The Effect of the MassHealth Hospital Pay-for-Performance Program on Quality

Andrew M. Ryan and Jan Blustein

Objective. To test the effect of Massachusetts Medicaid's (MassHealth) hospital-based pay-for-performance (P4P) program, implemented in 2008, on quality of care for pneumonia and surgical infection prevention (SIP).

Data. Hospital Compare process of care quality data from 2004 to 2009 for acute care hospitals in Massachusetts ($N = 62$) and other states ($N = 3,676$) and American Hospital Association data on hospital characteristics from 2005.

Study Design. Panel data models with hospital fixed effects and hospital-specific trends are estimated to test the effect of P4P on composite quality for pneumonia and SIP. This base model is extended to control for the completeness of measure reporting. Further sensitivity checks include estimation with propensity-score matched control hospitals, excluding hospitals in other P4P programs, varying the time period during which the program was assumed to have an effect, and testing the program effect across hospital characteristics.

Principal Findings. Estimates from our preferred specification, including hospital fixed effects, trends, and the control for measure completeness, indicate small and nonsignificant program effects for pneumonia (-0.67 percentage points, $p > .10$) and SIP (-0.12 percentage points, $p > .10$). Sensitivity checks indicate a similar pattern of findings across specifications.

Conclusions. Despite offering substantial financial incentives, the MassHealth P4P program did not improve quality in the first years of implementation.

Key Words. Econometrics, hospitals, incentives in health care, Medicaid

Pay-for-performance (P4P) in health care has been widely advanced as a means to improve the value of care. The majority of commercial HMOs now use P4P (Rosenthal et al. 2006), most state Medicaid programs use P4P (Kuhmerker and Hartman 2007), and Medicare has launched a number of P4P demonstrations (IOM 2006).

Under the Patient Protection and Affordable Care Act (PPACA 2010), hospital P4P is scheduled for nationwide implementation in 2013 as part of Medicare's Value-Based Purchasing Program. However, questions remain about the effectiveness of hospital-based P4P. As described in a recent systematic review (Mehrotra et al. 2009), only three hospital-based P4P programs have been evaluated. Much of the published research draws on experience of the Premier Quality Incentive Demonstration (PHQID), a nationwide P4P demonstration implemented jointly by Medicare and Premier Inc. in 2003. While early studies found evidence that the PHQID was effective (Grossbart 2006; Lindenauer et al. 2007), subsequent analysis has found limited evidence of an effect of the PQHID on process quality (Glickman et al. 2007) and casts doubt on the effect of the PHQID on quality and cost outcomes (Glickman et al. 2007; Bhattacharya et al. 2009; Ryan 2009a).

This study tests the early impact of the hospital P4P program recently implemented by the Massachusetts Medicaid program (also known as MassHealth).

The MassHealth Hospital P4P Program

The MassHealth P4P program was implemented in calendar year 2008, beginning with P4P for pneumonia and pay-for-reporting for surgical infection prevention (SIP) and transitioning to P4P for both conditions in 2009. The program measures and incentivizes hospital quality for a subset of MassHealth patients who are enrolled in plans that directly bill MassHealth: this is a population entirely under the age of 65 and skewed heavily toward mothers, newborns, and children. Hospitals receive incentive payments for quality delivered to this subset of patients, and they submit measure data through an electronic portal administered by MassHealth.

For pneumonia, performance was incentivized for the following measures in 2008 and 2009: oxygenation assessment, blood culture performed in emergency department before first antibiotic received in hospital, adult smoking cessation advice and counseling, initial antibiotic received within 6 hours of arrival, and appropriate antibiotic selection in immunocompetent patients. For SIP, performance was incentivized for the following measures in 2009:

Address correspondence to Andrew M. Ryan, M.A., Ph.D., Department of Public Health, Division of Outcomes and Effectiveness Research, Weill Medical College, Cornell University, 402 East 67th Street, LA-215 New York, NY 10065; e-mail: amr2015@med.cornell.edu. Jan Blustein, M.D., Ph.D., is with the Wagner Graduate School, New York University, and the Division of General Internal Medicine, NYU Medical School, New York, NY.

prophylactic antibiotic within 1 hour of surgical incision, appropriate antibiotic selection for surgical prophylaxis, and prophylactic antibiotic discontinued within 24 hours after surgery end time. All of the pneumonia and SIP measures incentivized in the MassHealth program are also publicly reported for U.S. hospitals as part of Hospital Compare, Medicare's public quality reporting program. Starting in 2004, Medicare made acute care hospitals' annual payment update conditional on reporting quality of care for 10 measures related to AMI, heart failure, and pneumonia. Thus, the MassHealth P4P program was overlaid on Medicare's existing public reporting program.

Under the MassHealth program, hospitals are rewarded based on composite measures calculated separately for each condition using the opportunities model, and, following a design proposed for Medicare's Value-Based Purchasing program (CMS 2007), quality is scored through a combination of quality attainment and improvement for clinical process scores. For each condition, payments are calculated as the product of a hospital's quality score, its number of eligible opportunities, and a predetermined dollar amount, which varies across incentivized conditions. The program's financial incentives are sizable: in 2008, up to U.S.$4.5 million in incentives were available statewide for pneumonia quality, with U.S.$2.6 million ultimately disbursed in payments to hospitals, averaging approximately U.S.$40,000 per hospital (25th percentile = U.S.$10,942; 50th percentile = U.S.$27,356; 75th percentile = U.S.$57,447).[1]

The two conditions studied here were incentivized early in the program. P4P was extended to heart attack, heart failure, and maternal and neonatal care in 2010, but these diagnoses are not evaluated because postimplementation data are not yet available.

METHODS

Data

We use 2004–2009 data on all-payer hospital process of care performance from Medicare's Hospital Compare program and data on hospital characteristics from Hospital Compare and the 2005 American Hospital Association Annual Survey. We exclude Critical Access Hospitals (CAHs), small, rural hospitals, from our analysis because CAHs are not obligated to report quality data for Hospital Compare. We also exclude hospital observation years with fewer than 25 opportunities in a given year and exclude hospitals with less

than 3 years of data, given that 3 years of data are required to estimate hospital-specific quadratic trends.

Identification Strategy

We do not observe the quality of care provided to *Medicaid* patients in Massachusetts and other states, and instead we observe the quality provided to patients from all payers. Our identification strategy assumes that the financial incentives of the MassHealth program, which are based on quality performance for only a subset of MassHealth patients, are reflected in the quality of care received by *all* patients. We believe this to be a reasonable assumption because (1) many, if not all, of hospitals' responses to MassHealth's financial incentives would likely entail improvements that would affect all patients (e.g., augmenting electronic health records, hiring quality improvement specialists); (2) it is likely not feasible for hospitals to specifically target MassHealth patients for quality improvement activities (i.e., staff are unlikely to be aware of payer status); and (3) even if feasible, clinicians would likely find such targeting to be unethical.

To estimate the effect of the MassHealth program on quality, it is critical to account for secular trends in quality improvement. A well-documented observation in P4P programs is that providers with lower initial quality tend to improve more (Rosenthal et al. 2005; Lindenauer et al. 2007). Standard difference-in-differences specifications result in biased estimates if treatment and control groups are on different outcome trajectories before the intervention begins (Abadie 2005). If we do not account for differential trends, Massachusetts hospitals, as a result of their higher initial quality, may appear to improve less than other states.

To account for heterogeneous levels and secular trends in quality for Massachusetts hospitals and other U.S. hospitals, our specification includes hospital fixed effects and hospital-specific time trends. For hospital j at time t we estimate

$$Quality_{jt} = b_0 + b_1 P4P_{jt} + b_2 u_j + b_3 (u_j \cdot Time_t) + e_{jt} \qquad (1)$$

where *Quality* is composite process quality, *P4P* is equal to 1 for hospitals in Massachusetts after the program was implemented and 0 otherwise, u is a vector of hospital fixed effects, and *Time* is a quadratic time trend (including a squared term). *Quality* is defined as a composite of the condition-specific process measures incentivized in the program. The composite is created for each condition using the "opportunities model" (Landrum, Bronskill, and Normand 2000), which is the sum of successfully achieved processes across

incentivized measures divided by the number of patients eligible to receive these processes for a given condition, for a given hospital, in a given year. Note that hospital fixed effects are interacted with the time trends, creating hospital-specific trends. We specify quadratic time trends to account for the expected attenuation of quality improvement trends as hospitals approach the upper bound of quality performance (see Figure 1). A positive coefficient on b_1 would indicate that the MassHealth program improved process quality. Equation 1 is estimated separately for pneumonia and SIP.

In public reporting under Hospital Compare, hospitals have discretion to exclude some patients as counting toward quality measures. This creates the potential for hospitals to selectively exclude patients in order to achieve higher scores (Ryan et al. 2009). If hospitals report performance for a higher pro-portion of measures that are easier to achieve, performance may be artificially inflated. Because Massachusetts hospitals may face different reporting incen-tives than hospitals in the nation at-large as a result of P4P, we create a *Difficulty index* for each hospital in each year, which is equal to the hospitals' expected performance on the process composite had they performed at the mean for each measure that they reported: a higher difficulty index indicates that the hospital's denominator consisted of *less* difficult to achieve measures.[2] We modify our base specification, including this control

$$Quality_{jt} = b_0 + b_1 P4P_{jt} + b_2 u_j + b_3(u_j \cdot Time_t)$$
$$+ b_4 Difficulty\ index_{jt} + e_{jt} \tag{2}$$

Sensitivity Checks

We estimate alternative specifications, including and excluding hospital-specific trends and the Difficulty index to examine the sensitivity of our results. We also perform sensitivity checks by excluding hospitals that participated in the PHQID, given that hospitals in this intervention faced similar financial incentives for quality improvement. Also, while the MassHealth began in 2008 for pneumonia and 2009 for SIP, in anticipation of the onset of the program, hospitals may have begun quality improvement before program implementation, particularly for SIP, which was subject to data validation in 2008. To test the sensitivity of our results to potential anticipation effects, we reestimate all models, assuming that the intervention began a year earlier for each condition. Further, we test the sensitivity of the results to estimation using linear regression models and generalized estimating equations (GEEs). While GEEs may better account for the bounding of quality scores between 0 and

Figure 1:   Process Quality in MA and Other States for Pneumonia and SIP, 2004–2009
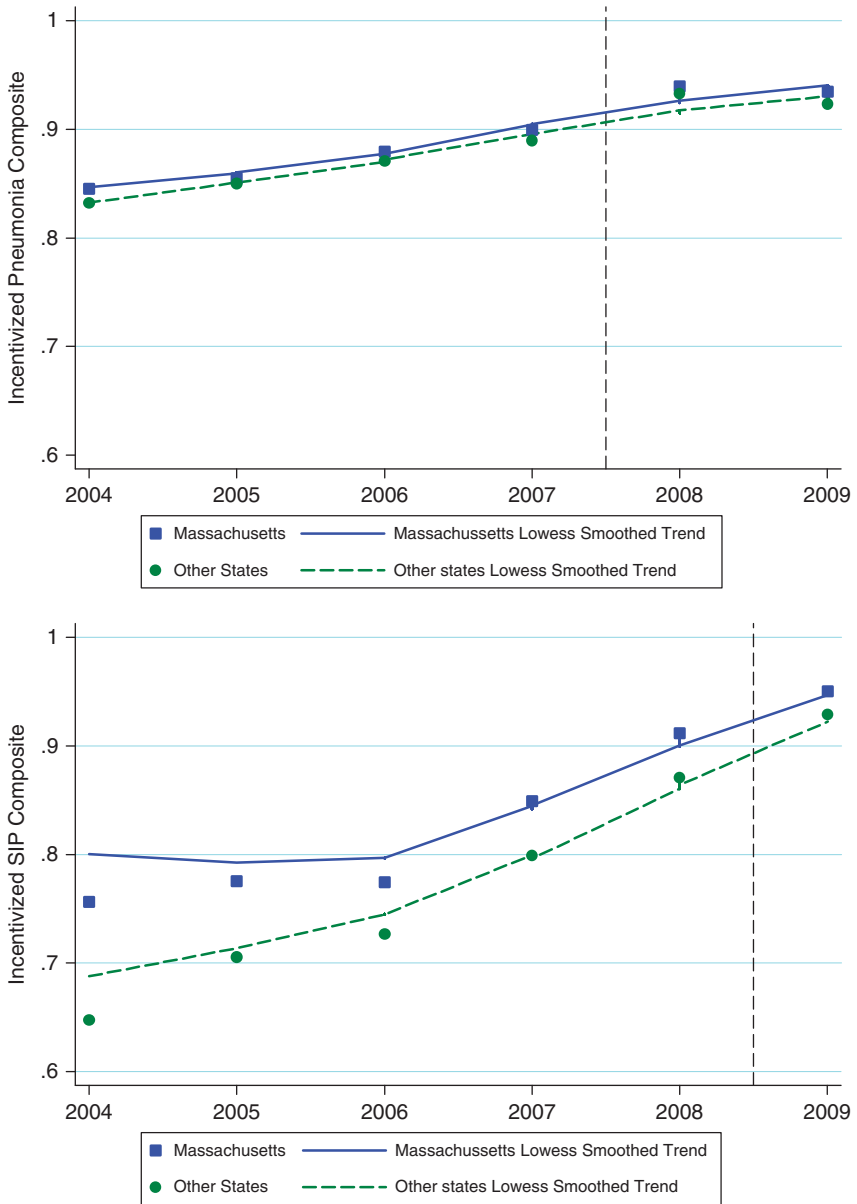
Note. Vertical line denotes period immediately preceding pay-for-performance.

100, and thus may address the attenuation of quality improvement trends as hospitals approach maximum quality scores, we prefer linear models with fixed effects given what we view to be the unreasonable assumption required for GEE estimation that model variables are uncorrelated with unobserved effects (Wooldridge 2002, p. 486). Nevertheless, we estimate models using both linear regression and GEE to examine the sensitivity of our results.

Out of concern that the substantial differences in hospital characteristics between Massachusetts and other states (see Table 1) may give rise to differential changes in quality that are not captured by our trend models, we use propensity score, nearest neighbor, one-to-one matching without replacement to create a sample of Massachusetts hospitals matched with other U.S. hospitals. We match Massachusetts hospitals to other hospitals on ownership, number of beds, urbanicity, teaching status, proportion of discharges from

Table 1:   Characteristics of Hospital Cohorts

	Massachusetts Hospitals	*Other State Hospitals*
N	62	3,676
Ownership		
Government run	0.05	0.20
For-profit	0.02	0.20
Not-for-profit	0.94	0.61
Number of beds		
1–99	0.23	0.33
100–399	0.65	0.60
400+	0.12	0.07
Urban	0.88	0.68
Medical school affiliation	0.77	0.29
Member of Council of Teaching Hospitals	0.22	0.08
Proportion of total discharges from Medicaid patients, (mean)	0.13	0.17
Proportion of total discharges from Medicare patients (mean)	0.49	0.45
Composite process performance, mean (between-hospital SD, within-hospital SD)		
Pneumonia	89.2 (2.8, 4.3)	88.4 (4.9, 5.0)
Surgical infection prevention	86.3 (6.2, 8.8)	81.1 (10.4, 11.0)
Difficulty index, mean (between-hospital SD, within-hospital SD)		
Pneumonia	88.1 (0.8, 3.3)	88.1 (1.0, 3.2)
Surgical infection prevention	81.1 (2.2, 7.2)	80.7 (3.1, 7.3)

Note. Table includes data from hospitals in at least one model.

Data on hospital characteristics come from the 2005 American Hospital Association Annual Survey.

Medicare patients, proportion of discharges from Medicaid patients, and pre-intervention mean levels of process performance, and estimate the previously described specifications in the matched sample.

We also estimate an alternative specification that tests the effect of MassHealth on hospitals' trend in quality improvement. This specification tests whether Massachusetts hospitals experienced a greater rate of quality improvement after the program was implemented and is estimated only for pneumonia, which has 2 years of postintervention data.

Because hospitals' payouts in the MassHealth program are proportional to the number of eligible Medicaid beneficiaries, hospitals with larger Medicaid populations may be more motivated to improve quality. To examine this issue, we test whether the effect of the program varies over hospitals' Medicaid share by interacting the indicator of program participation with a vector of hospital characteristics (X):

$$Quality_{jt} = b_0 + b_1\, P4P_{jt} + \alpha_1(P4P_{jt} \cdot X_j) + b_2\, u_j + b_3(u_j \cdot Time_t)$$
$$+ b_4\, Difficulty\ index_{jt} + e_{jt} \tag{3}$$

where X is a vector of hospital characteristics, including ownership, number of beds, urbanicity, teaching status, proportion of discharges from Medicare patients, and proportion of discharges from Medicaid patients. Additional sensitivity analysis reestimates the previously described specifications in the bottom and top quartiles, nationally, of Medicaid share: greater improvement among hospitals with a higher Medicaid share would suggest stronger program effects for hospitals with greater financial incentives.

Standard errors are cluster robust at the hospital level. Analysis is performed using *Stata* 11.0.

RESULTS

Table 1 shows the characteristics of hospitals in Massachusetts and other states. Massachusetts hospitals are more likely to be not-for-profit (94 versus 61 percent), more likely to have over 400 beds (12 versus 7 percent), more likely to have a medical school affiliation (77 versus 29 percent) and be a member of the Council of Teaching Hospitals (22 versus 8 percent), more likely to be located in urban areas (88 versus 68 percent), have a lower average proportion of discharges from Medicaid patients (13 versus 17 percent) and have higher average composite process performance for pneumonia and SIP.

Figure 1 shows the trends in process performance from 2004 to 2009 for the incentivized pneumonia and SIP composite measures. It shows that, for pneumonia, Massachusetts and other state hospitals had similar initial quality, which improved at approximately the same rate over time. For SIP, Massachusetts hospitals had composite quality that was approximately 12 percentage points higher in 2004 but had narrowed to near equivalence with other states by 2009.

Table 2 shows the regression estimates of the effect of the MassHealth program on process quality for pneumonia and SIP. The columns to the left describe the alternative specifications while the columns to the right show the point estimates, standard errors, and sample sizes from the models. Table 2 reinforces the inference from Figure 1 that MassHealth's P4P program had little effect on quality. For pneumonia, estimates of the effect of the program range from 3.42 percentage points ($p < .01$) in the specification including fixed effects and hospital trends, to -0.31 ($p > .10$) in the specification including fixed effects and the Difficulty index, to -0.67 ($p > .10$) in the specification including fixed effects, trends, and the Difficulty index, our preferred specification. The change in inference associated with the inclusion of the Difficulty index is a result of Massachusetts hospitals reporting quality for a higher proportion of easier to achieve measures in the postintervention period. For SIP, estimates of the MassHealth effect are negative in all but one specification, ranging from -3.45 ($p < .01$) percentage points in the specification including only fixed effects, to -1.89 percentage points ($p < .05$) in the model including fixed effects and hospital trends. The effect of MassHealth on SIP performance is -0.12 percentage points ($p > .10$) in our preferred specification. For both pneumonia and SIP, estimation using GEE with the logit link function results in null inference.

Sensitivity checks (see Appendix SA2) indicate that the exclusion of PHQID hospitals results in nearly identical estimates and that the assumption that the effects of the program began 1 year earlier than the actual implementation does not affect inference. Estimation using a matched sample shows a small, negative effect for both pneumonia and SIP. The specification testing whether the program impacted the trend in quality improvement for pneumonia finds no evidence that Massachusetts hospitals improved quality at a greater rate after the commencement of P4P. Further, the effect of P4P is not found to vary across teaching status, number of beds, ownership status, urbanicity, Medicaid share, and Medicare share for pneumonia or SIP, suggesting that hospitals with a higher proportion of Medicaid patients did not show greater quality improvement. Analysis evaluating program effects

Table 2: Estimates of the Effect of the MassHealth P4P Program on Process Quality

	Fixed Effects	Difficulty Index	Hospital-Specific Trends	GEE	Pneumonia			Surgical Infection Prevention		
					MassHealth Effect	N	N Hospitals	MassHealth Effect	N	N Hospitals
Model 1	X				0.70* (0.42)	19,627	3,385	− 3.45*** (0.87)	13,804	3,213
Model 2	X	X			− 0.31 (0.43)	19,627	3,385	− 2.92*** (0.86)	13,804	3,213
Model 3	X	X	X		3.42*** (0.05)	19,569	3,356	− 1.89** (0.75)	13,678	3,150
Model 4	X	X	X		− 0.67 (0.51)	19,569	3,356	− 0.12 (0.77)	13,678	3,150
Model 5		X		X	0.45 (0.54)	19,252	3,303	0.15 (1.04)	13,552	3,172

Notes. Standard errors are displayed in parentheses. Models 1–4 show cluster robust (at the hospital level) standard errors while Model 5 shows standard errors that are robust to arbitrary heteroskedasticity.

GEE models include hospital controls (ownership, size, urbanicity, teaching status, Medicare share, Medicaid share) that are excluded from other specifications given that the controls are time invariant. Also, in GEE models, hospital-specific trends are not specified, although controls are interacted with the quadratic time trends.

For GEE models, average marginal effects are presented.

Trends are quadratic hospital-specific time trends.

GEE, generalized estimating equation; P4P, pay for performance.

***p<.01,

**p<.05,

*p<.10.

among hospitals in the bottom and top quartiles of Medicaid share similarly found no evidence that Medicaid share moderates the effect of the program.

DISCUSSION

This is the first study to evaluate the effect of the MassHealth hospital-based P4P program. Further, in addition to the Premier Hospital Quality Incentive Demonstration, this is the only other hospital-based P4P program to be evaluated using a contemporaneous control group (Mehrotra et al. 2009). Our analysis finds no evidence that the MassHealth program improved quality of care for pneumonia or surgical infection prevention, conditions incentivized under the program. These findings are robust across a broad range of modeling specifications.

The question that emerges from our analysis is: Why didn't the program improve quality? We discuss several possible explanations.

The Financial Incentives Were Too Small

As noted previously, the MassHealth program disbursed U.S.$2.6 million in financial incentives for pneumonia process quality in 2008, amounting to an average of U.S.$40,000 per hospital. Hospitals were *eligible* to earn much more in incentives: Figure 2 shows that the 25th percentile of potential payouts was U.S.$27,000, the 75th percentile was U.S.$88,000, and the 99th percentile was U.S.$331,000. Even more generous incentives were available in 2009. Despite the magnitude of these incentives, they remained a small fraction of total hospital revenues: the ratio of potential pneumonia payouts to total hospital revenues was 0.00015 at the 25th percentile, 0.00029 at the 50th percentile, 0.00045 at the 75th percentile, and 0.00122 at the 99th percentile.[3] However, financial incentives for the Premier Hospital Quality Incentive Demonstration were approximately U.S.$6,533 per condition (pneumonia, heart failure, AMI, hip and knee replacement, and CABG) per hospital, per year from 2003 through 2006,[4] a period during which it has been suggested that the PHQID increased process quality for AMI, heart failure, and pneumonia (Lindenauer et al. 2007). Two alternative conclusions could be drawn from this contrast with the PHQID: (1) insufficient financial incentives in the MassHealth program *cannot* explain our results; (2) insufficient financial incentives *can* explain our results, and the effects observed in the PHQID were a result of selection into the program by hospitals motivated to improve quality, not an effect of financial incentives.

Figure 2:   Distribution of Pneumonia Pay-for-Performance Payouts for Which Hospitals Were Eligible, 2008
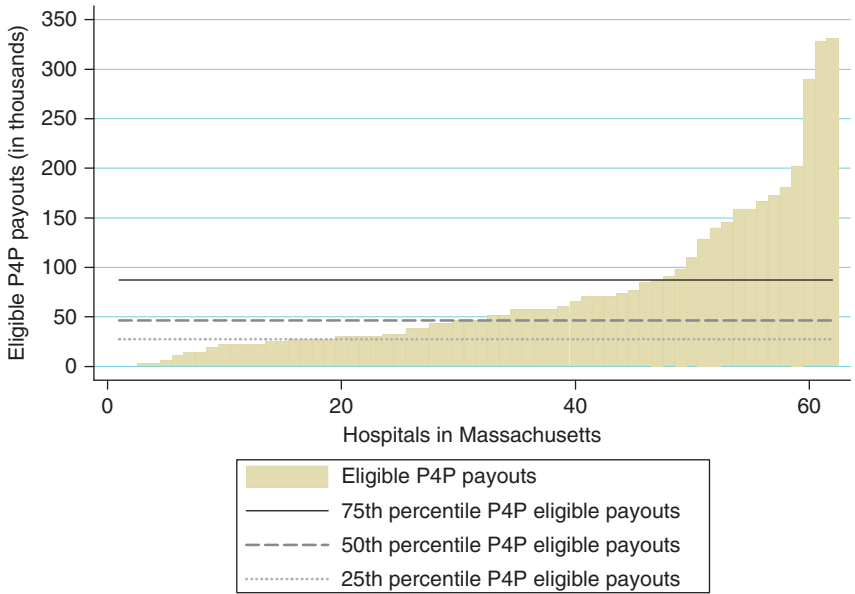
The Payout Rules Did Not Strongly Incentivize Quality Improvement

The payout rules for the MassHealth program were modeled on those proposed for Medicare's Value-Based Purchasing program (CMS 2007), whereby hospitals receive "points" for quality attainment and quality improvement, and the maximum of these point values is the quality score used to disburse incentives. In theory, this type of design should create stronger incentives for performance than designs based on quality attainment alone, such as the PHQID, especially for hospitals with lower initial quality. In practice, confusion over the payout rules among Massachusetts hospitals may have attenuated hospitals' response to incentives.

It Was Too Difficult for Massachusetts Hospitals to Improve Quality from a High Base

It is possible that hospitals experience dramatic diminishing marginal returns from quality improvement activities after achieving a high level of quality (approximately 90 percent process compliance). However, hospitals in the PHQID also started with a higher initial level of quality than non-PHQID

hospitals and experienced greater quality improvement, suggesting that high initial quality levels alone are unlikely to discourage quality improvement in response to P4P.

Hospitals Were Overwhelmed by Reforms and Other Reporting Requirements, and Did Not Attend Sufficiently to the Incentives under P4P

All U.S. hospitals report to a vast array of state, federal, peer review, and payer organizations and may have difficulty keeping current with prevailing incentives. During the period studied, Massachusetts hospitals faced an unusually complex and rapidly changing environment as a result of the implementation of the Massachusetts health reform act. This environment had already been shaped, in part, by Medicare's existing public reporting program. In addition, during this time MassHealth began to require hospitals to collect and report on newly developed measures of clinical quality for maternity care and measures of health disparities. Reporting on these new measures was incentivized under a pay-for-reporting program (MassHealth 2009). While we are unable to assess the impact of these additional requirements (due to a lack of "pre" data), it is plausible that attention to these other demands made hospitals less likely to respond to P4P incentives. In future years, as payouts continue, hospitals may grasp the magnitude of the financial incentives under the program and improve quality in response.

Several limitations of this study should be noted. First, we examined the effects of the MassHealth P4P program on all patients, not on the subset of Medicaid patients for which quality was incentivized. It is possible that hospitals focused their quality improvement efforts on these patients alone, and quality in the aggregate was minimally affected. However, we found no evidence that hospitals with a greater share of Medicaid patients showed greater quality improvement under the program, suggesting that Medicaid patients were not specifically targeted for quality improvement efforts.

Second, with only 62 hospitals in Massachusetts reporting quality for pneumonia (and 58 for SIP), our study may have been underpowered to detect small effects. However, point estimates of the effect of P4P tended to be *negative* across specifications and standard errors were small enough to detect program effects well under 2 percentage points with 95 percent confidence. Consequently, our findings of no positive program effects are not primarily a result of low power.

Third, the postimplementation period evaluated in this study was only 2 years for pneumonia and only 1 year for SIP. It is possible that it takes hospitals longer to respond to the financial incentives of P4P than could be observed in

this time window. Future studies should examine the longer term effect of the MassHealth program.

Fourth, we examined only the effect of the MassHealth program on process measures, which were the only type of measure incentivized in the program. While it is possible that the program had effects on outcomes, this seems unlikely given that outcome improvement is more difficult than process improvement and outcomes were not incentivized. Also, we would have little power to make inferences about program effects on outcomes given the small number of Massachusetts hospitals and the larger random variation in outcomes.

Fifth, although we accounted for potential hospital gaming by controlling for the mix of individual process measures that make up the composite scores, we did not account for gaming behavior, such as exception reporting, that could artificially inflate performance on individual measures. However, because the incentives of the MassHealth P4P program would be expected to increase such gaming, our results would be expected to be biased *away* from the null. Therefore, because we found no effect of the program on quality, our inference that the program did not improve quality is strengthened.

Sixth, our analysis could be confounded by time-varying factors that occurred simultaneously with the MassHealth program. An example would be P4P programs implemented in other states. We attempted to address this by excluding hospitals participating in the PHQID in sensitivity analysis, but we certainly have not excluded hospitals participating in all relevant quality incentive programs.

Implications. This study adds to the limited literature of hospital-based P4P, finding that the MassHealth P4P program has not yet improved quality. While we do not know why a program offering such generous financial incentives was unsuccessful at changing provider behavior, our work adds to a growing body of research that suggests that hospital P4P, as currently implemented, may be inadequate to improve value in health care (Ryan 2009b).

ACKNOWLEDGMENTS

Disclosures: None.
Disclaimers: None.

NOTES

1. Authors' analysis using MassHealth data on hospital payouts.
2. For a given condition with n process measures, the difficulty index is given by

$$Difficulty\ index_{it} = \sum_{m=1}^{n} \overline{Score}_{mt} \times p_{mit}$$

   where m indexes to process measures, i indexes to hospitals, and t indexes to years, $\overline{Score}$ is the national average score for a given measure in a given year, and p is the proportion of hospitals' composite denominator from the given measure in a given year.
3. Authors' analysis using MassHealth data on hospital payouts and data on hospital revenues from the Massachusetts Division of Health Care Finance and Policy report "Massachusetts Acute Care Financial Performance: Fiscal Year 2008" available at: www.mass.gov/Eeohhs2/docs/. . ./fy08_acute_hospital_financial.ppt
4. Calculation: U.S.$24.5 million/250 hospitals/5 conditions/3 years = U.S.$6,533 per hospital per condition per year (see Premier 2008).

REFERENCES

Abadie, A. 2005. "Semiparametric Difference-in-Difference Estimators." *Review of Economic Studies* 71: 1–19.

Bhattacharya, T., A. A. Freiberg, P. Mehta, J. N. Katz, and T. Ferris. 2009. "Measuring the Report Card: The Validity of Pay For-Performance Metrics in Orthopedic Surgery." *Health Affairs* 28 (2): 526–32.

Centers for Medicare and Medicaid Services (CMS). 2007. *Plan to Implement a Medicare Hospital Value-Based Purchasing Program*. Report to Congress.

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): 29S–48S.

Institute of Medicine (IOM). 2006. *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: National Academy Press.

Kuhmerker, K., and T. Hartman. 2007. "Pay-for-Performance in State Medicaid Programs: A Survey of State Medicaid Directors and Programs" [accessed on December 16, 2010]. Available at: http://www.commonwealthfund.org/Content/Publications/Fund-Reports/2007/Apr/Pay-for-Performance-in-State-

Medicaid-Programs–A-Survey-of-State-Medicaid-Directors-and-Programs. aspx.

Landrum, M. B., S. E. Bronskill, and S. T. Normand. 2000. "Analytic Methods for Constructing Cross-Sectional Profiles of Health Care Providers." *Health Services Research Methodology* 1 (1): 23–47.

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* 356: 486–96.

MassHealth Office of Acute and Ambulatory Care, Commonwealth of Massachusetts Executive Office of Health and Human Services. 2009. *RY 2008: MassHealth Hospital Pay-for-Performance Summary Report.* Provided by T. Yannetti, EOHSS.

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.

Patient Protection and Affordabable Care Act. 2010. H.R 3590. Public Law 111–148.

Premier Inc. 2008. "About the Hospital Quality Incentive Demonstration" [accessed on July 6, 2010]. Available at http://www.premierinc.com/p4p/hqi/year-3-results/HQID-FactsonlyPartic%20_2_.pdf

Rosenthal, M. B., B. E. Landon, S. T. Normand, R. G. Frank, and A. M. Epstein. 2006. "Pay for Performance in Commercial HMOs." *New England Journal of Medicine* 355: 1895–902.

Rosenthal, M. B., R. G. Frank, Z. H. Li, and A. M. Epstein. 2005. "Early Experience with Pay-for-Performance—from Concept to Practice." *Journal of the American Medical Association* 294: 1788–93.

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

———. 2009b. "Hospital-Based Pay-for-Performance in the United States." *Health Economics* 18 (10): 1109–13.

Ryan, A. M., C. Tompkins, J. Burgess, and S. Wallack. 2009. "The Relationship between Performance on Medicare's Process Quality Measures and Mortality: Evidence of Correlation, Not Causation." *Inquiry* 46 (3): 274–90.

Wooldridge, J. A. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.
Appendix SA2: Supplemental Results.

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.