Terrace (IL): Joint Commission; 2010 [cited 2011 May 4]. Available from: <http://www.jointcommission.org/assets/1/6/ARoadmapforHospitalsfinalversion727.pdf>
- 45 Department of Health and Human Services. Call for comments on the existing national standards for the culturally and linguistically appropriate services in health care. *Fed Regist* [serial on the Internet]. 2010;75(184):57957-8 [cited 2011 May 4]. Available from: <http://federalregister.gov/a/2010-23760>
- 46 National Quality Forum. Closing the disparities gap in healthcare quality with performance measurement and public reporting [Internet]. Washington (DC): NQF; 2008 Aug [cited 2011 May 4]. (Issue Brief No. 16). Available from: http://www.qualityforum.org/Publications/2008/08/Closing_the_Disparities_Gap_in_Healthcare_Quality_With_Performance_Measurement_and_Public_Reporting.aspx
- 47 Agency for Healthcare Research and Quality. Guide to prevention quality indicators: hospital admission for ambulatory care sensitive conditions [Internet]. Rockville (MD): AHRQ; 2007 Mar 12 [cited 2011 May 25]. Available from: http://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V31/pqi_guide_v31.pdf

ABOUT THE AUTHORS: JAN BLUSTEIN, JOEL S. WEISSMAN, ANDREW M. RYAN, TIM DORAN & ROMANA HASNAIN-WYNIA



Jan Blustein is a professor at New York University.

In this issue of *Health Affairs*, Jan Blustein and coauthors examine a Massachusetts program to link pay-for-performance incentives to the reduction of ethnic and racial disparities in health care. They conclude that for various reasons, the linkage was not an effective way to reduce those disparities. The findings provide “a bit of a reality check for policy makers, going forward,” says Blustein, a professor of health policy and medicine at New York University’s Wagner School of Public Service and School of Medicine.

The results were somewhat unexpected for Blustein and her coauthors, Joel Weissman, Andrew Ryan, Tim Doran, and Romana Hasnain-Wynia. The five came together on this project because they shared an interest in treatment disparities and a belief that pay-for-performance incentives might drive change. The authors now hope that their article will make policy makers aware of the practical and technical pitfalls that can undermine program effectiveness.

Blustein received her medical degree from Yale University’s School of Medicine and her doctorate in public administration from New York University.



Joel S. Weissman is an associate professor at Massachusetts General Hospital and Harvard Medical School.

Weissman is an associate professor of health policy at the Massachusetts General Hospital’s Mongan Institute for Health Policy and at Harvard Medical School. He received his doctorate in health policy from Brandeis University’s Florence Heller School of Social Welfare.



Andrew M. Ryan is an assistant professor at the Weill Cornell Medical College.

Ryan is an assistant professor at the Weill Cornell Medical College. He received his doctorate in social policy from Brandeis and won the

AcademyHealth Dissertation Award in 2009.



Tim Doran is a clinical research fellow at the University of Manchester.

Doran is a clinical research fellow in public health at the School of Health Sciences of the University of Manchester, in England. He received his medical degree from the University of Edinburgh.



Romana Hasnain-Wynia is the director of the Center for Healthcare Equity at Northwestern University.

Hasnain-Wynia is the director of the Center for Healthcare Equity at Northwestern University and an associate research professor at the university’s Feinberg School of Medicine. She received her doctorate of philosophy in social policy from Brandeis.