SPECIAL ARTICLE

Public Reporting and Pay for Performance in Hospital Quality Improvement

Peter K. Lindenauer, M.D., M.Sc., Denise Remus, Ph.D., R.N., Sheila Roman, M.D., M.P.H., Michael B. Rothberg, M.D., M.P.H., Evan M. Benjamin, M.D., Allen Ma, Ph.D., and Dale W. Bratzler, D.O., M.P.H.

ABSTRACT

BACKGROUND

Public reporting and pay for performance are intended to accelerate improvements in hospital care, yet little is known about the benefits of these methods of providing incentives for improving care.

METHODS

We measured changes in adherence to 10 individual and 4 composite measures of quality over a period of 2 years at 613 hospitals that voluntarily reported information about the quality of care through a national public-reporting initiative, including 207 facilities that simultaneously participated in a pay-for-performance demonstration project funded by the Centers for Medicare and Medicaid Services; we then compared the pay-for-performance hospitals with the 406 hospitals with public reporting only (control hospitals). We used multivariable modeling to estimate the improvement attributable to financial incentives after adjusting for baseline performance and other hospital characteristics.

RESULTS

As compared with the control group, pay-for-performance hospitals showed greater improvement in all composite measures of quality, including measures of care for heart failure, acute myocardial infarction, and pneumonia and a composite of 10 measures. Baseline performance was inversely associated with improvement; in pay-for-performance hospitals, the improvement in the composite of all 10 measures was 16.1% for hospitals in the lowest quintile of baseline performance and 1.9% for those in the highest quintile ($P<0.001$). After adjustments were made for differences in baseline performance and other hospital characteristics, pay for performance was associated with improvements ranging from 2.6 to 4.1% over the 2-year period.

CONCLUSIONS

Hospitals engaged in both public reporting and pay for performance achieved modestly greater improvements in quality than did hospitals engaged only in public reporting. Additional research is required to determine whether different incentives would stimulate more improvement and whether the benefits of these programs outweigh their costs.

From the Division of Healthcare Quality, Baystate Medical Center, Springfield, MA (P.K.L., M.B.R., E.M.B.); the Department of Medicine, Tufts University School of Medicine, Boston (P.K.L., M.B.R., E.M.B.); Premier Healthcare Informatics, Premier, Charlotte, NC (D.R.); the Centers for Medicare and Medicaid Services, Baltimore (S.R.); and the Oklahoma Foundation for Medical Quality, Oklahoma City (A.M., D.W.B.). Address reprint requests to Dr. Lindenauer at the Division of Healthcare Quality, Baystate Medical Center, 759 Chestnut St., P-5931, Springfield, MA 01199, or at peter.lindenauer@bhs.org.

N Engl J Med 2007;356:486-96.

The need to improve both the quality and the safety of health care in the United States is well documented.[1-5] Traditional strategies to stimulate improvement include regulation, measurement of performance and subsequent feedback, and marketplace competition.[6] Despite limited evidence, public reporting of hospital quality data and pay for performance have emerged as two of the most widely advocated strategies for accelerating quality improvement.[7-11] Public reporting stimulates interest in quality on the part of physicians and hospital leaders, perhaps by appealing to their professional ethos.[12] Pay-for-performance programs are intended to strengthen the business case for quality improvement by rewarding excellence and reversing what have been described as perverse financial incentives that can deter hospitals from investing in quality-improvement efforts.[9,13,14] In enacting the Deficit Reduction Act of 2005, Congress demonstrated its support for financial incentives by calling on the Centers for Medicare and Medicaid Services (CMS) to develop a plan for hospital "value based purchasing" by 2009.[15]

Despite the instinctive appeal of pay for performance and public reporting, little is known about the individual or combined benefits of such programs,[12,16,17] and both are the subject of ongoing debate.[18-23] In order to determine the incremental effect of pay for performance, we measured improvements in hospital quality that occurred when financial incentives were combined with public reporting and compared these improvements with gains associated with public reporting alone.

METHODS

HOSPITAL QUALITY ALLIANCE

In December 2002, the American Hospital Association, the Federation of American Hospitals, and the Association of American Medical Colleges launched the Hospital Quality Alliance (HQA), a national public–private collaboration to encourage hospitals to collect and report data regarding the quality of care on a voluntary basis.[24] The HQA was designed to provide information about the quality of hospital care to the public and to "invigorate efforts to improve quality." All acute care hospitals in the United States were invited to participate, and by linking participation in the program to the annual Medicare payment update, the CMS was able to achieve participation rates of more than 98%. Participating hospitals were required to collect and report data on a minimum of 10 quality measures regarding three clinical conditions: heart failure, acute myocardial infarction, and pneumonia (Table 1). Hospitals began submitting data in the fourth quarter of 2003, and this information was made available on the Hospital Compare Web site.[25] In order to provide stable rate estimates, data from hospitals that submitted information on fewer than 25 cases

Table 1. Quality Measures Shared by the Hospital Quality Alliance and Hospital Quality Incentive Demonstration.*

Acute myocardial infarction
Percentage of patients who were given aspirin on arrival
Percentage of patients who were given an ACE inhibitor or ARB for left ventricular systolic dysfunction
Percentage of patients for whom aspirin was prescribed at discharge
Percentage of patients who were given a beta-blocker on arrival
Percentage of patients for whom a beta-blocker was prescribed at discharge
Heart failure
Percentage of patients who were assessed for left ventricular function
Percentage of patients who were given an ACE inhibitor or ARB for left ventricular systolic dysfunction
Pneumonia
Percentage of patients who were assessed for oxygenation
Percentage of patients who were given initial antibiotics within 4 hours after arrival
Percentage of patients who were assessed and given pneumococcal vaccination

* ACE denotes angiotensin-converting enzyme, and ARB angiotensin-receptor blocker.

each year for a given condition were not reported online.

HOSPITAL QUALITY INCENTIVE DEMONSTRATION

In March 2003, hospitals subscribing to a quality-benchmarking database, known as Perspective, which is maintained by Premier Healthcare Informatics, were invited to participate in the CMS–Premier Hospital Quality Incentive Demonstration (HQID), a multiyear collaborative whose goal was "to determine if economic incentives are effective at improving the quality of inpatient care."[26] Hospitals that accepted the invitation collected and submitted data on 33 quality measures regarding five clinical conditions: heart failure, acute myocardial infarction, community-acquired pneumonia, coronary-artery bypass grafting, and hip and knee replacement. This set of conditions included the previously described 10 measures reported on the Hospital Compare Web site. The remaining 23 measures are described elsewhere.[26]

Hospitals needed to have a minimum of 30 cases per condition annually to be eligible for the demonstration. For each of the clinical conditions, hospitals performing in the top decile on a composite measure of quality for a given year received a 2% bonus payment in addition to the usual Medicare reimbursement rate. Hospitals in the second decile received a 1% bonus. Bonuses averaged $71,960 per year and ranged from $914 to $847,227. These additional payments are anticipated to be partially offset by financial penalties ranging from 1 to 2% of Medicare payments for hospitals that by the end of the third year of the program had failed to exceed the performance of hospitals in the lowest two deciles, as established during the program's first year.

Of 421 hospitals that were invited to participate, 266 (63%) initially accepted, 155 declined, and 11 later withdrew. In several instances, multiple hospitals shared the same Medicare provider number. These multihospital organizations submit billing claims and clinical quality data to the CMS as a single entity and were treated as a single hospital for the purpose of our analysis. The demonstration project began in the fourth quarter of 2003 and continued through the third quarter of 2006. Informed consent and institutional review board approval were not required because the data were collected for administration of the Medicare program, not for research, and access to these data is provided to the program by law. All the authors assume full responsibility for the accuracy and completeness of the data presented.

STATISTICAL ANALYSIS

The overlapping reporting requirements between the HQA and HQID allowed us to compare improvements in quality associated with public reporting with those achieved when financial incentives are combined with public reporting. Hospitals were eligible for our analysis if they participated in the HQA program and submitted data on a minimum of 30 cases for a single condition annually, including at least 8 cases in both the fourth quarter of 2003 and the third quarter of 2005. In our primary analyses, we matched each HQID participant with as many as two HQA hospitals on the basis of the number of beds (matched to within five beds), teaching status (teaching or nonteaching), region (Northeast, Midwest, South, or West), location (urban or rural), and ownership status (not-for-profit or for-profit). These analyses focused on HQID hospitals that had participated in the program throughout the entire 2-year study period. From the pool of HQA hospitals available for matching, we excluded those that had either declined participation in the HQID or started the demonstration and then withdrew, since these decisions may have reflected doubts about whether the hospitals would be successful or other confounding factors.

We treated the matched sets as the primary units of analysis. We calculated the change in adherence to each of the HQA quality measures over a period of eight quarters for each hospital. We then calculated the difference in the improvement for each HQA quality measure for pay-for-performance hospitals, as compared with the control group. In sets with two control hospitals, the improvements at the two facilities were averaged. A paired t-test was performed to evaluate the difference in improvement between pay-for-performance and control hospitals. In addition, we calculated the percentage change in adherence to two sets of compound measures for each of the clinical conditions. First, we calculated a "composite process score" by adding up the total number of opportunities for each condition for which correct care was provided and dividing this result by the sum of the number of correct care opportunities.[27] Using this same approach, we calculated a summary composite score that com-

bined all 10 individual measures. Second, we created an "appropriate care measure" by calculating the percentage of patients who received all recommended interventions for a given clinical condition. As compared with composite process measures, appropriate care measures may better represent the interests and likely desires of patients, are more sensitive to subtle improvements, and can help foster a system perspective in quality measurement.[28]

We performed a series of stratified analyses to evaluate the effects of baseline performance, teaching status, and number of beds on the response to these incentives and compared the improvement of pay-for-performance hospitals with that of control hospitals. To estimate the incremental effect of financial incentives, multiple linear regression was applied to the matched sample. The dependent variable in these analyses was the difference in improvement between pay-for-performance hospitals and control hospitals for each matched set. We controlled for baseline hospital performance and diagnosis-specific hospital volume. The four composite process measures were used for all stratified and multivariable analyses.

To provide additional validation of our results, we conducted another multiple linear regression using the entire set of HQA participants, not only those identified through matching, with the individual hospital as the unit of analysis. In this regression, the dependent variable was the improvement over the 2-year study period, and we adjusted for baseline hospital performance, diagnosis-specific hospital volume, and all other available hospital characteristics.

To evaluate the potential contribution of a "volunteer bias" among the pay-for-performance group, we repeated our multivariable analysis by grouping hospitals that had either declined participation in the HQID or had withdrawn, together with those that had accepted and had completed the 2 years. Finally, we repeated our multiple linear regression using the entire set of HQA participants, with the hospital as the unit of analysis, and added an interaction term to explore whether the effect of pay for performance varied across quintiles of baseline performance. All analyses were carried out with the use of SAS software, version 9.1 (SAS Institute). P values of less than 0.05 were considered to indicate statistical significance.

RESULTS

Of the 4691 hospitals that submitted data for the HQA between the fourth quarter of 2003 and the third quarter of 2005, 2490 met our enrollment criteria, including 266 participants in the HQID. Eleven hospitals withdrew from the HQID during the first 2 years, leaving 255 pay-for-performance hospitals eligible for our primary analysis. We successfully matched 207 of these 255 HQID hospitals with 406 HQA controls, including 199 with two matches and 8 with one match. The typical hospital included in the study was a small-to-midsize, nonteaching, not-for-profit facility serving an urban population in the South (Table 2). As compared with all hospitals participating in the HQA, study hospitals were larger, less likely to have for-profit ownership, more likely to be urban, and more likely to have house staff. Hospitals that declined participation in the HQID were on average smaller, more rural, and less engaged in housestaff training (see Table 2A of the Supplementary Appendix, available with the full text of this article at www.nejm.org).

IMPROVEMENTS IN QUALITY

Over the 2-year study period, both pay-for-performance hospitals and control hospitals showed evidence of improvement in each of the individual and compound measures of performance (Table 3 and Fig. 1). Pay-for-performance hospitals showed significantly greater improvement than did control hospitals in 7 of the 10 individual measures of performance, with absolute differences in improvement ranging from 0.6% for oxygen assessment among patients with pneumonia (P=0.09) to 10.9% for vaccination among patients with pneumonia (P<0.001) (Table 3). Pay-for-performance hospitals also achieved greater improvement in all the composite process measures, with differences ranging from 4.1% for pneumonia (P<0.001) to 5.2% for heart failure (P<0.001). For each of the conditions, differences in the composite measures of performance between the two hospital groups increased throughout the 2-year study period (Fig. 1). A similar pattern was observed for the appropriate care measures (i.e., percentages of patients who received all recommended treatments for the condition), with absolute differences in changes ranging from 6.0% for heart failure (P<0.001) to 7.5% for acute myocardial infarction (P<0.001) (Table 3). Hospitals that declined to participate in the pay-for-per-

Table 2. Characteristics of Hospitals Included in the Analysis.

Characteristic	Hospitals with Pay for Performance plus Public Reporting (N=207)	Hospitals with Public Reporting Only (Control Group) (N=406)	All Hospitals with Public Reporting (N=2490)*
		number (percent)	
Number of beds			
<200	63 (30.4)	128 (31.5)	1355 (54.4)
200–400	77 (37.2)	146 (36.0)	697 (28.0)
>400	67 (32.4)	132 (32.5)	438 (17.6)
Location			
Rural	36 (17.4)	69 (17.0)	873 (35.1)
Urban	171 (82.6)	337 (83.0)	1617 (64.9)
Teaching status			
Teaching	97 (46.9)	191 (47.0)	783 (31.4)
Nonteaching	110 (53.1)	215 (53.0)	1707 (68.6)
Region			
Midwest	45 (21.7)	90 (22.2)	662 (26.6)
Northeast	28 (13.5)	54 (13.3)	417 (16.7)
South	109 (52.7)	212 (52.2)	1032 (41.4)
West	25 (12.1)	50 (12.3)	379 (15.2)
Ownership			
For-profit	4 (1.9)	7 (1.7)	461 (18.5)
Not-for-profit	203 (98.1)	399 (98.3)	2029 (81.5)

* Listed are all hospitals that submitted performance data on a minimum of 30 cases per year and 8 cases in the fourth quarter of 2003 and the third quarter of 2005. Percentages may not total 100 because of rounding.

formance demonstration improved less than did participating hospitals (Table 3A of the Supplementary Appendix).

Stratified analyses showed an inverse relationship between baseline performance and improvement in both groups of hospitals (Table 4). This factor influenced comparisons on the basis of hospital size and teaching status (Table 4A of the Supplementary Appendix). The difference in improvement between pay-for-performance hospitals and control hospitals varied with baseline performance, ranging from 1.2% for the composite measure of care for heart failure among hospitals with the highest baseline performance to 9.6% for the same measure among hospitals with the poorest baseline performance (Table 4).

After adjustment for the effects of baseline performance and for differences in baseline performance and condition-specific volumes between pay-for-performance hospitals and control hospitals, the incremental effect of financial incentives decreased to 2.6% for the composite process measure of acute myocardial infarction (P<0.001) and 4.1% for heart failure (P<0.001) (Table 5). A second multivariable analysis, which included the entire pool of HQA participants (with adjustment for baseline performance, condition-specific volume, and all hospital characteristics), yielded similar findings. A third analysis, intended to account for a volunteer effect by including hospitals that declined to participate in pay for performance, showed a persistent, albeit smaller, effect of financial incentives (Table 5). A final multivariable analysis, which accounted for the effects of baseline performance and other hospital characteristics, showed that the effect of financial incentives varied according to baseline performance for the composite measure of care for heart failure, with the largest improvements observed among hospitals with the poorest baseline performance. In contrast, for the composite measures of acute myocardial infarction and pneumonia

Table 3. Improvements in Quality over a 2-Year Period among Hospitals Engaged in Pay for Performance and Public Reporting.*

Variable	Hospitals with Pay for Performance plus Public Reporting			Hospitals with Public Reporting Only (Control Group)			Absolute Difference between Hospital Groups	P Value†
	4th Quarter, 2003	3rd Quarter, 2005	Absolute Change	4th Quarter, 2003	3rd Quarter, 2005	Absolute Change		
	no. of patients (%)		%	*no. of patients (%)*		%	*% (95% CI)*	
Acute myocardial infarction								
Aspirin on arrival	8,293 (92.2)	6,776 (96.3)	4.1	12,780 (94.9)	10,370 (95.7)	0.8	3.3 (1.0 to 5.6)	0.006
Aspirin at discharge	10,204 (93.9)	8,358 (96.1)	2.2	16,660 (92.7)	13,564 (94.0)	1.3	0.9 (−1.4 to 3.1)	0.48
ACE inhibitor for LVSD	2,016 (78.2)	1,809 (88.9)	10.7	3,129 (82.9)	2,737 (83.7)	0.8	9.9 (5.5 to 14.1)	<0.001
Beta-blocker on arrival	10,122 (89.1)	8,607 (95.9)	6.8	16,524 (90.4)	13,673 (94.4)	4.0	2.8 (0.7 to 5.0)	0.009
Beta-blocker at discharge	7,224 (87.6)	5,345 (93.5)	5.9	11,187 (89.6)	8,432 (92.7)	3.1	2.8 (0.4 to 5.2)	0.02
Heart failure								
LV assessment	17,301 (83.2)	14,933 (93.3)	10.1	25,431 (85.1)	21,953 (90.1)	5.0	5.1 (2.5 to 7.6)	<0.001
ACE inhibitor for LVSD	5,728 (77.7)	5,970 (86.3)	8.6	8,606 (77.6)	8,857 (84.2)	6.6	2.0 (−2.1 to 6.1)	0.34
Pneumonia								
Antibiotic timing	17,280 (68.2)	7,877 (79.9)	11.7	25,944 (68.9)	11,521 (76.3)	7.4	4.3 (1.8 to 6.6)	<0.001
Vaccination	9,357 (42.7)	5,515 (72.7)	30.0	13,599 (44.6)	8,009 (63.7)	19.1	10.9 (5.4 to 16.4)	<0.001
Oxygen assessment	18,523 (97.9)	9,757 (99.6)	1.7	28,077 (98.3)	14,562 (99.4)	1.1	0.6 (−0.1 to 1.2)	0.09
Appropriate care measures								
Acute myocardial infarction	13,187 (77.0)	11,193 (88.6)	11.6	21,665 (81.6)	17,912 (85.7)	4.1	7.5 (4.2 to 10.9)	<0.001
Heart failure	17,569 (76.1)	15,264 (88.3)	12.2	25,578 (77.8)	22,223 (84.0)	6.2	6.0 (3.1 to 9.0)	<0.001
Pneumonia	18,531 (48.5)	9,758 (70.6)	22.1	28,077 (50.0)	14,562 (65.0)	15.0	7.1 (3.6 to 10.6)	<0.001
Composite process scores								
Acute myocardial infarction	39,543 (88.7)	32,247 (94.8)	6.1	62,944 (91.3)	51,194 (93.1)	1.8	4.3 (2.5 to 6.1)	<0.001
Heart failure	23,656 (81.2)	21,588 (91.5)	10.3	34,829 (82.9)	31,730 (88.0)	5.1	5.2 (2.8 to 7.7)	<0.001
Pneumonia	46,339 (75.2)	23,724 (86.4)	11.2	70,363 (76.2)	35,283 (83.3)	7.1	4.1 (2.3 to 5.9)	<0.001
All 10 measures	116,613 (81.0)	96,695 (90.5)	9.5	192,381 (82.9)	165,805 (88.1)	5.2	4.3 (3.0 to 5.7)	<0.001

* ACE denotes angiotensin-converting enzyme, LVSD left ventricular systolic dysfunction, and LV left ventricular.
† Student's t-test was performed to compare the absolute difference in the change between hospitals with pay for performance plus public reporting and those with public reporting only (control hospitals).

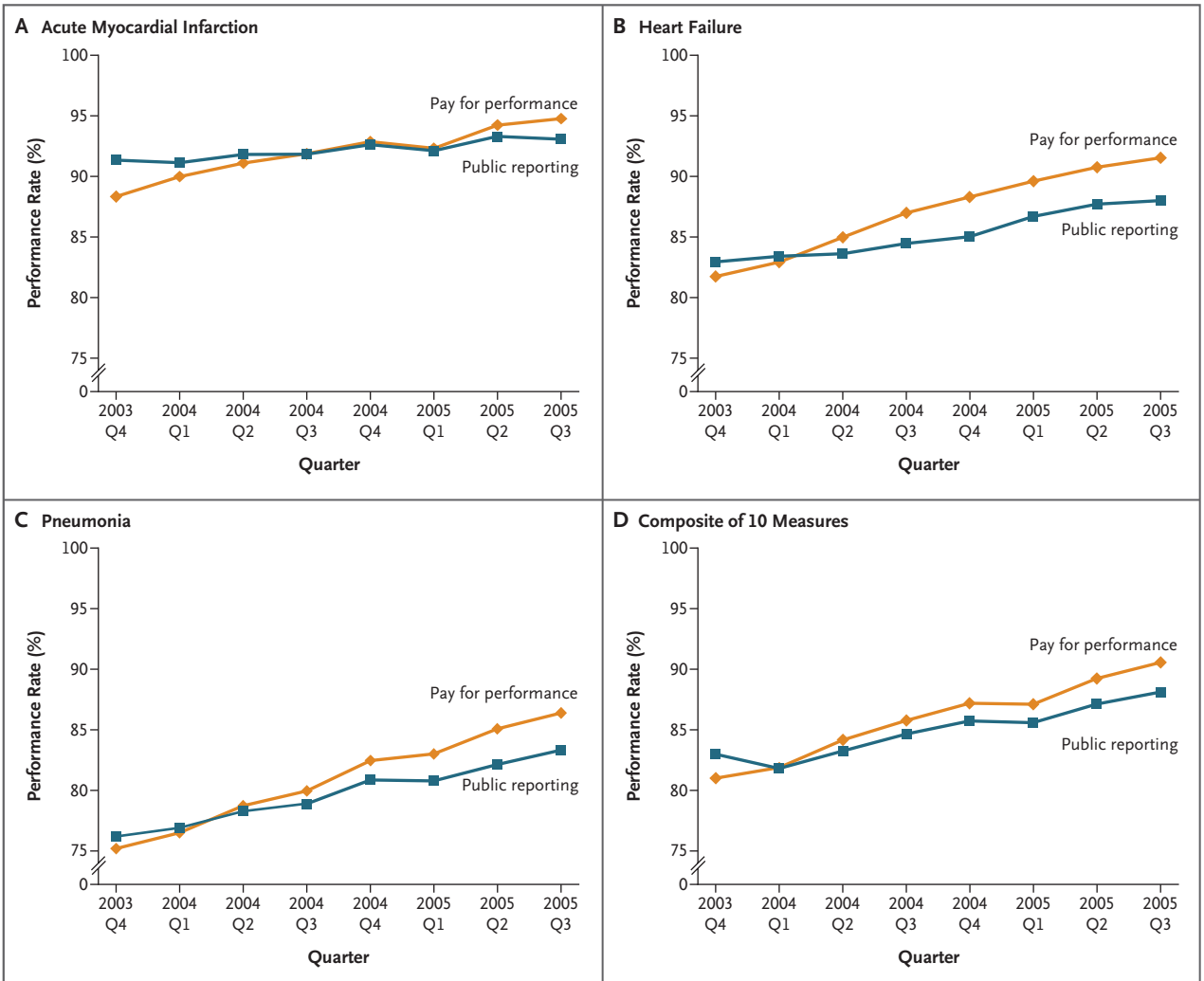

Figure 1. Improvement in Composite Process Measures among Hospitals Engaged in Both Pay for Performance and Public Reporting and Those Engaged Only in Public Reporting.

In an analysis matched for hospital characteristics, pay for performance was associated with improvements in composite process measures ranging from 4.1 to 5.2% over 2 years, including those in four key areas: acute myocardial infarction (Panel A), heart failure (Panel B), pneumonia (Panel C), and a composite of 10 measures (Panel D). The performance rate is the percentage of patients who were given the specified care for the condition. Q denotes quarter.

and all 10 measures combined, estimates of the improvement attributable to financial incentives were similar, regardless of baseline performance.

DISCUSSION

Public reporting and pay for performance are two of the most important methods that have been proposed to close persistent gaps in the quality and safety of health care. To evaluate whether combining pay for performance with public reporting results in more improvement than public reporting alone, we took advantage of a natural experiment involving several thousand hospitals engaged in a national public-reporting initiative, with more than 200 simultaneously participating in a pay-for-performance demonstration. We found that hospitals that were offered a 1 to 2% bonus for achieving high levels of performance relative to their peers had greater improvements in quality over a 2-year period than did those receiving no financial incentives. After adjustment for differences in baseline performance and other characteristics between the two groups of hospitals,

Table 4. Effect of Baseline Performance on Improvement in Quality among Hospitals Engaged in Pay for Performance and Public Reporting.

Composite Process Score	Hospitals with Pay for Performance plus Public Reporting			Hospitals with Public Reporting Only (Control Group)			Absolute Difference between Hospital Groups	P Value*
	4th Quarter, 2003	3rd Quarter, 2005	Absolute Change†	4th Quarter, 2003	3rd Quarter, 2005	Absolute Change†		
	no. of patients (%)		%	*no. of patients (%)*		%	%	
Acute myocardial infarction								
Quintile 5	5,678 (72.9)	5,151 (90.7)	17.8	9,728 (79.5)	8,141 (89.8)	10.3	7.5	0.007
Quintile 4	6,658 (86.5)	6,067 (93.2)	6.7	13,212 (89.5)	10,361 (92.1)	2.6	4.1	0.003
Quintile 3	9,596 (91.1)	7,508 (94.4)	3.3	17,055 (93.2)	13,759 (94.7)	1.5	1.8	0.05
Quintile 2	9,218 (94.0)	7,646 (96.9)	2.9	19,805 (95.8)	16,067 (94.7)	−1.1	4.0	<0.001
Quintile 1	9,934 (97.9)	7,335 (96.8)	−1.1	12,800 (98.7)	10,615 (95.3)	−3.4	2.4	0.03
Heart failure								
Quintile 5	2,412 (62.6)	2,247 (87.8)	25.2	6,038 (67.0)	6,012 (82.6)	15.6	9.6	<0.001
Quintile 4	5,202 (79.0)	4,511 (91.3)	12.3	7,902 (79.7)	6,897 (88.3)	8.6	3.7	0.02
Quintile 3	5,564 (83.6)	5,057 (93.2)	9.6	9,012 (85.2)	7,515 (88.6)	3.4	6.2	<0.001
Quintile 2	5,395 (88.1)	4,647 (91.5)	3.4	8,019 (89.3)	7,094 (89.8)	0.5	2.9	0.05
Quintile 1	6,003 (93.7)	5,943 (93.6)	−0.1	9,632 (95.2)	9,291 (93.9)	−1.3	1.2	0.31
Pneumonia								
Quintile 5	11,532 (62.8)	6,208 (82.4)	19.6	16,292 (64.4)	7,725 (77.5)	13.1	6.5	<0.001
Quintile 4	10,856 (70.3)	5,538 (84.0)	13.7	17,449 (70.7)	7,572 (80.1)	9.4	4.3	0.004
Quintile 3	10,742 (74.9)	5,403 (85.1)	10.2	17,566 (75.5)	8,765 (83.3)	7.8	2.4	0.21
Quintile 2	9,113 (79.8)	4,652 (89.1)	9.3	15,257 (80.3)	7,933 (85.2)	4.9	4.4	<0.001
Quintile 1	8,536 (87.4)	4,077 (90.8)	3.4	12,281 (87.9)	7,433 (89.0)	1.1	2.3	0.11
Composite of 10 measures								
Quintile 5	19,625 (69.7)	14,994 (85.8)	16.1	26,297 (70.5)	21,578 (83.7)	13.2	2.9	0.08
Quintile 4	21,969 (77.5)	16,555 (90.6)	13.1	39,711 (79.3)	29,466 (87.1)	7.8	5.3	<0.001
Quintile 3	23,513 (81.3)	18,504 (91.9)	10.6	45,420 (82.9)	34,886 (88.2)	5.3	5.3	<0.001
Quintile 2	26,984 (85.4)	22,316 (91.5)	6.1	45,170 (87.5)	38,093 (90.6)	3.1	3.0	0.003
Quintile 1	24,522 (91.2)	24,326 (93.1)	1.9	35,783 (94.2)	41,782 (91.2)	−3.0	5.0	<0.001

* Student's t-test was performed to compare the absolute difference in the change between hospitals with pay for performance plus public reporting and those with public reporting only (control hospitals) for each quintile.

† Analysis of variance was performed in each category to compare the change from the fourth quarter of 2003 to the third quarter of 2005 for hospitals with pay for performance plus public reporting and those with public reporting only (control hospitals). P<0.001 for all comparisons.

Table 5. Estimates of Incremental Effect of Pay for Performance.*

Analytic Approach	Incremental Effect of Pay for Performance			
	Acute Myocardial Infarction	Heart Failure	Pneumonia	Composite of 10 Measures
	% (95% CI)			
Matched for hospital characteristics	4.3 (2.5–6.1)	5.2 (2.8–7.7)	4.1 (2.3–5.9)	4.3 (3.0–5.7)
Matched for hospital characteristics and adjusted for baseline performance and condition-specific volume†	2.6 (1.3–3.9)	4.1 (2.6–5.5)	3.4 (1.9–4.9)	2.9 (1.9–3.9)
Unmatched and adjusted for baseline performance, condition-specific volume, and all hospital characteristics‡	1.9 (0.8–3.1)§	3.8 (2.1–5.5)	3.5 (2.3–4.7)	3.4 (2.3–4.5)
Unmatched and adjusted for baseline performance, condition-specific volume, and all hospital characteristics; hospitals that declined participation in pay for performance included and grouped with those that agreed to participate¶	1.8 (0.9–2.8)	2.8 (1.4–4.2)	2.7 (1.7–3.6)	2.8 (1.9–3.6)

* P<0.001 for all categories, unless otherwise noted.
† Multiple linear regression of matched pairs was adjusted for baseline performance and condition-specific hospital volume.
‡ Multiple linear regression of data for 2490 hospitals that engaged in pay for performance and public reporting was adjusted for hospital size, teaching status, region, location, ownership status, baseline performance, and condition-specific volume.
§ P=0.002.
¶ Multiple linear regression of data for 2490 hospitals that engaged in pay for performance and public reporting was adjusted for hospital size, teaching status, region, location, ownership status, baseline performance, and condition-specific volume; hospitals that declined to participate or withdrew from the Hospital Quality Incentive Demonstration were added to the pay-for-performance group to attempt to account for a volunteer effect.

the incremental effect of financial incentives was reduced, amounting to 2.6 to 4.1% over a period of 2 years.

Why are these findings important? Although the effect of the incentives was modest, our results suggest that financial incentives are capable of catalyzing quality-improvement efforts among hospitals already engaged in public reporting. And although the lion's share of bonus payments were made to hospitals with the highest baseline performance, participants across the entire spectrum responded similarly, perhaps equally motivated by the desire to avoid financial penalties.

However, before widespread application of financial incentives is considered, it should be acknowledged that pay for performance is more complex than public reporting in several ways. First, unless new money is infused into the payment system or savings are identified from improvements in quality, the size of any bonuses will need to be balanced by reductions in reimbursements across the entire system or to underperforming hospitals, creating significant concern about the possibility of harm to safety-net institutions.[23] Second, complex and politically charged judgments need to be made about fundamental system design. For example, should bonuses be paid to top-performing hospitals, to those with the greatest improvements, or to all those that meet a performance threshold? Third, the costs of administering pay-for-performance programs are likely to be higher than those for public-reporting programs. With these issues in mind, it will be important to determine not simply whether the addition of pay for performance results in more improvement than public reporting alone, but whether the benefits of such a program are worth the added cost and complexity.

Little is known about the effect of public reporting on the quality of care. In one well-documented case, rates of death after coronary bypass surgery in New York State were observed to fall after hospital-specific rates became public,[10] but the mechanisms through which this occurred have been debated. Similarly, according to a report on the QualityCounts program, run by the Employer Health Care Alliance Cooperative, hospitals with public-reporting programs engaged in more quality-improvement activities[29] and were more likely to have improved outcomes than were controls.[11]

Even less is known about the effect of pay for performance on quality or outcomes. In a recent analysis, Rosenthal et al.[30] showed that offering financial incentives to physician groups produced little gains in measures of the quality of ambulatory care and largely rewarded groups with high performance at baseline.

Our study has a number of limitations. First, it is unclear how either pay for performance or public reporting alone would have compared with no reporting, and previous studies have noted improvements over time associated with other quality-improvement efforts.[31,32] Second, hospitals that were involved in the pay-for-performance demonstration differed from the entire pool of HQA applicants, and our findings should be generalized with caution. Third, baseline performance for 5 of the 10 measures approached or exceeded 90%, thereby limiting our power to detect differences between the two groups. Fourth, although the HQID involved 33 measures spread across 5 conditions and procedures, we assessed the effect of financial incentives only on the 10 conditions that were shared by the HQA. Fifth, our attempt to adjust for volunteer bias may have underestimated the true effect of pay for performance. Sixth, the financial incentives offered to hospitals were modest, and larger bonuses might have led to more sizable improvements in quality. Finally, only the top 20% of participants were eligible for financial rewards, and interest in the program might decline over time, especially among hospitals that consistently fail to garner bonus payments. A choice to use alternative strategies, such as incentives for threshold achievement or absolute or relative improvement, might have led to different outcomes.

In conclusion, financial incentives can modestly increase improvements in quality among hospitals already engaged in public reporting. Additional research is required to determine whether larger incentives or the restructuring of payment models can stimulate more meaningful improvements and to evaluate whether the benefits of these programs outweigh their costs.

Supported by a contract (500-02-OK-03) from the CMS.

No potential conflict of interest relevant to this article was reported.

The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the Department of Health and Human Services.

We thank Robert Wachter, M.D., for his thoughtful comments on an earlier version of the manuscript, and Penelope Pekow, Ph.D., for her guidance with biostatistics.

REFERENCES

1. Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academies Press, 2001.
2. National Healthcare Quality Report, 2005. Rockville, MD: Agency for Healthcare Research and Quality, 2005. (Accessed January 8, 2007, at http://www.ahrq.gov/qual/nhqr05/nhqr05.htm.)
3. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003;348:2635-45.
4. Jha AK, Li Z, Orav EJ, Epstein AM. Care in U.S. hospitals — the Hospital Quality Alliance program. N Engl J Med 2005;353:265-74.
5. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868-74.
6. Chassin MR, Galvin RW. The urgent need to improve health care quality: Institute of Medicine National Roundtable on Health Care Quality. JAMA 1998;280:1000-5.
7. Institute of Medicine. Leadership by example: coordinating government roles in improving health care quality. Washington, DC: National Academies Press, October 2002.
8. Berwick DM, DeParle NA, Eddy DM, et al. Paying for performance: Medicare should lead. Health Aff (Millwood) 2003;22(6):8-10.
9. Millenson ML. Pay for performance: the best worst choice. Qual Saf Health Care 2004;13:323-4.
10. Chassin MR. Achieving and sustaining improved quality: lessons from New York State and cardiac surgery. Health Aff (Millwood) 2002;21(4):40-51.
11. Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. Health Aff (Millwood) 2005;24:1150-60.
12. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 2000;283:1866-74.
13. Dudley RA, Miller RH, Korenbrot TY, Luft HS. The impact of financial incentives on quality of health care. Milbank Q 1998;76:649-86.
14. Epstein AM, Lee TH, Hamel MB. Paying physicians for high-quality care. N Engl J Med 2004;350:406-10.
15. Deficit Reduction Act of 2005. S. 1932 Section 5001 Public Law No. 109-171.
16. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. Health Aff (Millwood) 2004;23:127-41.
17. Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. Strategies to support quality-based purchasing: a review of the evidence. Rockville, MD: Agency for Healthcare Research and Quality, July 2004. (AHRQ publication no. 04-0057.)
18. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA 2005;293:1239-44.
19. Schauffler HH, Mordavsky JK. Consumer reports in health care: do they make a difference? Annu Rev Public Health 2001;22:69-89.
20. Robinowitz DL, Dudley RA. Public reporting of provider performance: can its impact be made greater? Annu Rev Public Health 2006;27:517-36.
21. Krumholz HM, Rathore SS, Chen J, Wang Y, Radford MJ. Evaluation of a consumer-oriented Internet health care report card: the risk of quality ratings based on mortality data. JAMA 2002;287:1277-87.
22. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. JAMA 2004;291:2466-70.

23. Vladeck BC. Ineffective approach. Health Aff (Millwood) 2004;23:285-6.
24. Hospital Quality Alliance home page. (Accessed January 8, 2007, at http://www.aha.org/aha/key_issues/qualityalliance/index.html.)
25. Hospital Compare — a quality tool for adults, including people with Medicare. Washington, DC: Department of Health and Human Services, 2006. (Accessed January 8, 2007, at http://www.hospitalcompare.hhs.gov/.)
26. CMS/Premier Hospital Quality Incentive Demonstration. Charlotte, NC: Premier, 2006. (Accessed January 8, 2007, at http://www.premierinc.com/all/quality/hqi/.)
27. Landrum MB, Bronskill SE, Normand ST. Analytic methods for constructing cross-sectional profiles of health care providers. Health Serv Res Methodology 2000; 1:23-48.
28. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. JAMA 2006;295:1168-70.
29. Hibbard JH, Stockard J, Tusler M. Does publicizing hospital performance stimulate quality improvement efforts? Health Aff (Millwood) 2003;22(2):84-94.
30. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. JAMA 2005;294:1788-93.
31. Jencks SF, Huff ED, Cuerdon T. Change in the quality of care delivered to Medicare beneficiaries, 1998-1999 to 2000-2001. JAMA 2003;289:305-12. [Erratum, JAMA 2002;289:2649.]
32. Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. Quality of care in U.S. hospitals as reflected by standardized measures, 2002-2004. N Engl J Med 2005;353: 255-64.