Does Winning a Pay-for-Performance Bonus Improve Subsequent Quality Performance? Evidence from the Hospital Quality Incentive Demonstration

Andrew Ryan, Matthew Sutton, and Tim Doran

Objective. To test whether receiving a financial bonus for quality in the Premier Hospital Quality Incentive Demonstration (HQID) stimulated subsequent quality improvement.

Data. Hospital-level data on process-of-care quality from Hospital Compare for the treatment of acute myocardial infarction (AMI), heart failure, and pneumonia for 260 hospitals participating in the HQID from 2004 to 2006; receipt of quality bonuses in the first 3 years of HQID from the Premier Inc. website; and hospital characteristics from the 2005 American Hospital Association Annual Survey.

Study Design. Under the HQID, hospitals received a 1 percent bonus on Medicare payments for scoring between the 80th and 90th percentiles on a composite quality measure, and a 2 percent bonus for scoring at the 90th percentile or above. We used a regression discontinuity design to evaluate whether hospitals with quality scores just above these payment thresholds improved more in the subsequent year than hospitals with quality scores just below the thresholds. In alternative specifications, we examined samples of hospitals scoring within 3, 5, and 10 percentage point “bandwidths” of the thresholds. We used a Generalized Linear Model to estimate whether the relationship between quality and lagged quality was discontinuous at the lagged thresholds required for quality bonuses.

Principal Findings. There were no statistically significant associations between receipt of a bonus and subsequent quality performance, with the exception of the 2 percent bonus for AMI in 2006 using the 5 percentage point bandwidth (0.8 percentage point increase, $p < .01$), and the 1 percent bonus for pneumonia in 2005 using all bandwidths (3.7 percentage point increase using the 3 percentage point bandwidth, $p < .05$).

Conclusions. We found little evidence that hospitals’ receipt of quality bonuses was associated with subsequent improvement in performance. This raises questions about whether winning in pay-for-performance programs, such as Hospital Value-Based Purchasing, will lead to subsequent quality improvement.

Key Words. Pay-for-performance, incentive, hospitals, quality improvement

The use of financial incentives for high-quality care, known as pay-for-performance (P4P), is now widespread in the United States (Rosenthal et al. 2006; Kuhmerker and Hartman 2007; Alexander et al. 2012). Beginning in October 2012, the Centers for Medicare and Medicaid Services (CMS) rolled out Hospital Value-Based Purchasing to all acute care hospitals in the United States. Hospital Value-Based Purchasing is the largest implementation of P4P in the United States to date and as with prospective payment, it has the potential to be adopted by private payers, broadening the impact of the program.

The effects of P4P on provider behavior have been extensively studied (Petersen et al. 2006; Mehrotra et al. 2009; Van Herck et al. 2010; Flodgren et al. 2011). Studies of P4P have generally focused on whether incentives for quality induce providers to deliver higher quality in the short term, with a minority of others addressing unintended consequences of P4P. Underlying these investigations is the theory that the hope or expectation of future financial returns—in addition to reputational incentives that are often integrated into P4P programs through benchmarking or public reporting—will motivate current efforts to improve quality. However, independent of expectations for future payments (i.e., the possibility of receiving a bonus), the receipt of a financial incentive (i.e., the actual bonus), in and of itself, might generate quality improvement through reinvestment in further quality improvement activities. CMS has explicitly argued that incentive payments for quality are intended to drive additional quality improvements when justifying its P4P program for Medicare Advantage (Government Accountability Office 2012).

Recent studies have raised concerns that the national implementation of P4P might not improve quality of care. Evaluations of the Premier Hospital Quality Incentive Demonstration, the precursor to Hospital Value-based Purchasing, suggest that apparent early improvements were attributable to selection into the program rather than to financial incentives, and that the program did not generate sustained improvements in quality of care or better health outcomes (Jha et al. 2012; Ryan, Blustein, and Casalino 2012). Given that this evidence diminishes expectations as to the likely overall effect of Hospital Value-Based Purchasing, it is important to

Address correspondence to Andrew Ryan, Ph.D., Public Health Department, Weill Cornell Medical College, 402 East 67th Street, New York, NY 10065; e-mail: amr2015@med.cornell.edu. Matthew Sutton, Ph.D., is with the Centre for Health Economics, Institute of Population Health, University of Manchester, Manchester, England. Tim Doran, M.D., is with the Department of Health Sciences, University of York, York, England.

evaluate whether any component of P4P, such as the payout of bonuses, might generate future improvements for a subset of participating hospitals.

Behavioral Response to Performance Incentives in Health Care

Economic theory of hospital behavior holds that hospitals are sensitive to revenues and reputation (Newhouse 1970; Pauly 1987; Sloan 2000; Miller, Eggleston, and Zeckhauser 2006). While the specification of hospital utility functions vary with respect to whether revenue is considered part of the objective function or a constraint, standard theory indicates that utility is increasing in revenue and reputation, and decreasing in costs. Ex ante, by increasing both revenue and reputation through quality improvement activities, the incentives of P4P programs have the potential to increase hospital utility. However, quality improvement comes at some cost and hospitals must determine whether the expected benefits from bonus payments exceed the costs of quality improvement. Thus, theory suggests that hospitals determine a level of effort to dedicate to quality improvement activities which balances the expected benefits and costs of participation. Following this allocation, along with random variation, some level of quality performance is achieved, and bonuses are disbursed based on this level of performance.

Hospitals receiving bonuses will have the opportunity to hold increased reserves or to allocate the additional revenue toward achieving various aims. One possibility is that hospitals would invest some portion of the additional revenue toward quality improvement activities, which may lead to future improvements in quality of care. Hospitals that do not receive bonuses lack this opportunity for additional investment. This scenario is possible if hospitals' investment in quality of care in a prior year was not utility maximizing as a result of resource constraints, and additional revenue from winning bonuses would allow hospitals to make utility maximizing decisions. Evidence that hospitals in poorer financial condition are less likely to invest in quality of care supports this view (Bazzoli et al. 2007). It is also possible that hospitals have both formal and informal accounts for different types of expenses (e.g., labor, supplies, capital expenses, quality improvement activities) and that revenue generated through incentive payments in P4P programs would remain in the "quality improvement activities" account, contributing to the future quality improvement efforts, rather than simply contributing to general revenue.

The Current Investigation

We address the question of whether receiving a quality bonus leads to greater subsequent quality improvement in the context of the Premier Hospital Quality Incentive Demonstration (HQID). The first 3 years of the HQID provided bonus payments on the basis of whether hospitals' quality performance exceeded strict threshold values. We exploit these thresholds, using a regression discontinuity design—a powerful tool for estimating causal relationships in health policy (Almond and Doyle 2008; Card, Dobkin, and Maestas 2009)—to evaluate whether hospitals with quality scores just above payment thresholds improved quality more in subsequent periods than hospitals with quality scores just below the payment thresholds.

The Premier Hospital Quality Incentive Demonstration

The HQID was implemented in two phases. In the first phase, from Q4 2003 to Q3 2006, the program incentivized quality among 265 hospitals for acute myocardial infarction (AMI), heart failure, pneumonia, bypass surgery, and hip and knee replacement. Quality bonuses were disbursed based on relative thresholds: hospitals scoring at or above the 90th percentile on a condition-specific composite quality measure received a 2 percent bonus on condition-specific Medicare payments, and hospitals scoring at or above the 80th percentile but below the 90th percentile received a 1 percent bonus. Given that these payments were based on the distributions of performance at the end of the year, the specific threshold values required for payments were not known in advance by participating hospitals. Among hospitals receiving the 1 percent bonus, payments averaged \$30,000 for AMI (\$60 per discharge), \$20,000 for heart failure (\$46 per discharge), and \$29,000 for pneumonia (\$68 per discharge). Composite quality scores ranged from 0 to 100 and can be interpreted as the proportion of opportunities for providing recommended care that was actually provided for a given health condition. The second phase of the HQID, from Q4 2006 to Q3 2009, used a different incentive scheme which incentivized high quality attainment, moderate quality attainment, and quality improvement. For a given condition, the same hospital could receive payments for both high and moderate quality attainment, and moderate quality attainment and improvement but could not receive payments for both high quality attainment and improvement. The Phase 2 incentive scheme is not suitable to our regression discontinuity design, and therefore data from this phase were not considered in this analysis.

METHODS

Data

We used data from the Premier Inc. website (www.premierinc.com) to identify hospitals that received a 1 percent or 2 percent bonus payment for each condition in each year of Phase 1 of the HQID (2004–2006). Data on the quality scores themselves were not published by Premier, so we derived data on a set of measures very similar to that used to determine bonuses for the HQID from Hospital Compare (Medicare's public quality reporting website). Hospital Compare provides numerators (the number of patients successfully receiving a measure) and denominators (the number of patients eligible to receive the measure) for process-of-care measures for AMI, heart failure, and pneumonia (see Appendix A). We did not evaluate bypass surgery and hip and knee replacement, because quality data for these procedures were not available on Hospital Compare. Overall, quality data were available for 260 hospitals participating in the HQID, although data were not available for each measure for each hospital. Descriptive statistics for the hospitals are shown in Table 1.

Following the HQID approach for determining incentive payments (Premier Inc 2004), we use Hospital Compare data to calculate condition-specific composite measures using the opportunities model, which is the sum of successfully achieved processes divided by the number of patients eligible to receive these processes for a given condition, for a given hospital, in a given year. Using these composite scores, we calculated whether hospitals would be eligible for either the 1 percent or 2 percent bonuses, by performing in the second highest and the highest deciles of quality, respectively. We refer to these as the authors' estimates of quality bonuses. We then compared these estimates to the actual quality bonuses awarded under the HQID for each condition in 2004 and 2005. For each condition in each year, some fraction of the authors' estimates did not correspond to the actual bonuses (see Table 2).

There are three main reasons for discrepancies between our estimates of incentive payments and actual payments: (1) the time intervals for the Hospital Compare and the HQID data collection differed by one quarter; (2) denominator exclusions may have varied somewhat between data reported for Hospital Compare and for the HQID; and (3) the quality measures incentivized in the HQID did not exactly match the measure reported on Hospital Compare (e.g., mortality, which was incentivized for AMI, was not reported in Hospital Compare). These discrepant observations were excluded from all

Table 1: Descriptive Statistics on Hospitals Participating in the Hospital Quality Incentive Demonstration

<i>Characteristic</i>	<i>Statistic</i>
<i>N</i>	260
Ownership, %	
Not-for-profit	91.1
Government run	8.5
For-profit	0.4
Urban, %	78.7
Affiliated with teaching hospital, %	38.2
Member of Council of Teaching Hospitals, %	13.9
Number of beds, mean (SD)	300.5 (210.6)
Total admissions, mean (SD)	14,192.7 (10,784.5)
Proportion of admission from Medicare patients, mean (SD)	.45 (.10)
Proportion of admission from Medicaid patients, mean (SD)	.16 (.09)
Composite quality in 2004, mean (SD)	
AMI	.90 (.07)
Heart failure	.77 (.12)
Pneumonia	.79 (.07)
Composite quality in 2006, mean (SD)	
AMI	.94 (.05)
Heart failure	.85 (.10)
Pneumonia	.87 (.07)
Number of years hospital received either 1% or 2% bonus 2004–2006, %	
AMI	
0 years	67.6
1 year	16.2
2 years	10.3
3 years	5.9
Heart failure	
0 years	63.6
1 year	20.2
2 years	10.3
3 years	5.9
Pneumonia	
0 years	66.4
1 year	18.6
2 years	6.3
3 years	8.7

analysis, thereby creating a dataset in which scores on the condition-specific composite quality measures exactly determined actual bonus payments.

Participation in the HQID was voluntary, and hospitals were able to exit the program at will. We excluded observations from HQID hospitals after they left the program. We also integrated information on hospital characteristics from the 2005 American Hospital Association Annual Survey.

Table 2: Correspondence between Authors' Estimates of Quality Bonuses and Actual Bonuses, 2004–2005

	Authors' Estimates of Bonuses											
	AMI				Heart Failure				Pneumonia			
	Below Bonus Thresholds	1% Bonus Winners	2% Bonus Winners	Total	Below Bonus Thresholds	1% Bonus Winners	2% Bonus Winners	Total	Below Bonus Thresholds	1% Bonus Winners	2% Bonus Winners	Total
Actual bonuses	396 (77.0%)	16 (3.1%)	8 (1.6%)	420 (81.7%)	383 (74.5%)	19 (3.7%)	12 (2.3%)	414 (80.5%)	392 (76.3%)	17 (3.3%)	6 (1.2%)	415 (80.7%)
Below bonus thresholds												
1% bonus winners	17 (3.3%)	20 (3.9%)	10 (1.9%)	47 (9.1%)	27 (5.3%)	20 (3.9%)	4 (0.8%)	51 (9.9%)	19 (3.7%)	21 (4.1%)	11 (2.1%)	51 (9.9%)
2% bonus winners	2 (0.4%)	14 (2.7%)	31 (6.0%)	47 (9.1%)	2 (0.4%)	12 (2.3%)	35 (6.8%)	49 (9.5%)	1 (0.2%)	13 (2.5%)	34 (6.6%)	48 (9.4%)

Cells contain counts of hospital observations (% of total) from 514 observations in 2004 and 2005. The shaded cells represent correspondence between the authors' estimates of bonuses and the actual receipt of bonuses. The correspondence rate was 86.9% for AMI, 85.2% for heart failure, and 87.0% for pneumonia.

Regression Discontinuity Design

In our study, the regression discontinuity design is preferred to other quasi experimental methods, such as a difference-in-differences design, for two primary reasons. First, by definition, winners and losers in the HQID program will have different baseline levels of performance for the subsequent year. Comparing changes in performance among groups with different baseline performance can be problematic if there are diminishing marginal returns for quality improvement efforts, which is certainly the case as hospitals approach the maximum scores on the measures (Ryan et al. 2012). Second, when treatment and comparison groups are defined based on their prior performance on the dependent variable, there is the potential for confounding due to regression to the mean (Chay, McEwan, and Urquiola 2005). Regression discontinuity designs are a good way to combat regression to the mean effects, given that identification is based on differences in subsequent performance between treatment and comparison groups with equivalent baseline performance.

After the data exclusions, financial bonuses are a deterministic function of composite quality scores that are discontinuous at the thresholds for quality bonuses, which is known as a “sharp” regression discontinuity design (Angrist and Pischke 2008). There are two discontinuities in the relationship between composite quality and the receipt of bonuses, one at the threshold for the 1 percent bonus (the 80th percentile) and the other at the threshold for the 2 percent bonus (the 90th percentile). Figure 1 shows histograms of composite quality scores for each incentivized diagnosis in each year, as well as the threshold values required for bonuses. For AMI, the thresholds for both the 1 percent and 2 percent bonuses were very high in 2004 (96.0 and 97.0, respectively) and remained high in 2005. The thresholds for 1 percent and 2 percent bonuses were 87.0 and 91.0 for heart failure, and 85.0 and 88.0 for pneumonia in 2004, all of which increased in 2005.

When implementing regression discontinuity designs, there is a trade-off between generating more consistent but less efficient estimates by only using data that are close to the discontinuity thresholds, and less consistent but more efficient estimates by using data that are further from the discontinuity thresholds. In our study, the choice of bandwidth is unrelated to our exclusion of observations for which our estimates of the receipt of bonuses were at odds with the actual receipt of bonuses. To examine the sensitivity of our results to our choice of data inclusion criteria, we select several alternative bandwidths (specified intervals around the thresholds) to create alternative discontinuity samples. We define l as the lower threshold (the minimum performance

Figure 1: Receipt of Quality Bonuses by Quality Performance for Incentivized Diagnoses, 2004–2005

Note. Dotted blue vertical line denotes threshold for 1 percent bonus and dashed red vertical line denotes threshold for 2 percent bonus.

required for the 1 percent bonus), u as the upper threshold (the minimum performance required for the 2 percent bonus), b as the bandwidth, and c as the composite scores for each diagnosis in each year. We choose discontinuity samples for $b = 3, 5$, and 10 percentage points, where $l - b \leq c \leq u + b$. Using the example of heart failure in 2005, the 80th percentile of composite quality in the prior year was 87.0 and the 90th percentile was 91.0. Using a bandwidth of ± 5 percentage points, we would include hospitals in the analysis that had composite quality between 82.0 and 96.0 in the prior year.

Identification and Estimation

We estimate the effect of a hospital receiving the 1 percent bonus or the 2 percent bonus on the level of quality in the year following the receipt of the bonus, conditional on the level of quality in the prior year. For a given diagnosis in a given year, our estimating equation for hospital j at time t is:

$$Quality_{jt} = f(Quality_{jt-1}) + \alpha X_j + b_1 1\% \text{ bonus}_{jt-1} + b_2 2\% \text{ bonus}_{jt-1} + \epsilon_{jt} \quad (1)$$

where $Quality$ is diagnosis-specific composite quality, X is a vector of hospital characteristics (urban location, teaching status [member of the Council of Teaching Hospitals], affiliation with a medical school, number of beds, number of condition-specific admissions, proportion of admissions from Medicare patients, and proportion of admissions from Medicaid patients), and $1\% \text{ bonus}$ and $2\% \text{ bonus}$ are dummies for whether a hospital received either bonus. While the financial impact of receiving a 2 percent bonus is twice that of a 1 percent bonus, there is no expectation that receiving one bonus or the other would provide a greater quality improvement. This is because the difference between getting a 1 percent bonus and no bonus is 1 percent, while the difference between getting the 2 percent bonus and the 1 percent is also 1 percent. We specify $f(Quality_{jt-1})$ as a quadratic function, allowing the relationship between lagged quality and quality to be nonlinear.¹ Figure 2 shows this to be a reasonable approximation of the functional form.

Equation 1 was estimated separately for each diagnosis, in each year, for each discontinuity sample, providing maximum flexibility in modeling the functional form between lagged quality and quality. The equation was estimated parametrically using a Generalized Linear Model with a logit link, accounting for the bounding of quality scores between 0 and 100 and introducing additional nonlinearity as scores approach 100. From these models, we calculate the marginal effects of the 1 percent and 2 percent bonuses on the proportion scale (rather than the percentage scale), evaluated at the threshold

Figure 2: Quality Performance around Payout Thresholds for Incentivized Diagnoses, 2005–2006

Note. Dotted blue vertical line denotes threshold for 1 percent bonus and dashed red vertical line denotes threshold for 2 percent bonus. Ninety-five percent confidence intervals shown for observed data.

values of lagged quality in each case. We also present the effect size for each estimate, standardizing the marginal effects by dividing the standard deviation of composite quality for the given diagnosis in the given year (Cohen 1992). Standard errors account for clustering at the hospital level.

Sensitivity and Specification Checks

We performed a variety of specification checks to assess the validity of our regression discontinuity design. First, we assessed graphically the presence of any apparent change in the density of the distribution of lagged quality scores at the threshold points, potentially indicative of unobserved differences between hospitals on either side of the bonus thresholds (Imbens and Lemieux 2008). Figure 1 shows the density of lagged quality for each diagnosis in each year and provides no evidence of systematic changes in the density of lagged quality around the threshold values for payment, supporting the validity of our design.

Second, we assessed the association between the covariates and bonus threshold values. The presence of such an association indicates the potential for omitted variable bias, given that some factor other than the receipt of a quality bonus may be responsible for any observed effects. We tested for covariate discontinuities for each threshold value (80th and 90th percentile) for each condition for each year using Equation 1 above in which the covariate was substituted as the dependent variable. Appendix A shows the results from these tests, providing virtually no evidence of discontinuities in the covariates at the bonus thresholds for AMI and pneumonia, but some evidence that covariates were associated with threshold values for heart failure. This motivated our decision to control the covariates in our implementation of the regression discontinuity design (Equation 1).

We also estimated an additional model to test whether the receipt of a bonus 2 years prior to a given year was associated with the quality in that year. This analysis was motivated by the fact that there was some time lag between the end of a reporting period in which quality was assessed and when hospitals would actually receive quality bonuses. It is therefore possible that hospitals would delay the reinvestment of quality bonuses, and effects on quality would not be apparent for another year. Finally, we assessed whether receiving a bonus for one diagnosis (e.g. AMI) may affect future quality for a different diagnosis (e.g., heart failure). This analysis could not exploit the sharp regression discontinuity design but instead modeled the incremental effect of the

quality bonuses on subsequent quality, conditional on lagged quality for each diagnosis (see Appendix A for more detail).

RESULTS

Figure 2 shows the relationship between quality and lagged quality from the regression discontinuity analysis. Results are shown from models estimated using the alternative bandwidths in both 2005 and 2006. Figure 2 shows that the relationship between quality and lagged quality is generally smooth through the 1 percent and 2 percent bonus thresholds, suggesting that receiving a quality bonus did not affect quality in the subsequent year.

Table 3 shows estimates of the effects of receipt of 1 percent and 2 percent bonuses on subsequent quality performance for each diagnosis in 2005 and 2006. Results are shown for bandwidths of ± 3 percentage points, ± 5 percentage points, and ± 10 percentage points. As the bandwidths increase, the number of observations included in the model also increase as observations at a greater distance from the bonus thresholds are included in the models.

For AMI, there are no statistically significant associations between the receipt of a bonus and subsequent quality performance, with the exception of the 2 percent bonus in 2006 ($p < .01$). The magnitude of the effect, however, is small (marginal effect 0.8 percentage points, effect size 0.16 standard deviations). For heart failure, there are no significant associations (at $p < .05$) between the receipt of either quality bonus and subsequent quality performance in either year for any of the bandwidths. For pneumonia, the 1 percent bonus is associated with a quality increase in 2005 for each of the bandwidths, with the largest effect occurring for the bandwidth of ± 3 percentage points (marginal effect 3.7 percentage points, effect size 0.60 standard deviations). However, neither bonus is associated with quality performance in 2006 for any bandwidth.

Results of the sensitivity tests are shown in Appendix A. Models testing whether a bonus received in 2004 was associated with greater quality in 2006 found no evidence of this association for either the 1 percent or 2 percent bonus for any diagnosis, indicating that the receipt of a bonus 2 years prior to a given period was not associated with improved quality. Also, analysis of whether bonuses for one diagnosis improved quality for another diagnosis found no evidence of this effect for AMI. For heart failure, a 2 percent bonus received for another diagnosis was associated with a small increase in subsequent heart failure quality (1.7 percentage points, $p < .01$) in 2006, and for

Table 3: Effects of Pay-for-Performance Bonus on Quality from Regression Discontinuity Models

	AMI			Heart Failure			Pneumonia		
	2005		2006	2005		2006	2005		2006
	Marginal Effect	Effect Size	Marginal Effect	Marginal Effect	Effect Size	Marginal Effect	Marginal Effect	Effect Size	Marginal Effect
Bandwidth: ± 3 percentage points									
Receipt of bonus	0.3 (1.2)	0.05	-0.1 (0.3)	-0.01	-0.48	-0.6 (1.2)	3.7** (1.7)	.60	0.3 (1.0)
1% bonus _{t-1}									
Receipt of bonus	-0.3 (1.7)	-0.04	0.2 (0.3)	0.04	-0.97	-8.2 (6.6)	0.8 (2.2)	.13	0.7 (1.1)
2% bonus _{t-1}									
<i>n</i>	55		85		39	41	47		61
Bandwidth: ± 5 percentage points									
Receipt of bonus	0.2 (0.9)	0.03	0.5* (0.3)	0.10	-0.05	-2.2* (1.3)	2.9** (1.4)	.47	0.6 (0.8)
1% bonus _{t-1}									
Receipt of bonus	-0.5 (1.0)	-0.09	0.8*** (0.2)	0.16	-0.36	-5.6 (3.9)	1.7 (1.8)	.28	0.3 (1.0)
2% bonus _{t-1}									
<i>n</i>	104		129		57	73	74		91
Bandwidth: ± 10 percentage points									
Receipt of bonus	0.2 (0.9)	0.04	0.2 (0.3)	0.05	-0.07	-1.2 (1.5)	3.4** (1.3)	.54	0.0 (0.7)
1% bonus _{t-1}									
Receipt of bonus	-0.5 (1.2)	-0.09	0.4 (0.2)	0.08	-0.27	0.3 (2.3)	2.5 (1.6)	.40	0.6 (0.9)
2% bonus _{t-1}									
<i>n</i>	167		176		107	110	132		165

* $p < .10$, ** $p < .05$, *** $p < .01$.
Marginal effects are shown in units of percentage points. Effect size is shown in units of the standard deviation of composite quality.
Cluster-robust standard errors in parentheses.
Estimates are from Generalized Linear Models with logit link functions to account for bounding of quality scores between 0 and 100.
Effect size is calculated by dividing the marginal effect by the standard deviation of composite quality for a given condition in a given year.
AMI models at bandwidth ± 3 , ± 5 are estimated with linear lag function.

pneumonia, a 1 percent bonus for another condition was associated with an increase in pneumonia quality in 2005 (2.1 percentage points, $p < .05$). These significant effects, however, are not consistent across years and are not based on the regression discontinuity design, making inference less reliable.

DISCUSSION

This study exploited a sharp discontinuity in the probability that a hospital would receive a bonus payment to estimate the effect of quality bonuses on subsequent quality improvement. Overall, we found limited evidence supporting the idea that hospitals' receipt of quality bonuses facilitates subsequent quality improvement. Among our many specifications, while 20 of the 36 estimates were positive, only a few significant effects were observed: the 2 percent bonus for AMI in 2006 was significantly associated with increased quality improvement in the following year (0.8 percentage points, or 0.16 SD of composite AMI quality) using a bandwidth of ± 5 percentage points; and the 1 percent bonus for pneumonia in 2005 was associated with increased quality improvement (3.7 percentage points, or 0.60 SD of composite pneumonia quality) using a bandwidth of ± 3 percentage points. These effects are small in absolute terms but relatively large compared to the limited variance of the quality scores. However, the significant effect for AMI was sensitive to the choice of bandwidth, with effects disappearing with bandwidths of ± 3 percentage points and the significant effects for pneumonia were not consistent across the study years. The lack of robustness of the positive results across the specifications and the years assessed suggests that the relationship between bonus payments and subsequent quality improvement is not consistent and unlikely to reflect behavior change induced by the financial incentives.

Our study has a number of limitations. It could be hypothesized that hospitals that narrowly missed incentive thresholds would actually attempt to improve quality more in the following year, which is contrary to our conceptual model that the receipt of incentives would lead to greater quality improvement. However, if hospitals that just missed performance thresholds in fact improved quality more in the subsequent period, we would expect a negative effect of receiving a bonus on subsequent quality, which we did not observe in our study. We also did not have access to the actual quality scores used to determine incentive payments, and instead constructed these scores using an alternative data source. We excluded data from hospitals whose

actual payments did not correspond with payments estimated from this data source. Nonetheless, we have no reason to believe that the discrepancies between estimated and actual payments were systematic, rather than simply random error. Thus, the exclusion of discrepant observations decreased statistical power and external validity by reducing the number of hospitals included in the study by ~15 percent, but likely did not adversely affect the internal validity of the study.

Our inferences are also limited by not knowing precisely when hospitals in the HQID received incentive bonuses. Anecdotal evidence suggests that this was at least 6 months after a reporting period ended (Damberg et al. 2007). It is therefore possible that we did not find substantial effects of the receipt of quality bonuses on subsequent improvement because hospitals were not able to reinvest these bonuses in the period assessed in this study. However, we did test whether the receipt of a quality bonus affected quality 2 years later, and we found no evidence of any longer term effects. Hospitals may also have reinvested quality bonuses to improve quality in areas that were not measured in this study, such as surgical safety. It is also possible that in addition to allocating additional resources from quality bonuses toward quality improvement, the positive feedback signal from these bonuses may stimulate additional quality improvement. By rewarding hospitals for their quality performance, a bonus may reinforce a hospital's strategy to pursue high-quality care and encourage future high performance. Notification of a bonus could therefore lead to quality improvement before the bonus was actually received, which would diminish the effects of the time lag in receiving the bonus.

Given the relatively small number of hospitals in the Demonstration for which data was available in this study (260), and given that there were limited data around the payment threshold (because the payment thresholds were based on the 80th and 90th percentiles of performance), the study had limited statistical power to detect program effects. Nonetheless, most of our estimates were fairly precise, with standard errors around 1 percentage point, giving us the ability to detect relatively small effects. Finally, as hospitals approached the maximum performance scores for many of the measures, ceiling effects attenuate further quality improvement (Ryan et al. 2012). This may have resulted in downward bias in our estimates of the effects of receiving quality bonuses on subsequent improvement, particularly in 2006.

Overall, our study found that in most cases the receipt of quality bonuses was not associated with subsequent quality improvement for hospitals participating in the HQID. This could be due to the small absolute magnitude of bonuses (Ryan et al. 2012), because hospitals reinvested little or none of their

bonuses in quality improvement activities for the diagnoses evaluated in this study, or because reinvestment did not improve quality. There are also likely limits to what further investment can achieve for hospitals at the higher end of the performance scale. It is also possible that, early on in the HQID, hospitals achieved high quality scores on measures that were less costly to achieve, and that further quality improvement would involve improvement on the more costly measures (Nicholas, Dimick, and Iwashyna 2011).

This study has important implications for Hospital Value-Based Purchasing and other value-based purchasing programs codified in the Patient Protection and Affordable Care Act (Ryan and Damberg 2013). In addition to the recent evidence that hospitals were not responsive to the financial incentives in the HQID (Ryan et al. 2012) and that the HQID did not improve health outcomes (Ryan 2009; Jha et al. 2012), our study suggests that the winners in the program did not respond by improving quality of care to a greater extent than hospitals not receiving bonuses. For most hospitals, the financial incentives in Hospital Value-Based Purchasing will be of a similar magnitude to those in the HQID. Results from this study raise questions about whether winning in Hospital Value-Based Purchasing will lead to subsequent improvement. Going forward, research should assess the impact of quality bonuses on subsequent quality improvement in other value-based purchasing programs, particularly those with larger financial incentives such as the Medicare Advantage Quality Bonus Program (Ryan and Damberg 2013). Future research should also evaluate how providers spend bonus payments from incentive programs, and whether the threat of losing is a stronger motivator for quality improvement than actually winning.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to acknowledge Jayme Mendelsohn for research assistance. Funding for Dr. Ryan was supported by a career development award from the Agency for Healthcare Research and Quality (1 K01 HS018546-01A1). Tim Doran was supported by a National Institute for Health Research (NIHR) Career Development Fellowship. This report is independent research and the views expressed are those of the author and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Disclosures: None.

Disclaimers: None.

NOTE

1. For AMI with bandwidths of ± 3.0 and ± 5.0 , marginal effects were not estimable using the quadratic lag function because of the concentration of composite quality scores close to 100, and models were instead specified with a linear lag function.

REFERENCES

- Alexander, J. A., Maeng D., Casalino, L. P., and Rittenhouse, D. 2012. "Use of Care Management Practices in Small- and Medium-Sized Physician Groups: Does Public Reporting of Physician Quality and Financial Incentives Matter?" *Health Services Research* 48 (2): 376–97.
- Almond, D., and J. J. Doyle Jr. 2008. *After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays*. Cambridge, MA: National Bureau of Economic Research Working Paper Series No. 13877.
- Angrist, J. D., and J. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Bazzoli, G. J., J. P. Clement, R. C. Lindrooth, H. F. Chen, S. K. Aydede, B. I. Braun, and J. M. Loeb. 2007. "Hospital Financial Condition and Operational Decisions Related to the Quality of Hospital Care." *Medical Care Research and Review* 64 (2): 148–68.
- Card, D., C. Dobkin, and N. Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics* 124 (2): 597–636.
- Chay, K. Y., P. J. McEwan, and M. Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review* 95 (4): 1237–58.
- Cohen, J. 1992. "A Power Primer." *Psychological Bulletin* 112 (1): 155–159.
- Damberg, C. L., M. E. S. Sorbero, A. Mehrotra, S. S. Teleki, S. Lovejoy, and L. Bradley. 2007. *An Environmental Scan of Pay for Performance in the Hospital Setting: Final Report*. Santa Monica, CA: RAND. Available at <http://aspe.hhs.gov/health/reports/08/payperform/PayPerform07.html>.
- Flodgren, G., M. P. Eccles, S. Shepperd, A. Scott, E. Parmelli, and F. R. Beyer. 2011. "An Overview of Reviews Evaluating the Effectiveness of Financial Incentives in Changing Healthcare Professional Behaviours and Patient Outcomes." *Cochrane Database of Systematic Reviews*, issue 7.
- Government Accountability Office. 2012. *Medicare Advantage Quality Bonus Payment Demonstration*. Washington, DC: United States Government Accountability Office.
- Imbens, G. W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.
- 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.

- 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>.
- 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.
- Miller, N., K. Eggleston, and R. Zeckhauser. 2006. "Provider Choice of Quality and Surplus." *International Journal of Health Care Finance Economics* 6 (2): 103–17.
- Newhouse, J. P. 1970. "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital." *American Economic Review* 60 (1): 64–74.
- Nicholas, L. H., J. B. Dimick, and T. J. Iwashyna. 2011. "Do Hospitals Alter Patient Care Effort Allocations under Pay-for-performance?" *Health Services Research* 46 (1 Pt 1): 61–81.
- Pauly, M. V. 1987. "Nonprofit Firms in Medical Markets." *American Economic Review* 77 (2): 257–62.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.
- Rosenthal, M. B., B. E. Landon, S. L. Normand, R. G. Frank, and A. M. Epstein. 2006. "Pay for Performance in Commercial HMOs." *New England Journal of Medicine* 355 (18): 1895–1902.
- 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.
- Ryan, A. M., J. Blustein, and L. P. Casalino. 2012. "Medicare's Flagship Test of Pay-for-performance Did Not Spur More Rapid Quality Improvement among Low-Performing Hospitals." *Health Affairs (Millwood)* 31 (4): 797–805.
- Ryan, A. M., and C. L. Damberg. 2013. "What Can the Past of Pay-for-Performance Tell Us about the Future of Value-Based Purchasing in Medicare?" *Health Care: The Journal of Delivery Science and Innovation* 1 (1): 42–9.
- Ryan, A. M., J. Blustein, T. Doran, M. D. Michelow, and L. P. Casalino. 2012. "The Effect of Phase 2 of the Premier Hospital Quality Incentive Demonstration on Incentive Payments to Hospitals Caring for Disadvantaged Patients." *Health Services Research* 47 (4): 1418–36.
- Sloan, F. A. 2000. Not-for-Profit Ownership and Hospital Behavior. In *Handbook of Health Economics*, edited by A. J. Culyer & J. P. Newhouse, pp. 1141–74. Amsterdam: Elsevier.
- Van Herck, P., D. De Smedt, L. Annemans, R. Remmen, M. B. Rosenthal, and W. Sermeus. 2010. "Systematic Review: Effects, Design Choices, and Context of Pay-for-Performance in Health Care." *BMC Health Services Research* 10: 247.

SUPPORTING INFORMATION

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

Appendix SA1: Author Matrix.