

HealthAffairs

At the Intersection of Health, Health Care and Policy

Cite this article as:

Andrew M. Ryan, Jan Blustein and Lawrence P. Casalino
Medicare's Flagship Test Of Pay-For-Performance Did Not Spur More Rapid Quality
Improvement Among Low-Performing Hospitals
Health Affairs, 31, no.4 (2012):797-805

doi: 10.1377/hlthaff.2011.0626

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

<http://content.healthaffairs.org/content/31/4/797.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 © 2012 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 Andrew M. Ryan, Jan Blustein, and Lawrence P. Casalino

Medicare's Flagship Test Of Pay-For-Performance Did Not Spur More Rapid Quality Improvement Among Low-Performing Hospitals

DOI: 10.1377/hlthaff.2011.0626
 HEALTH AFFAIRS 31,
 NO. 4 (2012): 797-805
 ©2012 Project HOPE—
 The People-to-People Health
 Foundation, Inc.

ABSTRACT Medicare's flagship hospital pay-for-performance program, the Premier Hospital Quality Incentive Demonstration, began in 2003 but changed its incentive design in late 2006. The goals were to encourage greater quality improvement, particularly among lower-performing hospitals. However, we found no evidence that the change achieved these goals. Although the program changes were intended to provide strong incentives for improvement to the lowest-performing hospitals, we found that in practice the new incentive design resulted in the strongest incentives for hospitals that had already achieved quality performance ratings just above the median for the entire group of participating hospitals. Yet during the course of the program, these hospitals improved no more than others. Our findings raise questions about whether pay-for-performance strategies that reward improvement can generate greater improvement among lower performing providers. They also cast some doubt on the extent to which hospitals respond to the specific structure of economic incentives in pay-for-performance programs.

Andrew M. Ryan (amr2015@med.cornell.edu) is an assistant professor of public health at Weill Cornell Medical College, in New York City.

Jan Blustein is a professor of health policy and medicine at the Wagner Graduate School and the School of Medicine, New York University, in New York City.

Lawrence P. Casalino is chief of the Division of Outcomes and Effectiveness Research and the Livingston Farrand Associate Professor of Public Health at Weill Cornell Medical College.

Nationwide implementation of programs designed to reward hospitals for improved performance on measures of care, called hospital pay-for-performance, is scheduled for late 2012 as part of Medicare's Value-Based Purchasing Program.¹ This is part of a larger initiative from Medicare to integrate pay-for-performance into its payment system for physicians and nursing homes as well as hospitals.² Hospital pay-for-performance enjoys considerable support from policy makers and payers.²⁻⁵ Nonetheless, questions remain about how effective it will be.^{6,7} It is also unknown whether the design of pay-for-performance programs, such as whether they reward quality improvement along with quality attainment, affects how responsive hospitals are to them.⁵

Much of the published evidence on hospital pay-for-performance draws on experience from the Premier Hospital Quality Incentive Demon-

stration, a highly visible nationwide program jointly run by the Centers for Medicare and Medicaid Services and Premier since 2003.⁸⁻¹⁰ The demonstration was designed to "explore ways to define quality measures and reward top-performing hospitals based on the quality of care that they provide to patients."¹¹ In this program, 266 hospitals—all subscribers to Premier's "Perspective" hospital performance benchmarking service—agreed to collect their scores on a set of quality measures and make their performance subject to financial incentives. Participation in the demonstration was voluntary. Sixty-three percent of hospitals eligible to participate chose to do so.⁸

For the first three years of the demonstration, only the very-highest-performing hospitals received financial awards. However, in 2006 incentives were added for both quality improvement and good performance across an increased range of quality measures. The program's administra-

tors hoped that quality improvement would continue—particularly for hospitals at the low end of the performance spectrum, which had the greatest potential for improvement.¹²

Impact Of The Demonstration

LOGIC OF INCENTIVES Incentives for improvement are aimed at different providers than incentives for reaching performance levels, called “attainment” by analysts in the field.³ Rewards for attainment recognize the achievement of the highest-performing providers and give them incentives to continue their efforts. In contrast, rewards for improvement are aimed at lower-performing providers, encouraging them to do better and recognizing those that succeed in improving.⁵ Rewarding improvement can also enhance health care equity because low-performing hospitals disproportionately serve poor and minority patients¹³ and are located in disadvantaged areas.¹⁴

Providing incentives for both attainment and quality improvement in pay-for-performance has been recommended by the Institute of Medicine.⁵ It is also an essential feature of Medicare’s planned Hospital Value-based Purchasing Program.¹

PRIOR RESEARCH Studies of the first phase of the Premier Hospital Quality Incentive Demonstration, from the fourth quarter of 2003 through the third quarter of 2006, suggest that hospitals participating in the demonstration improved their scores on process measures of quality more than nonparticipating hospitals did.⁸ However, improvement patterns in phase 1 were inconsistent with the financial incentive structure. The lowest-performing hospitals showed the greatest improvement relative to the comparison group, despite having little chance of receiving payments under a program that rewarded only the very highest performers.¹⁵

A recent article examined quality improvement for hospitals during phase 1 and phase 2, which ran from the fourth quarter of 2006 through the third quarter of 2009. But the article did not examine whether changes in the financial incentives between the two phases led to differences in quality improvement among participating hospitals.¹⁶

PUBLIC TRANSPARENCY The demonstration’s effect on quality can be measured only against a backdrop of public reporting by hospitals. Starting in 2004, in a change nearly concurrent with the beginning of the demonstration, Medicare made annual payment updates for all acute care hospitals conditional on their submitting quality data for ten measures related to heart attack, heart failure, and pneumonia. In April

2005 Medicare began publicly reporting these data as part of Hospital Compare.

Moreover, in 2007 a high-profile report from Medicare proposed options for the extension of pay-for-performance to all acute care hospitals.³ According to Medicare’s publicly reported measures, hospitals’ quality of care has improved steadily since nationwide public reporting began—a period that spanned both phases of the demonstration—particularly for hospitals whose initial performance was low.^{8,14}

Qualitative evidence suggests that hospitals improved their quality either to enhance their reputations as providing high quality or to avoid being perceived as providing poor quality.¹⁷ The relatively greater improvement by providers that initially did poorly may reflect the concept of “low-hanging fruit”—in other words, these providers moved swiftly to take relatively easy steps to improve quality—and the progressively greater difficulty of improvement as hospitals approach a score of 100 percent on performance measures.

Thus, the demonstration was implemented concurrently with a national public reporting program in which quality of care was improving and the expectation of nationwide implementation of pay-for-performance was established.

IMPACT OF PHASE 2 The release of data from phase 2 of the Premier program provides an opportunity to assess hospitals’ responses to the new set of financial incentives. To evaluate the effect of the change in phase 2, we used a “difference in differences” approach that compared the difference in quality performance between groups of hospitals across phase 1 and phase 2.

We asked two main research questions. First, did hospitals participating in the demonstration improve more in phase 2 than in phase 1? We addressed this question by examining differences between the hospitals participating in the demonstration and a set of matched comparison hospitals in the rate of quality improvement across both phases of the intervention.

Second, did the participating hospitals that had lower quality at the beginning of the demonstration improve more in phase 2 than participating hospitals that had moderate quality at the beginning? We addressed this question by examining whether the demonstration hospitals that began phase 1 in the lowest quartile of quality performance improved more between phase 1 and phase 2, compared to matched comparison hospitals, than demonstration hospitals that began phase 1 in the second-lowest quartile of quality performance.

To examine whether financial incentives explain hospitals’ improvement patterns in phase 2, we also simulated payments that would have

been received by hospitals under two hypothetical quality improvement scenarios.

Study Data And Methods

FINANCIAL INCENTIVES UNDER THE DEMONSTRATION

As noted previously, phase 1 of the demonstration rewarded only high attainment. Medicare paid a 2 percent bonus on its reimbursement rates to hospitals performing in the top tenth of demonstration hospitals on a composite quality measure for each clinical diagnosis and procedure incentivized in the demonstration—heart attack, heart failure, pneumonia, bypass surgery, and hip and knee replacement—and a 1 percent bonus for hospitals performing in the second-highest decile. Beginning in the third year of the demonstration and continuing through phase 2, penalties were imposed for hospitals performing below the twentieth percentile of hospitals two years prior to the current year. For example, a hospital with performance in year 3 that was below the twentieth percentile in year 1 would be penalized.

During phase 2, hospitals were eligible to receive three types of rewards. First was an attainment award, given to hospitals whose composite scores in the current year exceeded the median of demonstration hospitals two years prior to the current year. Second was a top performer award, given to hospitals that scored in the top 20 percent of demonstration hospitals in the current year. Third was an improvement award, given to hospitals with scores above the median of demonstration hospitals in the current year that ranked in the top 20 percent of demonstration hospitals for quality improvement. Hospitals could receive both top performer and attainment awards or both improvement and attainment awards. However, they could not receive both top performer and improvement awards.

The amount of incentive payments increased from an average of \$8.2 million per year in phase 1 to \$12 million per year in phase 2.^{18,19} Of the phase 2 bonuses, 60 percent was allocated to top performer and improvement awards and 40 percent to attainment awards. The incentivized quality measures remained very similar across the two phases. Online Appendix A summarizes the demonstration's structure, timing, and incentives.²⁰

DATA SOURCES We used annual, hospital-level data on quality from Hospital Compare for discharges in calendar years 2004–09; data on hospital characteristics from the 2005 American Hospital Association Annual Survey; data on the receipt of incentive payments from the Premier website; and, to estimate incentive payments in phase 1, data on hospital revenues from

the 2004–06 Medicare Provider Analysis and Review files.

SAMPLE Because demonstration hospitals are systematically different from average US hospitals (see Appendix B),²⁰ we created a matched sample of US hospitals not participating in the demonstration to serve as the comparison group. This approach limits differences in quality performance between demonstration and comparison hospitals that would occur otherwise (Appendix C provides details of our propensity score matching procedure).²⁰

After matching, our analytic sample included 250 demonstration hospitals and 250 comparison hospitals. During the two phases of the demonstration, thirty-two hospitals withdrew from the program. We performed an intent-to-treat analysis, including exiting hospitals as participating hospitals throughout the study period.

MEASURES OF PERFORMANCE Our outcomes were composite process quality scores for heart attack, heart failure, and pneumonia (see Appendix C for details about the creation of composite scores).²⁰ Because quality data for bypass surgery and hip and knee replacement were not available, we did not evaluate hospitals' performance on these procedures.

ANALYSIS STRATEGY Public reporting of quality was in place for all hospitals, not just demonstration participants. Therefore, we examined changes in quality performance from phase 1 to phase 2 of the demonstration that could be attributed only to the financial incentives provided to hospitals in the program.

Our "difference in differences" approach and the details of our regression models are described in Appendix C.²⁰ Briefly, we relied on a series of contrasts, or differences in quality between groups of hospitals across the two phases. Using the matched sample of hospitals, we addressed our first research question ("Did demonstration hospitals improve more in phase 2 than in phase 1?") by testing whether differences in quality improvement between hospitals participating and not participating in the demonstration were greater in phase 2 than in phase 1.

Again using the matched sample, we addressed our second research question ("Did demonstration hospitals with initially lower performance improve more than those with initially moderate performance during phase 2?") by testing whether differences in improvement between hospitals participating and not participating in the demonstration were larger in phase 2 for the low performers than the moderate performers.

Both demonstration and matched comparison hospitals generally improved each year. In addition, some hospitals eventually reached an im-

provement “ceiling”: Other research^{8,21} has shown that hospitals with higher initial quality show less absolute improvement as a result of ceiling effects, whereas those with lower initial quality show more absolute improvement.

We accounted for these issues mathematically by controlling for the previous year’s performance score in our regression models. The resulting measure, “average adjusted annual quality improvement,” which took into account the ceiling effect and differences in starting point, formed the basis for our comparisons between groups of hospitals across time.

POTENTIAL FINANCIAL GAINS FROM QUALITY IMPROVEMENT

We explored whether our findings could be explained by different financial returns on quality improvement experienced by hospitals with different initial quality. To do this, we first estimated the average payout per discharge (in dollars) for each type of incentive award for each incentivized diagnosis in the last year of phase 1 and the first year of phase 2, using data from the Premier website and the Medicare Provider Analysis and Review files. Then we used composite quality data to estimate which hospitals would receive each type of incentive award in the last year of phase 1 and the first year of phase 2 under two alternative scenarios. Our methods are described in Appendix C.²⁰

LIMITATIONS Our study has several limitations. First, we did not evaluate all of the diagnoses and procedures included in the demonstration. We excluded bypass surgery and hip and knee replacement because process quality data for these procedures are not publicly reported on Hospital Compare.

Second, although we attempted to address the problem of comparing rates of quality improvement between hospitals with different levels of quality through matching and statistical adjustment for previous quality, the results of our study might have been different if hospitals were not subject to ceiling effects.

Third, because demonstration hospitals were very different from US hospitals overall, our findings might not be generalizable to a pay-for-performance program implemented nationally.

Fourth, when estimating the financial returns to quality improvement, we assumed that the same absolute amount of quality improvement would be as easy to achieve for low-performing hospitals as for high-performing hospitals. To the extent that quality improvement is easier for low-performing hospitals, our estimates would underestimate the returns to quality improvement for these hospitals.

Study Results

Demonstration and matched comparison hospitals were similar with respect to ownership, size, location, teaching status, composite quality performance in 2004, and other characteristics (see Appendix B).²⁰

Exhibits 1–3 show the trends in composite process quality scores for demonstration and matched comparison hospitals for three incentivized diagnoses. In every case, the quality of demonstration hospitals improved more than that of matched comparison hospitals in phase 1, but the demonstration hospitals experienced a weakening of quality improvement relative to matched comparison hospitals in phase 2. Also, demonstration hospitals starting in the lowest quartile of quality showed much more quality improvement than their matched comparison hospitals in phase 1. This was not the case in phase 2.

Average adjusted annual quality improvement was greater for demonstration hospitals than for matched comparison hospitals for each diagnosis in both phases of the demonstration (Exhibit 4). However, as indicated by the overall difference-in-differences estimates, demonstration hospitals improved less in phase 2 than in phase 1, compared to comparison hospitals. This difference was significant for heart failure and pneumonia, but not for heart attack.

In addition, for each diagnosis, the difference-in-differences estimates for hospitals in the lowest initial quality quartile were not significantly different than those for hospitals in the quartile 2. In other words, there is no evidence that hospitals in the lowest initial quartile of performance responded to the change in incentives in phase 2 with greater improvement in their performance.

A sensitivity analysis measuring quality of care using the incentivized process measure for each diagnosis with the lowest quality score—thus, the least affected by the ceiling effect—found a similar slowdown in quality improvement for demonstration hospitals in phase 2. It also found no evidence that demonstration hospitals initially performing at a low level did better in phase 2 (see Appendix D).²⁰

For each diagnosis, hospitals starting the demonstration in the third quartile of quality had the most to gain from high quality improvement after the financial incentives changed for phase 2, increasing to \$41.67 per discharge for heart attack, \$37.10 for heart failure, and \$37.70 for pneumonia (Exhibit 5). Hospitals with the lowest initial quality had the least to gain financially from quality improvement. This is because, for many of these hospitals, even high quality improvement would not put their quality score

above the median of all hospitals, making them ineligible for improvement awards.

Despite the fact that hospitals with higher initial quality had the most to gain financially from high improvement between the end of phase 1 and the beginning of phase 2, Exhibit 4 shows no evidence that these hospitals improved disproportionately more in phase 2 than hospitals in the other quartiles of quality.

Discussion

This study found that quality improvement relative to that in matched comparison hospitals was significantly less for demonstration hospitals in phase 2 than in phase 1 of the intervention for heart failure and pneumonia, and also less, although not significantly, for heart attack. In addition, the change in design in phase 2—which added an incentive for quality improvement to that for quality attainment and reduced the threshold attainment scores required for incentive payments—did not have the desired effect. It did not lead hospitals whose performance had been lower to achieve greater improvement in quality than hospitals whose initial performance had been higher.

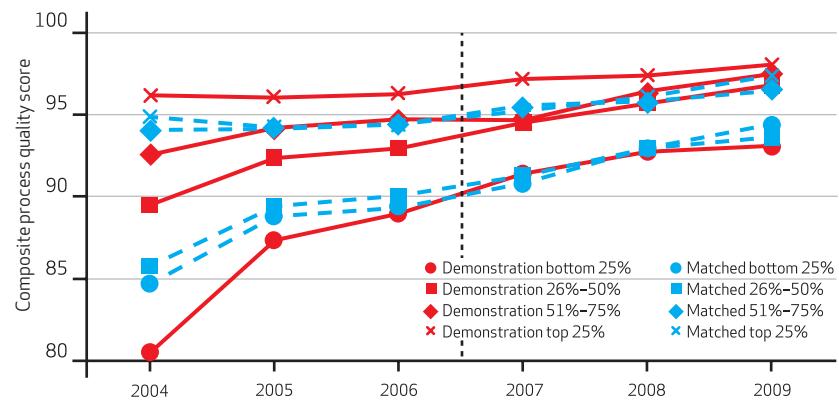
We estimate that the change in incentive design provided the strongest financial incentive not to hospitals with the lowest initial quality, but to hospitals whose initial quality was just above the median. However, these hospitals did not improve disproportionately more in phase 2 than hospitals in the other quartiles of initial quality. There are a variety of possible explanations for our findings.

REDUCED RESPONSE IN PHASE 2 Quality in demonstration hospitals improved less in phase 2 than in phase 1. One possible explanation is that hospitals approached the quality ceiling and had difficulty improving further. Both demonstration and matched comparison hospitals had increased their levels of quality by the start of phase 2, and particularly the highest-performing hospitals might have found it difficult to continue to make the same absolute gains through phase 2.

However, our analysis addressed this issue by holding previous quality constant. This allowed us to evaluate how much quality improvement occurred among demonstration and comparison hospitals if their quality had been equal in the prior year. Sensitivity analysis, using the performance measure with the lowest score, also reinforced our findings. Nonetheless, it is difficult to evaluate policy interventions in the presence of ceiling effects. Nor is it certain how demonstration hospitals would have fared in phase 2 in the absence of ceiling effects. Other researchers have

EXHIBIT 1

Demonstration And Matched Comparison Hospitals' Composite Process Quality Scores For Heart Attack, By Initial Quality Quartile, 2004-09



SOURCE Authors' analysis. **NOTE** The dashed vertical line denotes the period immediately preceding phase 2 of the Premier Hospital Quality Incentive Demonstration.

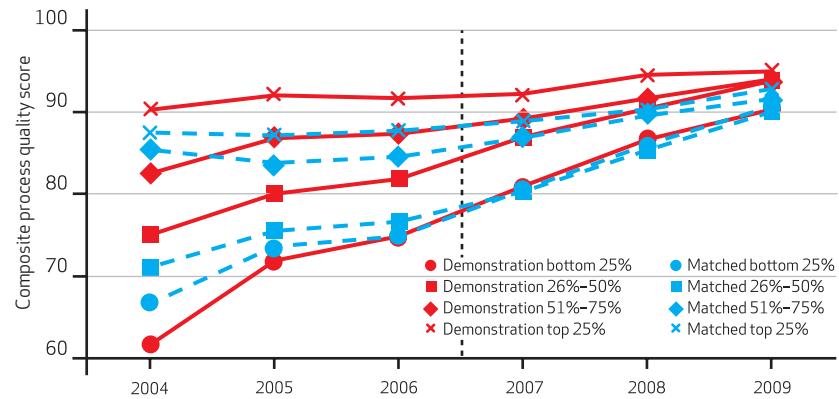
noted the difficulties associated with accurately rewarding quality attainment and evaluating improvement given the ceiling effects in the demonstration.¹⁷

Another possible explanation is that changes in the program design decreased the incentives for improvement. Although the incentives disbursed in the demonstration increased by almost 50 percent in phase 2, the change in incentive design might have been counterproductive, perhaps because of the complexity of the new design.

It is also possible that hospitals' enthusiasm for quality improvement wore off. Publicity about the demonstration as a model program

EXHIBIT 2

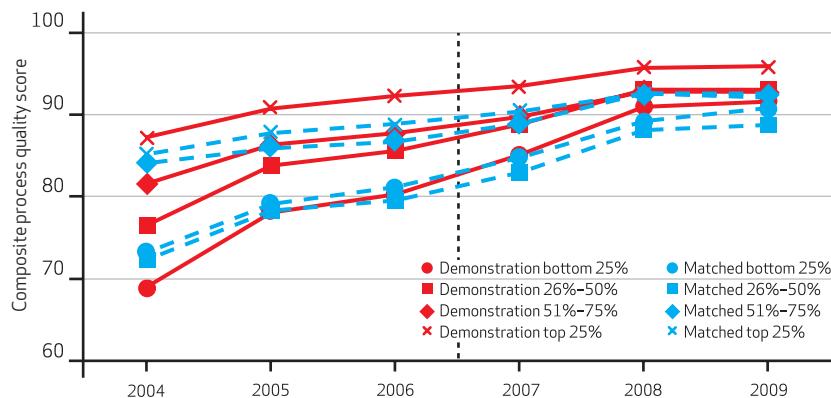
Demonstration And Matched Comparison Hospitals' Composite Process Quality Scores For Heart Failure, By Initial Quality Quartile, 2004-09



SOURCE Authors' analysis. **NOTE** The dashed vertical line denotes the period immediately preceding phase 2 of the Premier Hospital Quality Incentive Demonstration.

EXHIBIT 3

Demonstration And Matched Comparison Hospitals' Composite Process Quality Scores For Pneumonia, By Initial Quality Quartile, 2004-09



SOURCE Authors' analysis. **NOTE** The dashed vertical line denotes the period immediately preceding phase 2 of the Premier Hospital Quality Incentive Demonstration.

for health policy appeared to diminish over time, and hospitals' drive to improve quality as part of this high-profile program might have diminished as well.

LOWER PERFORMERS' FAILURE TO RESPOND
Hospitals with low initial performance on qual-

ity measures showed no response to incentives to improve. One possible explanation is that the phase 2 incentives did not succeed in targeting the lowest-performing hospitals. Despite the change in incentive design to reward quality improvement, our analysis suggests that the hospitals with the lowest initial performance in fact had the weakest incentives to improve quality between the last year of phase 1 and the first year of phase 2. This is because, even with high improvement, their quality scores would remain too low to receive improvement awards, which required performance above the median. As a result, incentives for quality improvement in phase 2 were out of reach of the lowest-performing hospitals.

It is also possible that low-performing hospitals did not have the capacity to maintain improvement at a rapid rate. After substantial improvement during phase 1, these hospitals may not have had the capacity to muster additional resources to respond to the stronger financial incentives, particularly because quality improvement became more difficult after hospitals had picked the "low-hanging fruit."

EXHIBIT 4

Average Adjusted Annual Quality Improvement For Demonstration And Matched Comparison Hospitals

Diagnosis	Phase 1		Phase 2		Difference in differences (95% CI) ^a
	Demonstration hospitals	Matched comparison hospitals	Demonstration hospitals	Matched comparison hospitals	
HEART ATTACK					
Overall	1.17	0.37	1.57	1.27	-0.50 (-1.12, 0.13)
Quartile 1	1.04	0.15	0.77	1.06	-1.19 ^b (-3.08, 0.70)
Quartile 2 (ref)	1.12	0.31	1.83	0.68	0.34 (-0.85, 1.53)
Quartile 3	1.32	0.64	1.62	1.52	-0.59 ^b (-1.43, 0.26)
Quartile 4	1.19	0.40	2.03	1.82	-0.58 ^b (-1.28, 0.11)
HEART FAILURE					
Overall	2.36	0.33	4.38	3.80	-1.45 (-2.60, -0.30)
Quartile 1	1.73	0.20	4.48	4.48	-1.53 ^b (-4.58, 1.52)
Quartile 2 (ref)	1.72	-0.20	4.94	3.73	-0.72 (-3.07, 1.64)
Quartile 3	2.88	-0.03	4.11	3.56	-2.35 ^b (-4.19, -0.50)
Quartile 4	3.11	1.34	4.01	3.48	-1.24 ^b (-2.93, 0.45)
PNEUMONIA					
Overall	3.11	1.61	3.57	2.97	-0.91 (-1.57, -0.25)
Quartile 1	2.88	1.77	3.86	3.27	-0.52 ^b (-2.06, 1.02)
Quartile 2 (ref)	3.26	1.24	3.62	2.78	-1.18 (-2.64, 0.27)
Quartile 3	2.74	1.24	3.05	3.24	-1.69 ^b (-2.98, -0.39)
Quartile 4	3.57	2.19	3.73	2.62	-0.27 ^b (-1.23, 0.68)

SOURCE Authors' analysis. **NOTES** Average adjusted annual quality improvement is calculated at the grand mean of lagged quality for all hospitals for a given diagnosis. Quartile is initial quality quartile. Quartile 1 is the lowest, and quartile 4 is the highest. Estimates control for lagged quality and its square. CI is confidence interval. ^aDifference in differences is the difference between demonstration hospitals and matched comparison hospitals in phase 2 minus the difference between demonstration hospitals and matched comparison hospitals in phase 1. ^bDifference-in-difference estimate not statistically different ($p < 0.05$) from that of the reference category (quartile 2).

Estimated Payouts Per Discharge For Demonstration Hospitals, By Amount Of Improvement

Quartile/ diagnosis	Last year of phase 1	Estimated payout per discharge (\$) ^a		
		First year of phase 2, ^b with no improvement	First year of phase 2, ^b with high improvement ^c	Difference between high and no improvement (\$)
HEART ATTACK				
Quartile 1	3.81	4.82	22.68	17.86
Quartile 2	12.02	11.62	38.81	27.19
Quartile 3	13.02	15.82	57.49	41.67
Quartile 4	40.74	28.49	66.14	37.65
HEART FAILURE				
Quartile 1	2.95	4.29	19.71	15.42
Quartile 2	5.93	10.66	37.50	26.84
Quartile 3	15.92	17.87	54.97	37.10
Quartile 4	32.82	31.31	68.02	36.71
PNEUMONIA				
Quartile 1	5.62	4.04	21.63	17.59
Quartile 2	7.81	9.72	36.68	26.96
Quartile 3	17.97	14.91	52.61	37.70
Quartile 4	67.38	35.01	68.34	33.33

SOURCE Authors' analysis. ^aEstimated payouts are based on process performance data publicly reported on the Medicare Hospital Compare website and estimates of payouts per discharge for hospitals receiving awards. Actual dollar amounts of awards received are not shown. ^bAssuming phase 2 incentives were in place. ^cHigh improvement is the minimum improvement required to receive an improvement award in the first year of phase 2. This value is equal to 2.46 percentage points for heart attack, 5.83 for heart failure, and 4.20 for pneumonia.

Implications For Value-Based Purchasing

The demonstration was intended to inform the design of Medicare's national Hospital Value-based Purchasing Program, which is scheduled to begin later in 2012. In the first year of value-based purchasing, Medicare will provide incentives for process and patient experience measures of quality, using a combination of payment for attainment and payment for improvement.¹ The design for value-based purchasing differs from phase 2 of the demonstration—notably by forgoing quality thresholds in place of a continuous points system. However, the two designs both emphasize quality attainment and improvement, and both create incentives that are achievable for most hospitals, rather than only the top performers.

For those hoping to preview the likely impact of Medicare's Hospital Value-based Purchasing Program in 2013, the experience of phase 2 of the demonstration is not encouraging. There are several reasons for this.

First, the slowdown in improvement raises questions about how to sustain the momentum

of quality improvement. Second, two of our findings—that the hospitals with the lowest initial performance had the least to gain from quality improvement at the beginning of phase 2, and that they did not improve substantially in that phase—suggest that incentives must be designed carefully if they are to stimulate improvement among lower performers.

Third, evidence that hospitals with the most to gain financially from quality improvement did not improve disproportionately more than other demonstration hospitals raises important questions about whether quality improvement in phase 1 of the demonstration was simply an artifact of unobserved selection, and whether financial incentives for quality in the Value-based Purchasing Program will be effective.

This study, along with recent evidence of the ineffectiveness of a nonvoluntary pay-for-performance program in Massachusetts,²¹ which had financial incentives much larger than those in the demonstration, raises questions about whether value-based purchasing will be able to stimulate quality improvement in hospitals. ■

NOTES

1 Centers for Medicare and Medicaid Services. Medicare program; hospital inpatient value-based purchasing program. Final rule. *Fed Regist*. 2011;76(88):26490–547.

2 Centers for Medicare and Medicaid Services. Roadmap for implementing value driven healthcare in the traditional Medicare fee-for-service program [Internet]. Baltimore (MD): CMS; [cited 2012 Feb 29]. Available from: https://www.cms.gov/QualityInitiativesGenInfo/downloads/VBPRoadmap_OEA_1-16_508.pdf

3 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 2012 Feb 29]. Available from: https://www.cms.gov/AcuteInpatientPPS/downloads/HospitalVBPPlanRTC_FINALSUBMITTED2007.pdf

4 Berwick DM, DeParle NA, Eddy DM, Ellwood PM, Enthoven AC, Halvorson GC, et al. Paying for performance: Medicare should lead. *Health Aff (Millwood)*. 2003;22(6):8–10.

5 Institute of Medicine. Rewarding provider performance: aligning incentives in Medicare. Washington (DC): National Academies Press; 2007.

6 Ryan A. Hospital-based pay-for-performance in the United States. *Health Econ*. 2009;18(10):1109–13.

7 Mehrotra A, Damberg CL, Sorbero ME, Teleki SS. Pay for performance in the hospital setting: what is the state of the evidence? *Am J Med Qual*. 2009;24(1):19–28.

8 Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007;356(5):486–96.

9 Glickman SW, Ou FS, DeLong ER, Roe MT, Lytle BL, Mulgund J, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297(21):2373–80.

10 Ryan AM. The effects of the Premier Hospital Quality Incentive Demonstration on mortality and hospital costs in Medicare. *Health Serv Res*. 2009;44(3):821–42.

11 Premier. FAQs: years 1 through 3 [Internet]. Charlotte (NC): Premier; [cited 2012 Feb 29]. Available from: <http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/faqs-year1-3.jsp>

12 Premier [Internet]. Charlotte (NC): Premier. Press release, Medicare value-based purchasing project shows any hospital can achieve positive results, regardless of size, location or patient-payer mix; 2009 Aug 6 [cited 2012 Feb 29]. Available from: <http://www.premierinc.com/about/news/09-aug/DSH-press-release080609.jsp>

13 Jha AK, Orav EJ, Epstein AM. The effect of financial incentives on hospitals that serve poor patients. *Ann Intern Med*. 2010;153(5):299–306.

14 Blustein J, Borden WB, Valentine M. Hospital performance, the local economy, and the local workforce: findings from a US national longitudinal study. *PloS Med*. 2010;7(6).

15 Epstein AM. Pay for performance at the tipping point. *N Engl J Med*. 2007;356(5):515–7.

16 Werner RM, Kolstad JT, Stuart EA, Polksky D. The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Aff (Millwood)*. 2011;30(4):690–8.

17 Damberg CL, Sorbero M, Mehrotra A, Teleki S, Lovejoy S, Bradley L. An environmental scan of pay for performance in the hospital setting: final report [Internet]. Santa Monica (CA): RAND; 2007 Nov [cited 2012 Feb 29]. Available from: <http://aspe.hhs.gov/health/reports/08/payerform/PayPerform07.html>

18 Premier. About the Hospital Quality Incentive Demonstration (HQID) [Internet]. Charlotte (NC): Premier; 2010 [cited 2012 Feb 29]. Available from: http://www.premierinc.com/p4p/hqi/year-3-results/HQID-FactsonlyPartic%20_2_.pdf

19 Premier. FAQs: Hospital Quality Incentive Demonstration (HQID) fact sheet: year 4 [Internet]. Charlotte (NC): Premier; 2010 [cited 2012 Feb 29]. Available from: <http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/faqs-year4.jsp>

20 To access the Appendix, click on the Appendix link in the box to the right of the article online.

21 Ryan AM, Blustein J. The effect of the MassHealth pay-for-performance program on quality. *Health Serv Res*. 2011;46(3):712–28.

ABOUT THE AUTHORS: ANDREW M. RYAN, JAN BLUSTEIN & LAWRENCE P. CASALINO



Andrew M. Ryan is an assistant professor of public health at Weill Cornell Medical College.

In this month's *Health Affairs*, Andrew Ryan and coauthors present their analysis of 2006 changes in the Medicare hospital pay-for-performance program, the Premier Hospital Quality Incentive Demonstration, that was intended to encourage quality improvement, particularly in lower-performing hospitals.

The authors found no evidence of success. The program's strongest incentives to improve were directed at hospitals that had already achieved quality performance ratings just above the median for the entire group of participating hospitals. Yet even those hospitals improved no more than the others. The authors' findings cast doubt that pay-for-performance can result in greater improvement among lower-performing providers.

"This study led us to question whether hospitals in this program have been responding to financial incentives at all," says Ryan, an

assistant professor of public health at Weill Cornell Medical College. He and his coauthors collaborate on research evaluating the effectiveness of pay-for-performance programs on quality of care as well as potential unintended consequences of pay-for-performance. This is their fourth paper in this field over the past few years.

Ryan is also the Walsh McDermott Scholar in the Division of Outcomes and Effectiveness Research at Weill Cornell Medical College. He received a doctorate in social policy with a concentration in health policy from the Heller School of Social Policy and Management at Brandeis University.



Jan Blustein is a professor of health policy and medicine at New York University.

Jan Blustein is a professor of health policy and medicine at the Wagner Graduate School and the School of Medicine at New York University. Her current research

examines the relationship between hospital performance measurement and health service equity. Blustein received her medical degree from the Yale School of Medicine and a doctorate in public administration from the Wagner Graduate School.



Lawrence P. Casalino is chief of the Division of Outcomes and Effectiveness Research at Weill Cornell Medical College.

Lawrence Casalino is chief of the Division of Outcomes and Effectiveness Research and the Livingston Farrand Associate Professor of Public Health in the Department of Public Health at Weill Cornell Medical College. He is the recipient of an Investigator Award in Health Policy Research from the Robert Wood Johnson Foundation. Casalino received his medical degree from the University of California, San Francisco, and both a master's degree in public health and a doctorate in health services research from the University of California, Berkeley.