DOI: 10.1111/1475-6773.12003
BEST OF THE 2012 ACADEMY HEALTH ANNUAL RESEARCH MEETING

The Impact of Hospital Pay-for-Performance on Hospital and Medicare Costs

Gregory B. Kruse, Daniel Polsky, Elizabeth A. Stuart, and Rachel M. Werner

Objective. To evaluate the effects of Medicare's hospital pay-for-performance demonstration project on hospital revenues, costs, and margins and on Medicare costs.

Data Sources/Study Setting. All health care utilization for Medicare beneficiaries hospitalized for acute myocardial infarction (AMI; ICD-9-CM code 410.x1) in fiscal years 2002–2005 from Medicare claims, containing 420,211 admissions with AMI.

Study Design. We test for changes in hospital costs and revenues and Medicare payments among 260 hospitals participating in the Medicare hospital pay-for-performance demonstration project and a group of 780 propensity-score-matched comparison hospitals. Effects were estimated using a difference-in-difference model with hospital fixed effects, testing for changes in costs among pay-for-performance hospitals above and beyond changes in comparison hospitals.

Principal Findings. We found no significant effect of pay-for-performance on hospital financials (revenues, costs, and margins) or Medicare payments (index hospitalization and 1 year after admission) for AMI patients.

Conclusions. Pay-for-performance in the CMS hospital demonstration project had minimal impact on hospital financials and Medicare payments to providers. As P4P extends to all hospitals under the Affordable Care Act, these results provide some estimates of the impact of P4P and emphasize our need for a better understanding of the financial implications of P4P on providers and payers if we want to create sustainable and effective programs to improve health care value.

Key Words. Pay-for-performance, health care costs, hospitals, Medicare

Transforming health care payment from a system that pays based on intensity (volume, duration, frequency, and type) of the services provided, to one that pays based on quality has been embraced as a way of reforming the U.S. health care system, and it has been widely adopted in the Affordable Care Act. This payment reform is commonly referred to as pay-for-performance (P4P).

Previous research on P4P has focused primarily on its impact on quality. Yet the effect of P4P on costs of care may be equally important to its adoption, sustainability, and success. Nonetheless, little attention has been given to the impact of P4P bonuses on provider and payer costs. Our objective is to test whether one P4P program in hospitals resulted in changes in costs for hospitals and for the payer, Medicare.

BACKGROUND

P4P has become an increasingly popular payment mechanism to improve quality of care in hospitals. In the private sector, more than 40 hospital P4P programs now exist (Mehrotra et al. 2009). In the public sector, the Centers for Medicare and Medicaid Services (CMS) began experimenting with P4P in hospitals in 2003 through the Premier Hospital Quality Incentive Demonstration project (CMS 2003). More recently, under the Affordable Care Act, P4P was implemented in all hospitals nationwide in October 2012 (Department of Health and Human Services and Centers for Medicare and Medicaid Services 2011).

Most research on hospital P4P has focused on the effect of the 2003 CMS P4P demonstration project on quality of care, with mixed results. Lindenauer et al. (2007) and Grossbart (2006) both found that quality improved at P4P hospitals compared with non-P4P hospitals after the demonstration project started. However, Glickman et al. (2007), studying the impact of P4P on acute myocardial infarction (AMI) outcomes, found that P4P had little effect. More recent research found that any initial quality improvements resulting from participating in P4P declined over time, suggesting no long-term impact of P4P on quality of care (Werner et al. 2011).

Quality improvement under P4P is essential for its success. However, quality improvement alone may not be enough. Given rapidly rising health care costs and the pressure to constrain costs, many stakeholders agree that P4P has the best chance of being successful if it does not reduce revenues or margins for providers and is cost neutral or cost saving for payers.

Address correspondence to Rachel M. Werner, M.D., Ph.D., Division of General Internal Medicine, Philadelphia VAMC and University of Pennsylvania, Blockley Hall, Room 1230, 423 Guardian Drive, Philadelphia, PA 19104; e-mail: rwerner@upenn.edu. Gregory B. Kruse is with The Wharton School, University of Pennsylvania, Philadelphia PA. Daniel Polsky, Ph.D., is with the Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA. Elizabeth A. Stuart is with the Department of Mental Health, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

The effect of P4P on costs depends in part on whose perspective is taken —the provider's or the payer's. For the P4P demonstration to be financially not harmful to a hospital, the hospital's costs of improving quality must either be less than any net increases in revenue from quality improvement such as through the P4P bonuses and/or any other revenues the hospital may get from improving quality, or the quality-improvement costs must be offset by a reduction in other costs, such as from preventing complications or streamlining processes.

For CMS, the P4P demonstration can be financially neutral if the cost of the bonus payments is offset by P4P-driven quality improvements that lead to fewer medical errors, hospital readmissions, or consumption of other health care resources. The cost of the P4P program can also be offset with penalties for low-performing hospitals, such as those used starting in the third year of the hospital P4P demonstration project.

Limited evidence exists on the impact of P4P on costs for hospitals and CMS. A 2008 report found that among hospitals participating in the CMS P4P demonstration, the median hospital cost per patient declined by over $1,000. The report concluded that "if all hospitals nationally were to achieve the 3-years cost and mortality improvements found among the...project participants for pneumonia, heart bypass, heart failure, heart attack (acute myocardial infarction), and hip and knee replacement patient populations, they could...reduce hospital costs by more than $4.5 billion annually" (Premier Inc. 2008). However, this analysis did not compare cost trends at P4P hospitals to a comparison group of hospitals not enrolled in P4P to account for other temporal trends, making it difficult to draw conclusions about the effect of P4P on hospital costs.

Ryan (2009) examined the effects of P4P in the CMS demonstration project on 60 days Medicare inpatient payments to hospitals and found no evidence that P4P had a significant effect on AMI, heart failure, pneumonia, or CABG costs. However, this study was limited to cost changes from Medicare's perspective, did not include all payments associated with hospitalizations (e.g., physician payments), and only examined Medicare's inpatient costs for the first 60 days after hospitalization though the effect of hospital P4P (such as reducing complications and readmissions) could extend beyond the initial hospitalization period.

We aim to fill this gap by investigating changes in the costs and revenues for hospitals participating in the CMS P4P demonstration. We examine changes in hospital revenue, costs, and margins for AMI admissions as well as changes in costs for the payer, CMS, for patients admitted for AMI to P4P and similar non-P4P hospitals.

CMS P4P Demonstration Project

CMS partnered with Premier Inc., a large provider-owned performance improvement alliance of hospitals, health systems, and providers, to implement the demonstration project. The CMS hospital P4P demonstration project was open to all 414 hospitals that were participating in the Premier Inc. hospital data system as of March 31, 2003 when enrollment opened. Participation was voluntary and 267 (64 percent) hospitals initially agreed to participate. Financial incentives began in fiscal year 2004 (starting October 2003) and were ongoing through September 2012.

The demonstration tracks hospitals' performance on measures related to the treatment of five conditions: acute myocardial infarction, heart failure, pneumonia, coronary artery bypass graft surgery, and hip and knee replacement. These measures are combined into condition-specific composite scores, which are used to determine bonus payments. Over the first 2 years of the demonstration project (fiscal years 2004 and 2005, which are the period of interest for this study), financial bonuses were distributed as add-ons to diagnosis-related group (DRG) base payments for each targeted clinical condition in hospitals with performance in the top 20 percent. CMS paid participating hospitals more than $17 million in rewards in the first 2 years of the demonstration.

Methods

Data Sources

The primary data source for this study was Medicare claims from fiscal years 2000 to 2005. We used 100 percent MedPAR files to identify AMI hospitalizations. We then tracked patient-level health care utilization over the 1 year after each AMI admission using the Standard Analytic Files containing claims for institutional outpatient providers, home health agencies, individual providers, and durable medical equipment and the 100 percent MedPAR file containing claims for all hospitalizations, skilled nursing facilities, and inpatient rehabilitation. These data were supplemented with the 100 percent Denominator File to identify HMO enrollment, patient date of birth, demographics, and death.

We supplemented Medicare claims data with hospital-level data from the annual Medicare Cost Reports (to obtain cost-center-specific cost-to-charge ratios), a variety of sources for hospital characteristics (including the Medicare Cost Reports, Impact file, and Provider of Services file), and

publicly available data on hospital performance from the CMS Website Hospital Compare for calculation of P4P bonus payments.

Empirical Approach

Our empirical approach exploits the natural experiment that occurred as a result of the policy changes initiating P4P in some hospitals in fiscal year 2004 (starting October 2003). This approach compares the changes in hospital revenues, hospital costs, and Medicare payments for hospitals subject to the policy change of P4P implementation with the changes in revenues, costs, and payments of a comparison group of hospitals not subject to P4P, from fiscal year 2002 through fiscal year 2005, spanning the initiation of P4P.

Patient Sample

We limited our sample of patient hospitalizations to those admitted with a primary diagnosis of AMI (ICD-9-CM code 410.x1). We chose to focus on the care of Medicare beneficiaries hospitalized with AMI for several important reasons. First, AMI is a common and costly condition associated with high morbidity and mortality. Thus, improving the quality of care and constraining costs in AMI should be a priority for both hospitals and Medicare. Second, AMI results in nearly universal hospitalization (Tay 2003). Third, nearly 50 percent of AMI patients arrive at the hospital by ambulance rather than choosing their hospital, and thus hospitals are less likely to be able to influence the characteristics of the AMI patients they treat (Dranove et al. 2003; Tay 2003), except through patient transfer.

Patient hospitalizations for AMI were transferred to another hospital about 12 percent of the time. We exclude all transfers, but to determine whether our results were sensitive to this decision, our sensitivity analysis ran additional models that included transfers in, out, and all transfers. We also excluded all hospitalizations discharged to hospice (less than 1 percent of patients).

We excluded hospitalizations for beneficiaries who were enrolled in Medicare managed care during any portion of the 1 year following admission for AMI because claims for managed care enrollees were not included in the data (approximately 2 percent of beneficiaries with AMI). We also excluded hospitalizations for beneficiaries with age less than 65 at the time of admission (approximately 10 percent of AMI beneficiaries), as these individuals are

enrolled in Medicare due to a disability or end-stage renal disease and excluding them creates a more homogenous patient group.

Hospital Sample

We included 260 of the 267 P4P hospitals (97 percent) that initially volunteered to participate in the demonstration. We excluded four critical-access hospitals because they receive payments based on each hospital's reported costs, rather than prospective payments. Therefore, their responses to P4P incentives might not have been typical. We excluded another three hospitals that were not listed in the datasets that we used to define variables for propensity-score matching, as described below.

As participation was limited to Premier hospitals and voluntary hospital selection into the demonstration makes finding an adequate comparison group more challenging, we used propensity-score matching to select a group of non-Premier comparison hospitals that were similar to the demonstration hospitals with respect to observed characteristics. We compared these 260 P4P hospitals to a group of comparison hospitals drawn from the pool of all acute care hospitals in the United States that did not participate in the P4P demonstration project (n = 3,335). Logistic regression was used to estimate the propensity score, defined for each hospital as the probability of being enrolled in the P4P demonstration project, where the dependent variable was an indicator of enrollment in P4P, and the independent variables were hospital characteristics in the 4 years prior to the start of the P4P project (fiscal years 2000–2003). To predict the propensity of enrollment in P4P, we used hospital characteristics known to be related to hospital quality and quality improvement: number of beds, ownership status, teaching status, accreditation by the Joint Commission, registered nurse- and licensed practical nurse-to-bed ratios, percentage of Medicare admissions, urban or rural location, the percentage of a hospital's patient days that are attributable to low-income patients, and level of market competition using the Herfindahl–Hirschman Index (defining markets as Hospital Service Area from the Dartmouth Atlas). To ensure that these two groups of hospitals were on similar cost and quality trajectories in the years prior to the initiation of P4P, we also included both the level of costs (i.e., total hospital costs for patients with AMI) and quality (i.e., average 30 days risk-standardized mortality rate for AMI) as well as the change in these two factors over the 4 years prior to the initiation of P4P. These temporal changes were calculated using a random-effects model with the cost and quality variables regressed on quarterly time dummies to calculate each hospital's intercept and slope of costs and mortality.

To form a matched comparison group from the pool of 3,335 non-P4P hospitals, we matched three comparison hospitals to each P4P hospital based on propensity scores. Other ratios of matches were also examined (e.g., 2 : 1, 4 : 1); we chose to use three comparison hospitals for each P4P hospital because it resulted in the best trade-off between finding high-quality matches for each P4P hospital and statistical power. Hospitals were matched within propensity score calipers (.25 standard deviations), combined with Mahalanobis distance matching to ensure particularly close matches on key covariates —levels and slopes of total hospital costs and risk-standardized mortality rates —and exact matching within geographic region (census division). Balance checks showed that the matching was adequate, with standardized mean differences less than .2 for all covariates and less than .1 for most (Rosenbaum and Rubin 1985; Stuart 2010) (see Table 1 for details).

The final hospital sample consisted of 260 P4P hospitals and 780 matched comparison hospitals that covered 420,211 admissions and 388,641 patients with AMI over the study period.

Dependent Variable for Hospital Perspective: Revenues, Costs, and Margins

Hospital revenues, costs, and margins were calculated for each AMI admission. Hospital revenues were defined as payments received by the hospital for the AMI hospitalization, which were obtained from four sources. From the MedPAR data we obtained three sources of hospital payment: (1) Medicare payment amount from the hospital claim, including DRG base, outlier, disproportionate share, indirect medical education, and total PPS capital payments; (2) beneficiary cost-sharing payments, including deductibles and coinsurance; (3) payments from the primary payer for beneficiaries whose primary payer is not Medicare. The fourth source of hospital payment was the P4P bonus payments from CMS. We approximate these bonus payments for each P4P hospital in our sample by reconstructing them using Medicare's Hospital Compare data (which closely approximates each hospital's performance upon which P4P bonuses are paid) and information on the bonus payments' structure from CMS and Premier. For example, based on the CMS/Premier-defined P4P bonus payment structure, if a hospital's AMI performance was in the top 10 percent of all P4P hospitals, that hospital's AMI-based bonus payment was 2 percent of the DRG base payments for AMI that hospital received in that year (where DRG base payments were calculated from MedPAR data). We distributed these hospital-level payments equally across all of a hospital's AMI admissions for that year. Thus, the entire bonus

Table 1: Summary of Baseline Characteristics and Propensity-Score Match

	P4P Hospitals (n = 260)	*All Non-P4P Hospitals (n = 3,335)*	*Matched Non-P4P Hospitals (n = 780)*	*Standardized Bias before Matching*	*Standardized Bias after Matching*
Number of beds, mean	260	167	234	0.52	0.15
Average daily census, mean	169.4	99.3	147.3	0.51	0.16
JCAHO accreditation, %	98	85	99	1.06	−0.05
Medical school affiliation, %	42.7	27.8	37.2	0.33	0.12
Resident-to-bed ratio, mean	0.07	0.05	0.06	0.09	0.09
Ownership, %					
For profit	1.9	19.7	2.2	−1.29	−0.02
Not for profit	86.5	60.4	85.0	0.76	0.05
Nurse staffing, mean					
RN FTEs	1.3	1.1	1.2	0.23	0.10
LPN FTEs	0.2	0.2	0.2	−0.15	−0.05
Medicare admissions, %	49.7	49.5	50.9	0.01	−0.09
Disproportionate share (DSH) payments, %	23.5	24.3	22.8	−0.05	0.05
Urban, %	36.5	25.2	44.0	0.16	0.02
Market concentration (HHI), mean	6,382	7,288	6,875	−0.27	−0.15
Case mix index, mean	1.42	1.29	1.39	0.58	0.12
Hospital performs open heart surgery, %	48.1	26.4	41.2	0.43	0.14
Hospital performs organ transplant, %	24.2	19.9	22.1	0.10	0.05
Total AMI admissions 2000–2003, mean	650	349	571	0.48	0.13
Total AMI patients transferred to another hospital, mean	67.6	56.4	75.1	0.15	−0.10
Total AMI patients transferred from another hospital, mean	110.9	45.9	86.7	0.24	0.09
Total AMI patients transferred to hospice, mean	2.3	1.2	1.7	0.24	0.13
Ln(cost)—intercept	9.29	9.07	9.24	0.60	0.14
Ln(cost)—slope	0.021	0.012	0.018	0.30	0.09
Mean 30 days risk-standardized mortality rate—intercept	0.160	0.178	0.163	−0.68	−0.18
Mean 30 days risk-standardized mortality rate—slope	0.000	0.001	0.000	−0.40	−0.06

Note. Variables included in propensity-score match, including summary of those variables for P4P hospitals, all non-P4P hospitals, and matched non-P4P hospitals, and standardized bias before and after matching. The standardized bias (or the difference in the mean of the two groups divided by the standard deviation of the non-P4P group) measures the difference between the two groups in standard deviation units. All characteristics for the propensity-score match are described for the 4 years prior to the implementation of P4P.

payment is captured across all patients with AMI and attributed fully to each hospital. Because the Hospital Compare data are similar, but not identical to the data used to calculate the P4P bonuses, the calculated P4P payment is a noisy approximation but should not be biased.

Hospital costs were calculated by converting MedPAR charges to costs using Medicare Cost Reports cost-center-specific charge-to-cost ratios and summing for each hospitalization. Hospital margins were calculated as the difference between hospital revenues (hospital inflows) and hospital costs (hospital outflows).

Dependent Variables for Medicare Perspective: Provider Payments

From Medicare's perspective we measure hospitalization payments, posthospitalization payments for the 1 year after hospital admission, and total 1-year payments (hospitalization and posthospitalization). The total Medicare payment for the index hospitalization has two main components: the hospital payment and the payment for physician fees during the hospitalization. The Medicare hospital payment was obtained from the MedPAR file; the physician payment was obtained from the physician claim file, including claims that occurred during the dates of hospitalization.

We calculated the Medicare posthospitalization payments as the sum of all Medicare payments between the date of discharge and 1 year following the admission to the index hospitalization. Using posthospitalization claims files, we captured Medicare payments for all institutional outpatient visits, individual provider fees (physician and nonphysician), home health care, durable medical equipment, skilled nursing facilities, inpatient rehabilitation facilities, rehospitalizations, and emergency room visits.

Medicare total 1-year payments were calculated as the sum of the Medicare payments for hospitalization and the Medicare payments for the posthospitalization services within 1 year of the AMI admission.

Model Specification

We used a difference-in-difference specification to examine the effects of P4P on hospital revenues, costs, and margins and Medicare payments. We estimated the following model using ordinary least squares and admission-level data:

$$Y_{i,j,t} = \beta_1 postP4P_t + \beta_2(P4P_j * postP4P_t) + \beta_3 X_{i,j,t} + \varphi_j + \varepsilon_{i,j,t}$$

where i indexes the individual patient, j indexes the hospital, and t indexes the fiscal year. The dependent variable, $Y_{i,j,t}$, measures revenues, costs, and margins from the hospital perspective and payments from the Medicare perspective. $Y_{i,j,t}$ is a function of a binary variable measuring whether P4P was in place (equaling 1 in 2004 and 2005 and zero otherwise), the interaction between a P4P-hospital indicator variable and the post-P4P indicator variable, a vector of individual- and hospital-level time-varying covariates ($X_{i,j,t}$), hospital fixed effects (ϕ_j), and a mean zero random error component ($\varepsilon_{i,j,t}$). Because the P4P indicator variable is hospital specific and time invariant, it is absorbed by the hospital fixed effects and drops out of the equation.

The interaction between the P4P and post-P4P indicator variables provides a difference-in-differences estimate of the effect of P4P on the dependent variable among patients in P4P hospitals after P4P was implemented, over and above the effect for patients in non-P4P hospitals at the same time. Thus, the coefficient on the interaction term, β_2, is the parameter of interest.

We test the difference-in-difference assumptions of nondiffering trends in the outcome variables in the pre-P4P period by interacting a linear time trend with a P4P indicator variable for the pre-P4P period for each of the six outcome variables. A coefficient on the interaction term that is statistically significantly different from zero would suggest that the equal trends assumption has been violated.

All dollar values were adjusted to 2007 dollars using the medical consumer price index. Patient characteristics that are known to influence quality and costs of care were included as covariates, including patient demographics (age, gender) and comorbidities defined using Elixhauser comorbidities (Elixhauser et al. 1998). Area-level characteristics such as market competition can influence hospital care so the Herfindahl–Hirschman Index (a measure of market concentration of hospital discharges) was included as a covariate, using the Dartmouth Atlas definition of Hospital Service Area as the market. All models use Huber-White standard errors to adjust for clustering within hospitals.

RESULTS

Our study included a total of 1,040 matched hospitals—260 P4P hospitals and 780 comparison hospitals. Table 1 summarizes the variables used in the propensity score to match P4P and non-P4P hospitals. Compared with the full group of non-P4P hospitals, the matched non-P4P hospitals had characteristics more similar to the P4P hospitals. However, the full group of non-P4P hospi-

tals were significantly smaller, less likely to be affiliated with a medical school, much more likely to be for profit, more likely to be rural in location, took care of fewer AMI admissions, and when they did were more likely to transfer them out. After the propensity-score matching, the two groups of hospitals were similar with respect to these characteristics. Although some differences remained, they were small.

Table 2 summarizes mean hospital financials (revenues, costs, and margins) and Medicare payments in the pre-P4P period. Hospitals were matched based in part on average hospital costs per AMI patient in the pre-P4P period. However, all hospital financial and Medicare payment variables improved after the match.

P4P bonuses are included in the hospital revenue and Medicare payment. Average bonuses over the 2-year post-P4P period were $41.32 per AMI admission at P4P hospitals (averaged across all P4P hospitals whether or not they received a P4P bonus). Among hospitals that did receive a bonus payment these bonuses were on average $118.75 per AMI admission.

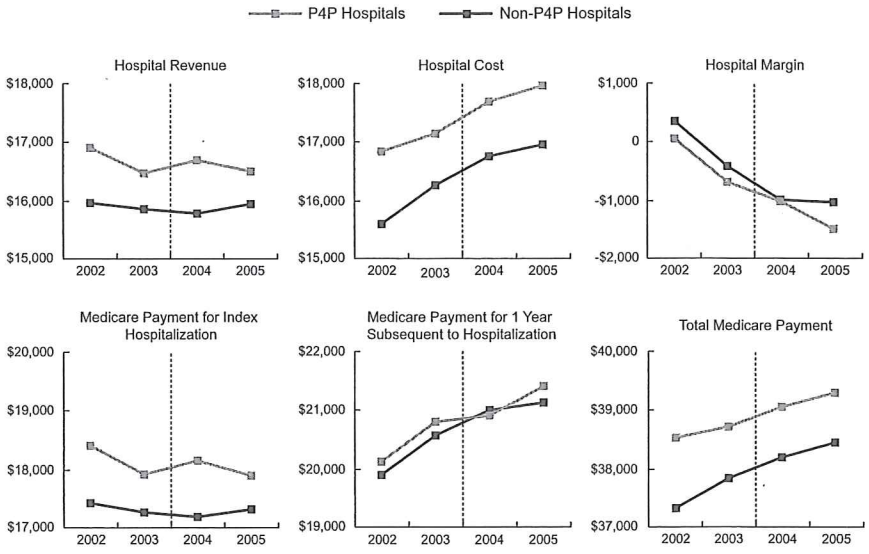
In testing the assumption of similar secular trends in the pre-P4P period for the six outcomes of interest, we found that the P4P and non-P4P trends did not significantly differ from one another for any of the six variables (see Figure 1 for descriptive results; statistical tests of trend not shown).

The difference-in-difference results are presented in Table 3. Changes in hospital financials associated with P4P were small and not statistically significant. Hospital revenues increased at P4P hospitals compared with non-P4P hospitals after the implementation of P4P, but these increases were small in absolute and relative terms and were not statistically significant—revenue increased by $29.65 per admission (or 0.2 percent) in the 2 years after the start

Table 2: Average Hospital Financials and Medicare Payments in the pre-P4P Period (2002–2003) for P4P and Non-P4P Hospitals

$ Per AMI Admission, Mean (SD)	*P4P Hospitals (n = 260)*	*Non-P4P Hospitals (n = 780)*
Hospital revenue	16,831 (16,762)	16,063 (26,677)
Hospital costs	16,982 (19,001)	15,932 (18,116)
Hospital margin	−151 (10,462)	131 (23,342)
Medicare payment for index hospitalization	17,988 (17,806)	17,161 (17,582)
Posthospitalization Medicare payment in 1 year after hospital admission	20,304 (30,299)	20,116 (29,959)
Total 1 year Medicare payments	38,292 (35,536)	37,278 (35,234)

Figure 1: Adjusted Changes over Time in Average Hospital Financials and Medicare Payments per AMI Admission at P4P and Non-P4P Hospitals



Note. The dashed vertical line represents the implementation of P4P at the beginning of fiscal year 2004.

of the P4P demonstration project compared with the 2 years prior to P4P. Similarly, the costs increased at P4P hospitals, but these changes were also small and not statistically significant. P4P hospital margins rose relative to non-P4P hospitals in the year after P4P was implemented (by $28.03 per admission or close to 5.0 percent), but these changes were statistically nonsignificant.

The difference-in-difference results for changes in Medicare payments associated with P4P were also small and not statistically significant. Medicare payment for the index hospitalization was $28.83 (0.2 percent) higher per admission to P4P hospitals compared with non-P4P hospitals in the year after P4P was implemented and the difference was not statistically significant. Medicare payments for posthospitalization care decreased by $98.27 per admission (−0.5 percent) over the 2 years after P4P was implemented, a non-statistically significant difference. We also found no statistically significant difference in total 1-year Medicare payments for patients seen at the two hospital groups.

Table 3: Difference-in-Difference Estimation of P4P on Hospital Financials and Medicare Payments

	Hospital			*Medicare*		
Variables	*Revenue*	*Costs*	*Margin*	*Payment during Hospitalization*	*1-Year Posthospitalization Payment*	*1-Year Total Payments*
P4P (omitted)						
post-P4P	13.60 (105.8)	925.7 (135.8)***	−912.1 (144.4)***	−34.62 (110.9)	722.4 (122.2)***	687.8 (170.2)***
P4P * post-P4P	29.65 (177.9)	1.63 (247.5)	28.03 (254.8)	28.83 (190.3)	−98.27 (222.3)	−69.45 (308.2)
HHI	−0.271 (0.156)*	−0.383 (0.239)	0.112 (0.228)	−0.323 (0.165)**	−0.0418 (0.237)	−0.365 (0.338)
Age	−284.8 (6.162)***	−328.8 (8.049)***	44.04 (4.944)***	−321.8 (6.739)***	−196.2 (8.808)***	−518.0 (10.03)***
Female	−1,098 (53.50)***	−689.6 (67.89)***	−408.1 (36.57)***	−1,149 (58.39)***	1,040 (112.80)***	−109.4 (118.20)
Constant	42,367 (1,068)***	46,420 (1,631)***	−4,054 (1,469)***	47,129 (1,140)***	30,832 (1,636)***	77,961 (2,196)***
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes
Patient comorbidities	Yes	Yes	Yes	Yes	Yes	Yes
Observations	420,211	420,211	420,211	420,211	420,211	420,211
R-squared	0.137	0.126	0.051	0.141	0.040	0.067
Number of hospitals	1,040	1,040	1,040	1,040	1,040	1,040
Relative change in dependent variable, %†	0.2	0.0	5.0	0.2	−0.5	−0.2

Note. Robust standard errors are in parentheses.
***$p < .01$, **$p < .05$, *$p < .1$
†The relative change is calculated as a percent change in the mean-dependent variable.

We conducted several sensitivity analyses. First, we included all transfers in and out of the hospitals. Second, we extended the post-P4P time frame out to 2007 to examine whether the effect of P4P could have been more gradual. Third, to test the sensitivity of our statistical model, we reran all models that had cost as an outcome using a generalized linear model specification with a gamma family and log link. All of the results of these analyses were consistent with the main results we present here.

DISCUSSION

P4P has become widely adopted as a method to improve health care quality and numerous prior studies have focused on the impact of P4P on quality. However, payers and policy makers have increasingly realized that for P4P to be successful and sustainable, it must be at worst cost neutral and at best cost saving. Thus, it is also important to evaluate the effect of P4P on the costs of care.

This study uses the CMS demonstration project in hospitals to examine the financial impact of P4P on both hospitals and Medicare. We examined changes in hospital financials—revenues, costs, and margins—associated with participation in the CMS P4P demonstration project and found no significant effect. A priori, the expected effect of P4P on hospital financials is ambiguous. P4P might increase hospital costs if hospitals respond to P4P by investing in organizational and structural changes to care. Prior work has found that hospitals try to improve quality by investing in information technology, reallocating or investing in human capital (more physicians and nurses), or adopting continuous monitoring and feedback systems to continually improve care (Bradley et al. 2001, 2003; Meyer et al. 2004). These costs from quality improvement may be offset by increased revenue through P4P bonus payments. It is also possible that a hospital's higher costs could be offset if quality improvement reduces complications that are costly to the hospital. Although hospitals are paid higher DRG rates by Medicare for complications, hospital managers and CEOs often view complications as costly and having a potentially negative impact on margins.

Given the spreading use of P4P in hospitals (Mehrotra et al. 2009) and the national P4P program beginning in 2012 (Department of Health and Human Services and Centers for Medicare and Medicaid Services 2011), our finding of no significant change in hospital revenue, costs, or margins is important, suggesting that hospitals' quality-improvement-investment decisions in the setting of P4P are paying off. Our results suggest that although P4P may

not represent a large revenue stream to hospitals, it might not negatively affect hospital costs or margins either.

Although we cannot directly observe investments in quality improvement, previously documented quality improvement under P4P (Lindenauer et al. 2007; Werner et al. 2011) does not appear to have a direct relationship to costs. Prior research has found a mixed relationship between cost and quality, with a positive relationship between costs and quality for congestive heart failure but a negative relationship for pneumonia (Chen et al. 2010). It is also possible that non-P4P hospitals changed their care in ways similar to P4P hospitals in anticipation of a national P4P program or in response to P4P programs funded by private insurers. This could have led to our finding of increasing costs for all hospitals and a minimal differential effect in P4P hospitals.

We also examined changes in Medicare costs and, similar to the hospital perspective, found no effect. From a payer's perspective, P4P programs often involve higher up-front payments due to the financial incentives provided to high-performing hospitals. As P4P has the potential to reduce complications and improve long-term outcomes, P4P could decrease Medicare's posthospitalization costs by either preventing hospital readmissions and other future health care utilization or lowering the costs of future visits as a result of fewer complications. These payment changes may result in lower Medicare payments in the year after hospitalization. However, this study found that P4P had no impact on Medicare payments for AMI care for the index hospitalization, in the 1 year following the hospital admission, or total 1-year payments. These findings call into question the notion that P4P will generate long-term cost offsets from fewer readmissions and lower future health care consumption.

It is interesting that there was no change in Medicare's hospital payments despite the payment of over $17 million in bonus payments over the study period. This suggests that P4P did generate offsets for Medicare. Our results suggest that after P4P was implemented, hospital-based payments to P4P hospitals increased slightly compared with before P4P and compared with the comparison hospitals, although the total increase in payments was less than the expected increase from bonuses. Our analysis does not investigate the mechanism for this, but one possibility is that these decreased payments could have occurred from decreased complications at P4P hospitals, which would have decreased DRG-based payments.

Newer hospital P4P programs are being funded through the redistribution of Medicare payments rather than supplementing existing payments with add-ons, as was the case for the CMS demonstration project. Using a redistribution of Medicare payments rather than add-ons to existing payments might

make it more likely for P4P programs to generate payment reductions for Medicare as CMS does not have higher up-front payments.

Hospital finances could also change under P4P in indirect ways, such as through reputational effects that could increase hospital volume and thus revenues, or through changes in negotiated rates with commercial insurers. We do not measure such changes on the extensive margin in this study.

Although we found no financial impact of P4P on hospital financials or Medicare payments, P4P may still increase the value of care from society's perspective. By itself, if P4P leads to higher quality with no impact on hospital margins or Medicare outlays, then it may be considered a successful program. Previous research has reported improved process-based performance in hospitals participating in P4P in the couple of years following the introduction (Lindenauer et al. 2007; Werner et al. 2011) but no change in mortality (Ryan 2009; Jha et al. 2012). Our own analyses (not reported here) found no change in 30 days risk-standardized mortality rates in P4P hospitals compared with control hospitals, suggesting that P4P had little impact on the value of care in this setting (or the cost-quality trade-off from a societal perspective).

There are limitations to this study. First, despite propensity-score matching, residual unobserved differences in hospital characteristics and patient populations may exist between hospitals that participated in P4P and those that did not. However, if hospitals that volunteered to participate in this program were more likely to achieve quality improvement and cost savings, we would expect an upward bias in our results. Our finding of no difference suggests that any residual selection bias did not significantly alter our findings. Second, we used claims data and Medicare Cost Reports to estimate the cost of hospitalization, based on the utilization of services during the hospital stay. Although this is a standard approach, it has potential limitations. We cannot directly estimate a hospital's investment in quality improvement. In addition, there are known limitations to the Medicare Cost Reports (Magnus and Smith 2000). However, this is the only method that allows for national estimates of patient-level costs of hospitalization. Nonetheless, the noise in these data prohibits us from definitively ruling out changes in hospital costs. Third, our estimates of how P4P changes the costs of care may underestimate the true effects of P4P for several reasons. Bonus payments in the CMS P4P demonstration project were small and thus may not induce hospitals to change the care they deliver. In addition, although the comparison group in this study was not subjected to P4P from Medicare, the vast majority of these hospitals did participate in Medicare's public

reporting initiative (Hospital Compare) and may have participated in P4P programs sponsored by private insurers. Finally, these results may not be generalizable due to the focus on the CMS demonstration project and restricting the sample to hospitalized AMI patients.

Pay-for-performance in the CMS hospital demonstration project had no significant effect on hospital financials and Medicare payments to providers. As P4P extends to all hospitals under the Affordable Care Act, these results provide a better understanding of how costs of care change when quality improves under P4P. Our finding of no significant changes in hospital financials or in costs for Medicare is encouraging in the face of escalating health care costs. Although our findings provide robust estimates of changes in costs under an early P4P program in hospitals, as we move forward with the expansion of P4P across more hospitals and more health care sectors, a better understanding of the financial impact of P4P to create sustainable and effective programs to improve health care value is needed.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was funded by a grant from the Agency for Healthcare Research and Quality (R01 HS018409-01). Rachel Werner was supported in part by a career development award from VA HSR&D.

Disclosures: None.

Disclaimers: None.

REFERENCES

Bradley, E. H., E. S. Holmboe, J. A. Mattera, S. A. Roumanis, M. J. Radford, and H. M. Krumholz. 2001. "A Qualitative Study of Increasing Beta-Blocker Use after Myocardial Infarction: Why Do Some Hospitals Succeed?" *Journal of the American Medical Association* 285 (20): 2604–11.

Bradley, E. H., E. S. Holmboe, Y. Wang, J. Herrin, P. D. Frederick, J. A. Mattera, S. A. Roumanis, M. J. Radford, and H. M. Krumholz. 2003. "What Are Hospitals Doing to Increase Beta-Blocker Use?" *Joint Commission Journal on Quality and Safety* 29 (8): 409–15.

Centers for Medicare and Medicaid Services. 2003. "Premier Hospital Quality Incentive Demonstration" [accessed on January 15, 2008]. Available at: http://www.cms.hhs.gov/HospitalQualityInits/35_HospitalPremier.asp

Chen, L. M., A. K. Jha, S. Guterman, A. B. Ridgway, E. J. Orav, and A. M. Epstein. 2010. "Hospital Cost of Care, Quality of Care, and Readmission Rates: Penny Wise and Pound Foolish?" *Archives of Internal Medicine* 170 (4): 340.

Department of Health and Human Services and Centers for Medicare and Medicaid Services. 2011. "Medicare Program; Hospital Inpatient Value-Based Purchasing Program; Final Rule." Available at http://www.gpo.gov/fdsys/pkg/FR-2011-05-06/pdf/2011-10568.pdf

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy* 111 (3): 555–88.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.

Glickman, S. W., F.-S. Ou, E. R. DeLong, M. T. Roe, B. L. Lytle, J. Mulgund, J. S. Rumsfeld, W. B. Gibler, E. M. Ohman, K. A. Schulman, and E. D. Peterson. 2007. "Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction." *Journal of the American Medical Association* 297 (21): 2373–80.

Grossbart, S. R. 2006. "What's the Return? Assessing the Effect of "Pay-For-Performance" Initiatives on the Quality of Care Delivery." *Medical Care Research and Review* 63 (1 suppl): 29S–48S.

Jha, A. K., K. E. Joynt, E. J. Orav, and A. M. Epstein. 2012. "The Long-Term Effect of Premier Pay for Performance on Patient Outcomes." *New England Journal of Medicine* 366 (17): 1606–15.

Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler. 2007. "Public Reporting and Pay for Performance in Hospital Quality Improvement." *New England Journal of Medicine:* 48: 6–96.

Magnus, S. A., and D. G. Smith. 2000. "Better Medicare Cost Report Data Are Needed to Help Hospitals Benchmark Costs and Performance." *Health Care Management Review* 25 (4): 65–76.

Mehrotra, A., C. L. Damberg, M. E. S. Sorbero, and S. S. Teleki. 2009. "Pay for Performance in the Hospital Setting: What Is the State of the Evidence?" *American Journal of Medical Quality* 24 (1): 19–28.

Meyer, J. A., S. Silow-Carrol, T. Kutyla, L. S. Stepnick, and L. S. Rybowski. 2004. "Hospital Quality: Ingredients for Success–Overview and Lessons Learned." The Commonwealth Fund. Available at http://www.commonwealthfund.org/programs/quality/761_Meyer_hospital_quality_overview.pdf

Premier Inc. 2008. "Quality and Cost Trend Analysis: Hospital Quality Improving, Cost, Mortality Rate Trends Declining for Participants in Medicare Pay-For-Performance Project" [accessed on September 26, 2012]. Available at: https://www.premierinc.com/about/events-education/newsletters/nexus/08-updates/P4P-II-Nexus020608.jsp

Rosenbaum, P. R., and D. B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39: 33–8.

Ryan, A. M. 2009. "Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost." *Health Services Research* 44 (3): 821–42.

Stuart, E. A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25 (1): 1.

Tay, A. 2003. "Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Differentiation." *The Rand Journal of Economics* 34 (4): 786.

Werner, R. M., J. T. Kolstad, E. A. Stuart, and D. Polsky. 2011. "The Effect of Pay-For-Performance in Hospitals: Lessons for Quality Improvement." *Health Affairs* 30 (4): 690–8.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.