Hospital Response to Pay-for-Performance Incentives

Lauren Hersch Nicholas
Institute for Social Research, University of Michigan

September 21, 2009

Mail: 426 Thompson Street, Rm. 3005, Ann Arbor, MI 48104
E-mail: lnichola@umich.edu
Phone: 734-764-2562

Abstract

Pay-for-Performance (P4P) mechanisms are increasingly common in health care despite the challenges in designing incentive-based compensation for multidimensional tasks and the potential for poor performance on unincentivized outcomes. I evaluate hospital response to a Medicare P4P demonstration program that offered hospitals incentives for high relative performance on composite measures of process and outcomes for five types of admissions. I test theoretical predictions that hospital response will be concentrated amongst those likely to receive bonus payments, that hospitals will improve scores by concentrating on easy rather than difficult tasks, and that patient outcomes for incentivized and unincentivized admissions will be unaffected by P4P incentives. Hospitals appear largely unmotivated by the P4P program, although participating hospitals increase efforts for some easy tasks. I find no evidence that P4P altered inpatient quality of care on targeted or unincentivized conditions. Bonuses were concentrated amongst already high-performing hospitals. Expansions of this type of P4P program are unlikely to improve patient outcomes.

1 Introduction

A series of recent research reports from the Institute of Medicine (IOM) has documented suboptimal quality of care and wide variation in patient outcomes at U.S. hospitals (IOM, 2000; 2001; 2007). Medicare, the federally funded health insurance program for the elderly and largest health care purchaser, spent nearly $160 billion on hospital care in 2006. Nearly 15 percent of this total represents spending on unplanned readmissions such as heart and kidney failure, hospital-acquired pneumonia, and post-operative infection or bleeding, reflecting poor quality of care (Jencks et al., 2009). Research suggests that hospitals could improve care delivery, hospitals fail to adhere to recommended process guidelines for common admissions including heart attack, hip fracture, and surgery nearly 50 percent of the time (McGlynn et al., 2003). Low quality care is costly to patients, who risk fatal complications and or extensive follow-up care, and to payers reimbursing the economic costs of additional services.

With hospital care accounting for nearly one-third of U.S health care spending and growing more rapidly than other health spending, payers are interested in emphasizing quality and efficiency of care (Centers for Medicare and Medicaid Services, 2009a). Volume-based hospital payments, which reimburse per admission regardless of outcome can encourage provision of low quality care while increasing spending on avoidable complications and readmissions (CMS, 2008; IOM, 2007). Accordingly, public and private payers are increasingly adopting pay-for-performance (P4P) strategies which align higher payments with better performance on targeted measures to stimulate quality improvement. P4P strategies are disseminating rapidly despite limited empirical evidence of their effectiveness.

Pay-for-performance is poised to play a large role in Medicare as policymakers look for cost-cutting measures to avoid insolvency projected to occur as early as 2019. Recognizing the business case for hospital quality, an IoM report recently recommended that Medicare should begin to implement pay-for-performance “as a stimulus to foster comprehensive and systemwide improvements in the quality of health care,” (IOM, 2007). The 2003 Medicare Modernization Act included Congressional mandates for the Centers for Medicare and Medicaid Services (CMS) to transition from a passive to a value-based purchaser, and current Congressional reform proposals envision a larger role for P4P in Medicare and Medicaid. In this paper, I consider the effects of an early CMS P4P initiative designed to improve quality of hospital care for all patients, the Premier Hosptial Quality Incentive Demonstration, to study hospital response to performance pay.

P4P strategies typically tie payments to measures of patient outcomes and or compliance with recommended processes of care such as use of beta blockers for heart attack patients. These strategies are believed to improve quality of care in at least two ways. First, hospitals are expected to improve performance on measured areas in order to qualify for bonus payments. Second, these efforts can have spillover effects for other patients as hospitals adopt their technology, staffing, and quality improvement efforts to meet P4P targets. Expanding use of electronic medication tracking systems to comply with incentives for timely pneumonia antibiotic use, for example, could improve adherence to medication protocols for patients with unincentivized admissions (Glazer et al., 2008).

Despite good intentions, whether P4P will improve quality and efficiency of health care is unknown (Petersen et al., 2006; IOM, 2007). Theoretical literature highlights the challenges in designing incentive-based compensation for multidimensional tasks such as health care delivery (Holmstrom, 1991; Prendergast, 1999). Lazear (2000) suggests that performance pay is ideal for tasks where output is easily measured, quality problems are easily detected and blame can be easily assigned; features that do not generally describe hospital care. The multitasking literature predicts that P4P programs concentrating incentives on a small number of admissions may adversely affect care in unincentivized areas if hospitals increase performance on measured areas at the cost of efforts on other important but unincentivized areas. Unintended consequences of P4P, such as "teaching to the test," directing resources only to measured types of care and "cherry-picking," providing care only to patients at low risk of adverse outcomes are possible. Several papers have demonstrated this response to the incentives created by public reporting in nursing homes (Lu, 2009) and hospital coronary artery bypass graft mortality (Dranove et al., 2003). Early public reporting initiatives, particularly the introduction of CABG report cards in New York and Pennsylvania, improved inpatient mortality rates, but also increased racial disparities in access as procedures were restricted to healthier patients at lower risk of adverse outcomes (Werner et al., 2005a,b; Dranove et al., 2003).

Relatively few studies have considered the effect of performance pay on health care quality, and the existing literature finds mixed evidence that P4P strategies actually improve quality of care. Mullen and colleagues (2009) and Rosenthal et al. (2005) find that several P4P efforts by commercial payers to improve outpatient care have little effect on targeted preventative services. In contrast, the introduction of P4P for primary care in England was associated with improvements in quality of care for asthma and diabetes (Campbell et al., 2007). Lindenauer et al. (2007) find that a subset of hospitals in the first two years of the Premier P4P demonstration showed larger gains in process compliance than hospitals only participating in public reporting, though Ryan (2009) did not find corresponding improvements in Medicare patient mortality on targeted conditions.

This paper makes several contributions to the existing literature. I first develop a conceptual framework to test hospital response to pay-for-performance incentives. The model shows that while payers hope to use P4P to improve quality of care, when bonuses are provided to a small number of hospitals based on relative rankings, the optimal response for most hospitals is unlikely to affect patient outcomes. I use Hospital Compare process compliance data to examine the effect of the Premier Demonstration on hospital efforts and State Inpatient Databases hospital discharge abstracts to study patient outcomes reflecting inpatient quality of care. During the study period, hospitals generally improve process compliance regardless of whether they participate in P4P. I show that when faced with incentive payments based on composite performance across different types of tasks, P4P hospitals modestly increase effort on easy, but not difficult targeted tasks. The increased effort is concentrated amongst hospitals with a high probability of receiving bonus payments or fines for failing to improve. The small responses observed suggest that hospitals perceive the bonus payments as weak incentives.

I extend Ryan's (2009) work to examine the quality effects of P4P on unincentivized as well as incentivized admissions amongst all patients hospitalized for targeted conditions. While Medicare patients comprise a majority of patients hospitalized for each of the incentivized conditions, between 25 and 45 percent of patients hospitalized for study conditions are not Medicare beneficiaries. Many clinical trials providing evidence-based support for the beneficial effects of compliance with the targeted process measures are conducted primarily in the under 65 (non-Medicare) population, so excluding this group may bias estimates of the P4P effect towards 0 if these measures are less salient for quality outcomes amongst the older population. I find little evidence that P4P affected inpatient outcomes for either targeted admissions or unincentivized medical and surgical admissions. though point estimates are consistent with performance pay adversely impacting patient outcomes. The absence of a spillover effect is consistent with hospitals viewing the bonus payments as weak incentives for targeted tasks.

2 Hospital Quality

Consider a simple model to analyze potential hospital responses to P4P incentives. Hospitals produce patient outcomes for n patients with multiple types of admission, i = 1, 2, ..., I with quality levels Q_i. Let higher levels of Q_i indicate higher quality.[1] Assuming that all hospitals treat identical patients, then quality of care for condition i is

$$Q_i = F(X_{1,2,...,J}, Z_{1,2,...,K}) \tag{1}$$

where X_j is a vector of J observable inputs and Z_k is a vector of K inputs that are unobservable to the payer or regulator. Elements of X_j and Z_k may be correlated.

The socially optimal quality level as the solution to the maximization $\sum_{i=1}^{I} n_i Q_i(X_{i,j}, Z_{i,k})$ subject to the budget constraint $c(X, Z) = C$. The socially optimal Q_i^* equalizes marginal benefits of spending on inputs across conditions. The hospital's choice of Q_i^h will fall below the socially efficient Q_i^* when the hospital faces additional costs of choosing a given combination of $(X_{i,j}, Z_{i,k})$. A higher level of Q_i^h may be costly to hospitals because of lost revenue from readmissions or complications or reduced total service volume (Horwitz and Nichols, 2007; Abelson, 2009).

Payers and regulators interested in motivating hospitals to produce Q_i^* face a mechanism design problem in offering hospitals a payment contract that aligns hospital and social objectives. Fee-for-service payments such as those used in the Medicare program fail to provide such incentives, since they are not directly tied to quality (i.e. higher payment rates to hospitals with lower 30-day mortality rates) or compensate hospitals for preventing repeat hospitalizations.[2]

Glazer et al. (2008) point out that if all elements of Q_i could be observed, hospitals could be motivated to choose an efficient level of output quality through a profile approach which scores hospital quality $S_h = \sum_{i=1}^{I} \alpha_i Q_i(X_j, Z_k)$ where weights on discharge types α_i can be specified by the regulator. When all elements of Q_i cannot be observed, but share common inputs, (X_j, Z_k), an optimal partial profile can be specified by develop-

[1] In subsequent applications, quality of care will be assessed using mortality and complication rates, where a lower level of Q_i indicates higher quality.

[2] Medicare pays hospitals according to a set fee schedule based on a patient's DRG (Diagnosis Related Group) code. Payments bundle hospital labor and services. Base payments are adjusted for factors such as graduate medical education, and whether the patient has a particularly long or resource-intensive stay triggering outlier payments.

ing weights α_j for a subset $Q_j \subset Q_i$ which reflect common factors of quality production across discharge types.

In practice, the set of hospital outcomes Q_i contains many elements that are difficult to measure. Ranking hospitals according to patient outcomes alone is challenging for many admission types i, such as hip or knee repair, where observable adverse events such as mortality are rare and subsequent improvements in patient mobility and quality of life are often unknown after discharge. Even when mortality is relatively common, many hospitals do not perform enough procedures for certain admission types for outcomes such as risk-adjusted inpatient mortality to reliably differentiate hospital quality from statistical noise (Staiger et al., 2009; Dimick et al., 2004). Payers attempting to use P4P are constrained by cost and feasibility of collecting data to monitor and rank hospital performance. One solution is to design incentives around process compliance, or inputs to the quality production function, under the assumption that patients will have better outcomes at hospitals with higher rates of compliance with evidence-based processes of care. Accordingly, the remainder of this paper focuses on the Premier Hospital Quality Incentive Demonstration, a process compliance-focused pay-for-performance program implemented by the CMS to improve hospital quality.

2.1 Premier Hospital Quality Incentive Demonstration

CMS introduced the Premier Hospital Quality Incentive Demonstration program in October 2003. At the same time, hospitals began voluntary public reporting of process compliance measures for patients admitted with three common medical admissions (acute myocardial infarction (AMI) or heart attack, heart failure, pneumonia). Measures were reported on the Hospital Compare website developed by a public-private alliance between hospitals and the CMS. CMS required that hospitals participate in public reporting by October 2004 in order to receive annual payment rate updates. Process compliance measures, defined as the number of times a hospital performed a recommended process divided by the number of opportunities, reflect inputs into a hospital's quality production function.

Prior to the introduction of public reporting, hospitals that already subscribed to Premier, a quality reporting and purchasing collective, were invited to participate in a P4P demonstration. Participants needed at least 30 annual admissions for targeted conditions. P4P participants were already subscribed to a quality reporting service and may be more motivated to improve quality of care than hospitals engaged in reporting only. 421 hospitals were invited to participate, and 255 completed the demonstration (Lindenauer et al., 2007).

Premier provides bonus payments B for participating hospitals in the top two deciles of each of 5 condition-specific composite measures comprised of a subset of observable inputs X_{ij} and outcome performance Q_i for the three public reporting medical conditions and two surgical procedures (coronary artery bypass graft (CABG) and hip/knee replacements). All composites are heavily weighted towards process compliance. Although risk-adjusted mortality is recognized as a quality indicator for all of the targeted admissions, it is included in the composites only for AMI and CABG. Figure 1 details the incentivized measures, each of which is given equal weight in the condition-specific composite score. In an effort to improve hospital quality for all patients, Medicare requires hospitals to report data on all patients treated for the targeted conditions, but pays bonuses only for Medicare-covered admissions.

Hospital quality scores are calculated using a two-stage process. The score for each component uses an opportunity model reflecting the number of times a task occurred relative to the number of opportunities for task completion, $s_x = \frac{1}{n}\sum_{i=1}^{n} success$ for each of x tasks and n eligible patients.[3] Scores for individual tasks are then averaged to generate the composite score for each condition, $S_i = \sum_{i=1}^{X} \frac{1}{X} \times s_x$. Thus, an individual AMI patient who died in the hospital would be included as an opportunity for beta-blockers at arrival, but not at discharge. Each of the nine AMI task measures contributes $\frac{1}{9}$ to the AMI composite score. Hospitals in the highest decile of composite score receive an annual bonus B_i equal to 2 percent of the DRG payment for all Medicare admissions with the incentivized condition, those in the second decile receive a 1 percent bonus. Hospitals that fail to improve above the lowest quintile of initial performance by the end of the third year face fines of 2 percent of DRG payments in the lowest decile and 1 percent in the second lowest decile of S_i. Composite scores and bonus payments are calculated separately for each condition.

2.2 Hospital Effort Under Pay-for-Performance

The goal of the Premier incentives is to motivate *additional* efforts on incentivized tasks relative to a hospital's initial efforts absent the demonstration ($x_1^o, ..., x_j^o, z_1^o, ...z_k^o$). Hos-

[3] For outcomes, "success" refers to risk-adjusted avoidance of the adverse event.

pitals decide how much effort to allocate to each incentivized condition by solving the problem

$$\overset{Max}{x_1, ..., x_j, z_1, ..., z_k} \sum_{i=1}^{I} n_i(Q_i(X_1, ..., X_j, Z_1, ..., Z_k) \quad s.t. \quad \sum_{i=1}^{J+K} c(X_j, Z_k) = C + E(B) \tag{2}$$

This formulation has several implications for hospital efforts.

Proposition 1: *Hospitals at most points in the initial quality distribution should not increase efforts under P4P incentives.* When bonus payments are based on relative rankings rather than absolute improvement or a fixed score cutpoint, E($B(S_i)$) = 0 for hospitals with average quality for condition i. For hospitals to increase efforts relative to their initial performance, the incentive compatibility constraint requires hospital utility (defined in this example as reimbursement for admission less cost of care) at higher levels of input to be greater than or equal to the initial allocation. Thus, most hospitals should not increase X_j in response to the P4P incentives unless they can catch up to already high-performers at costs at or below E(B). Hospitals near the bonus threshold face incentives to improve and high performers have weak incentives to maintain rankings through modest additional improvement in subsequent years. The poorest performing hospitals in the initial period face fines, F in subsequent years if composite scores remain in the lowest quintile of Year 1 performance. These hospitals should increase X_j when E(F) $> \sum_{i=1}^{x} C(X_j)$. Hospital response to the P4P incentives should be concentrated amongst the highest and lowest performing hospitals.

Proposition 2: *Performance pay should motivate greater effort on easier tasks.* Composite measures weight all incentivized tasks equally. Hospitals can minimize $C(X_1, ...X_j)$ by focusing on lower effort tasks. The expected bonus payment would be the same for an increase in the percentage of AMI patients receiving smoking cessation counseling (which can be accomplished by distributing an anti-smoking booklet during patient registration) or a same-sized improvement in inpatient survival, which may require changes on multiple tasks.

2.3 Hospital Quality and Patient Outcomes Under Pay for Performance

Proposition 3: *Quality of care for incentivized and unincentivized measures should not necessarily improve in response to performance pay.* Hospitals maximizing performance scores, and consequently bonus payments under Equation (2) will not necessarily improve quality of care on incentivized measures. The composite measures used to allocate bonus payments are comprised of elements which are independently associated with improved patient outcomes or recommended as practice guidelines; $\frac{\partial Q}{\partial X_j} > 0$. However, the elements of each composite include only a subset of factors which can affect patient mortality under Equation (1), (including the mortality rate itself for AMI and CABG).

If hospitals increase S_i by substituting effort across X_j and Z_k, the sign of the total derivative $\frac{dQ}{dP} = \frac{\partial Q}{\partial X} + \frac{\partial Q}{\partial Z}\frac{\partial Z}{\partial X}$ is ambiguous. The multitasking literature notes that incentives targeting an observable subset of job components can motivate agents to increase efforts measured tasks at the expense of the unmeasured efforts, particularly if incentives are strong (Holmstrom and Milgrom, 1991; Prendergast, 1999). In the hospital setting, reduced efforts on Z_t may offset the increases in P_x. Given gaps in scientific knowledge of how best to treat patients and variation in resources and technology across hospitals, it is unlikely that the simple P4P composite score profiles the socially optimal hospital quality production function.

The effect of performance pay on quality of care for unincentivized conditions is also ambiguous. If incentivized tasks are complementary across types of admission, improvements for incentivized admissions may result in quality spillovers for unincentivized conditions. Alternatively, hospitals may "teach to the test" and substitute efforts on unincentivized conditions for effort on incentivized tasks.

Generalizing from Proposition 1, the average teatment effect of performance pay on quality of care will be modest. In 1 and 2 we note the optimal hospital response to the Premier incentives for many hospitals is to do nothing because they will not receive bonus payments or to focus on easy tasks which may have little direct effect on patient outcomes. Thus, hospitals responding optimally are unlikely to improve quality for targeted conditions or adverse affect other types of admissions.

3 Data

This paper relies on data from several sources. I use measures of hospital compliance with process of care measures and number of admissions for AMI, Heart Failure and Pneumonia reported in Hospital Compare to compare hospital efforts under P4P and public reporting. Hospital Compare measures are collected first through the voluntary Hospital Quality Alliance and then under the CMS Reporting Hospital Quality Data for

Annual Payment Update initiative and cover fiscal years 2003 - 2006. Hospital Compare measures are posted on the Hospital Compare website (www.hospitalcompare.hhs.gov) with a 9 month lag Data are available for 243 Premier hospitals and 3,100 non-Premier hospitals.

Consistent with the demonstration participation guidelines, the analytic sample is restricted to 192 Premier and 1,596 comparison hospitals with at least 30 admissions for each of the incentivized medical conditions in Years 1 and 3 of P4P/public reporting. This inclusion criteria ensures that only hospitals voluntarily participating in their respective initiative are compared. Critical access hospitals, which tend to be small hospitals in rural areas and receive cost-based reimbursement from Medicare, are excluded from the sample. Hospitals were initially required to report compliance with 10 process compliance measures covering AMI, heart failure and pneumonia, other measures are added over time but not reliably reported during the P4P period. Since P4P and public reporting were introduced simultaneously, pre-P4P measures of process compliance are unavailable.

Table 1 summarizes rates of process compliance for Premier and control group hospitals. Both groups make large improvements in process compliance during the study period. On average, P4P hospitals outperform comparison hospitals at both baseline and Year 3, though both groups improve during the period.

Hospital characteristics from the 2003 American Hospital Association (AHA) Annual Survey are included as control variables to isolate the effect of participation in the Premier demonstration from other underlying differences between Premier and non-Premier hospitals. Hospital characteristics include profit status, teaching hospital, whether the hospital's mission includes community benefit, the number of full-time doctors, nurses and residents on staff, percentage of admissions reimbursed by Medicare and Medicaid, and whether the hospital is part of a larger network. Appendix Table A1 details hospital characteristics. Premier and comparison hospitals are similarly reliant on Medicare and Medicaid for reimbursement for nearly 60 percent of admissions. Premier hospitals are more likely to be teaching hospitals (44% vs. 39%), to be part of a hospital network (45% vs. 32%) and average higher numbers of nursing staff (392 vs. 329).

Outcome measures are calculated using the State Inpatient Databases (SID) from Florida and New York. The SID data are collected by the Healthcare Cost and Utilization Project and include the discharge abstract for all in-state inpatient hospitalizations. Discharge abstracts contain patient demographics, payer information, and ICD-9 procedure and Hospitals are observed for three fiscal years prior to the introduction of public reporting/P4P (October 2000 - September 2003) and for three years of these efforts (October 2003 - September 2006). The SID data include 39 hospitals in the Premier demonstration and 351 non-Premier hospitals. Although P4P bonuses were only paid for the care of Medicare patients, hospital scores were calculated across all admissions for eligible patients, so all-payer data is necessary to estimate the effect of P4P across all treated patients.

The main outcomes of interest are risk-adjusted mortality rates for incentivized conditions: AMI, Heart Failure, Pneumonia, CABG and Hip/Knee replacement, and unincentivized medical (gastrointestinal hemorrhage and stroke) and surgical (abdominal aortic aneurysm repair (AAAR), non-AMI angioplasty, carotid endarectomy (CEA), craniotomy, esophageal resection and pancreatic resection). The surgical procedures include higher risk, less common (AAAR, CABG, craniotomy, esophageal and pancreatic resection) and more common (angioplasty, CEA, hip/knee replacement) procedures. Commonalities in production may increase the likelihood of spillover effects from incentivized admissions for cardiac and vascular surgeries, (angioplasty, CEA and AAAR) relative to the resections, which are high-risk surgeries for cancer patients.

International Classification of Diseases (ICD-9) procedure and diagnostic codes are used to identify admission types of interest, following Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators guidelines (AHRQ, 2007). Each patient's probability of dying in-hospital is predicted using logistic regression to adjust for age, race, sex, payer, and dummy variables for comorbidities for risk-adjustment.[4] The hospital-level risk-adjusted mortality rate y_{ht} is calculated as: $y_{ht} = MR_t \times (\sum_{j=1}^{n} d_{jt}) / \sum_{j=1}^{n} p_{jt})$ where MR is the observed population mortality rate, d_{jt} is observed mortality and p_{jt} is expected mortality for patient j in period t controlling for the comorbidities and demographics listed above.

I also examine risk adjusted rates of hemorrhage and hematoma (post-surgical bleeding) and metabolic derangement rates (post-surgical diabetic complications or kidney failure) for CABG and hip and knee replacement, additional outcome measures included in P4P composite. These are calculated in the same manner as risk-adjusted mortality. For hip and knee replacement, outcomes are only incentivized for Medicare patients, allowing comparison between Medicare and non-Medicare patients. Adverse outcomes are

[4] These include chronic pulmonary disease, congestive heart failure, cerebrovascular disease, peripheral vascular disease, diabetes, liver disease, renal disease, AIDS, malignant and non-malignant tumors, alcoholism, deficiency anemia, obesity, ulcers, weight loss neurological impairment, paralysis, lymphoma, arthritis, lymphoma, hypertension, and hypothyroidism.

very rare with hip and knee replacement, so I exclude metabolic derangement. One P4P-targeted process measure, CABG using the internal mammary artery, has its own set of ICD-9 codes and is also identified in the SID data. The latter is a process measure, so higher rates indicate better performance, while lower rates indicate higher quality for the adverse events.

Table 2 describes quality of hospital care as measured by inpatient mortality for incentivized and unincentivized conditions. Patient outcomes were typically better (i.e. lower mortality) at Premier hospitals prior to the introduction of P4P and remained better after P4P relative to comparison hospitals. However, improvements in inpatient mortality rates during the P4P period were larger in magnitude in control hospitals for all of the incentivized admissions except hip and knee replacement. Differences in improvements in pneumonia mortality were particularly striking; P4P hospitals reduced average inpatient mortality by 1.4 percentage points while control hospitals posted reductions of 2.2 percentage points.

4 Empirical Approach

I examine hospital response to a pay-for-performance demonstration implemented between October 2003 and September 2006. Participating hospitals are compared to non-participants to understand the ways hospitals change inputs into their quality production functions and the impact of these responses on quality outcomes. During the P4P period, all hospitals are required to participate in public reporting of process compliance for certain admissions. The Medicare program continues to expand hospital public reporting requirements to receive full payment rates, so the policy-relevant P4P effect is relative to any improvements which are achieved through public reporting only.

4.1 Hospital Effort Under Pay-for-Performance

Hospital process compliance data from the three medical admissions, AMI, heart failure and pneumonia, during the P4P period facilitate tests of Propositions 1 and 2. Since hospital response to P4P is expected to vary based on bonus (fine) eligibility, hospitals are classified into quintiles of P4P Year 1 performance on each of the three medical conditions.[5]

[5] Fines will be assessed for hospitals that do not improve above Year 1 performance thresholds by the end of Year 3, so improvements for these hospitals should be concentrated in Years 2 and 3. Pre-demonstration process compliance data is unavailable, so this classification process necessarily misses any changes in

Comparison hospitals are ranked according to where they fall in the P4P distribution.

4.1.1 Hospital Efforts and Initial Ranking

I test for heterogenous response to P4P incentives by estimating:

$$\triangle S_{i,h} = \alpha + \beta(P4P_h) + \delta(Q^o_{i,h}) + \gamma(P4P_h * Q^o_{i,h}) + \kappa(H_h) + \varepsilon_{i,h} \tag{3}$$

where $\triangle S_{i,h}$ is the change in composite performance for hospital h on condition i from between Years 1 and 3 of the demonstration; $P4P_h$ indicates participation in the P4P demonstration; Q^o_h is a vector of dummy variables indicating quintile of initial performance score for condition i; $P4P_h * Q^o_h$ is an interaction term and H is a vector of hospital characteristics from the AHA survey described previously. The reference category is hospitals of average quality (quintile 3), who are expected to be unaffected by P4P incentives since $\mathrm{E}(B(S_i)) = 0$.

Observations are weighted by the average number of eligible admissions for the targeted conditions. Equation (3) is estimated jointly for the 3 medical admissions to allow for correlations in $\varepsilon_{i,h}$ across conditions using a feasible generalized least squares application of Zellner's (1962) seemingly unrelated regressions (SUR) model which allows for correlation in the error terms across conditions i.

4.1.2 Hospital Efforts and Task Difficulty

Since the Premier composite weights performance on all tasks equally, the model predicts that hospitals should increase effort on easier rather than difficult tasks to improve performance scores. The difficulty of composite component tasks which are reliably reported in all 3 years of data are classified as follows:[6]

ranking occurring between the start of the demonstration and the beginning of Year 1.

[6] I am grateful to Justin Dimick, Theodore Iwashyna and Brahmajee Nallamothu for guidance in developing this classification scheme.

Condition	**Easy**	**More Difficult**
AMI	Aspirin at Discharge β-Blocker at Discharge Aspirin at Arrival β-Blocker at Arrival	ACE Inhibitors for Left Ventricular Dysfunction
Heart Failure	Left Ventricular Assessment Smoking Cessation Counseling Discharge Instructions	ACE Inhibitors for Left Ventricular Dysfunction
Pneumonia	Pneumococcal Vaccination Status Oxygenation Assessment	Blood Culture before Antibiotics Given Timing of Initial Antibiotic Receipt

I reestimate Equation (3) using two observations per condition indicating performance change on easy and more difficult task performance:

$$\triangle S_{i,h,e} = \alpha + \beta(P4P_h) + \delta(Q^o_{i,h}) + \gamma(P4P_h * Q^o_{i,h}) + \kappa(H_h) + \varepsilon_{i,h} \tag{4}$$

where e indicates easy or more difficult tasks. While I cannot directly assess the welfare implications of allocating efforts to easier versus more difficult tasks because the marginal costs of these tasks are unknown, published results from the medical literature provide information about the relative benefits of some of these tasks. For AMI patients, the 1-year mortality reduction associated with ACE-inhibitors is 0.057, nearly double the return on β-Blocker at discharge (0.034), the most effective of the "easy" tasks and ten times the return of β-Blocker at admission (Werner et al., 2008). Marginal costs for easy tasks such as distributing anti-smoking brochures and aspirin at discharge can be assumed to be close to 0, more difficult measures requiring specialized equipment and physician time are higher. Allocating efforts to generate the highest composite score S_i does not guarantee that efforts are concentrated amongst tasks with the highest expected mortality reductions.

The $\varepsilon_{i,h}$ correlation matrix provides information about the relatedness of hospital response across conditions and effort types. Spillover effects of P4P to unincentivized admissions are unlikely to occur if there are not common underlying determinants of performance across conditions.

4.2 Quality of Care

I next consider whether hospital participation in P4P motivates improvement in patient outcomes. Simultaneous efforts to improve hospital quality, particularly public reporting, suggest that pre-post comparisons of inpatient mortality rates will overstate the effect of P4P on hospital quality. Table 2 shows that risk-adjusted mortality rates are lower in P4P hospitals than non-P4P hospitals prior to the demonstration for most incentivized and unincentivized outcomes, indicating that a cross-sectional comparison may also overstate the P4P effect.

I use a difference-in-difference (DID) estimation strategy to control for unobserved hospital characteristics related to quality outcomes and participation in the P4P demonstration. The effect of the P4P incentives is identified as the difference in outcomes between the P4P and non-P4P hospitals in the P4P period less the differences between participants and non-participants in the pre-P4P years by estimating

$$Y_{ht} = \alpha + \beta(Post_t * P4P_h) + \gamma P4P_h + \delta Post_t + \psi X_{ht} + \rho T_{ht} + \epsilon_{ht} \tag{5}$$

where Post is an indicator for the demonstration period, P4P identifies Premier hospitals, $X_{h,t}$ is a vector of time-varying hospital characteristics including payer mix, patient demographics, and number of admissions, and T_{ht} is a linear time trend. Standard errors are clustered at the hospital level to address intra-hospital correlations in ϵ_{ht} (Bertrand et al., 2004).

I estimate several variants of Equation (5), first comparing only Year 3, the immediate pre-period to Year 6, the end of demonstration, then use all 6 of data. My preferred specification includes a vector of hospital fixed effects in Equation (5) to control for unmeasured hospital-specific characteristics influencing participation and hospital quality. Since mortality rates are proportions bounded by 0 and 1, the linear relationship imposed by Equation 2 may be incorrect. As an alternative, I also use generalized estimating equations (GEE) with risk-adjusted mortality distributed binomial with a logistic link function and an autoregressive (AR 3) error structure. The GEE estimates provide population averaged estimates of the cross-sectional and time series variation across hospitals, which are inconsistent under settings where the fixed effect model is appropriate.

The validity of the DID identification strategy depends on time trends in quality to be the same in Premier and non-Premier hospitals prior to the introduction of P4P and public reporting. I test this assumption by regressing each outcome on a set of year dummy variables, the P4P indicator and P4P $\times$Year interaction terms and testing the joint significance of the preperiod interaction terms. I fail to reject the differential trends hypothesis for all study outcomes. There is also no differential trend in procedure volume across P4P and comparison hospitals. Differential changes in volume could directly affect quality through the volume-outcome relationship or if hospitals responded by admitting fewer patients with high risk of mortality.

Risk-adjusted mortality poses several limitations as a dependent variable. Death is a rare event following many procedures. While risk-adjustment helps to minimize the contribution of patient clinical factors influencing death, additional random variation typically remains, especially in hospitals with relatively low procedure (admissions) volume (Dimick et al., 2009). I address this in two ways. The first is to limit the analytic sample to hospitals that have a minimum number of cases annually (at least 100 of each of 2 of the incentivized conditions for medical admissions, and more than the procedure specific thresholds for very low volume identified in Birkmeyer et al. (2002). I also use analytic weights inversely proportional to the variance in each outcome for reliability adjustment. Appendix Table A2 includes hospital procedure volume and payer mix. On average, Premier hospitals have higher admission rates for most procedures and are less reliant on public payers (40 vs. 44 % Medicare and 14 vs. 18 % Medicaid). Payer mix in SID states is similar to national averages reported in Appendix 1.

5 Results

5.1 Hospital Efforts

The base analysis (Table 3) indicates that hospitals did not increase efforts on composite task performance in response to P4P incentives. The $\beta(P4P_h)$ coefficient is statistically insignificant and small in magnitude for AMI, heart failure and pneumonia admissions. Results are similar when estimated separately by quintile of initial compliance or by hospital size (results not reported). I find little evidence of either an overall response or that hospitals differentially respond to P4P based on initial composite score. Under these results, the majority of hospitals (i.e. those unlikely to qualify for bonus payments), are behaving as expected. Rosenthal et al. (2005) noted a similar phenomenon in physician response to P4P incentives, finding that bonus payments were rewarded those already performing at bonus thresholds rather than motivate improvement as providers compete for bonuses.

Hospitals with very low initial quality scores make large improvements during the P4P/public reporting period regardless of whether they are at risk for fines, which may reflect hospital response to public reporting or other secular trends in hospital quality. Thus, the P4P incentives are likely to be relevant only for the highest performing hospitals. To ensure bonus payments in subsequent years, high-ranking hospitals may be more concerned with preserving rankings rather than making new improvements. Figure 2 shows this graphically. Composite scores for highly ranked P4P hospitals stay at or above Year 1 levels, while the score distribution flattens for non-P4P hospitals. Raw Year 3 scores are actually lower for the highest unincentivized Year 1 performers for AMI and heart failure.

While imprecisely estimated, the magnitudes of the high-performer P4P interactions for heart failure (0.45, s.e. = 1.08) and pneumonia (0.72, s.e. = 1.07) are large enough to impact hospital movement between deciles of bonus eligibility. In Year 1, hospitals with pneumonia compliance scores greater than 83.52 received 2 percent bonuses and hospitals with scores less than 80.32 were bonus-ineligible. Hospitals participating only in public reporting face little economic consequence of small year-to-year changes in score.

Table 3 also shows a strong positive correlation (nearly 0.6) between pairs of $\varepsilon_{i,j}$. A Breusch-Pagan Lagrange multiplier test rejects error independence across the three equations ($\chi^2(3) = 1812.63$), indicating common underlying elements of hospital response across the three conditions as expected. Spillover effects of P4P-induced improvements could potentially operate through similar commonalities in production.

5.1.1 Task Difficulty

In addition to creating incentives for modest increases in effort for hospitals at the margin of bonus eligibility, the Premier quality score methodology potentially distorts hospitals' choices of inputs. While the socially optimum profile would motivate choice of inputs equalizing the hospital marginal costs and patient benefits, under the P4P scoring mechanism, benefits of all targeted tasks are equalized regardless of underlying difficulty. Under Equation (4), I test whether P4P hospitals shift efforts to easy from more difficult tasks during the P4P period. If this occurred, $\beta(P4P_h)$ would be negative for more difficult tasks and positive for easier tasks. Table 4 reports regression results from the SUR model. P4P hospitals show greater improvements on easy tasks for AMI, (0.31 points, s.e. = 0.63), heart failure (3.27, s.e. = 1.74) and pneumonia (4.38, s.e. = 1.58), though only the pneumonia effect is statistically significant. Despite concerns that hospitals would shift efforts away from more difficult tasks, process compliance rates remain unchanged in P4P hospitals relative to control hospitals for AMI and heart failure and increase insignificantly for pneumonia. Amongst the initially highest performing hospitals, where we expect the P4P response to be concentrated, $\beta(P4P_h * Q5)$ remains insignificant though point estimates are consistent with modest increases in ACE-inhibitor use for left ventricular disfunction for both AMI and heart failure patients, the more difficult task. Note that hospitals have room to improve on both easy and difficult tasks, so results are not artifacts of ceilings on observable performance.

Hospitals in the lowest heart failure quintile also make larger gains in ACE-inhibitor use (7.27 percentage points, s.e. = 2.62); this appears to be at the cost of gains in easier tasks ($\beta(P4P_h * Q5)$ = -2.11, s.e. = 2.52). There is no relationship between unobserved determinants of efforts on easy and more difficult tasks for heart failure ($\rho_{e,d} = 0.02$) or pneumonia ($\rho_{e,d} = 0.08$). Performance on the more difficult measures may be less responsive to P4P because they reflect physician decisions (and set practice patterns) rather than tasks that hospital managers can easily delegate as part of the intake or discharge process.

Results provide little evidence that hospitals strategically alter task performance to improve performance scores. However, results should be interpreted cautiously in light of several limitations. Hospital Compare data do not include all of the incentivized measures including two easy (smoking cessation counseling for AMI patients and flu shots for pneumonia patients) and three more difficult measures (thrombolytic administered within 30 minutes of arrival and time to primary primary coronary intervention for AMI and appropriate antibiotic selection for pneumonia[7]), so results may understate substitution away from difficult tasks. Case studies on the Premier website highlight low-effort strategies for hospitals to improve performance scores such as distributing smoking cessation brochures at registration and implementing standing orders for aspirin at admission for all patients with chest pain (Premier, 2008).

Hospital Compare data also do not include unincentivized tasks, so it is unknown whether hospitals instead redirected efforts from untargeted tasks. Using data from another AMI quality improvement program including a subset of Premier and control hospitals, Glickman et al. (2007) find no difference in improvements over time on 7 unincentivized process measures in P4P hospitals relative to non-P4P hospitals and a modest

[7] Many of these variables are now reported on Hospital Compare, but were either not collected or reliably reported during the study period.

differential gain for only one (easy) measure; lipid-lowering medication at discharge.

5.2 Patient Outcomes

Regression estimates in Table 5 confirm the descriptive evidence that inpatient mortality rates did not improve under P4P. The P4P effect on incentivized and unincentivized quality outcomes is statistically indistinguishable from 0 for all study measures. The direction of $\beta(Post_t * P4P_h)$ is consistent with a decline in quality outcomes for most incentivized admissions. Hip/knee replacement mortality is the lone exception ($\beta(Post_t * P4P_h) = -0.47, s.e. = 0.22$), though the improvement is only statistically significant in the least restrictive model which uses a single year of data for the pre and post period, the coefficient increases to -0.18 deaths per 100 in the complete model and -0.04 in the GEE specification.

Across all admissions, i, specifications using all 6 periods of data with and without hospital fixed effects are very similar. There is little difference between the population-averaged GEE marginal effects and the fixed effect OLS coefficients which account for non-random participation in the demonstration.[8] Generally the estimated magnitudes of the P4P effect are also small; the 0.37 percentage point increase (s.e. = 0.43) for AMI is considerably smaller than published mortality rate reductions of 2.5 percentage points from giving aspirin and 0.6 percentage points from beta-blockers at admission. The pneumonia effect of 0.81 (s.e. = .51) is closer in magnitude to the effect of initial antibiotic timing (1.1), but less than a third of the size of vaccinating with the pneumococcal vaccine at admission (2.9) (Werner et al., 2008).

Estimates of $\beta(Post_t * P4P_h)$ for the unincentivized conditions are also statistically insignificant and small in magnitude. These results provide a partial validity test for the DID identification strategy. Given estimates of $\beta(Post_t * P4P_h)$ = 0 for incentivized conditions, it is unlikely that there are positive or negative spillover effects of P4P on unincentivized outcomes. Thus, nonzero estimates of $\beta(Post_t * P4P_i)$ for unincentivized conditions would be driven by another shock differentially affecting Premier hospitals during the treatment period. Estimates for the cancer surgeries, esophageal and pancreatic cancer, are large in magnitude but imprecisely measured. These are relatively rare, high-risk procedures, so the $\beta(Post_t * P4P_h)$ coefficient may be partially driven by statistical noise.

Table 5b examines correlations between the residuals in hospital mortality rates for the

[8] Pancreatic resection, one of the relatively rare unincentivized admissions, is an exception, while the estimate is always statistically indistinguishable from 0, the OLS estimate is 100 times larger.

subset of hospitals performing all procedures, revealing modest underlying commonalities across some but not all of the incentivized and unincentivized admissions.[9] Results from the models omitting hospital fixed effects (not shown) reflect a similar lack of correlation across underlying determinants of inpatient mortality. These findings highlight the challenge in selecting admission targets for performance pay that will plausibly generate positive spillovers for other types of admission. CABG, which shares unobserved determinants of quality with all of the incentivized conditions except for heart failure and many of the unincentivized surgical conditions (three vascular procedures, angioplasty, AAA repair and carotid enderectomy as well as esophageal resection), is one such example. Some states already use CABG mortality for hospital report cards as a sole indicator of surgical or cardiac quality of care.

The inpatient data also includes two adverse intermediate surgical outcomes indicating quality of care included in the Premier composite and one process of care measure. Table 6 presents the DID estimates of the P4P effect. The process measure, CABG using the internal mammary artery, reflects choice of operation, and under the effort allocation framework, classified as an easier task for P4P hospitals to achieve than reductions in risk-adjusted mortality or complication rates. Hospitals can set standing orders to always use IMA unless counter-indicated, for example. P4P exhibit large gains in process compliance with this measure, ($\beta(Post_t * P4P_h)$ = 6.64, s.e. = 2.53).

Similar to inpatient mortality, rates of major complications are unresponsive to P4P incentives. Targeted complications include hemorrhage/hematomas; internal bleeding resulting from surgery and post-surgical physical and metabolic derangement resulting from improper management of a patient's diabetes or other chronic conditions during surgery can be fatal or extend patient stays. Rates of derangement are common for CABG patients during the study period, averaging 9 per 100 operations and increasing by 0.62 under P4P incentives. Hemorrhages and hematomas also increase nonsignificantly for both CABG and hip/knee replacement patients in response to P4P incentives. The Premier incentives only include hip and knee complication rates for Medicare patients, who are older and possibly sicker than other hip replacement patients. Amongst this targeted population, rates complication rates significantly increase ($\beta(Post_t * P4P_h)$ = 0.61, s.e. = 0.22). Since complications are one potential determinant of inpatient mortality, the lack of improvement on complication rates is consistent with mortality results.

[9] Patient outcome regressions are estimated separately in order to use observations from hospitals treating patients for some but not all admissions of interest and to weight by condition-specific admissions volume.

5.2.1 Robustness: CABG for Medicare Patients

In the preceding analysis of patient outcomes, I focused primarily on inpatient mortality as an indicator of quality of care, both because it is widely accepted as a quality indicator and can be easily identified in all-payer data. However, P4P may have improved hospital quality on other dimensions which cannot be observed in the SID data, such as readmission. I test the robustness of the SID findings using all Medicare-covered admissions where CABG is the primary procedure for patients admitted to hospitals performing at least 30 procedures annually during the study period; 766,336 patients in 1,059 hospitals nationwide. I restimate Equation (5) for Medicare inpatient mortality rates as well as postoperative surgical site infection rates (a complication targeted by three of the incentivized CABG process measures), and 30, 90 and 365 day readmission rates for patients who survive the initial admission. All outcomes are risk adjusted as described previously.

The Medicare results support earlier inferences that the Premier Demonstration did not affect patient outcomes (Table 7). $\beta(Post_t * P4P_h)$ is statistically insignificant but positive in magnitude for all six of the CABG outcomes (consistent with an adverse effect of P4P on quality). The inpatient mortality effect is larger in magnitude (0.16, s.e. = 0.17) in the Medicare relative to all-payer data (0.07, s.e. = 0.24), but both are substantively small and imprecisely estimated. Despite several measures targeting infection reduction in Premier hospitals through appropriate antibiotic use, I find no improvements in rates of postoperative surgical site infection ($\beta(Post_t * P4P_h) = 0.11, s.e. = .10)$, which may reflect general hospital non-response to the Premier incentives or a lack of concordance between incentivized process measures and desired patient outcomes.

6 Conclusion

Performance pay is viewed as a promising strategy to foster improved health care quality by many policymakers and payers. This paper considers hospital response to a national Medicare demonstration program, the Premier Hospital Quality Incentive Demonstration, finding little empirical evidence that hospitals responded to program incentives to change allocation of effort or improve quality of care. Nearly 26 million dollars of bonus payments were allocated to already high-performing hospitals. Expansion of this approach would allow CMS and other payers to allocate higher payments to providers with higher rates of process compliance and lower risk-adjusted mortality in cross-sectional data, but is unlikely to stimulate further quality improvement. Results from the current study have important implications for the design of future P4P incentive schemes.

First, non-response to P4P incentives is the optimal response for many hospitals when incentives are based on relative performance rankings. Hospitals with average composite scores are unlikely to qualify for bonus payments, so there is no expected return on investments in improved process compliance. Furthermore, the relationship between hospital inputs and patient outcomes specified by the Premier scoring methodology is not necessarily the socially optimal resource allocation. I cannot rule out the possibility that hospitals do not change effort allocations to maximize bonus scores because they know that quality of care or another hospital objective would suffer.

Alternatively, hospital efforts and patient outcomes may not respond to incentives because bonus payments are too small. Two percent of DRG payments, the maximum bonus, is between $300 and $500 for Medicare patients. Although P4P reporting requirements cover all patients with targeted conditions, bonus payments are only made for the 55 to 70 percent of each class of admissions experienced by Medicare beneficiaries, so total bonus payments will be small even for hospitals eligible for the maximum bonus percentage. In contrast, the financial incentives associated with public reporting involve a two percentage point reduction in the annual payment rate update across all conditions for noncompliance. CMS provides implicit incentive payments for lower quality outcomes through the outlier payment system, which reimburses additional hospital costs for the sickest and longest-staying patients (including those triggered by hospital-acquired conditions and complications), and the potential for readmissions.

Hospitals could raise comparable levels of revenue by modestly increasing patient volume by attracting new patients (possibly by signalling high quality) or readmitting patients post-discharge (particularly amongst lower quality hospitals). Even when hospitals can improve P4P scores at very low marginal cost, the response is modest, suggesting that hospitals do not perceive the Premier payments as substantial enough to motivate changes in care delivery. While it is beyond the scope of this paper to assess the level of bonus payment that would motivate hospital response, it is clear that the Premier payments were inadequate to generate changes on intended or unintended dimensions.

As discussed in the theoretical framework, the heavy representation of process compliance measures in the scoring methodology provides incentives for hospitals to improve their scores without necessarily improving quality of care. Process compliance can be easily tracked, so that hospitals know where they stand relative to other hospitals in pursuit of bonus payments. Since process measures are used as quality indicators because of their association with patient outcomes in clinical trials and other studies, hospitals may expect outcomes to improve in response to increased compliance. However, the relationship between process compliance and outcomes is still poorly understood, as evidenced by the large changes in process compliance during the study period accompanied by little improvement in mortality. In addition to finding appropriate incentive payment amounts to trigger hospital response, payers and policymakers need to ensure that the tasks they incentivize correlate with anticipated patient outcomes.

If hospital process compliance continues to improve over time, bonus payments would be primarily allocated based on patient outcomes. This could make payments to hospitals more variable over time, particularly at smaller hospitals, where mortality and complication rates are noisy signals of hospital quality. Despite concerns (hopes) that negative (positive) spillovers from incentivized conditions will significantly alter quality of care for unincentivized admissions, empirical evidence in this paper suggests that this concern may be exaggerated.

My findings are similar to Mullen et al.'s (2009) and Rosenthal et al.'s (2005) results for doctors; P4P does not appear to noticeably improve quality on incentivized outcomes nor motivate providers to radically alter practice patterns. Absent anticipated main effects, concerns about unintended consequences are also unfounded. Consistent with Ryan's study of Medicare beneficiaries, I find that the Premier demonstration did not improve inpatient mortality rates for all patients hospitalized with incentivized conditions, nor did it alter quality of care for unincentivized conditions.

These results raise questions about the role of P4P in value-based purchasing. Payers and policymakers are unlikely to improve patient outcomes through pay-for-performance without addressing several mechanism design challenges. The first is to be relevant for all points in the quality distribution. This could be accomplished for example by offering bonus payments tied to overall rank as well as greatest improvement, or for crossing a certain score threshold. Requirements could vary based on initial hospital quality. Relevant research from other applications of performance pay, such as Neal and Schanzenbach's (2007) evaluation of No Child Left Behind, highlights the pitfalls of targeting high-powered incentives to a single threshold- schools appear to concentrate efforts only on students close to the accountability thresholds.

A second area is to align scientific knowledge about the quality production function for incentivized and unincentivized conditions with economic incentives to improve or maintain high quality. Prescribing the same quality formula for all hospitals by scoring composite process and outcome measures may not be the most effective way to motivate improvements in individual hospitals. Finally, bonus payments need to be large enough to motivate hospital response without increasing reimbursement rates above the value of lives saved or hospitalizations avoided. Alternative payment approaches, such as episode payment bundling that create financial incentives to limit readmission, for example, could create variable incentive amounts for hospitals and allow hospitals to identify the changes in production generating the necessary improvements in patient outcomes.

REFERENCES

1. Abelson, R. “Hospitals Pay for Cutting Costly Readmissions,” *The New York Times*, 9 May 2009, B1.

2. Agency for Healthcare Research and Quality. “Guide to Inpatient Quality Indicators.” 2007: Agency for Healthcare Research and Quality.

3. Asch SM, McGlynn EA, Hogan MM, Hayward RA, Shekelle P, Rubenstein L, Keesey J, Adams J, Kerr EA. “Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample,” Annals of Internal Medicine. 2004;141(12):938-45.

4. Bertrand M, Duflo E, Mullainathan S. “How Much Should We Trust Differences-in-Differences Estimates?” Quarterly Journal of Economics, 2004; 119(1):249-75.

5. Berwick D, DeParle NA, Eddy DM, Ellwood PM, Enthoven AC, Halvorson GC, Kizer KW, McGlynn EA, Reinhardt UE, Reischauer RD, Roper WL, Rowe JW, Schaeffer LD, Wennberg JE, Wilensky GR. “Paying for Performance: Medicare Should Lead,” Health Affairs. 2003; 22(6) 8-10.

6. Birkmeyer JB, Siewers AE Finlayson EV, Stukel TA, Lucas FL, Batista I, Welch G, Wennberg D. “Hospital Volume and Surgical Mortality in the United States,” The New England Journal of Medicine, 2002;(346):1128-37.

7. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. “Quality of Primary Care in England with the Introduction of Pay for Performance,” N Engl J Med, 2007; 357(2): 181-190.

8. Centers for Medicare and Medicaid Services. ”National Health Expenditures by Type of Service and Souce of Funds”. 2009: CMS.

9. Centers for Medicare and Medicaid Services. Roadmap for Implementing Value Driven Healtcare in the Traditional Medicare Fee-for-Serice Program. 2009b.

10. Dimick JB, Staiger DO, Birkmeyer JD. “Are we being fooled by randomness? Using reliability adjustment to improve hospital rankings,” Association for Academic Surgery and Society of University Surgeons- Abstracts. 2009; 297.

11. Dimick JB, Welch HG, Birkmeyer JD. “Surgical mortality as an indicator of hospital quality: the problem with small sample size,” Journal of the American Medical Association, 2004; 292(7):847-51.

12. Glazer J, McGuire T, Normand SL. “Mitigating the Problem of Unmeasured Outcomes in Quality Reports,” B.E. Journal of Economic Analysis and Policy, 2008; 8(2): Article 7.

13. Glickman SW, Ou F, DeLong ER, Roe MT, Lytle BL, Mulgund J, Rumsfeld JS, Gibler WB, Ohman EM, Schulman KA, Peterson ED. "Pay for Performance, Quality of Care and Outcomes in Acute Myocardial Infarction," Journal of the American Medical Association, 2007; 297(21): 2373-2380.

14. Holmstrom B, Milgrom P. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," Journal of Law, Economics, and Organization, 1991; 7(Special Issue): 24-52.

15. Horwitz J, Nichols A. "What Do Non-Profits Maximize? Non-Profit Hospital Service Provision and Market Ownership Mix," NBER Working Paper 13246, 2007.

16. Institute of Medicine. To Err is Human. Washington, DC: National Academies Press, 2000.

17. Institute of Medicine. Crossing the Quality Chasm. Washington, DC: National Academies Press, 2001.

18. Institute of Medicine. Rewarding Provider Performance. Washington, DC: National Academy Press, 2007.

19. Jencks S, Williams, MV and Coleman, EA. "Rehospitalizations among patients in the Medicare Fee-for-service program," The New England Journal of Medicine, 2009; 360:1418-1428.

20. Lazear EP. "Performance Pay and Productivity," The American Economic Review, 2000; 90(5): 1346-1361.

21. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. "Public Reporting and Pay for Performance in Hospital Quality Improvement," New England Journal of Medicine, 2007; 356(5): 486-496.

22. Lu SF. "Multitasking, Information Disclosure and Product Quality: Evidence from Nursing Homes," 2009: Working Paper.

23. Mullen KJ, Frank RG, Rosenthal MB. "Can You Get what You Pay For? pay-for-Performance and the Quality of Healthcare Providers," NBER Working Paper 14886, 2009.

24. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. "The quality of health care delivered to adults in the United States," N Engl J Med. 2003; 348(26):2635-45.

25. Neal D and Schanzenbach DW. "Left Behind by Design: Proficiency Counts and Test-Based Accountability," NBER Working Paper 13293, 2007.

26. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. “Does pay-for-performance improve the quality of health care?” Ann Intern Med. 2006;145(4):265-72.

27. Premier Inc. “Promted by CEO Mandate, Bon Secours Zooms Ahead: Rapid Improvement through Focused, Persistent Effort,” Case Study available at http://www.premierinc.com/quality-safety/case-studies/bon-secours-rip0106.pdf. Accessed 15 May 2009.

28. Prendergast C. “The Provision of Incentives in Firms,” Journal of Economic Literature, 1999; 37(1): 7-63.

29. Rosenthal MB, Frank RG, Li Z, Epstein AM. “Early Experience with Pay-for-Performance from Concept to Practice,” Journal of the American Medical Association, 2005; 294(14):1788-1793.

30. Ryan A. “Effects of the Premier Hospital Quality Incentive Demonstration on Medicare Patient Mortality and Cost,” Health Services Research, 2009: 44(3): 821-842.

31. Staiger DO, Dimick JB, Baser O, Fan Z, Birkmeyer JD. “Empirically derived composite measures of surgical performance,” Medical Care, 2009: 47(2): 226-33

32. Werner RM. Asch, DA. “The unintended consequences of publicly reporting quality information”, JAMA. 2005; 293(10):1239-1244.

33. Werner RM. Asch DA, Polsky D. “Racial Profiling: The Unintended Consequences of coronary artery bypass graft report cards,” Circulation, 2005; 111(10)1257-63.

34. Werner RM, Bradlow ET, Asch DA. “Does Hospital Performance on Process Measures Directly Measure High Quality Care or Is It a Marker of Unmeasured Care?” Health Services Research. 2008; 43(5): 1464-1484.

35. Zellner A. “An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias,” Journal of the American Statistical Association, 1962; 57(298): 507-530.

Figure 1: **Incentivized Admissions and Components of the Premier Hospital Quality Incentive Demonstration Composite Scores**

MEDICAL
Heart Attack (Acute Myocardial Infarction)
Aspirin at arrival
Aspirin prescribed at discharge
ACE inhibitor or Angiotensin II Receptor Blocker for left ventricular systolic dysfunction
Beta blocker at arrival
Beta blocker prescribed at discharge
Smoking cessation counseling
Time to primary coronary intervention
Thrombolytic within 30 minutes of arrival
Risk-adjusted mortality rate
Heart Failure
Left ventricular function assessment
ACE inhibitor (ACE-I) or Angiotensin II Receptor Blocker (ARB) for LVSD
Smoking Cessation Counseling
Detailed discharge instructions
Pneumonia
Timing of receipt of initial antibiotic following hospital arrival
Pneumococcal vaccination status
Blood culture performed before first antibiotic received in hospital
Oxygenation assessment
Influenza screening
Appropriate antibiotic selection
SURGICAL
Coronary Artery Bypass Graft
Aspirin at discharge
CABG using internal mammary artery
Prophylactic antibiotic pre-surgery
Appropriate antibiotic selection
Antibiotics discontinued post-surgery
Risk-adjusted metabolic derangement rate
Risk-adjusted hemorrhage/hematoma rate
Risk-adjusted mortality rate
Hip/Knee Replacement (Medicare beneficiaries only)
Prophylactic antibiotic pre-surgery
Appropriate antibiotic selection
Antibiotics discontinued post-surgery
Risk-adjusted metabolic derangement rate
Risk-adjusted hemorrhage/hematoma rate
30-day readmission

Figure 2: **Trends in Process Compliance in P4P and Comparison Hospitals by Quintile of Initial Performance, Demonstration Years 1-3**

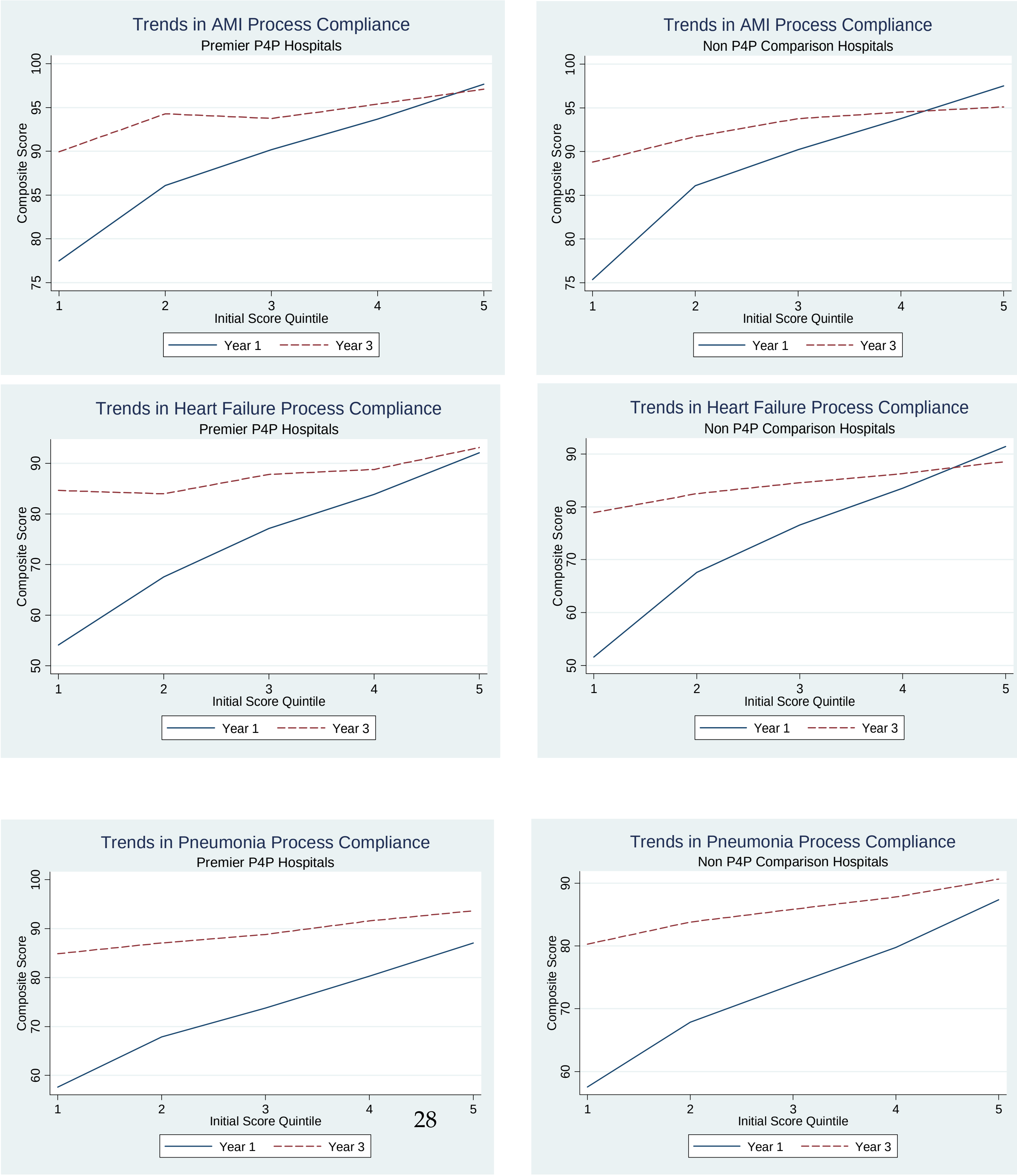

Table 1: Hospital Rates of Process Compliance, Premier P4P and Control Hospitals, FY 2003 - 2006

	Premier			Control			
	Year 1	Year 3	Change	Year 1	Year 3	Change	Diff-in-Change
AMI Composite	89.1	94.1	5.0	88.1	92.6	4.5	0.5
	(7.2)	(4.8)		(8.7)	(5.6)		
Aspirin at Arrival	94.3	96.9	2.6	94.4	96.2	1.9	0.8
	(4.8)	(3.2)		(5.1)	(3.7)		
Aspirin at Discharge	93.2	95.9	2.7	90.9	94.4	3.5	-0.8
	(7.5)	(5.4)		(10.7)	(7.0)		
B-Blocker at Arrival	89.3	94.3	5.0	88.7	93.0	4.3	0.7
	(8.2)	(4.8)		(9.9)	(6.1)		
B-Blocker at Discharge	90.9	95.9	5.0	89.3	94.6	5.3	-0.3
	(8.8)	(4.9)		(10.6)	(6.1)		
Smoking Cessation	81.3	96.5	15.2	77.5	92.8	15.3	-0.1
	(23.7)	(8.3)		(27.8)	(13.6)		
ACE-Inhibitors for LVD	77.8	87.5	9.7	77.1	84.9	7.8	2.0
	(16.2)	(12.1)		(19.1)	(12.3)		
Heart Failure Composite	74.8	87.6	12.8	71.3	83.5	12.2	0.5
	(13.3)	(8.1)		(14.8)	(9.7)		
Left Ventricular Assess	87.0	94.9	7.9	86.5	92.9	6.4	1.6
	(9.3)	(5.2)		(10.3)	(6.8)		
Smoking Cessation	74.7	94.3	19.6	67.8	90.0	22.2	-2.6
	(22.7)	(9.5)		(27.2)	(13.2)		
Discharge Instructions	55.9	73.2	17.3	47.4	66.5	19.0	-1.7
	(27.2)	(18.2)		(28.6)	(21.1)		
ACE-Inhibitors for LVD	76.8	87.8	11.0	75.1	84.6	9.5	1.5
	(13.8)	(8.4)		(14.5)	(9.6)		
Pneumonia Composite	72.0	88.8	16.8	69.8	84.5	14.7	2.1
	(10.4)	(5.4)		(10.7)	(7.8)		
Oxygenation Assess.	98.7	99.7	1.0	98.6	99.6	1.0	0.1
	(2.5)	(0.7)		(3.1)	(1.2)		
Pneumonia Vac. Status	47.1	79.4	32.3	40.3	70.1	29.9	2.4
	(26.3)	(13.8)		(25.5)	(19.7)		
Blood Culture	82.9	92.1	9.2	82.0	89.5	7.5	1.7
	(8.2)	(5.3)		(9.9)	(6.7)		
Antibiotic Timing	68.2	79.9	11.7	68.4	77.0	8.6	3.1
	(10.4)	(8.0)		(11.8)	(10.6)		
Observations	192			1596			

Centers for Medicare and Medicaid ServicesHospital Compare data from FY 2003 and 2006. Hospital compliance rates with evidence-based process of care measures (100* # times task completed/# eligible patients). Following Premier scoring guidelines, composite measures are unweighted averages of compliance on each component task. Standard deviations in parentheses.

Table 2: Risk-Adjusted Mortality & Complication Rates,
Premier P4P and Control Hospitals, FY 2000 - 2006

	Premier			Control			
	Pre	**Post**	**Diff**	**Pre**	**Post**	**Diff**	**DID**
Incentivized							
Acute Myocardial Infarc.	7.8	6.8	-1	8.9	7.4	-1.5	0.5
	(1.8)	(1.8)		(2.4)	(2.3)		
Heart Failure	4.99	4.32	-0.67	5.1	4.2	-0.9	0.23
	(1.6)	(1.7)		(1.9)	(1.9)		
Pneumonia	7.6	6.2	-1.4	9	6.8	-2.2	0.8
	(4.8)	(4.6)		(9.1)	(9.3)		
Coronary Artery Bypass	3.4	3.1	-0.3	3.6	3.2	-0.4	0.1
	(1.0)	(0.9)		(1.2)	(1.3)		
Hip/Knee Replacement	0.94	0.48	-0.46	0.75	0.48	-0.27	-0.19
	(1.8)	(0.5)		(0.8)	(0.6)		
Derangements: CABG	5.8	9.2	3.4	6.6	9.1	2.5	0.9
	(2.4)	(3.3)		(2.4)	(3.6)		
Hemorrhage: CABG	5.3	5.6	0.3	5.7	5.8	0.1	0.2
	(2.2)	(2.4)		(2.3)	(2.2)		
CABG:Internal Mam Artery	80.1	84.4	4.3	79.8	81.4	1.6	2.7
	(7.7)	(5.5)		(7.5)	(9.0)		
Hemorrhage: Hip/Knee	0.9	1	0.1	0.88	0.79	-0.09	0.19
	(1.0)	(0.9)		(2.4)	(1.1)		
Unincentivized							
Gastrointestinal Hemor.	3.3	3	-0.3	3.6	3	-0.6	0.3
	(1.4)	(1.4)		(1.8)	(1.7)		
Stroke	12.2	11.2	-1	12.2	11.3	-0.9	-0.1
	(4.4)	(3.3)		(4.2)	(4.2)		
AAA Repair	8.3	6.1	-2.2	8.4	5.8	-2.6	0.4
	(7.8)	(8.0)		(9.4)	(8.1)		
Angioplasty	0.86	0.79	-0.07	1	0.88	-0.12	0.05
	(0.4)	(0.6)		(0.6)	(0.5)		
Carotid Endarectomy	0.78	0.58	-0.2	0.68	0.56	-0.12	-0.08
	(1.3)	(0.8)		(1.1)	(1.1)		
Craniotomy	7.8	6.2	-1.6	7	6	-1	-0.6
	(4.4)	(3.1)		(4.5)	(3.8)		
Esophageal Resection	14.5	5.4	-9.1	9.6	7.6	-2	-7.1
	(20.0)	(17.2)		(35.4)	(18.1)		
Pancreatic Resection	13.9	7.4	-6.5	7.7	5.5	-2.2	-4.3
	(28.3)	(14.1)		(24.4)	(16.0)		

Annual hospital-level rates per 100 admissions risk-adjusted for patient demographics and comorbidities.
HCUP State Inpatient Data from Florida and New York, October 2000 - September 2006.
39 P4P and 351 control hospitals. Standard deviations in parentheses.

Table 3: Improvements in Process Compliance in Premier P4P and Control Hospitals, Oct 2003 - Sept 2006

	AMI	Heart Failure	Pneumonia
P4P	1.17	-0.38	0.77
	(0.62)	(0.78)	(0.71)
Quintile 1	12.1**	23.0**	14.7**
	(0.30)	(0.34)	(0.30)
Quintile 2	3.48**	8.05**	5.61**
	(0.29)	(0.34)	(0.32)
Quintile 4	-3.19**	-6.94**	-5.56**
	(0.29)	(0.37)	(0.33)
Quintile 5	-6.21**	-14.0**	-12.6**
	(0.30)	(0.38)	(0.41)
P4P*Q1	-1.88*	-0.47	0.46
	(0.90)	(1.07)	(0.90)
P4P*Q2	-0.17	2.63**	-0.18
	(0.82)	(1.00)	(0.90)
P4P*Q4	-0.14	1.11	-0.42
	(0.80)	(1.03)	(0.95)
P4P*Q5	0.14	0.45	0.72
	(0.85)	(1.08)	(1.07)

3b. Residual Correlation Matrix

AMI Difficult	1.00		
Heart Failure	0.59	1.00	
Pneumonia	0.59	0.56	1.00

Seemingly unrelated regressions of change in hospital process compliance on P4P participation and quintiles of Year 1 performance ranked by the P4P hospital distribution. Quintile 1 has lowest performance, average hospitals (quintile 3) are the reference category. CMS Hospital Compare data. Standard errors in parentheses. * Significant at 5%, ** significant at 1%. Changes in composite performance scores between 2003 and 2006. 192 Premier P4P hospitals and 1,596 control hospitals reporting data in all years. Models control for hospital ownership, payer mix, teaching status and mission. 3b represents the residual correlation from the SUR model.

Table 4: Changes in Hospital Effort on Easy and More Diffcult Incentivized Tasks: Premier P4P and Control Hospitals, Oct 2003 - Sept 2006

	AMI		**Heart Failure**		**Pneumonia**	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
P4P	0.31	0.01	3.27	0.2	4.38**	1.65
	(0.63)	(2.17)	(1.74)	(1.83)	(1.58)	(1.27)
Quintile 1	8.17**	17.30**	25.6**	6.49**	12.1**	3.95**
	(0.33)	(1.14)	(0.80)	(0.83)	(0.70)	(0.56)
Quintile 2	2.32**	3.66**	9.07**	2.39**	6.19**	0.41
	(0.32)	(1.11)	(0.82)	(0.85)	(0.76)	(0.61)
Quintile 4	-1.80**	-7.42**	-6.12**	-4.01**	-3.88**	-1.93**
	(0.32)	(1.11)	(0.88)	(0.92)	(0.79)	(0.64)
Quintile 5	-3.13**	-13.49**	-11.56**	-8.85**	-9.20**	-4.85**
	(0.33)	(1.14)	(0.91)	(0.94)	(0.96)	(0.77)
P4P*Q1	-0.82	0.01	-2.11	7.27**	-0.36	2.2
	(0.99)	(3.39)	(2.52)	(2.62)	(2.11)	(1.70)
P4P*Q2	0.69	4.64	0.03	-1.22	-4.18*	2
	(0.90)	(3.11)	(2.37)	(2.46)	(2.13)	(1.72)
P4P*Q4	0.22	2.92	-1.41	0.85	-3.88	1.18
	(0.88)	(3.01)	(2.43)	(2.52)	(2.24)	(1.81)
P4P*Q5	0.27	3.46	0.98	2.05	-2.46	0.49
	(0.93)	(3.19)	(2.53)	(2.64)	(2.49)	(2.01)

4b. Residual Correlation Matrix

AMI Easy	1.00					
AMI Difficult	0.15	1.00				
Heart Failure Easy	0.23	0.18	1.00			
Heart Failure Difficult	0.31	0.31	0.02	1.00		
Pneumonia Easy	0.17	0.11	0.21	0.19	1.00	
Pneumonia Difficult	0.12	0.06	0.03	0.16	0.08	1.00

Seemingly unrelated regressions of change in hospital process compliance by task difficulty on P4P participation and quintiles of Year 1 performance ranked by the P4P hospital distribution. CMS Hospital Compare data. Standard errors in parentheses. * Significant at 5%, ** significant at 1%. Changes in composite performance scores between 2003 and 2006. 192 Premier P4P hospitals and 1,596 control hospitals reporting data in all years. Models control for hospital ownership, payer mix, teaching status and mission. 4b represents the residual correlation from the SUR model.

Table 5: Pay for Performance and Risk-Adjusted Inpatient Mortality
Florida and New York, Oct 2000 - Sept 2006

	Pre-Post Comparison			6 Year FE DID		GEE
	P4P	**Post**	**P4P*Post**	**Post**	**P4P*Post**	**P4P*Post**
Incentivized						
Acute Myocardial Infarc.	-0.92**	-1.76**	0.37	0.36	0.37	0.23
	(0.31)	(0.17)	(0.43)	(0.18)	(0.21)	(0.21)
Heart Failure	0.08	-1.16**	0.001	0.11	0.12	0.12
	(0.24)	(0.12)	(0.32)	(0.10)	(0.18)	(0.12)
Pneumonia	-0.33	-2.7**	0.92	-0.18	0.81	0.27
	(1.51)	(0.77)	(2.09)	(0.32)	(0.51)	(0.15)
Coronary Artery Bypass	-0.04	-0.47*	-0.02	-0.14	0.07	0.07
	(0.33)	(0.18)	(0.46)	(0.18)	(0.24)	(0.21)
Hip/Knee Replacement	0.57**	-0.29**	-0.47*	-0.20**	-0.18	-0.004
	(0.16)	(0.08)	(0.22)	(0.08)	(0.19)	(0.08)
Unincentivized						
Gastrointestinal Hemor.	1.05	-1.69**	-0.51	-0.04	0.24	0.20
	(0.62)	(0.32)	(0.85)	(0.12)	(0.17)	(0.16)
Stroke	0.07	-0.70**	0.17	0.41	0.04	-0.15
	(0.25)	(0.13)	(0.34)	(0.27)	(0.56)	(0.26)
AAA Repair	1.85	-0.2.6**	-0.45	-1.83**	-0.06	0.02
	(1.72)	(0.91)	(2.29)	(0.63)	(0.89)	(0.72)
Angioplasty	-0.23	-0.09	0.24	-0.07	0.04	0.03
	(0.13)	(0.07)	(0.16)	(0.06)	(0.07)	(0.06)
Carotid Endarectomy	0.05	-0.09	-0.12	-0.33**	-0.13	-0.11
	(0.20)	(0.10)	(0.28)	(0.10)	(0.15)	(0.11)
Craniotomy	2.78**	-0.88*	-1.50	-0.48	-0.73	-0.72
	(0.90)	(0.42)	(1.77)	(0.39)	(0.86)	(0.41)
Esophageal Resection	26.97	-0.11	-23.24	2.27	5.39	3.28
	(17.8)	(4.6)	(21.8)	(5.8)	(10.2)	(3.51)
Pancreatic Resection	4.27	-1.85	-3.46	-0.97	-6.57	0.17
	(5.6)	(2.6)	(7.9)	(2.8)	(5.4)	(0.28)

Regressions of risk-standardized mortality rates per hundred admissions (procedures), on participation in the Premier Pay-for-Performance demonstration. P4P*Post is the difference-in-difference estimate of the P4P effect. Specification 1 is an OLS model comparing mortality rates at the end of the demonstration (2006) to the year prior to its start (2003). Specification 2 adds hospital fixed effects and uses data from 2000-2006. Specification 3 reports marginal effects from the population-averaged generalized estimating equations analog of 2. Clustered standard errors in parenthesis. *Significant at 5%, ** significant at 1%. All models control for hospital payer mix and patient demographics and weight by admissions. State Inpatient Data from Florida and New York, 2000 - 2006 includes 39 hospitals in the Premier Pay-for-Performance demonstration and 351 non-Premier hospitals.

Table 5b: Error Correlation Matrix, Risk-Adjusted Mortality Regressions

	AMI	HF	Pnemonia	CABG	Hip/Knee
AMI	1.00				
Heart Failure	0.06*	1.00			
Pneumonia	0.09*	-0.04	1.00		
CABG	0.13*	0.002	0.11*	1.00	
Hip/Knee Replacement	-0.02	0.04	0.03	0.10*	1.00
GI Hemorrhage	0.04	-0.01	0.05*	0.05	0.03
Stroke	0.07*	-0.01	0.07*	0.03	0.10*
Angioplasty	0.20*	-0.02	0.08*	0.10*	0.05
AAA Repair	-0.04	0.03	0.01	0.11*	0.03
Carotid Endarectomy	-0.001	0.04	0.02	0.11*	0.06*
Craniotomy	0.10*	0.08*	0.10*	-0.02	0.03
Esophageal Resection	-0.07	0.001	0.03	0.15*	-0.05
Pancreatic Resection	0.01	0.03	0.03	-0.04	-0.05

Residual correlation matrix from hospitals performing all of the incentivized and unincentivized procedures under Specification 2 (fixed effect DID models) regressions of risk-adjusted mortality on P4P and control variables. HCUP SID data, 2000-2006. * Statistically significant at 5%.

Table 6: Surgical Outcomes under Pay-for-Performance

	P4P
CABG	
CABG using Internal mammary artery	6.64*
	(2.53)
Hematoma/Hemorrhage	0.23
	(0.63)
Physical/Metabolic Derangement	0.62
	(0.74)
Hip/Knee Replacement	
All-Patient Mortality	-0.18
	(0.19)
Medicare Mortality*	-0.05
	(0.13)
All-Patient Hematoma/Hemorrhage	0.40
	(0.21)
Medicare Hematoma/Hemorrhage*	0.61**
	(0.22)

Difference-in-difference regressions of risk-standardized mortality rates per hundred admissions (procedures), on participation in the Premier Pay-for-Performance demonstration. CABG using the internal mammary artery with higher rates reflecting a better outcome. Clustered standard errors in parenthesis. * Significant at 5%, ** significant at 1%. State Inpatient Data from Florida and New York, 2000 - 2006 includes 39 hospitals in the Premier Pay-for-Performance demonstration and 351 non-Premier hospitals.

Table 7: Pay-for-Performance and Adverse Surgical Outcomes
Medicare Coronary Artery Bypass Grafts, 2000 - 2006

	P4P*Post
Inpatient Mortality	0.16
	(0.17)
30-Day Mortality	0.22
	(0.17)
30-Day Readmission	0.38
	(0.50)
90-Day Readmission	0.57
	(0.62)
365-Day Readmission	0.76
	(0.76)
Surgical Site Infection	0.11
	(0.10)

Difference-in-difference regressions of risk-adjusted rates of adverse surgical outcomes (mortality, readmission conditional on live discharge, and surgical site infection) on hospital participation in the Premier demonstration. 100% sample of Medicare CABG operations amongst patients 65-99, October 2000 - September 2006. Clustered standard errors in parentheses, all results are statistically insignificant.

Appendix A1: Baseline Hospital Characteristics

	Premier	Control
Government-Owned	0.13	0.14
	(0.33)	(0.34)
For-Profit	0.03	0.17
	(0.18)	(0.38)
Medicare Admissions	0.44	0.43
	(0.10)	(0.11)
Medicaid Admissions	0.17	0.16
	(0.12)	(0.11)
Network	0.45	0.32
	(0.50)	(0.47)
Teaching	0.44	0.39
	(0.50)	(0.49)
Full-time MDs	23.5	24.1
	(54.1)	(85.0)
Full-time Residents	36.1	36.7
	(88.9)	(121.7)
Full-Time RNs	391.5	329.1
	(314.0)	(308.4)
Community Mission	0.88	0.87
	(0.32)	(0.34)
Observations	192	1596

Notes: Baseline hospital characteristics reported in the 2003 American Hospital Association Annual Survey by Premier P4P and control group hospitals.

Appendix A2: Hospital Volume and Payer Mix, Florida and New York

	Premier		Control	
	All Volume	**Medicare**	**All Volume**	**Medicare**
AMI	335.0	0.61	192.3	0.64
	(395.8)	(0.49)	(295.3)	(0.48)
Heart Failure	527.7	0.76	363.5	0.76
	(330.3)	(0.42)	(364.4)	(0.43)
Pneumonia	429.5	0.68	297.6	0.68
	(232.4)	(0.47)	(271.3)	(0.47)
CABG	160.7	0.58	100.2	0.60
	(257.0)	(0.49)	(238.6)	(0.49)
Hip/Knee Replacement	84.2	0.67	62.3	0.67
	(85.1)	(0.01)	(120.2)	(0.01)
Gastrointestinal Hemorrhage	250.0	0.70	159.9	0.69
	(151.0)	(0.46)	(150.0)	(0.46)
Stroke	250.4	0.68	162.6	0.68
	(187.0)	(0.47)	(178.0)	(0.47)
Angioplasty	509.0	0.53	241.7	0.53
	(845.9)	(0.50)	(564.7)	(0.50)
AAA Repair	23.2	0.80	13.4	0.80
	(28.3)	(0.50)	(26.2)	(0.40)
Carotid Endarectomy	44.9	0.74	67.6	0.76
	(67.0)	(0.44)	(66.2)	(0.43)
Craniotomy	51.8	0.41	38.0	0.41
	(92.2)	(0.49)	(108.5)	(0.49)
Esophageal Resection	0.4	0.37	0.7	0.37
	(1.0)	(0.49)	(2.9)	(0.49)
Pancreatic Resection	2.4	0.54	2.0	0.54
	(3.2)	(0.50)	(8.0)	(0.50)
Total Admissions	18,761.3		11,792.0	
	(14072.6)		(12441.7)	
Medicare admissions	0.40		0.44	
	(0.12)		(0.17)	
Medicaid Admissions	0.14		0.18	
	(0.11)		(0.14)	
Observations	42		415	

Hospital volume including the percent of each admission covered by Medicare and total hospital payer mix. State Inpatient Data, Oct 2003 - Sept. 2006. Standard deviations in parentheses.