Published in final edited form as:

Ann Surg. 2014 April ; 259(4): 677–681. doi:10.1097/SLA.0000000000000425.

Does Pay-for-Performance Improve Surgical Outcomes? An Evaluation of Phase 2 of the Premier Hospital Quality Incentive Demonstration

Terry Shih, MD, Lauren H. Nicholas, PhD, Jyothi R. Thumma, MPH, John D. Birkmeyer, MD, and Justin B. Dimick, MD, MPH

Center for Healthcare Outcomes and Policy (TS, LHN, JRT, JDB, JD), University of Michigan, Ann Arbor, MI

Abstract

Objective—We sought to determine whether the changes in incentive design in Phase 2 of Medicare’s flagship Pay-for-Performance program, the Premier Hospital Quality Incentive Demonstration (HQID), reduced surgical mortality or complication rates at participating hospitals.

Background—The Premier HQID was initiated in 2003 to reward high-performing hospitals. The program redesigned its incentive structure in 2006 to also reward hospitals that achieved significant improvement. The impact of the change in incentive structure on outcomes in surgical populations is unknown.

Methods—We examined discharge data for patients who underwent coronary artery bypass (CABG), hip replacement, and knee replacement at Premier hospitals and non-Premier hospitals in Hospital Compare from 2003–2009 in 12 states (n=861,411). We assessed the impact of incentive structural changes in 2006 on serious complications and 30-day mortality. In these analyses, we adjusted for patient characteristics using multiple logistic regression models. To account for improvement in outcomes over time, we used difference-in-difference techniques that compare trends in Premier vs. non-Premier hospitals. We repeated our analyses after stratifying hospitals into quintiles according to risk-adjusted mortality and serious complication rates.

Results—After restructuring incentives in 2006 in Premier hospitals, there were lower risk-adjusted mortality and complication rates for both cardiac and orthopedic patients. However, after accounting for temporal trends in non-Premier hospitals, there were no significant improvements in mortality for CABG (OR 1.09, 95% CI 0.92 to 1.28) or joint replacement (OR 0.81, 95% CI 0.58 to 1.12). Similarly, there were no significant improvements in serious complications for CABG (OR 1.05, 95% CI 0.97 to 1.14) or joint replacement (OR 1.12, 95% CI 1.01 to 1.23). Analysis of the “worst” quintile hospitals that were targeted in the incentive structural changes also did not reveal a change in mortality (OR 1.01, 95% CI 0.78 to 1.32 for CABG and OR 0.96, 95% CI 0.22 to 4.26 for joint replacement) or serious complication rates (OR 1.08, 95% CI 0.88 to 1.34 for CABG and OR 0.92, 95% CI 0.67 to 1.28).

Conclusions—Despite recent enhancements to incentive structures, the Premier HQID did not improve surgical outcomes at participating hospitals. Unless significantly redesigned, pay-for-performance may not be a successful strategy to improve outcomes in surgery.

Keywords

Pay-for-performance; Health Services Research; Coronary Artery Bypass; Joint Replacement; Policy evaluation

INTRODUCTION

Pay-for-performance (P4P) has been embraced by policymakers as a strategy for improving quality in healthcare. Over the next few years, the Affordable Care Act will dramatically increase the scale of existing pay-for-performance through the Medicare Value-Based Purchasing program.¹ Evidence on the effectiveness of pay-for-performance for inpatient surgery comes from Medicare's flagship inpatient P4P program, the Premier Hospital Quality Incentive Demonstration (HQID). Evaluations of this first phase of the program showed limited improvement in processes of care and no significant improvements in patient outcomes, including surgery.²⁻⁴ Because the program initially only rewarded hospitals in the top two deciles of performance, there was little to motivate poorly performing hospitals to improve. The incentive structure of this program was redesigned in the fourth quarter of 2006. In phase 2, hospitals could qualify for financial incentives by ranking in the top 20% of hospitals for performance or ranking in the top 20% of hospitals for improvement and performing above the median level of hospitals.⁵

Whether these changes in incentive structure of P4P will improve outcomes in surgical populations is not known. Prior evaluations of the Premier HQID have been conducted for inpatient medical care and found no improvement in mortality rates.^{6, 7} However, evaluations of surgical populations are limited. Existing studies evaluate the impact of P4P on mortality for coronary artery bypass surgery but exclude hip and knee replacement.⁶ Moreover, previous studies have focused solely on mortality and have not considered the impact on overall or serious complications. Finally, previous evaluation focuses on the entire period since the program started in 2003 and does not specifically test the change in incentive structure in 2006.^{2, 6}

This study seeks to examine the effect of the incentive structural changes of Phase 2 of the Premier HQID on perioperative mortality and complication rates. Ideally, a delivery system innovation could be tested with randomized trials. However, this is not feasible for large-scale policy interventions. Therefore, we used a natural experiment evaluating the impact of the change in incentives in Premier hospitals compared to a control group of non-participating hospitals. This allowed us to adjust for secular trends and differences in baseline hospital performance. Using the State Inpatient Database for twelve states, we examined the incentivized procedures in the Premier HQID: coronary artery bypass graft (CABG) surgery, hip replacement, and knee replacement from 2003 to 2009.

METHODS

Data Source and study population

We used data from the State Inpatient Database (SID) of twelve states (Arizona, California, Florida, Iowa, Massachusetts, Maryland, North Carolina, Nebraska, New Jersey, New York, Washington, and Wisconsin) from 2003 to 2009. This data set is maintained and distributed as part of the Healthcare Cost and Utilization Project of the Agency for Healthcare Research and Quality, and contains all inpatient discharges from short-term, acute-care, nonfederal, general, and other specialty hospitals in participating states.⁸ We chose these 12 states because they (1) were geographically dispersed across the United States (allowing for diversity in our sample), (2) were available for the period we were studying, and (3) had

relatively large sample sizes. The discharge records from these databases contain information collected as part of billing records, including patient demographics, *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) procedures, diagnoses, expected payer, admission and discharge dates, and disposition. Data on hospital characteristics was obtained from the American Hospital Association Annual Survey.

Using appropriate ICD-9-CM procedure codes, we identified all adult patients undergoing CABG (36.10–19), hip replacement (81.51–52), and total knee replacement (81.54) in these twelve states. We excluded CABG patients with procedure codes indicating that other operations were simultaneously performed (i.e. valve replacement) (35.00–99, 36.2, 37.32, 37.34, 37.35). Joint replacement patients were excluded for revision procedures or trauma diagnoses (*ICD-9-CM* procedure codes 00.70–73, 00.80–85, 81.53, 81.55 and *ICD-9-CM* diagnosis codes 800–959).

Study Overview

In 2003, the Center for Medicare and Medicaid Services (CMS) created the largest experiment in pay-for-performance to date: the Premier Hospital Quality Incentive Demonstration.⁵ 216 hospitals agreed to provide data on process and quality indicators for three medical conditions (acute myocardial infarction, congestive heart failure, and pneumonia) and two surgical procedures (CABG and total knee or hip replacement) with additional indicators for risk-adjusted mortality for acute myocardial infarction and CABG and 30-day readmissions for total knee or hip replacement. Recruitment of all participating hospitals was completed in March 2013. Hospitals were required to participate in each of the five clinical areas only if they provided care for that condition. The project also required a minimum patient volume of 30 cases to be included. In Phase 1 of the Premier HQID (2003 to 2006), the top 20% of hospitals received 1 to 2% bonuses in Medicare reimbursements. The incentive structure was criticized for rewarding only high-performing hospitals, with little incentive for poor-performing hospitals to improve.⁹ Therefore, the Premier HQID restructured its incentive structure in 2006. In phase 2 (2006 to 2009), financial bonuses were additionally given to hospitals that significantly improved on their performance. Hospitals could now qualify for bonuses in three ways: (1) performing in the top 20% of hospitals (“Top Performance Award”) (2) performing above the median level of performance in the current year and ranking in the top 20% in terms of improvement (“Improvement Award”) and (3) performing above the median level of performance for a composite quality score benchmark from two years prior (“Attainment Award”). Over the six years of the demonstration, CMS awarded more than \$60 million in financial bonuses with almost \$12 million in incentive payments in the final year.⁵

The goal of this analysis was to examine the impact of the HQID incentive structure changes on adverse events after fully taking into account temporal trends towards improved outcomes in cardiac and orthopedic surgery. We used an econometric technique, the difference-in-difference approach, which is commonly used to evaluate the impact of policy change.^{10–12} This approach isolates the impact of the policy change on outcomes above and beyond any changes seen in a control group that were not exposed to the policy change. In our analysis, we chose non-Premier hospitals in the SID as the control group because they are exposed to all other factors driving improved outcomes over time except participation in the Premier HQID.

Quality Measures

Our outcomes of interest were risk-adjusted inpatient mortality, inpatient complication, and serious inpatient complications. We used specific ICD-9-CM codes to identify

complications as previously validated by chart review in The Complications Screening Program. The following postoperative complications were identified in our study: pulmonary failure (518.81, 518.4, 518.5, 518.8) pneumonia (481, 482.0–9, 483, 484, 485, 507.0), deep venous thrombosis/pulmonary embolism (415.1, 451.11, 451.19, 451.2, 451.81, 453.8), acute renal failure (584), hemorrhage (998.1), surgical site infection (958.3, 998.3, 998.5, 998.59, 998.51), gastrointestinal hemorrhage (530.82, 531.00–21, 531.40–41, 531.60–61, 532.00–21, 532.40–41, 532.60–61, 533.00–21, 533.40–41, 533.60–61, 534.00–21, 534.40–41, 534.60–61, 535.01, 535.11, 535.21, 535.31, 535.41, 535.51, 535.61, 578.9), and myocardial infarction (410.00–91) for joint replacement patients only. The coding of surgical and medical complications, including those identified in our study has been shown by others to be in good agreement when ICD-9-CM codes and the medical record were compared.^{13, 14} Myocardial infarction was not considered a complication following CABG because of the inability to assess the temporal relationship of an acute myocardial infarction to the operation. Serious complications were noted as any of the above complication with a length-of-stay above the 75th percentile. This addition of the extended length of stay criterion was intended to increase the specificity of the outcome variable.

Statistical Analysis

All participating Premier HQID hospitals through the six years of the demonstration in the above-mentioned 12 states were identified from the Premier website and included in the analysis.⁵ To perform the difference-in-difference analysis, we used the following logistic regression model to evaluate the relationship between patient outcomes Y_{it} (inpatient mortality, complications, and serious complications) and the HQID incentive structure changes:

$$Y_{it} = \alpha \text{Premier}_{it} + \beta \text{Post}_{it} + \delta \text{Premier}_{it} * \text{Post}_{it} + \theta X_{it} + \text{Time} + \varepsilon_{it}$$

We included categorical variables indicating whether the patient was treated at a Premier hospital (Premier) and whether this treatment was before or after (i.e., pre-post) the coverage decision (Post). Because the SID does not have dates of surgery, we used admission quarter to define whether patients had surgery in phase 1 (January 2003 to September 2006) or phase 2 (October 2006 to December 2009) of the program. To adjust for secular trends, we included a continuous time variable that effectively takes into account linear time trends. In all models we adjusted for patient characteristics (θX_{it}) by entering the 29 Elixhauser comorbid diseases as individual covariates, a widely used and previously validated approach for risk-adjustment in administrative data.^{15, 16} Finally, we added an interaction term of the Premier (vs. non-Premier) variable and the pre-post incentive change variable (Premier*Post). The coefficient from this interaction term, i.e., the difference-in-difference estimator, can be interpreted as the impact of the independent impact of the Premier HQID incentive changes.^{2, 11, 17}

We performed a sensitivity analysis examining the impact of the incentive change on the bottom 20% of hospitals. The incentive redesign in 2006 was aimed at motivating poor-performing hospitals to improve their performance. Hospitals in the bottom 20% of performance had the most room to improve and were the perceived targets of the expanded “Improvement Award.” To establish which hospitals were in the bottom 20%, we used logistic regression to predict the probability of inpatient mortality, any complication, and serious complication for each patient incorporating patient and hospital characteristics for both CABG and joint replacement. We aggregated these data at the hospital level to calculate six distinct ratios of observed to expected outcome rates (inpatient mortality, complications, and serious complications for both procedures). The bottom 20% of hospitals for each adverse outcome of interest was isolated and similar analysis was performed as above.

All statistical analyses were performed using STATA 11.0 (College Station, Texas).

RESULTS

Characteristics of Premier and non-Premier patients before and after the incentive structure changes are shown in Table 1. Patient characteristics were similar at Premier and non-Premier hospitals. With respect to the HQID incentive changes, patients treated after the policy changes had slightly more comorbidities, but these differences were similar for both Premier and non-Premier hospitals (Table 1). Characteristics of Premier and non-Premier hospitals are shown in Table 2. In the 12 states examined, there were 44 participating Premier hospitals and 321 non-participating hospitals for CABG. For hip and knee replacement, there were 93 participating Premier hospitals and 1046 non-participating hospitals. Hospitals participating in the Premier HQID were larger hospitals with more licensed beds and employees. These hospitals also were busier with more total annual admissions and surgical operations. Participating hospitals were more likely to be trauma centers, and for CABG, participating hospitals were more likely to be teaching hospitals.

In simple pre-post analysis after adjusting for patient characteristics, there were apparent declines in rates of inpatient mortality for both procedures and complications for joint replacement after the HQID incentive structure changes (Table 3). The odds ratio of mortality risk-adjusted for patient characteristics after the incentive structural changes was 0.70 for CABG (95% CI 0.66 to 0.75) and 0.78 for joint replacement (95% CI 0.61 to 1.00) (Table 3). However, after taking into account secular trends in other hospitals, there were no measurable improvements in outcomes after (vs. before) incentive changes for either procedure. There were no significant improvements in mortality for CABG (OR 1.09, 95% CI 0.90 to 1.32) or joint replacement (OR 0.85, 95% CI 0.54 to 1.32). Similarly, there were no improvements in any complication or serious complication rates.

Stratified sensitivity analysis focusing on the lowest performing (bottom 20%) hospitals demonstrated analogous results (Table 3). Simple pre-post analysis demonstrated improved mortality for CABG (OR 0.78, 95% CI 0.69 to 0.87) and joint replacement (OR 0.77, 95% CI 0.68 to 0.88). After adjusting for trends in non-participating hospitals, there were no significant improvements for mortality in CABG (OR 1.22, 95% CI 0.88 to 1.70) and joint replacement (OR 0.85, 95% CI 0.59 to 1.21). Again, there were no improvements in any complication or serious complications in the bottom 20% of hospitals.

DISCUSSION

In this study we evaluated Phase 2 of Medicare's flagship pay-for-performance pilot program, the Premier Hospital Quality Incentive Demonstration. In simple pre-post analysis, we found improvements in outcomes over the six years of the Premier HQID. However, after adequate adjustment with a control group of non-participating hospitals, there was no improvement due to the HQID incentive structure changes. The expansion of financial bonuses to hospitals that demonstrated improvement rather than just excellence did not spur a decline in adverse outcomes. Moreover, we found no improvement among the poorest-performing hospitals, the intended beneficiaries of the policy.

These conclusions should be considered in the context of prior studies. Previous reports demonstrated improved compliance with process measures in pay-for-performance hospitals.^{3, 4} Werner and colleagues demonstrated these improvements in process measures waned after 5 years.¹⁸ Expanding beyond adherence to process measures, Ryan demonstrated that participation in the Premier HQID did not effect 30-day mortality rates for the targeted three medical conditions (acute myocardial infarction, congestive heart failure,

and pneumonia) or CABG.² Jha and colleagues explored the long-term impact of the Premier HQID on mortality for the same conditions.⁶ This study did not find any evidence that the program led to a decrease in mortality for CABG. We have expanded on these previous studies by investigating orthopedic as well as cardiac procedures with no independent effect of the policy change on mortality. We have also extended beyond mortality to investigate other non-fatal complications. Finally, we also specifically evaluated the change in incentive structure rather than the entirety of the pilot program since its inception.

Without careful consideration of study design, it is possible to generate misleading results when evaluating health care policy changes. Most studies in the surgical literature that evaluate implementation of such changes simply compare outcomes before (pre) versus after (post) the onset of the policy change.¹⁹ However, this completely ignores the fact that surgical outcomes may be improving over time for other reasons, such as enhanced efforts to improve surgical safety, new technologies, and better training.

In our study, such a naïve pre-post study would have found a positive effect on outcomes over time that could erroneously be attributed as a benefit of the Premier HQID policy change. However when using appropriate techniques, adjusting for trends in non-participating hospitals towards improved outcomes overall, there is no independent effect towards improved outcomes due to this policy change. This highlights the importance of considering background trends towards improvement using econometric methods when evaluating health care policy changes. The natural experiment approach used in our study, where we compared an intervention group with a control group, allowed us to subtract out these background changes and generate an accurate estimate of the independent effect of the policy change.

There are limitations to our study. We used administrative data, which has well-known limitations in risk-adjustment.²⁰ Comorbidities and complications are derived from discharge codes and are not as reliable as those derived directly from the medical record. As surgical outcomes become more prominent, it is conceivable that billing practices change to “overcode” comorbidities and “undercode” complications. As illustrated by Table 1, patients at later time points have increased comorbidities in both surgical populations. However, our study is designed to address these changes in patient risk profiles. Because we incorporate a control group of non-Premier hospitals that also experienced an increase in the risk profile of patients, we effectively adjust for these differences in patient characteristics across hospitals.¹² In addition, complications ascertained from administrative data are not as accurate as those determined from the medical record. To address this problem we limited our complications to a previously validated subset of codes demonstrated to have the highest sensitivity and specificity.^{13, 14} In addition, we have added a prolonged length of stay (greater than the 75th percentile) to our definition of serious complications. Length of stay is very accurate in administrative data, and prolonged hospitalization is a reliable marker of postoperative complications. Furthermore, although one might expect billing practices change to “undercode” complications, crude and risk-adjusted complication rates have actually increased over the course of the Premier HQID for CABG patients and remained unchanged for joint replacement patients.

Our study has important implications pertaining to the new Value-Based Purchasing (VBP) program being implemented by the Center for Medicare and Medicaid Services through mandates put forth by the Affordable Care Act.²¹ The VBP was modeled largely after the Premier HQID evaluated in this study. Financial incentives are awarded via attainment of or improvement towards a high level of quality. Under this incentive structure, our analyses suggest this may not result in substantial improvements in mortality or complication rates in

the surgical population. The answer may be in larger incentives. Werner et al demonstrated that under the current structure of the Value Based Purchasing program modeled after the Premier HQID, the size of incentive redistribution would constitute only a fraction of one percent of hospital costs.²² Incentives of this size may not be a large enough motivation to change hospital behaviors.

However, the evidence that larger incentive bonuses can effect meaningful change is mixed. An analysis of a P4P program implemented by Medicaid in Massachusetts with incentives approximately five times the size of incentives in the Premier HQID also showed no improvements in quality.²³ Conversely, a recent analysis of the P4P program Advancing Quality in England modeled off of the Premier HQID but with much larger bonuses demonstrated significant improvements in mortality for three medical conditions (pneumonia, heart failure and acute myocardial infarction).²⁴ While both of these studies had significantly larger bonuses, one showed improvements in outcomes while the other did not. Further research is required to better understand how financial incentives influence the behavior of hospitals.

In summary, our study conclusively demonstrates that incentivizing poor performers had no salutary effects on overall performance, even among poor-performing hospitals. Despite financial enhancements targeted at poor performers, participating hospitals were insensitive to changes in financial incentives. Pay-for-performance under CMS' value based purchasing program may only have limited effectiveness in surgery unless significantly redesigned.

Acknowledgments

This study was supported by grants to Dr. Shih from the National Institutes of Health (5T32HL07612309) and Drs. Dimick, Nicholas, and Birkmeyer from the National Institute of Aging (R01AG039434). Dr. Dimick and Dr. Nicholas are also supported by career development awards from the Agency for Healthcare Research and Quality (K08HS017765) and the National Institute on Aging (K01AG041763), respectively. The views expressed herein do not necessarily represent the views of the United States Government.

Acknowledgements/Disclosures: Dr. Dimick and Dr. Birkmeyer are consultants and have an equity interest in ArborMetrix, Inc, which provides software and analytics for measuring hospital quality and efficiency. The company had no role in the study herein.

References

- Centers for Medicare and Medicaid Services. Medicare Program; Hospital Inpatient Value-Based Purchasing Program. Final Rule. Vol. 76: Federal Register. 2011:26490–26547.
- Ryan AM. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health services research*. 2009; 44(3):821–842. [PubMed: 19674427]
- Glickman SW, Ou F-S, Delong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA : the journal of the American Medical Association*. 2007; 297(21):2373–2380. [PubMed: 17551130]
- Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *The New England journal of medicine*. 2007; 356(5):486–496. [PubMed: 17259444]
- Premier. CMS/Premier Hospital Quality Incentive Demonstration 2013. Available at: <https://http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/index.jsp>.
- Jha AK, Joynt KE, Orav EJ, et al. The long-term effect of premier pay for performance on patient outcomes. *The New England journal of medicine*. 2012; 366(17):1606–1615. [PubMed: 22455751]
- Ryan AM, Blustein J, Casalino LP. Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals. *Health affairs (Project Hope)*. 2012; 31(4):797–805. [PubMed: 22492897]

8. Healthcare Cost and Utilization Project (HCUP). [Accessed May 31, 2013] Overview of the State Inpatient Databases. Available at: <http://www.hcup-us.ahrq.gov/sidoverview.jsp>
9. Casalino LP, Elster A, Eisenberg A, et al. Will pay-for-performance and quality reporting affect health care disparities? *Health affairs (Project Hope)*. 2007; 26(3):w405–W414. [PubMed: 17426053]
10. Colla CH, Wennberg DE, Meara E, et al. Spending differences associated with the Medicare Physician Group Practice Demonstration. *JAMA : the journal of the American Medical Association*. 2012; 308(10):1015–1023. [PubMed: 22968890]
11. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. *JAMA : the journal of the American Medical Association*. 2007; 298(9):975–983. [PubMed: 17785642]
12. Wooldridge, JM. *Introductory econometrics : a modern approach*. 5th ed.. Mason, OH: South-Western Cengage Learning; 2012.
13. Iezzoni L, Daley J, Heeren T, et al. Identifying Complications of Care Using Administrative Data. *Medical care*. 1994; 32(7):700–715. [PubMed: 8028405]
14. Weingart S, Iezzoni L, Davis R, et al. Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program. *Medical care*. 2000; 38(8):796–806. [PubMed: 10929992]
15. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Medical care*. 1998; 36(1):8–27. [PubMed: 9431328]
16. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Medical Care*. 2004; 42(4):355–360. [PubMed: 15076812]
17. Donald SG, Lang K. Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*. 2007; 89(2):221–233.
18. Werner RM, Kolstad JT, Stuart EA, et al. The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health affairs (Project Hope)*. 2011; 30(4):690–698. [PubMed: 21471490]
19. Nguyen NT, Hohmann S, Slone J, et al. Improved bariatric surgery outcomes for Medicare beneficiaries after implementation of the medicare national coverage determination. *Arch Surg*. 2010; 145(1):72–78. [PubMed: 20083757]
20. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med*. 1997; 127(8 Pt 2):666–674. [PubMed: 9382378]
21. Centers for Medicare and Medicaid Services. [Accessed December 10, 2012] Hospital Value-Based Purchasing Program. Available at: <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html>
22. Werner RM, Dudley RA. Medicare's new hospital value-based purchasing program is likely to have only a small impact on hospital payments. *Health affairs (Project Hope)*. 2012; 31(9):1932–1940. [PubMed: 22949441]
23. Ryan AM, Blustein J. The effect of the MassHealth hospital pay-for-performance program on quality. *Health services research*. 2011; 46(3):712–728. [PubMed: 21210796]
24. Sutton M, Nikolova S, Boaden R, et al. Reduced mortality with hospital pay for performance in England. *The New England journal of medicine*. 2012; 367(19):1821–1828. [PubMed: 23134382]

Table 1

Characteristics of Medicare patients undergoing surgery at Premier vs. non-Premier hospitals before and after the changes to the Premier Hospital Quality Incentive Demonstration Program.

Patient Characteristics	Premier			Non-Premier			p value (Premier vs. non-Premier)
	Before	After	p value (after vs. before)	Before	After	p value (after vs. before)	
Coronary artery bypass surgery							
Age (mean)	72.7	72.2	*	72.7	72.3	**	*
Gender (% Female)	31.4%	30.0%		31.6%	30.8%	*	
Race (% Caucasian)	74.6%	68.9%	**	71.1%	72.2%	*	
Low socioeconomic status	15.8%	17.6%	**	18.2%	17.8%	*	**
Non-elective admission	60.7%	57.7%		56.8%	54.0%		*
3 Comorbidities	29.4%	33.0%	**	30.8%	34.3%	**	*
Hip and knee replacement							
Age (mean)	73.2	72.9	**	73.4	73.0	**	*
Gender (% Female)	65.5%	64.7%		65.1%	64.5%	*	
Race (% Caucasian)	73.0%	67.0%	**	74.3%	75.9%	**	**
Low socioeconomic status	17.4%	15.3%	**	12.7%	12.4%	**	**
Non-elective admission	7.9%	5.5%	**	7.5%	5.7%	**	
3 Comorbidities	23.9%	31.3%	**	23.6%	29.4%	**	**

* p<0.05

** p<0.0001

Table 2

Characteristics of Premier vs. non-Premier hospitals.

Hospital Characteristics	Premier	Non-Premier	p-value
Coronary artery bypass surgery			
Number of hospitals	44	321	
Licensed Beds	624	563	**
Total Admissions	30,794	25,974	**
Trauma Hospital	63.9%	49.8%	**
Teaching Hospital	30.6%	24.6%	**
Full-Time Equivalent Employees	3,609	3,235	**
Total Annual Surgical Operations	23,429	17,489	**
Number of Operating Rooms	23.9	21.2	**
Urban Location	97.2%	96.8%	**
Hip and knee replacement			
Number of hospitals	93	1046	
Licensed Beds	484	399	**
Total Admissions	23,307	17,974	**
Trauma Hospital	49.6%	40.9%	**
Teaching Hospital	14.8%	16.5%	**
Full-Time Equivalent Employees	2,658	2,186	**
Total Annual Surgical Operations	17,913	13,258	**
Number of Operating Rooms	18.7	16.2	**
Urban Location	93.5%	93.4%	**

**
p<0.0001

Table 3

Impact of the 2006 changes in incentive structure to Premier Hospital Quality Incentive Demonstration

			Odds ratio for adverse outcome, After vs. Before the change in incentive design (95% CI)	
Adverse outcomes	Initial risk-adjusted rate for Premier hospitals	Risk-adjusted rate after incentive expansion for Premier hospitals	Simple pre-post analysis (adjusting for patient characteristics)	Independent effect of HQID incentive changes*
Coronary artery bypass surgery				
Mortality	3.1%	2.4%	0.70 (0.66, 0.75)	1.09 (0.90, 1.32)
Any complication	21.7%	23.2%	1.01 (0.94, 1.08)	1.13 (0.98, 1.29)
Serious complications	13.5%	13.6%	0.96 (0.92, 1.01)	1.06 (0.95, 1.19)
Hip and knee replacement				
Mortality	0.2%	0.2%	0.78 (0.61, 1.00)	0.85 (0.54, 1.32)
Any complication	4.2%	4.2%	0.89 (0.84, 0.95)	1.11 (0.91, 1.36)
Serious complications	2.8%	2.6%	0.79 (0.74, 0.84)	1.12 (0.95, 1.31)
Poor-performance hospitals (bottom 20%)				
Coronary artery bypass surgery				
Mortality	5.0%	4.6%	0.78 (0.69, 0.87)	1.22 (0.88, 1.70)
Any complication	28.0%	30.0%	1.13 (1.04, 1.23)	0.98 (0.80, 1.20)
Serious complications	18.6%	19.0%	0.99 (0.93, 1.05)	1.05 (0.82, 1.33)
Hip and knee replacement				
Mortality	0.7%	0.4%	0.77 (0.68, 0.88)	0.85 (0.59, 1.21)
Any complication	7.4%	6.0%	0.95 (0.87, 1.05)	0.83 (0.53, 1.29)
Serious complications	4.3%	4.7%	0.90 (0.85, 0.95)	1.30 (1.06, 1.60)

* The independent effect of the P4P incentive changes is derived from a difference-in-difference model that represents improvement in outcomes before and after the policy change in Premier hospitals compared to non-Premier hospitals, the control group not exposed to the P4P incentive changes. These logistic regression models also controls for patients characteristics and any differences in pre-implementation trends in outcomes between Premier and non-Premier hospitals.