Heterogeneous effects of incentive contracts in healthcare: evidence from Medicare reform

Galina Besstremyannaya[*] Sergei Golovan[†]

February 13, 2021

Abstract

The paper analyzes heterogeneous hospital behavior in response to an incentive scheme with linear increase of the reward function for measured quality. The main contribution of the paper is to reassess a stylized fact in the literature, which asserts that pay-for-performance systems lead to greater improvements from providers with lower baseline quality. We suggest that this "fact" is the result of measurement error, and we develop a structural model and an estimation approach to recover the effects of the linear incentive scheme on quality. Compared to threshold-based incentive schemes that pay a flat bonus for meeting a quality target, where incentives operate mostly for providers below and near to the target, linear incentives may have large effects on providers with any initial level of quality. The predictions of our theoretical model, which are verified empirically using the longitudinal data for US Medicare hospitals involved in value-based purchasing, show a direct association between prior quality and quality improvement owing to the incentive scheme. We argue that the differences between true quality and its measurable proxy may cause regression towards the mean, which needs to be excluded in empirical estimations of the impact of the incentive scheme. Our theoretical model suggests that the effect of the inaccuracy in quality measurement can also be diminished by increasing the share of hospital funds at risk.

An extension of the model considers fixed cost of hospital's investment into quality improvement. The hospitals split into two groups, and the effect of incentive on quality increase in the group that chooses to bear the fixed cost is larger than in the group that refuses the burden of the fixed cost.

JEL Classification Codes: D21, D22, D64, D86, C23, C33

Keywords: incentive contract, firm behavior, intertemporal optimization, dynamic panel data, hospitals

[*]National Research University Higher School of Economics, Russian Federation, Moscow, Pokrovsky Blvd 11, gbesstremyannaya@hse.ru. The paper was prepared in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).

[†]New Economic School, Moscow, Skolkovskoe shosse 45, sgolovan@nes.ru

1 Introduction

Public contracting with firms in conditions of asymmetric information about their technology presents an agency problem where the government, acting as the principal, can use price regulation to induce a socially efficient level of cost reduction (Laffont and Tirole, 1993; Shleifer, 1985). But reduction of firm costs entails risks for product quality (Chalkley and Malcomson, 1998a; Ellis and McGuire, 1996; Hölmstrom and Milgrom, 1991). One way of addressing this problem is a mechanism of extrinsic motivation, which uses quality indicators to define the performance level of each firm and relates firm remuneration to performance. Such a pay-for-performance mechanism is widely employed in the public sector (in civil service, education and social work) and is particularly valuable in healthcare, since healthcare is the classic example of an industry with asymmetric information where sustained quality of service is extremely important.

Ideally, pay-for-performance should create incentives for each healthcare provider, shifting the whole distribution of quality (Cashin, 2014c). However, the most widely used scheme – threshold-based incentives, which reward performance above a certain target value of a quality indicator – fails to provide sufficient stimuli for agents who are far below the threshold (Mullen et al., 2010; Siciliani, 2009). A more promising incentive mechanism makes the reward proportional to the measured quality. An example is Medicare's value-based purchasing, implemented at national level in the US in 2013 on the basis of a reward function that increases in linear fashion and relates the aggregate measure of hospital performance to remuneration. Pay-for-performance bonuses proportional to hospital performance, practised in Maryland, offer a precedent at state level (Murray, 2014). Other instances of an increasing reward function are per-patient bonuses to physician groups in Australia, France, Germany, New Zealand and the US, where quality is estimated as the percentage of patients covered by a certain type of primary care services (Cashin, 2014a; Cashin and Chi, 2014; Bousquet et al., 2014; Coleman et al., 2007; Busse, 2004). Bonuses proportional to quality levels or quality improvements, but provided only to agents with performance above a certain threshold level of quality, represent partial implementation of the linear reward function (Rosenthal, 2014; Cashin, 2014c).

The introduction of a reward function that increases in linear fashion is based on the plausible expectation that such a mechanism will be effective in stimulating healthcare providers with any initial level of quality. However, the degree of quality improvement may be different for different quality groups. The existing theoretical literature on real-world implementation of incentive contracts in healthcare, as well as the empirical research on heterogeneous effect with respect to quality groups, tends to focus on threshold-based schemes (Mullen et al., 2010; Siciliani, 2009; Doran et al., 2008). To the best of our knowledge, research on the linear reward function has been limited to investigation of the mean effect of per-patient bonuses in primary care (Kristensen et al., 2016). We discovered only one empirical paper on heterogeneous impact of per-patient bonuses (Coleman et al., 2007) but the inverse relationship between quality improvement and prior levels of quality, which it discovers, may not be regarded as the effect of the quality incentives, since the relationship holds for both incentivized and non-incentivized physicians.

The purpose of this paper is to analyze heterogeneity in response to a pay-for-performance mechanism in healthcare with a reward function proportional to quality. We start by building a theoretical model of hospital behavior, the key features of which may be summarized as follows. First, a hospital's objective function includes an altruistic component which is proportionate to the quality of provided services. This assumption is supported by abundant evidence about the intrinsic motivation of healthcare providers and accounts for the fact that some hospitals pay more attention to quality than others. Second, the model differentiates between unobserved quality and its measurable proxy, since the incentive payment is a function of observed quality that increases in linear fashion. We show that differences between true quality and its measured proxy may cause regression towards the mean ("mean reversion")[1] which needs to be excluded in empirical evaluation of the impact of pay-for-performance. Third, there are dynamic aspects induced by the behavior of the regulator and the hospital. The behavior of the regulator adds a dynamic to the hospital's task since the incentive payment lags performance: it is computed according to quality measured in the previous period.

Additional dynamics arise from the hospital's intertemporal incentive when the quality payments are expected to continue over a long term: the hospital understands that its current policies towards quality of care will influence future reimbursement. Empirical support for the existence of this intertemporal incentive is provided by interviews with hospital executives about the impact of per-patient bonuses in primary care (Bokhour et al., 2006; Conrad et al., 2006) and the effect of the linear rule in Medicare's value-based purchasing on the behavior of hospital executives and physicians (Smith, 2017; Jones, 2014). Hospital officials explicitly state that pay-for-performance impacts the way their hospital "is going to be paid in future" (Jones, 2014, p. 120), that "the stakes were high" and that each year their hospital could risk a large sum of money in case of poor performance (Smith, 2017, p. 145). Accordingly, the hospital executives "appear capable of creating an internal environment of high energy and high expectations" (Conrad et al., 2006, p. 449).

The model forecasts a positive effect on measured quality from the introduction of a linear incentive on measured quality with quality increase proportional to baseline performance. The prediction of a direct association between the prior level of quality and quality improvement is tested empirically using nationwide data for 2968 acute-care Medicare hospitals in 2011–2018. Our empirical results use variation in the size of quality incentives in order to estimate the effect of pay-for-performance cleansed of mean reversion. We control for other potential channels of quality improvement by Medicare hospitals, using data on the Hospital Readmissions Reduction Program and on the meaningful use of Electronic Health Records. Another theoretical prediction of the model which is verified empirically is positive relationship between the effect of pay-for-performance and the strength of the quality incentives, measured in terms of the share of hospital funds that are "at risk" in the scheme.

[1] Exceptionally low or high values of the variable in initial measurement tend to be closer to the center of the distribution in subsequent measurements (Davis, 1976).

We find that the higher the quintile of the composite quality measure at Medicare hospitals, the larger the estimated effect of the reform. The heterogeneity of the effect corresponds to findings of the health policy literature about stronger emphasis on quality-improving activities at high-quality hospitals or among high-quality physicians in comparison with low-quality hospitals and physicians (Damberg et al., 2009; Vina et al., 2009; Grossbart, 2006). Specific measures for quality improvement may include establishing best practices for each condition and dissemination of these practices at professional conferences, using clinical pathways and clinical guidelines, and providing feedback to physicians by reporting internal data (for instance, by letting the name of a successful heart surgeon or surgery group be known (Damberg et al., 2014, 2010, 2009; Vina et al., 2009; Bentley and Nash, 1998)). The predictions of our theoretical model and our empirical results suggest that the stylized fact of the inverse relationship between improvement owing to the incentives scheme and the baseline performance should be revisited. This inverse relationship has been found by most empirical assessments of the impact of incentive contracts on healthcare quality and seems to hold both for the linear payment rule and for other designs of pay-for-performance: it is observed for general practitioners in the UK, physician groups in California, Chicago and Ontario, US hospitals in Michigan, New York and Wisconsin, and hospitals involved in Medicare's pilot project for quality improvement (Li et al., 2014; Vaghela et al., 2009; Doran et al., 2008; Glickman et al., 2007; Coleman et al., 2007; Lindenauer et al., 2007; Hibbard et al., 2005; Rosenthal et al., 2004; Pai et al., 2002; Hannan et al., 1994).

However, we argue that the finding of an inverse relationship may be incorrect when the empirical approach does not account for imprecise measurement of quality and the stochastic nature of the imprecision. As we establish in our theoretical model regarding the effect on observed quality of linear increases of payment, the damaging effect of imprecision in quality measurement can be mitigated by increase of the share of hospital funds at risk in the scheme. However, the share cannot be very high, as the steady state conditions require an upper limit on the incentive size.

The remainder of the paper is structured as follows. Section 2 reviews the design of quality incentives in healthcare. A theoretical model of quality incentives with linear increase of the reward function and the predictions of heterogeneous effect for quality groups of hospitals are given in Section 3. Section 4 outlines the empirical methodology and describes the data for Medicare hospitals. The results of the empirical analysis are presented in Section 5. Section 6 contains a discussion of channels used for quality improvement and of the size of quality incentives.

2 Pay-for-performance in health care

2.1 Background

The origins of incentive regulation in conditions of asymmetric information can be traced to the approach used by Baron and Myerson (1982) and the yardstick competition model by

Shleifer (1985), which establishes the price for each firm depending on costs of comparable firms. Applied to healthcare, yardstick competition requires the identification of a hospital's products and determination of a reasonable cost for each product. This is done by assigning patients to a limited number of medically justified diagnosis-related groups (DRGs), with a statistically stable distribution of resource consumption within each group (Thompson et al., 1979). This assignment is the core of a prospective payment system, which is a reimbursement method that provides fixed payments for a patient with a given DRG. Piloted in New Jersey in the 1980s and then applied to all Medicare hospitals in the United States, prospective payment has now been adopted in most healthcare systems around the world.

Prospective payment aims to make product and service provision more efficient, but it may adversely affect the quality of product if quality and output are interrelated objectives of the firm (Hölmstrom and Milgrom, 1991). In this regard, Ma (1994) and Ma and Mak (2015) show that prospective payment can lead to efficient levels of costs and quality when these are the only two objectives of a hospital and when quality is verifiable. If quality is observable but non-verifiable, prospective payment causes underprovision of quality when quality and quantity are net substitutes (Laffont and Tirole, 1993). Actual implementation of prospective payment systems in various countries has indeed been accompanied by deterioration of healthcare quality measured, for instance, as intensity of care, mortality or readmission (Eggleston and Hsieh, 2004; Ellis and McGuire, 1996).

Theoretical approaches to designing a contract for an efficient level of costs and quality often assume that the social planner observes the true quality or at least knows the response of patient demand or hospital costs to quality (Ma and Mak, 2015; Chalkley and Malcomson, 1998b; Ma, 1994). Some papers do acknowledge the unobservable character of healthcare quality, and attempts have been made to use patient demand as a proxy for quality when designing an incentive contract (Chalkley and Malcomson, 1998b).

Practical implementation of an incentive contract for quality uses a number of verifiable performance measures as proxies for the unobserved quality of healthcare. The mechanism is called pay-for-performance and dates from the early 1980s when various performance targets were used for enhancing the quality of natural monopolies and telecommunications provision (Kridel et al., 1996; Joskow and Schmalensee, 1986).

Numerous programs for monitoring the value of various quality indicators were launched in healthcare in the US and Europe in the 1980s and 1990s (Christianson et al., 2008; Wagner et al., 2006), including one of the first examples of nationwide implementation of pay-for-performance for family practices in the UK in 2004 (Campbell et al., 2009). Pay-for-performance is currently used in hospitals in Brazil, Korea, the Netherlands, the UK and the US (Cashin et al., 2014; Dückers et al., 2009), in Germany's sickness funds (de Bruin et al., 2011; Busse, 2004) and in primary care in Australia, Canada, Estonia, France, New Zealand, Spain, Sweden, Turkey, the UK and the US (Practice Assist, 2017; Cashin et al., 2014; Li et al., 2014; Ödesjö et al., 2015; Buetow, 2008; Gené-Badia et al., 2007).

Pay-for-performance mechanisms in healthcare have several features in common. Firstly, the verifiable performance measures, which approximate true quality, cover several aspects of care: clinical quality (for instance, prescription of a certain drug or administration of a certain procedure), patient experience (subjective assessment of services that are received) and care outcomes (mortality or morbidity). Secondly, the incentive mechanisms aim to stimulate quality improvement. Several designs may be used for this purpose, and we classify them below as threshold-based with a flat bonus, threshold-based with a quality-related bonus, and continuously increasing reward function for quality. The theoretical model in this paper and the empirical analysis concerns the latter mechanism.

2.2 Design of quality incentive schemes

2.2.1 Threshold-based with flat bonus

This scheme provides a flat bonus to agents whose performance is above a certain target value of a quality indicator. The value can be set as an absolute standard or as a relative standard, related to the empirical distribution of agents' quality. A variant of the scheme considers several thresholds and uses a stepwise reward function. The flat bonus approach is the earliest quality-based reimbursement scheme and is used in primary care in the UK, Canada, Estonia and Spain. Thresholds are commonly established for clinical indicators, which describe the percentage of patients who have undergone immunization and screening, or who have good health outcomes (for example in terms of cholesterol level or blood pressure).

Theoretical analysis of threshold-based incentive schemes forecasts undesired effects for agents in the highest percentiles of quality, whose performance may deteriorate owing a crowding-out of motivation by extrinsic incentives (Bénabou and Tirole, 2006, 2003; Kreps, 1997), conformism (Murdock, 2002) or due to lessening of effort in tournaments with other health care providers (Casas-Arce and Martínez-Jerez, 2009; Prendergast, 1999; Radner, 1985).

Low quality agents who are very far from the threshold also lack incentives for improvement, since their cost of enhancing quality to the target value is less than the quality bonus (Mullen et al., 2010). Moreover, a threshold-based incentive scheme may cause various unintended effects, such as artificial enhancement of the quality indicator. An example of this would be artificial increase of the share of patients who undergo necessary procedures, achieved by underreporting the number of eligible patients (Gravelle et al., 2010).

2.2.2 Threshold-based with continuous bonus

This approach provides a continuous reward for performance above the threshold. Per-discharge awards for top performing hospitals, organized by Premier Inc. for the Hospital Quality Incentive Demonstration (HQID), offer an example of the approach. This voluntary program for Medicare hospitals was implemented in 2003–2008 and can be regarded as a pilot for the subsequent nationwide introduction of value-based purchasing. The HQID program used quality measures of the clinical process of care for five health conditions: acute myocardial infarction, heart failure, pneumonia, coronary artery bypass grafting, and knee or hip replacement. Main features of the HQID incentive scheme have been described in the literature (Cashin, 2014c; Ryan et al., 2012b). They included a reputational incentive: reporting of hospitals, where quality was above the median level. Financial incentives in the first phase of the program (2003–2006) offered a bonus of, respectively, 2% and 1% of their Medicare revenue to hospitals in the top first and second decile of quality measures for each health condition. The second phase of the program (2006–2008) incentivized achievement and improvement through three types of stimuli: 1) an attainment award for exceeding the median score that existed two years prior to the pilot, 2) a top-performance award for being in the top two quality deciles in the current year, 3) a top-improvement award for exceeding the median score in the current year and for being in the top two deciles for quality improvement. Although financial rewards to hospitals in the second stage were allocated on a per-patient basis, estimated values of the top-performance and top-improvement awards suggest that they were roughly comparable to a 1% bonus per condition, while the attainment award was equivalent to a 0.25% bonus over and above Medicare revenue per condition (Ryan et al., 2012a). Hospitals in the tenth and ninth deciles (the bottom deciles) for all quality measures by the end of the second phase of the program suffered a 2% and 1% reduction in Medicare payments, respectively.

HQID improved the mean level of hospital quality (Ryan, 2009; Christianson et al., 2008; Lindenauer et al., 2007; Glickman et al., 2007; Grossbart, 2006) and its success led to nationwide implementation. The US pilot served as a model for the Advancing Quality hospital program in the UK and for Korea's hospital incentives program, where bonuses are provided to top-performing hospitals based on the improvements, which they achieve (Bisiaux and Chi, 2014; Sutton et al., 2012).

As regards empirical estimates of the effect of the pilot on hospitals in different quality groups, most studies find an inverse association between the quality increase and the prior quality level, but they focus exclusively on incentivized hospitals (Glickman et al., 2007; Hibbard et al., 2005). However, papers that use policy evaluation techniques report that hospitals in the top two deciles showed the fastest improvement, while hospitals in the lowest deciles raised their quality to a much lesser extent or may even have failed to improve (Ryan et al., 2012b; Werner et al., 2011). These results go in line with the shortcomings of threshold-based schemes, manifested in the US pilot program. Only hospitals above the upper threshold (the 80th percentile of quality score) and below the bottom threshold (20th percentile) experience a clear stimulus. Other hospitals have little or no incentive for improvement. Indeed, hospitals in the $(0.2, 0.5)$ quantile interval of the quality score did not obtain rewards and did not suffer penalties, while hospitals in the $(0.5, 0.8)$ quality interval had to show very rapid improvement in order to qualify for a financial bonus.

2.2.3 Linear reward function and Medicare's value-based purchasing

This mechanism provides an incentive proportional to measured quality and has been applied to discharges in the inpatient prospective payment system at acute-care Medicare hospitals since 2013.[2] The scheme reduced Medicare's DRG-based payment to each hospital by a factor α which equaled 0.01 in 2013, was increased annually by 0.0025 in 2014–2017 and has remained flat at 0.02 since 2017. The accumulated saving is redistributed across hospitals according to the adjustment coefficient, which is computed as a linear function of the composite quality measure: $1 + \left(\kappa \frac{TPS_i}{100} - 1\right) \cdot \alpha$, where i is the index of a hospital and TPS_i is the hospital's *total performance score* ($0 \leq TPS_i \leq 100$). Hospitals are rewarded if the adjustment coefficient is above one and suffer financial loss otherwise. While the pilot program required additional financial resources, the nationwide quality incentives scheme is budget-neutral and the value of the slope κ is chosen to ensure budget neutrality.

The *total performance score* is a weighted sum of scores for measures in several domains: timely implementation of recommended medical interventions (*clinical process* of care), quality of healthcare as perceived by patients (*patient experience* of care), survival rates for AMI, heart failure and pneumonia patients and other proxies for *outcome* of care, healthcare-associated infections and other measures of *safety* of care, and spending per beneficiary as a measure of *efficiency* of care. The domain score is the sum of the scores for its measures, and measure scores are computed according to a discrete scale of 0 to 10. Higher score reflects higher position of the hospital in the empirical distribution of the quality measure in a given year or higher improvement of the quality measure relative to the baseline period.[3] (See details on domain scores in Appendix A).

A similar reward function with linear increase for healthcare quality is found in the quality incentive program in Maryland where hospitals are grouped according to the value of their composite quality indicator relative to the mean for the state: hospitals above the mean receive a bonus proportional to their quality, while hospitals below the mean suffer a proportional loss (Murray, 2014). Other examples are per-patient awards in incentive mechanisms for primary care quality in Australia, France, Germany, New Zealand and the US, where a bonus payment is provided for each patient who has undergone immunization, screening or other quality-related care.

3 Model

3.1 Setup

Under the principal-agent approach on the healthcare market, a principal (a government or a social planner) contracts agents (physicians or hospitals) on behalf of consumers (patients). Pay-for-performance incentive schemes, such as Medicare's value-based purchasing, target hospitals rather than physicians. However, in our model we equate incentives for the hospital with

[2] Two US states are exceptions to the rule: Puerto Rico, which only started innovating its healthcare system in 2015 and Maryland, which has a unique model for hospital financing.

[3] Specifically, achievement points are computed for each measure evaluating a hospital's performance relative to other hospitals in a given year, and improvement points for each measure are computed to assess change in the hospital's own performance in the given year relative to the baseline period. Then, for each measure, the highest of the two (achievement points or improvement points) is used as the hospital's score for that measure.

incentives for its physicians and consider the hospital as an aggregate agent. This approach matches changes in the management of Medicare hospitals, designed to bridge the potential gap between the interests of hospitals and of physicians (Centers for Medicare and Medicaid Services, 2007b).

Based on the results of numerous experimental and empirical studies,[4] we take account of the existence of altruism on the healthcare market. We assume that hospitals (physicians) have a type-specific altruism θ, where θ is a random variable with expected value $\bar{\theta}$. It should be noted that hospital-specific altruism becomes a source of hospital heterogeneity in our model. An alternative approach to modeling hospital heterogeneity, which leads to similar quantitative results, considers hospital-specific marginal costs as in Laffont and Tirole (1993).

We denote the quality of healthcare in period t as q_t. The quality depends on the hospital's efforts e_t, and we assume $q_t = e_t + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ (Eggleston, 2005). We abstract from multi-dimensional efforts and multi-dimensional quality in our model. The linear relationship between efforts and quality is another simplification but we relax it in Appendix C, which contains a numerical solution of a more general model.

The hospital has an additively separable objective function consisting of three parts: the benefit B from altruistic behavior (i.e. from the expected change in the patient's health due to treatment), wealth (net profit) π, and disutility of quality-enhancing efforts $C(e_t)$ (Blomqvist, 1997; Eggleston, 2005; Oxholm et al., 2018):

$$U = B + \pi - C.$$

The benefit from altruism depends positively on quality and we let $B = \theta q_t$. The hospital's net profit π is the difference between revenue and the financial cost of providing services. Revenue is the demand for healthcare multiplied by the per-patient price. Demand depends positively on quality (Ma and Mak, 2015; Siciliani et al., 2013; Chalkley and Malcomson, 1998b; Ellis and McGuire, 1996; Ma, 1994) and we assume a linear form of demand aq_t, where $a > 0$.

The per-patient price is the sum of R_t (approximates reimbursement for outlier cases in the prospective payment system or may be regarded as part of a cost-sharing tariff) and prospective payment p_t, which is subject to quality adjustment.

Based on the design of most pay-for-performance schemes, we assume that the principal computes a measure of hospital's quality m_t as a function of q_{t-1}, q_{t-2} etc. $m_t = M(q_{t-1}, q_{t-2}, \dots)$. The value of m_t is used for assigning additional stimulus to the hospital. Note that the hospital does not know m_t at $t-1$. The initial value m_0 is assumed to be known.

We model a linear rule in Medicare's value-based purchasing, so the principal adjusts the hospital's prospective payment through multiplication of the unit price p_t by the quality-adjustment coefficient $1 + (\kappa m_t - 1)\alpha$. Here $0 < \alpha < 1$ is the share of the hospital's revenue, which goes to the nationally accumulated fund and is then redistributed according to a linear function with slope κ. The quality incentives scheme is budget-neutral, and both κ and α are exogenous to

[4]See empirical literature on quantification of altruism (Li et al., 2014; Gruber and Owings, 1996) and various experimental works, such as Brosig-Koch et al. (2016).

a hospital. The hospital's costs are proportional to the volume of healthcare services provided with the coefficient $d_t > 0$, which may be taken as a proxy for the existing technology and is regarded as exogenous by the hospital.

The hospital's net profit is then $\pi = aq_t(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t)$.

The final part of the hospital's objective function is the disutility of efforts $C(e_t) = ce_t^2/2$, where $c > 0$.

Note that the marginal cost of quality improvement is zero at zero effort. In other words our main model assumes that all costs related to quality are variable. As an extension, we consider a version of the model where the fixed cost of quality improvement enters the hospital problem. Examples of such cost include the expense of several million USD to install an Electronic Health Records system that enables the hospital to collect and analyse data on patients and their diagnoses. In the extended model we assume that the fixed cost born in period one enables the hospital to maintain higher quality at lower variable cost in period 2. As a result, the hospitals split into two groups. A group with higher θ chooses to bear the fixed cost of quality improvements while the group of hospitals with lower θ decides to refuse the burden of the fixed cost.

The hospital's objective function is

$$U_t(m_t, q_t, e_t) = \theta q_t + aq_t(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - ce_t^2/2.$$

In period t the hospital chooses the level of efforts e_t, given the technology d_t, prices R_t, p_t and the quality-adjustment coefficient (all are set by the principal). Effort e_t determines the level of quality q_t, while q_t influences the estimated value of quality m_{t+1} and, hence, the value of the hospital's objective function in period $t + 1$. The discount factor for the utility of future periods is $\beta \in (0, 1)$.

The hospital's intertemporal maximization problem is:

$$\begin{aligned} &\max_{\{e_t, q_t, m_t\}} E \sum_{t=0}^{\infty} \beta^t U_t(m_t, q_t, e_t) \\ \text{s.t.} \quad & q_t = e_t + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \text{ i.i.d., } t = 0, 1, \dots, \\ & m_t = M(q_{t-1}, q_{t-2}, \dots), \ t = 1, 2, \dots, \text{ and } m_0 \text{ is given.} \end{aligned} \tag{1}$$

Note that the model considers two types of incentives for quality improvement. One incentive stems from the fact that the hospital increases the demand for its services by raising quality. Another incentive is intertemporal and is associated with pay-for-performance: the hospital realizes that if quality goes up in the current period, the payoff from the regulator will be higher in the next period.

3.2 Solution of the model

3.2.1 Assumptions

The analysis is based on assumptions about the redistributive character of the incentive scheme (Assumption 1), time-invariance of unit prices and costs of hospitals (Assumption 2) and sufficiently high disutility of quality-enhancing efforts (Assumption 3). Note that Assumption 1 may be relaxed, as is argued in the section below on extensions of the model. However, Assumptions 2 and 3 are necessary conditions for the existence of a stable solution and a stationary steady state for the hospital problem (1).

There are several essential features of the model which enable the closed-form solution. Firstly, the private utility of physician θq is linear in quality. Secondly, the quality-adjustment coefficient is linear in the measured quality (this corresponds to Medicare's incentive scheme). Finally, demand for hospital services is linear in quality and the marginal cost is linear in the quality-enhancing efforts. The model, in the general case with monotonically increasing physician utility, use of any convex cost function and any concave profit function cannot be solved analytically and requires a numerical solution. An example of a numerical solution of the model is given in Appendix C. It illustrates that the effect of pay-for-performance, albeit non-linear, still has all the implied properties discussed below: the reform effect is positive and increases monotonically in α and in θ.

Assumption 1. *The principal relates parameters κ and α through the condition*

$$E((\kappa m_t - 1)\alpha) = 0.$$

This assumption means that the expected value of the adjustment coefficient equals one, which corresponds to budget neutrality of the quality incentive scheme.

Assumption 2. *Prices and unit costs are fixed, so $R_t \equiv R$, $p_t \equiv p$, $d_t \equiv d$.*

Assumption 3. *The disutility of effort is sufficiently high: $ap\kappa A\alpha(1+\beta) < c$.*

If Assumption 3 does not hold then the stationary steady state for the measured quality m_t either does not exist or is unstable. This means that if the power of the pay-for-performance scheme is too high relative to the hospital's costs, then the incentives of the best hospitals are too strong and the incentives of the worst hospitals are too weak, which leads to divergence in quality.

We let $m_t = Aq_{t-1}$, where $A > 0$. This simplified framework corresponds to remunerating the achievement of a hospital and enables to obtain an analytical solution of the model.[5]

[5]The extension of the model which links incentive to the maximum of achievement and improvement is solved through the simulation approach.

3.2.2 Solution of the hospital problem

In this section we show that the solution of the hospital utility maximization problem under assumptions 2 and 3 converges to a stationary process and we find this process. The first proposition describes solution of the problem for a single hospital.[6]

Proposition 1. *Under assumptions 2 and 3 the solution of the hospital utility maximization problem for a hospital of type θ constitutes a stable autoregressive process*

$$e_t = \mu(\theta) + \lambda(e_{t-1} - \mu(\theta)) + \lambda\varepsilon_{t-1}, \tag{2}$$

where

$$\mu(\theta) = \eta_0 + \eta_1\theta, \tag{3}$$

for

$$\eta_0 = \frac{a(R-d) + ap(1-\alpha)}{c - ap\kappa A\alpha(1+\beta)}, \quad \eta_1 = \frac{1}{c - ap\kappa A\alpha(1+\beta)}, \tag{4}$$

and

$$\lambda = \frac{c - \sqrt{c^2 - 4\beta(ap\kappa A\alpha)^2}}{2\beta ap\kappa A\alpha}. \tag{5}$$

Next, since $q_t = e_t + \varepsilon_t$, the following corollary describes the process for quality:

Corollary 2. *The quality level q_t of a hospital with type θ forms a stable autoregressive process*

$$q_t = \mu(\theta) + \lambda(q_{t-1} - \mu(\theta)) + \varepsilon_t, \tag{6}$$

where $\mu(\theta)$ and λ are defined by (3) and (5) respectively.

The autoregressive process in Corollary 2 is stable, and this ensures existence and stability of a solution and a stationary steady state for the hospital's quality.

Corollary 3. *Under the conditions of assumptions 2 and 3, the quality of a hospital with type θ converges to a stationary steady state process for which*

$$E(q_t \mid q_{t-1}, \theta) = \mu(\theta) + \lambda(q_{t-1} - \mu(\theta)), \tag{7}$$

so the long-term expected value of q_t conditional only on θ is

$$E(q_t \mid \theta) = \mu(\theta) = \frac{a(R-d) + \theta + ap(1-\alpha)}{c - ap\kappa A\alpha(1+\beta)}. \tag{8}$$

3.2.3 Stationary equilibrium

Next, consider the steady state level of quality for all hospitals and add the budget neutrality assumption 1. By taking expectation of (8) we get the long-term unconditional expected value of the quality for all the hospitals.

[6] Proofs for all the statements can be found in Appendix B.

Corollary 4. *The unconditional expected value of the quality equals*

$$\mu = E(q_t) = \frac{a(R-d) + \bar{\theta} + ap(1-\alpha)}{c - ap\kappa A\alpha(1+\beta)}. \tag{9}$$

This allows us to rewrite the condition from Assumption 3 for a budget-neutral scheme, where κ is endogeneously determined by α:

Corollary 5. *The technical condition from Assumption 3 under the budget neutrality assumption 1 becomes:*

$$\alpha < (a(R-d) + \bar{\theta})/ap + 1, \tag{10}$$

which means that very high value of policy parameter α may break the stability of the system.

Proposition 6 describes hospital response to pay-for-performance: the mean value of measured quality m_t and the parameter of convergence λ increase in α.

Proposition 6. *Suppose that assumptions 1, 2, and 3 hold. Then the stationary steady state parameters are:*

$$\mu = \frac{a(R-d) + \bar{\theta} + ap(1+\beta\alpha)}{c}, \tag{11}$$

$$\eta_0 = \frac{(a(R-d) + \bar{\theta} + ap(1+\beta\alpha))(a(R-d) + ap(1-\alpha))}{c(a(R-d) + \bar{\theta} + ap(1-\alpha))}, \tag{12}$$

$$\eta_1 = \frac{a(R-d) + \bar{\theta} + ap(1+\beta\alpha)}{c(a(R-d) + \bar{\theta} + ap(1-\alpha))}, \tag{13}$$

$$\lambda = \frac{a(R-d) + \bar{\theta} + ap(1+\beta\alpha) - \sqrt{(a(R-d) + \bar{\theta} + ap(1+\beta\alpha))^2 - 4\beta(ap\alpha)^2}}{2\beta ap\alpha}, \tag{14}$$

$0 < \lambda < 1$ and μ, η_1, λ increase in α (η_0 can be an increasing or decreasing function of α, depending on $\bar{\theta}$).

3.2.4 Effect of pay-for-performance

Putting together equations (3), (12) and (13), we get the corollary on the effect of pay-for-performance:

Corollary 7. *The long-term mean of the quality of a hospital of type θ equals*

$$E(q_t \mid \theta) = \frac{a(R-d) + \bar{\theta} + ap(1+\beta\alpha)}{c} \cdot \frac{a(R-d) + \theta + ap(1-\alpha)}{a(R-d) + \bar{\theta} + ap(1-\alpha)}. \tag{15}$$

The mean function (15) increases in θ and its second mixed derivative in θ and α is positive.

Corollary 7 shows that the effect of pay-for-performance increases with respect to the hospital-specific parameter θ. However, θ is unobservable. Therefore, to verify this heterogeneous effect we employ the fact that on average θ is related to the quality level q_t. This link affects the conditional expectation $E(q_t \mid q_{t-1})$, given in Proposition 8. This allows to study effect of pay-for-performance at hospitals which are classified into groups according to their prior quality level. Specifically, we estimate the effect for each quintile of q_{t-1}, using the fact that the higher q_{t-1}, the higher the mean value of θ.

Proposition 8. *If the parameter of altruism θ is distributed normally $N(\bar{\theta}, \sigma^2)$ then the conditional mean $E(q_t \mid q_{t-1})$ is a function with positive mixed derivative with respect to q_{t-1} and α. This means that the higher the lagged value of quality q_{t-1}, the larger the effect of pay-for-performance.*

3.2.5 Implications

Remark 9. We obtain that the mean value of measured quality increases in the share of the hospital budget, which is at risk under pay-for-performance (equation (11)). Also, Corollary 7 shows that the type of altruism positively affects the impact of the reform, and that the type of altruism positively influences the hospital's choice of quality.

Remark 10. As $0 < \lambda < 1$ in (14), hospitals with the highest values of measured quality m_t show a reduction of quality in the next period: $E(m_{t+1} \mid m_t) < m_t$. The effect is opposite for hospitals with the lowest values of measured quality. This phenomenon of "regression towards the mean" is explained by the imperfection of quality measurement. The regression is less pronounced for higher values of λ. Accordingly, since λ increases in α, the pay-for-performance scheme reduces the phenomenon of regression towards the mean and makes the measured quality more persistent.

Remark 11. The model becomes static if we assume that $\beta \approx 0$. In this case the hospital does not take account of the future effect of the quality improvement and maximizes only the current value of its utility function. As may be seen from (11)–(14), the size of the pay-for-performance stimulus under budget neutrality still affects the hospital's quality (η_1 and λ depend positively on α even if $\beta \approx 0$), but the unconditional mean μ becomes constant and does not depend on α.

3.3 Autoregressive process and quality convergence

Equation (7) in the theoretical model and its empirical analogue used in our estimations may be interpreted as an autoregressive process for the measured quality m_t if the coefficient λ of the lagged dependent variable m_{t-1} is positive and less than one. The autoregressive specification in (7) can be taken equivalent to convergence of the measured quality towards the value $\mu(\theta)$[7] and λ is associated with the speed of quality convergence.

Using definitions in Friedman (1992), we can disentangle a permanent component in m_t, which is related to economic impact of pay-for-performance from a transient component (a

[7]Indeed, rewriting Equation (7) as $E(m_t \mid m_{t-1}, \theta) - \mu(\theta) = \lambda(m_{t-1} - \mu(\theta))$ and assuming $\lambda < 1$, we can see that the expected value of the current measured quality Em_t is closer to the mean value $\mu(\theta)$ than is the value of the measured quality in the previous period, i.e. m_{t-1}.

pure dynamic effect), which may be referred to as "mean reversion" or "regression towards the mean" (Galton and Dickson, 1886).[8]

The main reason for the phenomenon of mean reversion is the imprecision of quality measurement, namely, the existence of the random error ε_t in problem (1). Combined with the fact that hospitals make an intertemporal decision within the quality-based reimbursement, the random error causes the autoregressive form of measured quality m_t in (6) and (7).[9]

Research in health economics is often vulnerable to imprecise measurements.[10] However, to the best of our knowledge, only one paper discusses the impact of imprecise quality measurement on pay-for-performance (Oxholm et al., 2018) and only a few papers point to the need to estimate the impact of pay-for-performance cleansed of the potential effect of mean reversion (Gupta, 2017; Morton and Torgerson, 2003; Pai et al., 2002).

Mean reversion makes it incorrect to estimate the effect of pay-for-performance as the net change in the (fitted) value of measured quality at incentivized hospitals. But just this approach is employed in most empirical works that find an inverse relationship between quality improvement and the prior level of measured quality. Our sample is limited to incentivized hospitals only, but we use variation in α over time and evaluate the effect of pay-for-performance cleansed of the impact of mean reversion: we model a stationary process and estimate its long-term mean. An alternative approach for eliminating the effect of mean reversion proposes taking mean values of the dependent variable over a given period of time (e.g. before and after the reform or the intervention) for each observation (Kane and Staiger, 2002). This approach and other ways of estimating the effect of treatment are inapplicable for our data since the intensity of treatment (i.e. the size of α) changes in the course of the reform.

3.4 Extensions

3.4.1 Heterogeneity in hospital production

The model introduced heterogeneity through the hospital-specific parameter θ in the benefit function. An alternative approach incorporates heterogeneity into the cost function of hospitals (Laffont and Tirole, 1993; Mullen et al., 2010), instead of attributing it to altruistic behavior. This alternative formulation does not change the predictions of the model.

Suppose, however, that all components of the utility function are not hospital-specific. Then $\theta \equiv \bar{\theta}$ and the effect of pay-for-performance is homogeneous across hospitals. The differences

[8]Mean reversion generally implies that if the error is lower than average in period t, it is likely to be higher in period $t+1$ than in period t. Similarly, observations with high errors in period t tend to be followed by lower errors in $t+1$.

[9]There may be other causes of the dynamic effect apart from the effect of mean reversion owing to imprecision in quality measurements.

[10]The examples of mean reversion in health economics relate to blood pressure and cholesterol level of patients (Barnett et al., 2004) and birth weight in successive pregnancies (Wilcox et al., 1996), while occurrences of mean reversion in other economic fields may be linked to productivity of countries (Friedman, 1992), tax avoidance by a company's corporate board (Armstrong et al., 2015), or the height of a family member (Galton and Dickson, 1886).

in the dynamics of m_t for high-quality and low-quality hospitals are only due to imprecision in the quality measurement.

3.4.2 Budget neutrality of the quality incentive scheme

The formulation of the budget neutrality condition in Assumption 1 is close to the true condition of budget neutrality $E(aq_t(\kappa m_t - 1)\alpha) = 0$ in case of small values of α. Use of the simplified formulation of the budget neutrality condition avoids the unnecessary complexity of the model's solutions. It only negligibly affects the quantitative results for small values of α and does not change our results qualitatively, even for large values of α.

Suppose, however, that the incentive scheme is not budget-neutral. Consider equation (8), which describes the long-term mean value of the quality level of a hospital. In the absence of budget neutrality, κ becomes the varying parameter of policy intensity. The mean effect of pay-for-performance can be shown to increase in κ, while the effect of imprecision in the quality measurement will weaken with the rise in κ. The effect of pay-for-performance is heterogeneous: it is higher for hospitals with higher θ. So absence of budget neutrality of the incentive scheme does not affect the predictions of the model.

3.4.3 Imprecisely measured quality

We assumed that the difference between true quality and its measurable proxy is due to a random error. What if the measurement error is systematic? With the premise that the systematic part of the measurement error is the same for all hospitals, this additional effect would not qualitatively affect the impact of the incentive mechanism.

3.4.4 Accounting for fixed cost of quality improvement

Hospital utility function and the hospital problem

Consider the following augmentation in the hospital's utility function: the hospital can bear a specific amount of fixed cost in order to decrease the variable cost of quality efforts. The utility function can take one of two forms:

$$U_t(m_t, q_t, e_t) = \begin{cases} \theta q_t + aq_t(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - C - e_t^2/(2c_1) & \text{if hospital decides} \\ & \text{to bear fixed cost,} \\ \theta q_t + aq_t(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - e_t^2/(2c_0) & \text{otherwise.} \end{cases} \tag{16}$$

Here, $C > 0$ and $0 < c_0 < c_1$, so fixed cost alleviates the provision of quality. To concentrate on the effect of fixed cost, we consider a simple one-period model

$$\begin{aligned} &\max_{e_t, q_t, m_t} EU_t(m_t, q_t, e_t) \\ \text{s.t. } & q_t = e_t + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ & m_t \text{ is given.} \end{aligned} \tag{17}$$

The solution

The first order conditions for the two cases are, correspondingly:

$$\begin{aligned}
&\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - e_t/c_1 = 0 \quad \text{if costs are payed,}\\
&\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - e_t/c_0 = 0 \quad \text{otherwise.}
\end{aligned}$$

Accordingly, the optimal values of effort e_t are:

$$\begin{aligned}
e_t^{(1)} &= c_1(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t)) \quad \text{if costs are payed,}\\
e_t^{(0)} &= c_2(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t)) \quad \text{otherwise.}
\end{aligned}$$

To select between two strategies: pay the fixed cost and choose $e_t^{(1)}$, or refuse to bear the fixed cost and choose $e_t^{(0)}$ – the hospital has to compare the expected outcomes across the strategies. The expected utility in case of the first strategy is

$$\begin{aligned}
EU_t^{(1)} &= e_t^{(1)}(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t)) - C - (e_t^{(1)})^2/(2c_1)\\
&= c_1(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t))/2 - C.
\end{aligned}$$

The expected utility under the second strategy equals

$$\begin{aligned}
EU_t^{(0)} &= e_t^{(0)}(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t)) - (e_t^{(0)})^2/(2c_0)\\
&= c_0(\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t))/2.
\end{aligned}$$

So the first strategy is preferred if $EU_t^{(1)} > EU_t^{(0)}$, or if

$$\theta + a(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) > \frac{2C}{c_1 - c_0}.$$

Therefore, the effect of the reform (higher α) can be summarized in the following propositions.

Proposition 12. 1. *Hospitals split into two groups in the static model with fixed cost. Hospitals pay fixed cost in the group with sufficiently high θ or m_t, and hospitals do not pay fixed cost in another group.*

2. *The effect of the reform α is higher for hospitals which pay fixed cost and choose $e_t^{(1)}$ than for hospitals which refuse to bear fixed cost and choose $e_t^{(0)}$. In other words, $\partial e_t^{(1)}/\partial\alpha > \partial e_t^{(0)}/\partial\alpha$.*

Proposition 13. 1. *The higher the value of α, the more likely is the choice of fixed cost by hospitals with $m_t > 1/\kappa$.*

2. *The higher the value of α, the less likely is tehe choice of fixed cost by hospitals with $m_t < 1/\kappa$.*

The empirical part of the paper employs the data of the *Promoting Interoperability Program* and of the *American Hospital Association IT survey* to establish whether hospital incurred the fixed cost of installing the software for maintaining electronic health records. It should be noted that more than 90% of the hospitals in our sample have the EHR software installed by the beginning of the analyzed period. Moreover, the Program provides ample compensations to certified hospitals, so the fact of bearing fixed cost may not be considered as revealing differential patterns of quality dynamics at Medicare hospitals under value-based purchasing.

3.5 Testable hypotheses

3.5.1 Mean effect of the quality incentive

The first set of predictions concerns the mean effect of pay-for-performance with a linear rule. According to Proposition 6, the mean measured quality μ increases in α. So the model predicts a positive mean effect of the introduction of pay-for-performance and a positive mean effect of the increase in α within the implementation of pay-for-performance (Hypothesis $H1a$). Proposition 6 demonstrates that the derivative of λ in α is positive. Accordingly, the convergence of measured quality becomes weaker when larger share of the hospital budget is at risk (Hypothesis $H1b$).

$H1a$. The introduction of pay-for-performance and increase of the share of hospital funds at risk in pay-for-performance raise the mean level of measured quality.

$H1b$. Increase of the share of hospital funds at risk in pay-for-performance weakens the effect of convergence of the measured quality to the mean value.

As may be inferred from Remark 11, the empirical proof of $H1a$ (i.e. the fact that the mean level of measured quality increases in α) implies that hospitals can be treated as agents who take their future payments into account.

3.5.2 Heterogeneous effects of the quality incentive

$H2$. The introduction of pay-for-performance and the increase of the share of hospital funds at risk in pay-for-performance raise measured quality at high-quality hospitals more than at low-quality hospitals.

This hypothesis follows from Corollary 7, Proposition 8, and Remark 9.

3.5.3 Net total effect

It should be noted that the presence of mean reversion implies a differential time profile of measured quality: measured quality increases at hospitals in low percentiles of the quality distribution and decreases at hospitals in high percentiles. Combined with the positive effect of pay-for-performance on the mean value of measured quality μ, mean reversion results in heterogeneous net total effect of change in measured quality over time.

H3a. High-quality hospitals experience decrease of measured quality owing to regression towards the mean. However, the introduction of pay-for-performance and increase of the share of hospital funds at risk in pay-for-performance lead to improvements in measured quality at these hospitals. The net total effect may vary.

H3b. Low-quality hospitals increase their measured quality owing to regression towards the mean. The introduction of pay-for-performance and increase of the share of hospital funds at risk in pay-for-performance also cause a rise in measured quality, so the net total effect at these hospitals is positive.

Finally, the value of coefficient λ depends positively on α. If α is gradually raised in the course of implementation of the incentive scheme then, according to *H1b*, convergence of measured quality weakens over time. The net total effect at high-quality hospitals is the sum of the positive effect of the quality incentive and negative effect of the quality convergence. With increase in α, the number of hospitals where the positive effect outweighs the negative becomes larger.

H3c. The increase of hospital funds at risk under pay-for-performance weakens the effect of convergence of measured quality, so the number of high-quality hospitals with negative net total effect decreases.

4 Empirical approach

4.1 Data

4.1.1 Data sources and variables

The analysis uses data for Medicare hospitals from several sources. We use *Hospital Compare* data archives (January 2020 update) for quality measures, hospital ownership and geographic location. The medical school affiliation of a hospital, the number of hospital beds, nurses and physicians come from *Provider of Service* files. Other hospital control variables are taken from the Final Rules, which are Medicare's annual documents on reimbursement rates in the inpatient prospective payment system. Specifically, we use information from the *Impact Files*, which accompany the Final Rules and estimate the impact of the reimbursement mechanism on hospital characteristics. The variables taken from the *Impact Files* are the share of Medicare's discharges, ownership and urban location.

Patient characteristics are also taken from the *Impact Files*. The casemix variable reflects the relative weight of each DRG in financial terms and is adjusted for transfers of patients between hospitals.[11] Casemix makes it possible to control for the composition of patient cases

[11] If a patient was transferred to/from a hospital, then the transfer-adjustment factor is the lesser of one and the value of the patient's length of stay relative to the geometric mean of the national length of stay for this DRG. See Federal Register 2011, 42 CFR, Part 412.

taking account of the objective link between severity of illness and hospital resources. The disproportionate-share index accounts for the share of low-income patients and makes it possible to proxy a patient's income.

To account for other major channels of quality improvement by Medicare hospitals over the observed time period, we use the data for two programs run by the Centers for Medicare and Medicaid Services. One of them is the Hospital Readmissions Reduction Program, which applies to Medicare hospitals since fiscal year 2013 and penalizes them for excess readmissions. Specifically, the payment reduction which may equal from 0 to 3% is applied to hospital's Medicare remuneration, higher values of the percentage for the penalty represent more excess readmissions at the hospital. Using the *Hospital Readmissions Reduction Program (HRRP) Supplemental data files*, which accompany annual Final Rules on acute inpatient PPS (November 2019 update), we obtain values of the HRRP penalty for 2013–2018 and use them as one of the control variables in the empirical analysis.

We also consider the Electronic Health Records (EHR) Incentive Program, which is in force since 2011. The program establishes hospital attestation on the use of electronic health records. The adoption of quality-improving information technology requires substantial fixed cost, so the binary variable for hospital attestation within EHR allows to control for the fixed cost in the empirical analysis. The EHR promotion program consists of three stages (sequentially introduced in 2011, 2014 and 2017). Using data from *The Eligible Hospitals Public Use Files* on the EHR incentive program (February 2020 update), we set the EHR attestation dummy equal to one if the hospital passed its attestation for the given year at any stage. Owing to non-availability of data on the third stage of the program, we extend the second stage data from year 2016 to years 2017 and 2018. The use of attestation dummy enables to control for the fact of incurring the fixed cost of quality-improvement efforts. Owing to the small size of the non-EHR group (only 8-10% of sample), we do not analyze whether quality goes up faster in the group of the EHR hospitals (i.e. for instance, we do not interact the attestation dummy with α).

4.1.2 Sample and the flow of quality

The non-anonymous character of the data sources allows us to merge them by year and hospital name. Our analysis focuses on Medicare's acute-care hospitals, as the pay-for-performance incentives contract applies exclusively to this group. We restrict the sample by considering only hospitals with share of Medicare cases greater than 5%. There are 2968 such hospitals in our sample which make 15971 observations within 2011–2018 (Table 1). The aggregate quality measure TPS_{it} is the major variable for our analysis, so we begin with 2011 (the first year for which the corresponding data are available). The total performance score is calculated on a calendar-year basis, so we analyze annual data, and 2018 is the last year for which data have been released. The mean value of TPS over the analyzed period is 37.852 and the range of the variable is $(2.727, 98.182)$.

The descriptive analysis of the values of *TPS* offers suggestive evidence in support of some of the main hypotheses generated by the model. Specifically, we focus on the flow of hospitals between quintiles of *TPS* in different years. The Sankey diagram in Figure 1 uses the width of arrows as the intensity of flow rates and demonstrates how hospitals change their position in quintiles of the composite quality measure after the introduction of pay-for-performance (from 2012 to 2013).

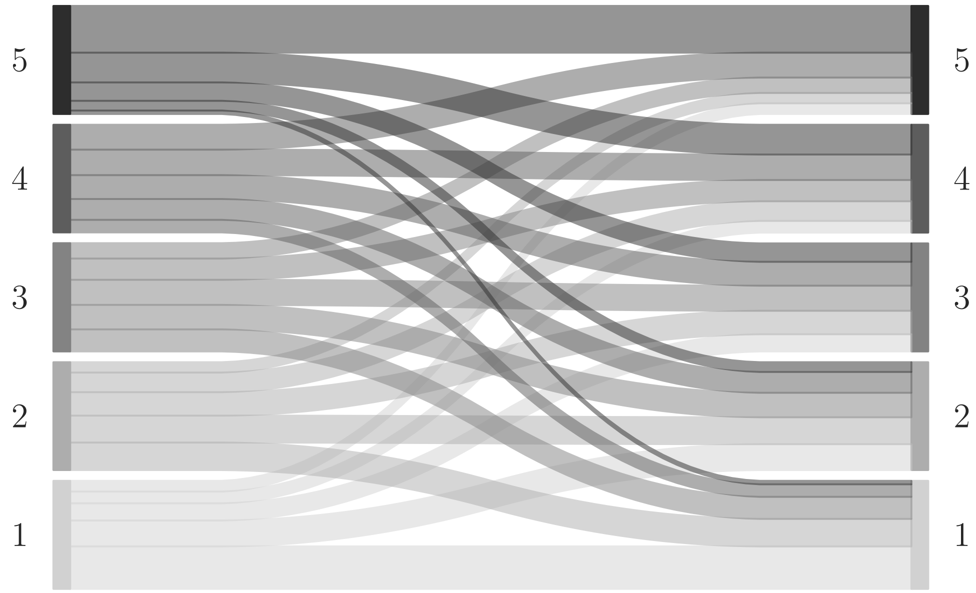

Figure 1: Flow of hospitals between *TPS* quintiles from 2012 to 2013

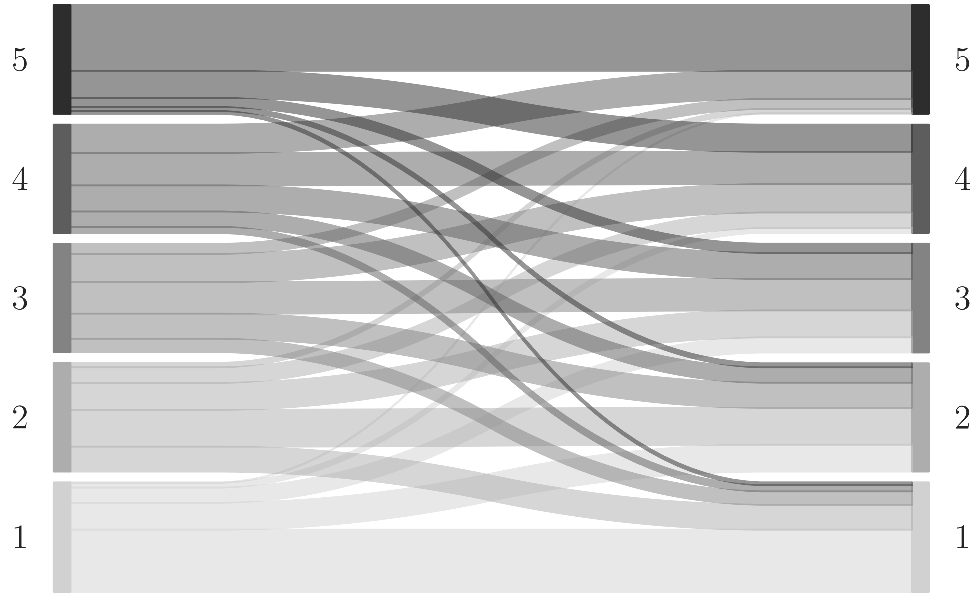

Figure 2: Flow of hospitals between *TPS* quintiles from 2016 to 2017

As can be inferred from Figure 1 there is considerable movement of hospitals between quintiles. This can be regarded as graphic support for the phenomenon of mean reversion since hospitals would rarely change their quintile from year to year in the absence of mean reversion.

It should be noted that pure mean reversion would cause uniform flow of hospitals from any quintile to all five quintiles. However, the flows are not in fact uniform and this may be explained by additional heterogeneity among hospitals (i.e. heterogeneous hospital response to the incentive scheme over the analyzed period).

Note that our model predicts that the mean reversion becomes weaker when there is an increase of α. Figure 2 supports this prediction. It shows the flow of hospitals between quintiles of TPS from 2016 to 2017, when the value of α was the highest in the analyzed period.[12] Compared to Figure 1, the flows are much weaker, so hospitals change their position in quintiles less often.

[12]The Sankey diagram for 2017–2018 when α remained at its highest value demonstrates similar pattern.

Table 1: Descriptive statistics for Medicare's acute-care hospitals in 2007–2018

Variable	Definition	Obs	Mean	St.Dev	Min	Max
Hospital performance						
TPS	Hospital total performance score	15971	37.852	11.382	2.727	98.182
Patient characteristics						
casemix	Transfer-adjusted casemix index	15971	1.588	0.296	0.834	3.972
dsh	Disproportionate share index, reflecting the prevalence of low-income patients	15971	0.306	0.165	0	1.232
Hospital characteristics						
nurses/beds	nurse-to-bed ratio	15971	1.310	3.909	0	170.479
physicians/beds	physician-to-bed ratio	15971	0.099	1.013	0	70.992
beds	Number of beds	15971	271.297	240.673	3	2449
log(*beds*)	Number of beds (in logs)	15971	5.280	0.821	1.099	7.803
medicare share	Share of Medicare cases	15971	0.381	0.119	0.050	0.983
HRRP penalty	Percentage reduction of the Medicare payments under Hospital Readmissions Reduction Program	15971	0.480	0.577	0	3.000
MUEHR	=1 if passed attestation for meaninful usage of electronic health records	15971	0.921	0.269	0	1
urban	=1 if an urban hospital	15971	0.729	0.444	0	1
public	=1 if managed by federal, state or local government, or hospital district or authority	15971	0.149	0.356	0	1
teaching	=1 if hospital has medical school affiliation	15971	0.362	0.481	0	1
Hospital location						
New England	=1 if located in Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, or Vermont	15971	0.046	0.210	0	1
Mid-Atlantic	=1 if located in New Jersey, New York, or Pennsylvania	15971	0.123	0.329	0	1
East North Central	=1 if located in Illinois, Indiana, Michigan, Ohio, or Wisconsin	15971	0.168	0.373	0	1
West North Central	=1 if located in Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, or South Dakota	15971	0.081	0.272	0	1
South Atlantic	=1 if located in Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, District of Columbia, or West Virginia	15971	0.176	0.381	0	1
East South Central	=1 if located in Alabama, Kentucky, Mississippi, or Tennessee	15971	0.087	0.282	0	1
West South Central	=1 if located in Arkansas, Louisiana, Oklahoma, or Texas	15971	0.129	0.336	0	1
Mountain	=1 if located in Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, or Wyoming	15971	0.068	0.252	0	1
Pacific	=1 if located in California, Oregon, Washington, Alaska, or Hawaii	15971	0.114	0.318	0	1

Note: Section 401 hospitals are treated as rural hospitals.

4.2 Specification

For the purpose of the empirical analysis, we analyze the comparative statics of different steady states of the stochastic process in our theoretical model.[13] The dependent variable y_{it} is the total performance score (TPS) of hospital i in year t.

While quality-related efforts of a hospital and the TPS composite quality measure are multi-dimensional, we do not touch upon multi-tasking in our theoretical model and in the empirical estimations. Our approach considers a one-dimensional effort, a one-dimensional true quality and its measurable proxy.[14] So the empirical specification describes a linear rule applied to the composite quality measure.

The value of y_{it} is used for remuneration of Medicare hospitals at time $t+2$, so we employ the second order dynamic panel,

$$y_{it} = \phi_0 + \phi_1 y_{it-1} + \phi_2 y_{it-2} + \phi_3 \alpha_t s_{it} + \phi_4 \alpha_t s_{it} y_{it-1} \\ + \phi_5 \alpha_t s_{it} y_{it-2} + \delta_0 s_{it} + z'_{it}\delta_1 + \alpha_t s_{it} \cdot z'_{it}\delta_2 + d'_t \delta_3 + u_i + \varepsilon_{it}, \quad (18)$$

where z_{it} are hospital time-varying characteristics, u_i are fixed effects (in particular, they incorporate the altruistic effects), the size of quality incentives α_t varies in different years and enters the equation multiplied by the share of Medicare discharges s_{it}, which indicates that the quality incentives apply only to treatment of Medicare patients, and d_t is a set of dummy variables which capture external time effects (effects unrelated to hospital decisions), the sum of the coefficients for components of d_t is normalized to zero. Hospital time-varying characteristics are disproportionate share index, casemix index, number of hospital beds, physician-to-bed ratio and nurse-to-bed ratio. The posterior analysis of the effect of quality incentives deals with hospital grouping according to the time-invariant characteristics, which could not be incorporated in the empirical specification with fixed effects: geographic region for hospital location, public ownership, urban location, teaching status.

We employ two hospital control variables which affect quality improvement and allow us to mitigate potential biases, which might occur if the pay-for-performance effect is identified based

[13]Panel data estimations assume that the stochastic process changes the steady state from year to year. The assumption is a simplification of real-world quality dynamics but it enables the empirical analysis of annual quality changes.

[14]Note that in case of Medicare's formula, the true multi-dimensional quality of hospitals (and hence quality-related efforts) is transformed into measured quality (i.e. TPS) in a non-linear manner, owing to the step-wise scale used for computing the points for each measure. We might nonetheless assume that quality is transformed into TPS monotonically and can be linearized in the empirical part of the paper. Several arguments may be listed to support the conjecture. First, the data for Medicare hospitals show that no hospital has the highest possible step-wise values for all its measures. So even the best hospitals have an incentive to work to increase at least one of their measure scores in order to improve TPS. Second, we can neglect disincentives within the step-wise scale used for aggregating measure scores, which may cause deterioration of quality for hospitals that are already positioned at the highest step. Such hospitals could afford only a slight decrease of their quality (due to slacking efforts) and yet remain at this step, and the impact of a fall in quality in only one quality measure on the value of TPS will be negligible. Third, interviews with executives of hospitals using value-based purchasing show that a hospital rarely gives special attention to a given subset of measures or shifts its administerial and other efforts across measures. All dimensions of TPS are monitored and actions to improve each dimension are implemented (Smith, 2017; Jones, 2014).

only on the variation of α in time. The HRRP penalty captures the impact of a simultaneously adopted incentives program with similar incentives. Moreover, the readmissions reduction program targets improvement of quality measures which are components of TPS.[15] The binary variable for successful attestation of meaningful usage of electronic health records accounts for the effect of another compulsory program, which provides bonuses to attested hospitals. The variable controls for the fixed cost incurred by a hospital to improve its quality through installing and using health information technology systems.

Casemix and the disproportionate share index are assumed to be endogenous: we rely on the empirical evidence of manipulation by hospitals of patient diagnoses (i.e. with casemix) and reluctance to admit low-income patients under quality-incentive schemes (Damberg et al., 2014; Eijkenaar et al., 2013; Centers for Medicare and Medicaid Services, 2007a). We assume that the Medicare share is endogenous, too: the fact may be explained by demand-side response from Medicare patients to publicly reported hospital quality (Ma and Mak, 2015; Siciliani et al., 2013; Hillman et al., 1991).

We estimate the model using the the Arellano and Bover (1995) and Blundell and Bond (1998) methodology for dynamic panel data, with Windmeijer (2005) robust standard errors. Here lagged levels and lagged differences of the endogenous variables are used as instruments, and validity of instruments is assessed through the Arellano-Bond test statistics.[16]

Lagged value of TPS and other hospital control variables in z_{it} (beds, physician-to-bed and nurse-to-bed ratios, HRRP penalty and the binary variable for hospital EHR attestation) are taken as predetermined and do not require the use of instruments in estimations.

We note the limitations of our approach. Both the theoretical and empirical parts of our analysis deal with the composite quality measure. We do not touch on the rules for computing the scores of each dimension of the composite measure or on aggregation of dimension scores. It is important to note that Medicare uses whichever is highest, improvement points or achievement points, as the score for each dimension. The choice between achievement and improvement points stimulates low-performing hospitals, and the uniform formula assumes that all groups of hospitals have equal margin for improvement.[17] The approach used by Medicare is in contrast with the methodology used in France, where all providers are stimulated according to improvement while only providers with quality above the mean value are also rewarded for their achievement (Bousquet et al., 2014).

Weighting of scores across domains is another feature of the design of the incentive mechanism which is not analyzed in our paper. So the dichotomous variables for annual periods in the empirical specification capture time effects unrelated to Medicare's value-based purchasing as well as time effects not associated with the size of incentives but potentially linked to changes in other elements of the reform design (i.e. changes in weights).

[15] 30-day unplanned readmission rates for acute myocardial infarction, hearth failure and pneumonia.

[16] The Sargan statistic may be used in dynamic panels for assessing validity of instruments under the homoskedasticity assumption. But it is not applicable to our specification with robust standard errors.

[17] A minor exception is protection of hospitals above the benchmark value of the 95th percentile of a corresponding measure score: these hospitals receive 10 points for their achievement on a $[0, 10]$ scale, while the maximum number of points for improvement by any hospital is 9.

Conventional policy evaluation using a control group of hospitals is not possible because quality measures for non-Medicare hospitals are not available.[18] The empirical part of the paper therefore focuses solely on pay-for-performance hospitals and identifies the effect of quality incentives based on variation in α_t.

It should be noted that our analysis does not touch on the effect of quality incentives on the true quality of hospital care. Mortality indicators are the variables which are commonly used as best proxies for true quality (Sutton et al., 2012), but the time-series data available for mortality at Medicare hospitals did not enable estimation of the autoregressive process.

4.3 Testing of hypotheses

4.3.1 Identification of the long-term mean

We interpret the second order dynamic panel (18) as a second order autoregressive process. The coefficients for the first and the second lags of y_{it} in this AR(2) process equal $\phi_1 + \phi_4\alpha_t s_{it}$ and $\phi_2 + \phi_5\alpha_t s_{it}$ respectively. Note that both coefficients are linear functions of α_t. While the standard form of the AR(2) process contains only the lags of the dependent variable, the right-hand side of our empirical equation includes various hospital characteristics and control variables.

To test the predictions of the model which concern the long-term mean value of the measured quality μ as a function of α, we measure the mean fitted value of y_{it} as follows.

For a fixed value of α we take the unconditional expected values of both sides of (18) and denote $\mu(\alpha) = E(y_{it})$:

$$\begin{aligned}\mu(\alpha) = {} & \phi_0 + \phi_1\mu(\alpha) + \phi_2\mu(\alpha) + \phi_3\alpha E(s_{it}) + \phi_4\alpha E(s_{it})\mu(\alpha) + \phi_5\alpha E(s_{it})\mu(\alpha) \\ & + \phi_4\alpha \operatorname{cov}(s_{it}, y_{it-1}) + \phi_5\alpha \operatorname{cov}(s_{it}, y_{it-2}) + \delta_0 E(s_{it}) + E(z_{it})'\delta_1 + \alpha E(s_{it}z_{it})'\delta_2 + E(d_t)'\delta_3,\end{aligned} \tag{19}$$

where $E(d_t)'\delta_3 = 0$ because of the normalization of coefficients δ_3 in (18). After collecting the terms with μ and rearranging them, we obtain:

$$\mu(\alpha) = \frac{\phi_0 + \phi_3\alpha E(s_{it}) + \delta_0 E(s_{it}) + E(z_{it})'\delta_1 + \alpha E(s_{it}z_{it})'\delta_2 + \phi_4\alpha \operatorname{cov}(s_{it}, y_{it-1}) + \phi_5\alpha \operatorname{cov}(s_{it}, y_{it-2})}{1 - \phi_1 - \phi_2 - \phi_4\alpha E(s_{it}) - \phi_5\alpha E(s_{it})}.$$

Since α differs across t, we use sample means across the hospitals for fixed t to obtain estimates of expectations.

[18]The total performance score or all its components are only available for hospitals in the Hospital Compare database. The Hospital Compare database does include a small group of non-incentivized hospitals together with value-based purchasing hospitals. These are children's hospitals and critical-access hospitals. But both groups offer a special type of healthcare and are not comparable with acute-care hospitals. Moreover, critical-access hospitals usually have no more than 20 beds, which makes it impossible to find a close match with acute-care hospitals. See Ryan et al. (2017) for an attempt of matching acute-care and critical-access hospitals.

The estimate of the long-term mean value $\mu(\alpha)$ is constructed by replacing the expected values and covariances by corresponding sample means and sample covariances:

$$\mu(\alpha) = \frac{\phi_0 + \phi_3\alpha\overline{s} + \delta_0\overline{s} + \overline{z}'\delta_1 + \alpha\overline{s}\overline{z}'\delta_2 + \phi_4\alpha\,\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha\,\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha\overline{s} - \phi_5\alpha\overline{s}}.$$

Note that the expression for $\mu(\alpha)$ does not contain the time effects $d_t'\delta_3$, as they represent shifts in quality which are common to all the hospitals and are caused by external circumstances.

4.3.2 Empirical tests

The policy parameter α increases in 2013–2017 and remains unchanged in 2017–2018. As follows from the hypothesis $H1a$ introduced in subsection 3.5.1, the value of $\mu(\alpha_t)$ should increase through 2013–2017 and should become flat in 2017–2018.

Accordingly, we examine the difference between $\mu(\alpha_t)$ and $\mu(\alpha_{t-1})$:

$$\begin{aligned}\mu(\alpha_t) - \mu(\alpha_{t-1}) = {} & \frac{\phi_0 + \phi_3\alpha_t\overline{s} + \delta_0\overline{s} + \overline{z}'\delta_1 + \alpha_t\overline{s}\overline{z}'\delta_2 + \phi_4\alpha_t\,\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha_t\,\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha_t\overline{s} - \phi_5\alpha_t\overline{s}} \\ & - \frac{\phi_0 + \phi_3\alpha_{t-1}\overline{s} + \delta_0\overline{s} + \overline{z}'\delta_1 + \alpha_{t-1}\overline{s}\overline{z}'\delta_2 + \phi_4\alpha_{t-1}\,\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha_{t-1}\,\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha_{t-1}\overline{s} - \phi_5\alpha_{t-1}\overline{s}}.\end{aligned}$$

The null hypothesis is:

$$H_0\colon \mu(\alpha_t) - \mu(\alpha_{t-1}) = 0,$$

and it is tested against the positive alternative.

Equivalently, we compute the difference between $\mu(\alpha)$ and $\mu(0)$:

$$\begin{aligned}\mu(\alpha) - \mu(0) = {} & \frac{\phi_0 + \phi_3\alpha\overline{s} + \delta_0\overline{s} + \overline{z}'\delta_1 + \alpha\overline{s}\overline{z}'\delta_2 + \phi_4\alpha\,\widehat{\text{cov}}(s, L(y)) + \phi_5\alpha\,\widehat{\text{cov}}(s, L^2(y))}{1 - \phi_1 - \phi_2 - \phi_4\alpha\overline{s} - \phi_5\alpha\overline{s}} \\ & - \frac{\phi_0 + \overline{z}'\delta_1}{1 - \phi_1 - \phi_2}.\end{aligned}$$

Note that $\mu(0)$ represents the value of the long-term mean in the pre-reform years when $\alpha = 0$ and is obtained analytically by plugging $\alpha = 0$ into the expression for $\mu(\alpha)$.

The null hypothesis is:

$$H_0\colon \mu(\alpha_t) - \mu(0) = 0,$$

and it is tested against the positive alternative.

In conjunction with hypothesis $H1a$, $\mu(\alpha_t) - \mu(\alpha_{t-1})$ should be positive in 2013–2017 and should be close to zero in 2017–2018. Equivalently, $\mu(\alpha_t) - \mu(\alpha_0)$ should be positive in 2013–2018 and should increase over the period 2013–2017.

Now consider hypothesis $H1b$. The persistence parameter $\lambda(\alpha)$ describes how quickly the effect of an error in measuring quality fades over time. For a second order autoregressive process the rate of convergence of the conditional expected value of y_{it} decays exponentially at a rate equal to the reciprocal value of the smallest root of the characteristic equation for the AR(2) process:

$$1 - (\phi_1 + \phi_4 \alpha_t s_{it})\lambda - (\phi_2 + \phi_5 \alpha_t s_{it})\lambda^2 = 0$$

(Hamilton, 1994, Section 2.3). Again, for a fixed value of α we take expectations

$$1 - (\phi_1 + \phi_4 \alpha E(s_{it}))\lambda - (\phi_2 + \phi_5 \alpha E(s_{it}))\lambda^2 = 0.$$

Then we replace the expected values by sample means and solve this quadratic equation to obtain the following formula for $\lambda(\alpha)$:

$$\lambda(\alpha) = \frac{\phi_1 + \phi_4 \alpha \overline{s} + \sqrt{(\phi_1 + \phi_4 \alpha \overline{s})^2 + 4(\phi_2 + \phi_5 \alpha \overline{s})}}{2},$$

where $\overline{s}$ is the the mean value of the share of Medicare cases for a given year.

An alternative approach considers the value of the autocorrelation function ACF(1) (the correlation coefficient between y_{it} and y_{it-1}) as the persistence parameter λ. Specifically, for the second order autoregressive process (18) the estimated value of ACF(1) becomes

$$\lambda(\alpha) = \frac{\phi_1 + \phi_4 \alpha \overline{s}}{1 - \phi_2 - \phi_5 \alpha \overline{s}}$$

(Hamilton, 1994, Section 3.4).

Testing $H1b$ implies analyzing whether $\lambda(\alpha)$ is an increasing function of α. So, similarly to $H1a$, the null hypothesis:

$$H_0\colon \lambda(\alpha_t) - \lambda(\alpha_{t-1}) = 0$$

is tested against the positive alternative.

Alternatively, we assess whether $\lambda(\alpha) - \lambda(0)$ is positive, whether it increases in 2013–2017 and changes only negligibly in 2017–2018.

To assess $H2$ we compute the effect of pay-for-performance as $\mu(\alpha_t) - \mu(0)$ or $\mu(\alpha_t) - \mu(\alpha_{t-1})$ at different quintiles of the lagged TPS_{it}, where quintile 1 denotes the lowest quality and quintile 5 denotes the highest. We investigate whether the effect is positive for $\mu(\alpha_t) - \mu(0)$ in 2013–2018 (and for $\mu(\alpha_t) - \mu(\alpha_{t-1})$ in 2013–2017) and whether the effect increases by quintile.

Testing $H3a$ and $H3b$ involves computing the predicted values of TPS_{it} at the mean value of each covariate for different quintiles of the lagged TPS_{it} and examining whether in 2013–2018 they change from positive in the lowest quintiles to negative or insignificant in the highest quintiles. Average difference between predicted TPS and lagged TPS shows the expected change in quality in consecutive years (the net total effect) which is the sum of the effect of pay-for-performance and the impact of mean reversion.

Evaluation of $H3c$ involves calculating annual values of the net total effect for quintiles of the lagged TPS and examining whether the negative values of the net total effect become less frequent across highest quintiles during the period 2013–2017 and stays constant in 2017–2018.

5 Empirical results

The specification with the second order lag enables estimation of the fitted values of TPS_t and the values of the long-term mean μ_t only starting 2013. Accordingly, we can analyze how the increase of α within the implemented linear incentive impacts quality improvement. We can also use the mean values $\mu(\alpha)$ for $\alpha = 0$ and measure the difference between $\mu(\alpha_t)$ and $\mu(0)$ to assess the quality changes relative to the pre-reform years 2011–2012. (In the absence of quality incentive $\alpha = 0$ in 2011–2012.)

The first set of our results is reported in Table 2 and concerns the mean effect of pay-for-performance at Medicare hospitals.

Table 2: Effect of pay-for-performance on the mean quality

	2013	2014	2015	2016	2017	2018
α_t	1.0	1.25	1.5	1.75	2.0	2.0
$\mu(\alpha_t)$	31.777***	33.421***	36.596***	38.497***	41.469***	41.691***
	(0.946)	(0.628)	(0.350)	(0.604)	(1.311)	(1.286)
$\mu(\alpha_t) - \mu(0)$	2.928***	4.160***	6.143***	8.388***	11.171***	11.181***
	(0.760)	(1.062)	(1.453)	(1.990)	(2.702)	(2.642)
$\mu(\alpha_t) - \mu(\alpha_{t-1})$	2.928***	1.644***	3.175***	1.901***	2.972***	0.222
	(0.760)	(0.399)	(0.517)	(0.621)	(0.791)	(0.257)

Notes: Standard errors calculated using the delta-method are in parentheses.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.

Measured as $\mu(\alpha_t) - \mu(0)$, the mean effect of pay-for-performance is positive in 2013–2018. The value of the effect increases in α_t in 2013–2017. (The increase in 2017–2018 is negligible and is in line with the fact that α has remained flat since 2017.) Similarly, the change in the effect of pay-for-performance in consecutive years, defined as $\mu(\alpha_t) - \mu(\alpha_{t-1})$, is positive for 2013–2017 and is extremely small in 2018 in comparison with the previous years. The finding corresponds to our hypothesis $H1a$ of improvement in mean quality owing to the introduction of pay-for-performance (i.e. the increase of α from 0 to 1) or with linear increase of the reward function (from 1 to 2 in 2013–2017).

Note that the mean value of $\mu(\alpha_t)$ increases in α_t, which supports our supposition that hospital managers take account of future benefits from improving current values of hospital quality. In other words, the intertemporal discount factor β differs from zero in the hospital problem, so hospital behavior is dynamic.

Table 3 shows the second set of results for heterogeneity of hospital response to pay-for-performance. The parameter λ is estimated as the inverse of the smaller root of AR(2) or as ACF(1). The values are significant and less than unity under both approaches. This points to mean reversion, so quality decreases towards the mean at high-quality hospitals and goes up towards the mean at hospitals with low quality. The values of λ rise with an increase in the size of incentives α, which implies that the persistence of the dynamic process increases, and hence the effect of mean reversion becomes weaker. Similarly, the values of $\lambda(\alpha_t) - \lambda(0)$ are

Table 3: Effect of pay-for-performance on mean reversion

	2013	2014	2015	2016	2017	2018
α_t	1.0	1.25	1.5	1.75	2.0	2.0
$\lambda(\alpha_t)$	0.263	0.441***	0.545***	0.619***	0.684***	0.675***
	(0.175)	(0.033)	(0.021)	(0.017)	(0.017)	(0.017)
$\lambda(\alpha_t) - \lambda(0)$	−0.208	−0.029	0.075	0.148***	0.214***	0.205***
	(0.212)	(0.068)	(0.053)	(0.047)	(0.043)	(0.044)
$\lambda(\alpha_t) - \lambda(\alpha_{t-1})$	−0.208	0.178	0.104***	0.073***	0.066***	−0.009*
	(0.212)	(0.145)	(0.017)	(0.009)	(0.008)	(0.005)
$\lambda(\alpha_t)$ (alternative)	0.414***	0.453***	0.499***	0.543***	0.592***	0.584***
	(0.019)	(0.016)	(0.013)	(0.013)	(0.015)	(0.015)
$\lambda(\alpha_t) - \lambda(0)$ (alternative)	0.152***	0.192***	0.238***	0.282***	0.330***	0.323***
	(0.021)	(0.026)	(0.031)	(0.037)	(0.042)	(0.041)
$\lambda(\alpha_t) - \lambda(\alpha_{t-1})$ (alternative)	0.152***	0.040***	0.046***	0.044***	0.048***	−0.007*
	(0.021)	(0.005)	(0.006)	(0.006)	(0.007)	(0.004)

Notes: Standard errors calculated using the delta-method are in parentheses.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.
The persistence parameter $\lambda(\alpha_t)$ is estimated as the inverse of the smaller root of AR(2) or as ACF(1), the latter is denoted as "alternative".

positive and increase in α_t. The time change in the convergence parameter: $\lambda(\alpha_t) - \lambda(\alpha_{t-1})$ is positive for 2013–2017. The value of $\lambda(\alpha_t) - \lambda(\alpha_{t-1})$ becomes negative in 2018 but is very small in absolute terms. The results support hypothesis $H1b$ of weakening of quality convergence to the mean value with a rise in α.

The heterogeneous changes in hospital quality owing to pay-for-performance are given in Tables 4– 5 where hospitals are divided into quintiles according to the values of their *TPS*. Note that the change in hospital quality is the function of regression coefficient and the mean values of covariates. So its standard error consists of two parts: the error of the estimated regression coefficient and the error of the mean values of covariates. Only the second part of this error depends on sample size and should go up approximately $\sqrt{5}$ times due to analysis by quintiles. However, the weight of this second part proves to be relatively small in case of our data, so the standard errors in Tables 4– 5 are only slightly larger than standard errors in Table 2.

The estimates support hypothesis $H2$: the higher the quintile of the quality distribution in the previous period, the larger the effect of pay-for-performance in terms of $\mu(\alpha_t) - \mu(0)$. With the exception of quintile 5, the estimate for the previous quintile is less than the estimate for the current quintile, differences in the effect of pay-for-performance across consecutive quintiles are mostly statistically significant. The change of the effect of pay-for-performance over time $\mu(\alpha_t) - \mu(\alpha_{t-1})$ goes up with a rise of the quality incentive α but almost stops increasing in 2018 when α becomes constant, as shown in Table 5. So pay-for-performance stimulates quality increase in all groups of Medicare's hospitals, and the impact of pay-for-performance is greater at higher-quality hospitals. Similarly to the results in Table 4, with the exception of quintile 5,

Table 4: Effect of pay-for-performance as $\mu(\alpha_t) - \mu(0)$ for quintiles of TPS_{t-1}

	2013	2014	2015	2016	2017	2018
quintile 1	2.411***	3.696***	5.206***	7.315***	9.818***	9.501***
	(0.762)	(1.055)	(1.421)	(1.882)	(2.545)	(2.416)
quintile 2	2.784***	3.777***	5.650***	7.625***	10.748***	10.359***
	(0.767)	(1.062)	(1.455)	(1.980)	(2.705)	(2.595)
the difference	−0.373***	−0.081	−0.444***	−0.310	−0.930**	−0.858**
	(0.142)	(0.131)	(0.183)	(0.269)	(0.412)	(0.407)
quintile 3	2.871***	4.047***	5.575***	8.186***	10.548***	11.004***
	(0.763)	(1.075)	(1.477)	(2.040)	(2.669)	(2.711)
the difference	−0.087	−0.270**	0.075	−0.561**	0.200	−0.645
	(0.077)	(0.120)	(0.167)	(0.266)	(0.398)	(0.437)
quintile 4	3.002***	4.259***	6.211***	8.512***	11.083***	11.131***
	(0.764)	(1.081)	(1.501)	(2.051)	(2.753)	(2.689)
the difference	−0.131*	−0.212*	−0.635***	−0.326	−0.535	−0.127
	(0.069)	(0.109)	(0.238)	(0.265)	(0.446)	(0.492)
quintile 5	3.192***	4.385***	6.785***	8.852***	11.854***	11.734***
	(0.762)	(1.085)	(1.528)	(2.065)	(2.951)	(2.912)
the difference	−0.190**	−0.126	−0.574	−0.340	−0.771	−0.603
	(0.097)	(0.250)	(0.512)	(0.457)	(0.670)	(0.611)

Notes: Quintile 1 denotes the lowest quality and quintile 5 – the highest. The differences between the previous and the current quintile are reported.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.
Standard errors (calculated using the delta-method for the difference of the reform effects across corresponding two categories of each time-invariant hospital characteristic) are in parentheses.
There are two sources of errors in the estimates shown in the Table: the error of regression coefficient and the error of the mean values of covariates. The first part of the error does not vary across all result tables, while the second part of the error depends on the group size and is approximately $\sqrt{5}$ times larger than its counterpart in Table 2. However, the errors of regression coefficient are considerably bigger than the errors of mean values of covariates, so the increase in the standard errors in this table relative to the standard error in Table 2 is only minor.

the differences in the change of the effect across consecutive quintiles are mostly negative and statistically significant.

Table 6 gives estimates of the net total effect, i.e. the expected change in hospital quality over time, measured as the difference between the predicted TPS and the lagged TPS. The net total effect is the sum of the impact of mean reversion and the effect of pay-for-performance.

Note that the estimation of the fitted value of TPS includes time effects which account both for time trend and for important changes in the incentive mechanism not captured by variation in α. An example of such change occurred in 2015 and temporarily decreased the value of TPS for each hospital.[19] Accordingly, Table 6 shows that the values of predicted TPS minus lagged TPS go down in 2015 for each quintile.

[19]Specifically, the pneumonia cohort was expanded and it caused a rise in pneumonia readmission rates in 2015. Additionally, the safety domain with relatively low scores in comparison to measures of other domains was added to the list of measures which constitute TPS.

Table 5: Effect of pay-for-performance as $\mu(\alpha_t) - \mu(\alpha_{t-1})$ by quintiles of TPS_{t-1}

	2013	2014	2015	2016	2017	2018
quintile 1	2.411***	−0.007	1.267***	2.883***	2.304***	0.355
	(0.762)	(0.530)	(0.517)	(0.631)	(0.774)	(0.467)
quintile 2	2.784***	0.982*	2.374***	2.181***	3.163***	−0.274
	(0.767)	(0.502)	(0.571)	(0.688)	(0.876)	(0.491)
the difference	−0.373***	−0.989*	−1.107**	0.701	−0.859	0.629
	(0.142)	(0.512)	(0.481)	(0.542)	(0.610)	(0.660)
quintile 3	2.871***	1.201***	1.742***	3.446***	2.601***	0.571
	(0.763)	(0.469)	(0.564)	(0.742)	(0.842)	(0.531)
the difference	−0.087	−0.219	0.633	−1.265***	0.562	−0.845
	(0.077)	(0.476)	(0.503)	(0.542)	(0.633)	(0.717)
quintile 4	3.002***	1.908***	3.421***	1.454*	3.205***	0.677
	(0.764)	(0.476)	(0.655)	(0.768)	(0.926)	(0.573)
the difference	−0.131*	−0.706	−1.679***	1.992***	−0.604	−0.107
	(0.069)	(0.488)	(0.553)	(0.621)	(0.673)	(0.784)
quintile 5	3.192***	3.889***	6.450***	−0.643	3.241***	−0.601
	(0.762)	(0.630)	(0.879)	(0.916)	(1.077)	(0.604)
the difference	−0.190**	−1.981***	−3.029***	2.097***	−0.035	1.278
	(0.097)	(0.584)	(0.802)	(0.850)	(0.793)	(0.834)

Notes: Quintile 1 denotes the lowest quality and quintile 5 – the highest. The differences between the previous and the current quintile are reported.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.
Standard errors calculated using the delta-method are in parentheses.
There are two sources of errors in the estimates shown in the Table: the error of regression coefficient and the error of the mean values of covariates. The first part of the error does not vary across all result tables, while the second part of the error depends on the group size and is approximately $\sqrt{5}$ times larger than its counterpart in Table 2. However, the errors of regression coefficient are considerably bigger than the errors of mean values of covariates, so the increase in the standard errors in this table relative to the standard error in Table 2 is only minor.

The values of net total effect reveal an increase of quality in the groups of low-quality hospitals, while quality deteriorates at high-quality groups, confirming hypotheses $H3a$ and $H3b$. Negative total effect is observed at fewer groups of high-quality hospitals in later years of the implementation of pay-for-performance. The result confirms hypothesis $H3c$ and can be attributed to the weakening of mean reversion with increase in α.

Finally, we focus on the effect of pay-for-performance for groups of Medicare hospitals according to their ownership, teaching status, urban location and geographic region. The mean effect increases in α for public and private hospitals, for urban and rural hospitals, for teaching and non-teaching hospitals, and for hospitals in each geographic region (Tables 7 and 8).

The effect of pay-for-performance is greater for private hospitals than for public hospitals, which corresponds to findings in Borah et al. (2012) and Werner et al. (2011). The result can be explained by a greater emphasis on financial incentives at these healthcare institutions. These profit constraints, combined with the altruistic character of healthcare services, induce more effective quality competition at non-public hospitals (Brekke et al., 2012).

Table 6: Net total effect by quintiles of TPS_{t-1}: Predicted TPS minus lagged TPS

	2013	2014	2015	2016	2017	2018
quintile 1	2.561***	4.847***	0.983***	7.579***	5.397***	5.077***
	(0.461)	(0.320)	(0.291)	(0.263)	(0.261)	(0.268)
quintile 2	−3.768***	0.753***	−2.267***	4.329***	2.517***	2.062***
	(0.283)	(0.238)	(0.228)	(0.216)	(0.206)	(0.217)
the difference	6.330***	4.093***	3.250***	3.250***	2.880***	3.015***
	(0.362)	(0.273)	(0.241)	(0.205)	(0.198)	(0.209)
quintile 3	−7.132***	−1.802***	−4.969***	2.448***	0.713***	0.343
	(0.271)	(0.228)	(0.229)	(0.212)	(0.212)	(0.219)
the difference	3.363***	2.556***	2.702***	1.882***	1.804***	1.719***
	(0.265)	(0.232)	(0.198)	(0.188)	(0.182)	(0.200)
quintile 4	−10.624***	−4.004***	−6.638***	0.063	−1.055***	−1.511***
	(0.317)	(0.262)	(0.270)	(0.242)	(0.249)	(0.266)
the difference	3.492***	2.202***	1.669***	2.385***	1.768***	1.854***
	(0.275)	(0.226)	(0.229)	(0.215)	(0.212)	(0.253)
quintile 5	−15.298***	−8.033***	−10.348***	−3.256***	−4.127***	−4.934***
	(0.433)	(0.390)	(0.453)	(0.378)	(0.378)	(0.385)
the difference	4.674***	4.028***	3.710***	3.318***	3.072***	3.423***
	(0.337)	(0.307)	(0.377)	(0.312)	(0.312)	(0.310)

Notes: Quintile 1 denotes the lowest quality and quintile 5 – the highest. The differences between the previous and the current quintile are reported.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.
Standard errors calculated using the delta-method are in parentheses.
There are two sources of errors in the estimates shown in the Table: the error of regression coefficient and the error of the mean values of covariates. The first part of the error does not vary across all result tables, while the second part of the error depends on the group size and is approximately $\sqrt{5}$ times larger than its counterpart in Table 2. However, the errors of regression coefficient are considerably bigger than the errors of mean values of covariates, so the increase in the standard errors in this table relative to the standard error in Table 2 is only minor.

As for teaching status, quality improvement owing to the incentive scheme is higher at non-teaching hospitals, which may be because they can devote all of their labor resources to patient treatment, while teaching hospitals lose some efficiency due to their educational activities (Pauly, 1980). Also, teaching hospitals may be treating more difficult cases. This complexity could not be fully captured by the casemix variable in our analysis and may cause a downwards bias of the estimated effect at teaching hospitals.

As regards geographic location, the mean effect of pay-for-performance is greater in New England than in several other geographic regions: South Atlantic, East South Central and Pacific.

The differences in the effect of pay-for-performance across urban and rural hospitals are insignificant for most years.

Table 7: Effect of pay-for-performance as $\mu(\alpha_t) - \mu(0)$ by hospital ownership, teaching status and urban location

	2013	2014	2015	2016	2017	2018
public	2.721***	3.842***	5.505***	7.601***	9.825***	9.678***
	(0.746)	(1.034)	(1.401)	(1.914)	(2.564)	(2.497)
private	2.954***	4.211***	6.258***	8.526***	11.400***	11.436***
	(0.764)	(1.069)	(1.466)	(2.008)	(2.733)	(2.674)
the difference	−0.233*	−0.369**	−0.753***	−0.925***	−1.575***	−1.758***
	(0.124)	(0.182)	(0.286)	(0.378)	(0.540)	(0.544)
urban	3.049***	4.240***	6.063***	7.999***	10.492***	10.396***
	(0.765)	(1.057)	(1.418)	(1.889)	(2.491)	(2.419)
rural	2.737***	4.100***	5.804***	8.891***	12.580***	12.547***
	(0.757)	(1.119)	(1.614)	(2.321)	(3.358)	(3.217)
the difference	0.312	0.140	0.259	−0.892	−2.089*	−2.151**
	(0.235)	(0.350)	(0.543)	(0.707)	(1.116)	(1.004)
teaching	2.943***	4.120***	5.772***	7.807***	10.209***	10.275***
	(0.778)	(1.062)	(1.424)	(1.890)	(2.488)	(2.456)
non-teaching	2.967***	4.257***	6.343***	8.724***	11.865***	11.818***
	(0.755)	(1.073)	(1.485)	(2.069)	(2.883)	(2.801)
the difference	−0.024	−0.136	−0.572	−0.917*	−1.656**	−1.543**
	(0.176)	(0.235)	(0.371)	(0.472)	(0.741)	(0.695)

Notes: Standard errors (calculated using the delta-method for the difference of the reform effects across corresponding two categories of each time-invariant hospital characteristic) are in parentheses.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.

Table 8: Effect of pay-for-performance as $\mu(\alpha_t) - \mu(0)$ for hospitals at different geographic regions

	2013	2014	2015	2016	2017	2018
New England	2.654***	3.207***	5.278***	9.446***	14.445***	14.025***
	(0.831)	(1.320)	(1.888)	(2.399)	(3.832)	(3.643)
Mid-Atlantic	2.575***	3.633***	5.105***	7.109***	9.595***	9.701***
	(0.766)	(1.056)	(1.426)	(1.917)	(2.563)	(2.520)
the difference	0.079	−0.426	0.173	2.337***	4.850***	4.324***
	(0.314)	(0.716)	(1.034)	(0.705)	(1.586)	(1.412)
East North Central	2.823***	4.184***	6.259***	8.681***	11.123***	11.169***
	(0.775)	(1.093)	(1.516)	(2.077)	(2.764)	(2.757)
the difference	−0.169	−0.977	−0.980	0.765	3.321***	2.857***
	(0.299)	(0.706)	(1.001)	(0.539)	(1.340)	(1.138)
West North Central	3.060***	4.392***	6.696***	9.635***	13.783***	13.785***
	(0.762)	(1.103)	(1.544)	(2.279)	(3.422)	(3.360)
the difference	−0.406	−1.185	−1.418	−0.189	0.661	0.240
	(0.316)	(0.721)	(1.047)	(0.587)	(1.072)	(0.990)
South Atlantic	2.871***	4.201***	6.474***	9.040***	12.150***	12.047***
	(0.780)	(1.094)	(1.538)	(2.109)	(2.904)	(2.801)
the difference	−0.217	−0.994	−1.196	0.406	2.295*	1.979*
	(0.310)	(0.712)	(0.994)	(0.513)	(1.207)	(1.100)
East South Central	2.826***	3.958***	6.288***	8.901***	12.585***	11.807***
	(0.795)	(1.113)	(1.606)	(2.290)	(3.335)	(3.046)
the difference	−0.172	−0.752	−1.010	0.545	1.860*	2.218**
	(0.343)	(0.737)	(1.039)	(0.584)	(0.997)	(0.957)
West South Central	2.850***	4.111***	6.566***	8.149***	10.496***	10.940***
	(0.755)	(1.041)	(1.426)	(1.931)	(2.553)	(2.534)
the difference	−0.196	−0.904	−1.288	1.297*	3.949***	3.085**
	(0.335)	(0.739)	(1.074)	(0.705)	(1.543)	(1.355)
Mountain	2.541***	3.848***	5.589***	7.066***	9.293***	9.672***
	(0.678)	(0.919)	(1.204)	(1.540)	(2.010)	(2.020)
the difference	0.113	−0.641	−0.311	2.380**	5.152***	4.353***
	(0.360)	(0.772)	(1.138)	(1.049)	(2.059)	(1.840)
Pacific	2.992***	3.941***	5.017***	6.932***	9.288***	9.251***
	(0.739)	(1.001)	(1.311)	(1.735)	(2.312)	(2.274)
the difference	−0.338	−0.735	0.261	2.515***	5.157***	4.774***
	(0.382)	(0.787)	(1.142)	(1.001)	(1.892)	(1.725)

Notes: Standard errors (calculated using the delta-method for the difference of the reform effects across New England hospitals and hospitals in each corresponding geographic region) are in parentheses.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.

6 Discussion

Our theoretical and empirical analysis shows that an incentive contract with linear increase of the reward function leads to an increase in the mean level of the quality measure. Arguably, this reflects the fact that the Medicare quality incentive scheme induces effective quality improvement activities by hospital management,[20] physicians and collaborative groups. Indeed, the results of qualitative surveys similarly reveal that hospital leadership responds to Medicare's pay-for-performance incentive scheme by investment in quality improvements (Smith, 2017, p. 145). Specifically, administrators and medical directors strive to understand "what actions might improve their low scores" (Conrad et al., 2006, p. 447) and admit that without changing the process of care "we would continue to get the same results that we always have" (Jones, 2014, p. 120).

The quality-enhancing efforts at the high-quality US hospitals under Medicare's pilot program and under Medicare's value-based purchasing ensured early diagnosis and timely care, helped to maintain accurate patient records and encouraged frequent analysis of data in order to assess performance relative to other hospitals (Smith, 2017; Jones, 2014; Grossbart, 2006). Patient satisfaction and clinical outcomes were targeted through establishing a special support team to address pain control and expedite the response to nurse bells, promoting quietness at hospitals by lowering the amount of noise from telephones and stopping the use of pagers at night, educating nurses to use opening and closing phrases to reduce patient anxiety and inform the patient when nurse will be back (Smith, 2017). Other quality improvement activities have been focused on perfecting the quality management system at hospitals by allocating more funds to data coding and information technology (Smith, 2017; Damberg et al., 2009; Wagner et al., 2006; Bentley and Nash, 1998).

Our results reveal that increase in the quality measure owing to pay-for-performance is greater at hospitals with higher levels of quality. The finding suggests stronger emphasis on quality activities at high-quality hospitals, and this is indeed discovered in a number of works. For instance, top-performing hospitals in the US pilot program paid more attention to quality enhancement than bottom-performing hospitals (Vina et al., 2009). Under the proportional pay-for-performance mechanism in California, high-quality physicians similarly placed more emphasis on an organizational culture of quality and demonstrate stronger dedication to addressing quality issues than low-quality physicians (Damberg et al., 2009). The desire of high-quality hospitals, which have reached top deciles of hospital performance, to pursue quality improvement by means additional to those proposed by the policy regulator is further evidence in support of our research (Grossbart, 2006).

As well as concentrating on the effect of a pay-for-performance mechanism and its heterogeneity across groups of hospitals of different quality, our theoretical model and empirical analysis focused on the power of the incentive scheme measured as the share of hospital revenue. We discover that higher values of this share (in terms of hospital funds at risk in a

[20] E.g. care plan management, complaints registration, incident and infection committees (Wagner et al., 2006).

budget-neutral scheme) intensify the quality improvement. The finding corresponds to greater effectiveness of larger incentives in comparison with smaller ones, which is found in real-world applications of pay-for-performance (Ogundeji et al., 2016; de Brantes and d'Andrea, 2009; Beaulieu and Horrigan, 2005). Also, if pay-for-performance schemes are voluntary, greater potential rewards encourage participation (de Brantes and d'Andrea, 2009).

There is no general agreement about the optimal size of the incentive, nor is there a clear empirical pattern of the "dose-response relationship", linking financial incentive and quality improvement. The actual share of affected revenue in pay-for-performance schemes varies from 2 to 20% of physician income (Cashin, 2014b; de Brantes and d'Andrea, 2009; Scott, 2007) and from 1 to 9% of hospital income (Bisiaux and Chi, 2014; Sutton et al., 2012; Conrad and Perry, 2009; Rosenthal et al., 2007; Scott, 2007). As regards desirable size of the incentives that would influence behavior of physicians, a survey of HMO managers suggests that the optimal share is in the interval 5–15% of a physician's income (Hillman et al., 1991).

Small incentives may fail to have impact on quality (Ogundeji et al., 2016; Glasziou et al., 2012; Conrad and Perry, 2009; Petersen et al., 2006; Beaulieu and Horrigan, 2005). On the other hand, the power of the incentive scheme must not be excessive. Redistributive programs, which are budget-neutral for the regulator, put a large share of hospital budgets at risk. This brings a danger of serious financial loss and potential damage for low-performing hospitals (Damberg et al., 2014). But when the regulator raises external funds to finance pay-for-performance mechanisms, high power of the incentives scheme may cause other methods of quality improvement to be overlooked (Glasziou et al., 2012). So it is important to evaluate opportunity costs of pay-for-performance through a comparative assessment of alternative designs of incentive stimuli (Meacock et al., 2014; Nahra et al., 2006; Kahn et al., 2006). Such alternatives include various regulatory and managerial initiatives, such as audit, reminders, collaboration and feedback through opinion leaders (Glasziou et al., 2012).

The search for the optimal price for quality of healthcare in terms of parameter κ in the reward function for quality is largely equivalent to finding an optimal size of incentives α. In fact, as we mention in the extensions to our model, the concentration of policy literature on the values of α stems largely from the need of the social planner to keep the incentive scheme budget-neutral. So a more general model of a non-budget-neutral pay-for-performance scheme would regard κ as a policy parameter. The approach is implemented in the literature on pay-for-performance in the UK (Kristensen et al., 2016; Sutton et al., 2012).

7 Conclusion

Studies of incentive contracts for healthcare quality usually focus on the mean tendency and give scant attention to potentially heterogeneous response to pay-for-performance by hospitals or physicians at different percentiles of quality distribution. But insufficient theoretical and empirical analysis of such heterogeneity may lead to speculation on the ceiling effects and belief that there are no ways of further improving performance by healthcare providers with better quality.

This paper considered an incentive mechanism with a linear increase of the quality-reward function and provided a theoretical model of dynamic hospital behavior under such remuneration. The predictions of the model show that the incentive mechanism stimulates all groups of hospitals and that there is a direct association between observed quality and its increase in the next period. Larger quality incentives in terms of hospital revenue at risk cause greater increase of observed quality.

The empirical part of the paper used the longitudinal data for Medicare's acute-care hospitals taking part in the nationwide quality incentive mechanism ("value-based purchasing") with a linear rule applied to the composite quality measure. We found that the model's predictions are supported. The higher the quintile of quality in the prior period, the larger the increase in quality owing to the introduction of pay-for-performance with the linear rule. Improvement of the composite quality measure in each quintile increases with the increase in size of the quality incentive.

The empirical part of the paper highlights the importance of disentangling the effect of the regression towards the mean from the policy effect when there is imprecision in verifying true quality through its measurable proxy. The obvious way to reduce the impact of such imprecision is to design better proxies for healthcare quality. The approach was implemented in Medicare's revision of the dimensions of the composite quality measure: for instance, stronger emphasis was placed on indicators of in-hospital mortality and safety, while indicators linked to the observance of clinical guidelines were gradually abolished.

Our theoretical model suggests another way of tackling the problem: the effect of mean reversion can be diminished through raising the power of the incentive scheme. However, theoretical considerations linked to stability conditions in the intertemporal hospital problem as well as lessons learnt from policy implementation in various countries caution against excessive increase of the share of hospital funds at risk.

Directions for future work may include analysis of heterogeneous hospital response to quality incentives by considering different dimensions of the composite quality measure. The empirical literature shows evidence for "trade-offs" between the amount of effort expended to raise the quality of particular dimensions: on average, Medicare hospitals are likely to focus on improving those dimensions where quality enhancement gives the largest marginal payoff (Norton et al., 2018). A related field of research is the study of potential sacrifice of quality of non-incentivized measures in favor of measures incentivized by the linear rule. This has been analyzed at the mean level (Kaarbøe and Siciliani, 2011; Eggleston, 2005) and may be expanded to account for different behavior by high-quality and low-quality hospitals.

References

Acemoglu, D. (2009). *Introduction to modern economic growth.* Princeton University Press, Princeton and Oxford.

Arellano, M. and O. Bover (1995). "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68(1), 29–51.

Armstrong, C. S., J. L. Blouin, A. D. Jagolinzer, and D. F. Larcker (2015). "Corporate governance, incentives, and tax avoidance." *Journal of Accounting and Economics*, 60(1), 1–17.

Barnett, A. G., J. C. van der Pols, and A. J. Dobson (2004). "Regression to the mean: What it is and how to deal with it." *International Journal of Epidemiology*, 34(1), 215–220.

Baron, D. P. and R. B. Myerson (1982). "Regulating a monopolist with unknown costs." *Econometrica: Journal of the Econometric Society*, 50(4), 911–930.

Beaulieu, N. D. and D. R. Horrigan (2005). "Putting smart money to work for quality improvement." *Health Services Research*, 40(5p1), 1318–1334.

Bénabou, R. and J. Tirole (2003). "Intrinsic and extrinsic motivation." *The Review of Economic Studies*, 70(3), 489–520.

Bénabou, R. and J. Tirole (2006). "Incentives and prosocial behavior." *American Economic Review*, 96(5), 1652–1678.

Bentley, J. M. and D. B. Nash (1998). "How Pennsylvania hospitals have responded to publicly released reports on coronary artery bypass graft surgery." *The Joint Commission Journal on Quality Improvement*, 24(1), 40–49.

Bisiaux, R. and Y.-L. Chi (2014). "Republic of Korea: value incentive programme." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 251–264. World Health Organization.

Blomqvist, Å. (1997). "Optimal non-linear health insurance." *Journal of Health Economics*, 16(3), 303–321.

Blundell, R. and S. Bond (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1), 115–143.

Bokhour, B. G., J. F. Burgess, J. M. Hook, B. White, D. Berlowitz, M. R. Guldin, M. Meterko, and G. J. Young (2006). "Incentive implementation in physician practices: A qualitative study of practice executive perspectives on pay for performance." *Medical Care Research and Review*, 63(1 Suppl), 73S–95S.

Borah, B. J., M. G. Rock, D. L. Wood, D. L. Roellinger, M. G. Johnson, and J. M. Naessens (2012). "Association between value-based purchasing score and hospital characteristics." *BMC health services research*, 12(1), 464.

Bousquet, F., R. Bisiaux, and Y.-L. Chi (2014). "France: Payment for public health services." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 141–156. World Health Organization.

Brekke, K. R., L. Siciliani, and O. R. Straume (2012). "Quality competition with profit constraints." *Journal of Economic Behavior & Organization*, 84(2), 642–659.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (2016). "Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision." *Journal of Economic Behavior & Organization*, 131, 17–23.

Buetow, S. (2008). "Pay-for-performance in New Zealand primary health care." *Journal of Health Organization and Management*, 22(1), 36–47.

Busse, R. (2004). "Disease management programs in Germany's statutory health insurance system." *Health Affairs*, 23(3), 56–67.

Campbell, S. M., D. Reeves, E. Kontopantelis, B. Sibbald, and M. Roland (2009). "Effects of pay for performance on the quality of primary care in England." *New England Journal of Medicine*, 361(4), 368–378.

Casas-Arce, P. and F. A. Martínez-Jerez (2009). "Relative performance compensation, contests, and dynamic incentives." *Management Science*, 55(8), 1306–1320.

Cashin, C. (2014a). "New Zealand: primary health organization performance programme." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 265–288. World Health Organization.

Cashin, C. (2014b). "P4P programme design." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 23–39. World Health Organization.

Cashin, C. (2014c). "United States: hospital quality incentive demonstration." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 287–300. World Health Organization.

Cashin, C. and Y.-L. Chi (2014). "Australia: practice incentives programme." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 109–126. World Health Organization.

Cashin, C., Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson (eds.) (2014). *Paying for performance in health care*. World Health Organization.

Centers for Medicare and Medicaid Services (2007a). "Medicare hospital value-based purchasing plan development. 1st public listening session, January 17, 2007." Issues paper, U.S. Department of Health and Human Services, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/Hospital_VBP_Plan_Issues_Paper.pdf.

Centers for Medicare and Medicaid Services (2007b). "Report to Congress: plan to implement a Medicare hospital value-based purchasing program." Tech. rep., U.S. Department of Health and Human Services, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/downloads/hospitalVBPplanRTCFinalsubmitted2007.pdf.

Chalkley, M. and J. M. Malcomson (1998a). "Contracting for health services when patient demand does not reflect quality." *Journal of Health Economics*, 17(1), 1–19.

Chalkley, M. and J. M. Malcomson (1998b). "Contracting for health services with unmonitored quality." *Economic Journal*, 108(449), 1093–1110.

Christianson, J. B., S. Leatherman, and K. Sutherland (2008). "Lessons from evaluations of purchaser pay-for-performance programs." *Medical Care Research and Review*, 65(6 Suppl), 5S–35S.

Coleman, K., K. L. Reiter, and D. Fulwiler (2007). "The impact of pay-for-performance on diabetes care in a large network of community health centers." *Journal of Health Care for the Poor and Underserved*, 18(5), 966–983.

Conrad, D. A. and L. Perry (2009). "Quality-based financial incentives in health care: Can we improve quality by paying for it?" *Annual Review of Public Health*, 30, 357–371.

Conrad, D. A., B. G. Saver, B. Court, and S. Heath (2006). "Paying physicians for quality: Evidence and themes from the field." *The Joint Commission Journal on Quality and Patient Safety*, 32(8), 443–451.

Damberg, C. L., K. Raube, S. S. Teleki, and E. dela Cruz (2009). "Taking stock of pay-for-performance: A candid assessment from the front lines." *Health Affairs*, 28(2), 517–525.

Damberg, C. L., S. M. Shortell, K. Raube, R. R. Gillies, D. Rittenhouse, R. K. McCurdy, L. P. Casalino, and J. Adams (2010). "Relationship between quality improvement processes and clinical performance." *The American Journal of Managed Care*, 16(8), 601–606.

Damberg, C. L., M. E. Sorbero, S. L. Lovejoy, G. R. Martsolf, L. Raaen, and D. Mandel (2014). "Measuring success in health care value-based purchasing programs: Findings from an environmental scan, literature review, and expert panel discussions." Research report, RAND Corporation, Santa Monica, CA, http://www.rand.org/pubs/research_reports/RR306.html.

Davis, C. E. (1976). “The effect of regression to the mean in epidemiologic and clinical studies.” *American Journal of Epidemiology*, 104(5), 493–498.

de Brantes, F. S. and B. G. d’Andrea (2009). “Physicians respond to pay-for-performance incentives: Larger incentives yield greater participation.” *American Journal of Managed Care*, 15(5), 305–10.

de Bruin, S. R., C. A. Baan, and J. N. Struijs (2011). “Pay-for-performance in disease management: A systematic review of the literature.” *BMC Health Services Research*, 11(1), 272.

Doran, T., C. Fullwood, E. Kontopantelis, and D. Reeves (2008). “Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework.” *The Lancet*, 372(9640), 728–736.

Dückers, M., P. Makai, L. Vos, P. Groenewegen, and C. Wagner (2009). “Longitudinal analysis on the development of hospital quality management systems in the Netherlands.” *International Journal for Quality in Health Care*, 21(5), 330–340.

Eggleston, K. (2005). “Multitasking and mixed systems for provider payment.” *Journal of Health Economics*, 24(1), 211–223.

Eggleston, K. and C.-R. Hsieh (2004). “Healthcare payment incentives.” *Applied Health Economics and Health Policy*, 3(1), 47–56.

Eijkenaar, F., M. Emmert, M. Scheppach, and O. Schöffski (2013). “Effects of pay for performance in health care: A systematic review of systematic reviews.” *Health Policy*, 110(2), 115–130.

Ellis, R. P. and T. G. McGuire (1996). “Hospital response to prospective payment: Moral hazard, selection, and practice-style effects.” *Journal of Health Economics*, 15(3), 257–277.

Friedman, M. (1992). “Do old fallacies ever die?” *Journal of Economic Literature*, 30(4), 2129–2132.

Galton, F. and H. Dickson (1886). “Family likeness in stature.” *Proceedings of the Royal Society of London*, 40, 42–73.

Gené-Badia, J., G. Escaramis-Babiano, M. Sans-Corrales, L. Sampietro-Colom, F. Aguado-Menguy, C. Cabezas-Peña, and P. G. de Puelles (2007). “Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care.” *Health Policy*, 80(1), 2–10.

Glasziou, P. P., H. Buchan, C. Del Mar, J. Doust, M. Harris, R. Knight, A. Scott, I. A. Scott, and A. Stockwell (2012). “When financial incentives do more good than harm: A checklist.” *BMJ*, 345, e5047.

Glickman, S. W., F.-S. Ou, E. R. DeLong, M. T. Roe, B. L. Lytle, J. Mulgund, J. S. Rumsfeld, W. B. Gibler, E. M. Ohman, K. A. Schulman, and E. D. Peterson (2007). “Pay for performance, quality of care, and outcomes in acute myocardial infarction.” *JAMA*, 297(21), 2373–2380.

Gravelle, H., M. Sutton, and A. Ma (2010). “Doctor behaviour under a pay for performance contract: Treating, cheating and case finding?” *The Economic Journal*, 120(542), F129–F156.

Greene, W. H. (2018). *Econometric analysis.* 8 ed., Pearson, New York.

Grossbart, S. R. (2006). “What’s the return? Assessing the effect of ‘pay-for-performance’ initiatives on the quality of care delivery.” *Medical Care Research and Review*, 63(1 Suppl), 29S–48S.

Gruber, J. and M. Owings (1996). “Physician financial incentives and cesarean section delivery.” *RAND Journal of Economics*, 27(1), 99–123.

Gupta, A. (2017). “Impacts of performance pay for hospitals: The readmissions reduction program.” Working Paper 2017-07, Becker Friedman Institute for Research in Economics, The University of Chicago, https://bfi.uchicago.edu/wp-content/uploads/HES_No.2017-07.pdf.

Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press, Princeton.

Hannan, E. L., D. Kumar, M. Racz, A. L. Siu, and M. R. Chassin (1994). “New York State’s cardiac surgery reporting system: Four years later.” *The Annals of Thoracic Surgery*, 58(6), 1852–1857.

Hibbard, J. H., J. Stockard, and M. Tusler (2005). “Hospital performance reports: Impact on quality, market share, and reputation.” *Health Affairs*, 24(4), 1150–1160.

Hillman, A. L., M. V. Pauly, K. Kerman, and C. R. Martinek (1991). “HMO managers’ views on financial incentives and quality.” *Health Affairs*, 10(4), 207–219.

Hölmstrom, B. and P. Milgrom (1991). "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, & Organization*, 7, 24–52.

Jones, N. B. (2014). *Health care executives participating in value-based purchasing: A qualitative phenomenological study*. Ph.D. thesis, University of Phoenix.

Joskow, P. L. and R. Schmalensee (1986). "Incentive regulation for electric utilities." *Yale Journal on Regulation*, 4, 1–49.

Kaarbøe, O. M. and L. Siciliani (2011). "Multi-tasking, quality and pay for performance." *Health Economics*, 20(2), 225–238.

Kahn, C. N., T. Ault, H. Isenstein, L. Potetz, and S. van Gelder (2006). "Snapshot of hospital quality reporting and pay-for-performance under medicare." *Health Affairs*, 25(1), 148–162.

Kane, T. J. and D. O. Staiger (2002). "The promise and pitfalls of using imprecise school accountability measures." *Journal of Economic Perspectives*, 16(4), 91–114.

Kreps, D. M. (1997). "Intrinsic motivation and extrinsic incentives." *American Economic Review*, 87(2), 359–364.

Kridel, D. J., D. E. M. Sappington, and D. L. Weisman (1996). "The effects of incentive regulation in the telecommunications industry: A survey." *Journal of Regulatory Economics*, 9(3), 269–306.

Kristensen, S. R., L. Siciliani, and M. Sutton (2016). "Optimal price-setting in pay for performance schemes in health care." *Journal of Economic Behavior & Organization*, 123, 57–77.

Laffont, J.-J. and J. Tirole (1993). *A theory of incentives in procurement and regulation*. MIT Press.

Li, J., J. Hurley, P. DeCicca, and G. Buckley (2014). "Physician response to pay-for-performance: Evidence from a natural experiment." *Health Economics*, 23(8), 962–978.

Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler (2007). "Public reporting and pay for performance in hospital quality improvement." *New England Journal of Medicine*, 356(5), 486–496.

Ma, C. A. (1994). "Health care payment systems: Cost and quality incentives." *Journal of Economics & Management Strategy*, 3(1), 93–112.

Ma, C.-t. A. and H. Y. Mak (2015). "Information disclosure and the equivalence of prospective payment and cost reimbursement." *Journal of Economic Behavior & Organization*, 117, 439–452.

Meacock, R., S. R. Kristensen, and M. Sutton (2014). "The cost-effectiveness of using financial incentives to improve provider quality: A framework and application." *Health Economics*, 23(1), 1–13.

Morton, V. and D. J. Torgerson (2003). "Effect of regression to the mean on decision making in health care." *British Medical Journal*, 326(7398), 1083–1084.

Mullen, K. J., R. G. Frank, and M. B. Rosenthal (2010). "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers." *RAND Journal of Economics*, 41(1), 64–91.

Murdock, K. (2002). "Intrinsic motivation and optimal incentive contracts." *RAND Journal of Economics*, 33(4), 650–671.

Murray, R. (2014). "United States: maryland hospital acquired conditions programme." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 265–288. World Health Organization.

Nahra, T. A., K. L. Reiter, R. A. Hirth, J. E. Shermer, and J. R. Wheeler (2006). "Cost-effectiveness of hospital pay-for-performance incentives." *Medical Care Research and Review*, 63(1 Suppl), 49S–72S.

Norton, E. C., J. Li, A. Das, and L. M. Chen (2018). "Moneyball in Medicare." *Journal of Health Economics*, 61, 259–273.

Ödesjö, H., A. Anell, S. Gudbjörnsdottir, J. Thorn, and S. Björck (2015). "Short-term effects of a pay-for-performance programme for diabetes in a primary care setting: An observational study." *Scandinavian Journal of Primary Health Care*, 33(4), 291–297.

Ogundeji, Y. K., J. M. Bland, and T. A. Sheldon (2016). "The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes." *Health Policy*, 120(10), 1141–1150.

Oxholm, A. S., S. R. Kristensen, and M. Sutton (2018). "Uncertainty about the effort–performance relationship in threshold-based payment schemes." *Journal of Health Economics*, 62, 69–83.

Pai, C.-W., G. K. Finnegan, and M. J. Satwicz (2002). "The combined effect of public profiling and quality improvement efforts on heart failure management." *The Joint Commission Journal on Quality Improvement*, 28(11), 614–624.

Pauly, M. (1980). *Doctors and their workshops: Economic models of physician behavior*. A National Bureau of Economic Research Monograph, The University of Chicago Press, Chicago and London.

Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan (2006). "Does pay-for-performance improve the quality of health care?" *Annals of Internal Medicine*, 145(4), 265–272.

Practice Assist (2017). *Practice Incentives Program. Overview.* http://www.practiceassist.com.au/PracticeAssist/media/ResourceLibrary/Practice%20Incentives%20Program/PIP_Overview_User-Guide_V2-1.pdf.

Prendergast, C. (1999). "The provision of incentives in firms." *Journal of economic literature*, 37, 7–63.

Radner, R. (1985). "Repeated principal-agent games with discounting." *Econometrica*, 53(5), 1173–1198.

Rosenthal, M. B. (2014). "United States: california integrated healthcare association physician incentive programme." In *Paying for performance in health care*, edited by C. Cashin, Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson, pp. 224–250. World Health Organization.

Rosenthal, M. B., R. Fernandopulle, H. R. Song, and B. Landon (2004). "Paying for quality: Providers' incentives for quality improvement." *Health Affairs*, 23(2), 127–141.

Rosenthal, M. B., B. E. Landon, K. Howitt, H. R. Song, and A. M. Epstein (2007). "Climbing up the pay-for-performance learning curve: Where are the early adopters now?" *Health Affairs*, 26(6), 1674–1682.

Ryan, A. M. (2009). "Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost." *Health Services Research*, 44(3), 821–842.

Ryan, A. M., J. Blustein, and L. P. Casalino (2012a). "Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals." *Health Affairs*, 31(4), 797–805.

Ryan, A. M., J. Blustein, M. D. Michelow, and L. P. Casalino (2012b). "The effect of phase 2 of the Premier Hospital Quality Incentive Demonstration on incentive payments to hospitals caring for disadvantaged patients." *Health Services Research*, 47(4), 1418–1436.

Ryan, A. M., S. Krinsky, K. A. Maurer, and J. B. Dimick (2017). "Changes in hospital quality associated with hospital value-based purchasing." *New England Journal of Medicine*, 376(24), 2358–2366.

Scott, I. A. (2007). "Pay for performance in health care: Strategic issues for Australian experiments." *Medical Journal of Australia*, 187(1), 31–35.

Shleifer, A. (1985). "A theory of yardstick competition." *RAND Journal of Economics*, 16, 319–327.

Siciliani, L. (2009). "Paying for performance and motivation crowding out." *Economics letters*, 103(2), 68–71.

Siciliani, L., O. R. Straume, and R. Cellini (2013). "Quality competition with motivated providers and sluggish demand." *Journal of Economic Dynamics and Control*, 37(10), 2041–2061.

Smith, L. R. (2017). *A case study: The executive leadership response at a community hospital to the value-based purchasing requirements on the Patient Protection and Affordable Care Act*. Ph.D. thesis, University of North Florida, College of Education and Human Services, https://digitalcommons.unf.edu/cgi/viewcontent.cgi?article=1791&context=etd.

Sutton, M., S. Nikolova, R. Boaden, H. Lester, R. McDonald, and M. Roland (2012). "Reduced mortality with hospital pay for performance in England." *New England Journal of Medicine*, 367(19), 1821–1828.

Thompson, J. D., R. F. Averill, and R. B. Fetter (1979). "Planning, budgeting, and controlling–one look at the future: Case-mix cost accounting." *Health Services Research*, 14(2), 111–125.

Vaghela, P., M. Ashworth, P. Schofield, and M. C. Gulliford (2009). "Population intermediate outcomes of diabetes under pay-for-performance incentives in England from 2004 to 2008." *Diabetes Care*, 32(3), 427–429.

Vina, E. R., D. C. Rhew, S. R. Weingarten, J. B. Weingarten, and J. T. Chang (2009). "Relationship between organizational factors and performance among pay-for-performance hospitals." *Journal of General Internal Medicine*, 24(7), 833.

Wagner, C., L. Gulácsi, E. Takács, and M. Outinen (2006). "The implementation of quality management systems in hospitals: A comparison between three countries." *BMC Health Services Research*, 6(1), 50.

Werner, R. M., J. T. Kolstad, E. A. Stuart, and D. Polsky (2011). "The effect of pay-for-performance in hospitals: Lessons for quality improvement." *Health Affairs*, 30(4), 690–698.

Wilcox, M. A., A. M. Chang, and I. R. Johnson (1996). "The effects of parity on birthweight using successive pregnancies." *Acta Obstetricia et Gynecologica Scandinavica*, 75(5), 459–463.

Windmeijer, F. (2005). "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126, 25–51.

Appendix A Price-setting in US Medicare's value-based purchasing

The aggregation of scores within domains is conducted as follows. For each hospital i and each measure m in the clinical care, patient experience of care, safety and efficiency domains *achievement points* a_i^m ($0 \leq a_i^m \leq 10$) are calculated as:

$$a_i^m = \begin{cases} 10, & \text{if } y_i^m \geq m_b, \\ \text{Round}\left[\dfrac{9(y_i^m - m_a)}{m_b - m_a} + 0.5\right], & \text{if } m_a \leq y_i^m < m_b, \\ 0, & \text{if } y_i^m < m_a, \end{cases}$$

where y_i^m is the value of measure m for hospital i in the current period, m_b is the benchmark and m_a is the achievement threshold for measure m. The benchmark and achievement threshold are respectively set as the mean of the decile at the best-performing hospital and the median in the empirical distribution of y^m, according to the survey in the baseline period. (The means of the *top* deciles are used as benchmarks for measures of patient experience of care along with survival rate measures of clinical care. The means of the *bottom* deciles are employed for complication/infection measures of safety and spending per beneficiary).

Improvement points p_i^m ($0 \leq p_i^m \leq 9$) for all measures are computed as the difference between the value of the measure in the current period and the baseline period, normalized by the hospital's distance from the benchmark in the baseline period:

$$p_i^m = \begin{cases} 9, & \text{if } y_i^m > m_b, \\ \text{Round}\left[\dfrac{10(y_i^m - y_{i0}^m)}{m_b - y_{i0}^m} - 0.5\right], & \text{if } y_{i0}^m < y_i^m \leq m_b, \\ 0, & \text{if } y_i^m \leq y_{i0}^m, \end{cases}$$

where y_{i0}^m is the score for measure m for hospital i in the baseline period. Note that incentives for improvement apply only to hospitals below the benchmark.

The score for each measure is the maximum of improvement and achievement points: $\max\{a_i^m, p_i^m\}$.

The use of the round function is explained by the desire of Centers for Medicare and Medicaid Services to robustly estimate the point score for each measure and to compare the point scores across the measures with different ranges of their original continuous values.[21]

Additionally, *consistency points* c_i for the patient-experience-of-care domain are calculated as the lowest of the M_P dimension scores d_i^m:

$$c_i = \text{Round}\left[20 \min_m \{d_i^m\} - 0.5\right],$$

[21] *Federal Register*, Vol.76, No.88. Friday, May 6, 2011. *Rules and Regulations*, p.26518.

where $d_i^m = \dfrac{y_i^m - m_f}{m_a - m_f}$, m_f is the floor for measure (the minimal value across all hospitals) and $m = 1, \ldots, M_P$.

The scores for the clinical care and safety domains are the sum of the values for all quality measures within the domain, divided by the total potential score and translated into percentage points: $d_i^C = \dfrac{\sum_{m=1}^{M_C} \max\{a_i^m, p_i^m\}}{10 M_C} \cdot 100$ for clinical care and $d_i^S = \dfrac{\sum_{m=1}^{M_S} \max\{a_i^m, p_i^m\}}{10 M_S} \cdot 100$ for safety. The score for the efficiency domain is $d_i^E = \max\{a_i^1, p_i^1\} \cdot 100$, where its only measure (spending per beneficiary) is used.

In case of patient experience of care, the domain score is the sum of the values for each measure plus consistency points, divided by the total potential score for quality measures plus the maximum value of consistency points (percentage points): $d_i^P = c_i + \dfrac{\sum_{m=1}^{M_P} \max\{a_i^m, p_i^m\}}{10 M_P} \cdot 80$.

The values of the threshold, floor and benchmark are re-estimated annually, based on the empirical distribution of hospital-level quality measures.

The total performance score of each hospital is a weighted sum of its domain scores: $TPS_i = \sum_{k=1}^{K} w_k d_{ik}$, where K is the number of domains in a given year and weights w_k are established uniform across hospitals (Table 9).

Table 9: Domain weights

Domain	2011	2012	2013	2014	2015	2016	2017	2018
Clinical process of care	0.70	0.45	0.20	0.10	0.05	–	–	–
Patient experience of care	0.30	0.30	0.30	0.25	0.25	0.25	0.25	0.25
Outcome of care (Clinical care from 2016)	–	0.25	0.30	0.40	0.25	0.25	0.25	0.25
Safety	–	–	–	–	0.20	0.25	0.25	0.25
Efficiency	–	–	0.20	0.25	0.25	0.25	0.25	0.25

Note: "–" indicates that a domain is not used in calculation of the total performance score (*TPS*).
Source: https://www.qualitynet.org/inpatient/hvbp/participation#tab4.

It should be noted that domain weights serve as a tool for placing emphasis on particular groups of measures: greater weight given to a domain implies that the policy-maker is attempting to foster quality increase of measures within this domain, see (Centers for Medicare and Medicaid Services, 2007b). For instance, lower weight for the patient-experience-of-care domain is explained by the subjective character of measures in this domain. The regulator explains reduction of the weight of the clinical-process-of-care domain by the fact that most measures in this domain are already "topped-up", i.e. have reached high threshold and benchmark values (no statistical difference between the 75th and 90th percentiles). Moreover, medical practitioners believe that some clinical-process-of-care measures are not strongly correlated with adverse outcomes for patients. Accordingly, giving more weight to the outcome-of-care domain (with survival rates and complication/infection rates) becomes an attempt at more reasonable approximation of medical quality.

Appendix B Proofs of the model propositions

Proof of Proposition 1. Substituting $e_t + \varepsilon_t$ for q_t and differentiating with respect to e_t gives the first-order condition:

$$E_t[a(R_t - d_t + p_t(1 + (\kappa m_t - 1)\alpha)) + \theta - ce_t + \beta a p_{t+1}\kappa A\alpha(e_{t+1} + \varepsilon_{t+1})] = 0. \tag{20}$$

Under fixed prices and unit costs (2), and given that $E_t[\varepsilon_{t+1}] = 0$, we can simplify the first-order condition to

$$a(R - d) + ap(1 + (\kappa m_t - 1)\alpha) + \theta - ce_t + \beta a p\kappa A\alpha E_t e_{t+1} = 0. \tag{21}$$

Or, after substituting $m_t = Aq_{t-1} = A(e_{t-1} + \varepsilon_{t-1})$, to

$$ce_t - ap\kappa A\alpha e_{t-1} - \beta ap\kappa A\alpha E_t e_{t+1} = a(R - d) + \theta + ap(1 + (\kappa A\varepsilon_{t-1} - 1)\alpha). \tag{22}$$

The characteristic equation of the difference equation (22) is

$$\beta ap\kappa A\alpha\lambda^2 - c\lambda + ap\kappa A\alpha = 0. \tag{23}$$

Under Assumption 3 it has two roots

$$\lambda = \frac{c \pm \sqrt{c^2 - 4\beta(ap\kappa A\alpha)^2}}{2\beta ap\kappa A\alpha},$$

one of which is larger than $1/\beta$ and therefore should be discarded because of the transversality condition

$$\lim_{t\to\infty} \beta^t E\left[\frac{\partial U_t}{\partial m_t} m_t\right] = \lim_{t\to\infty} \beta^t E[ae_t p\kappa A\alpha m_t] = 0,$$

see Acemoglu (2009), Theorem 16.8. Another root has absolute value less than one, so the solution is stable.

Now, factorize the left-hand side of equation (22):

$$\frac{ap\kappa A\alpha}{\lambda}(1 - \beta\lambda L^{-1})(1 - \lambda L)e_t = a(R - d) + \theta + ap(1 + (\kappa A\varepsilon_{t-1} - 1)\alpha). \tag{24}$$

Here $Lx_t = x_{t-1}$ is the lag operator and $L^{-1}x_t = E_t x_{t+1}$ is its left inverse. After factorization, the smaller root of the characteristic equation equation (23) is

$$\lambda = \frac{c - \sqrt{c^2 - 4\beta(ap\kappa A\alpha)^2}}{2\beta ap\kappa A\alpha}. \tag{25}$$

Applying the operator $(1 - \beta\lambda L^{-1})^{-1} = 1 + \beta\lambda L^{-1} + (\beta\lambda)^2 L^{-2} + \ldots$ to (24) leads to the equation

$$\frac{ap\kappa A\alpha}{\lambda}(1 - \lambda L)e_t = \frac{a(R - d) + \theta + ap(1 - \alpha)}{1 - \beta\lambda} + ap\kappa A\alpha\varepsilon_{t-1}. \tag{26}$$

After multiplying both sides of (26) by $\dfrac{\lambda}{ap\kappa A\alpha}$ and noticing that

$$\frac{\lambda}{ap\kappa A\alpha(1-\beta\lambda)} = \frac{1-\lambda}{c - ap\kappa A\alpha(1+\beta)}$$

we get the statement of Proposition 1. □

Proof of Corollary 2. By substituting $e_t = q_t - \varepsilon_t$ into (2), we get

$$q_t = \mu(\theta) + \lambda(q_{t-1} - \mu(\theta)) + \varepsilon_t,$$

which is the equation (6). □

Proof of Corollary 3. Since the autoregressive process in (6) is stable, it converges to a stationary autoregressive process defined by the same equation.

Given that $E(\varepsilon_t \mid q_{t-1}, \theta) = 0$, we immediately get (7) for the expectation conditional on θ. Equation (8) immediately follows from the law of iterated expectation $E(q_t \mid \theta) = E(E(q_t \mid q_{t-1}, \theta) \mid \theta)$, the fact that $E(q_{t-1} \mid \theta) = E(q_t \mid \theta)$ for a stationary process, and equation (7). □

Proof of Corollary 4. The statement immediately follows from the law of iterated expectation $E(q_t) = E(E(q_t \mid \theta))$, and equation (8). □

Proof of Corollary 5. The equation (10) follows directly from assumptions 3, 1, and Corollary 4. □

Proof of Proposition 6. The condition $E((\kappa m_t - 1)\alpha) = 0$ from Assumption 1 implies that $\kappa = 1/A\mu$. Substituting $\kappa = 1/A\mu$ into equation (9), we obtain (11), into equation (4), we obtain (12) and (13), into equation (25), we obtain (14). The inequalities $0 < \lambda < 1$ directly follow from (25) and Assumption 3. Straightforward differentiation with respect to α shows that μ, η_1 and λ are increasing in α. As for η_1, it decreases in α for sufficiently large values of $\bar{\theta}$. □

Proof of Corollary 7. The equation (15) directly follows from (3) in Proposition 1 after substituting (12) and (13) there. The statement about derivatives is the direct implication of the fact that the derivative in θ equals η_1, and the latter increases in α due to Proposition 6. □

Proof of Proposition 8. It follows from (6) and (15) that the altruism level θ and the quality q_{t-1} form the bivariate normal distribution with covariance matrix

$$\begin{pmatrix} \sigma^2 & \eta_1\sigma^2 \\ \eta_1\sigma^2 & \eta_1^2\sigma^2 + \sigma_\varepsilon^2/(1-\lambda^2) \end{pmatrix},$$

where η_1 is defined in (13). By the properties of multivariate normal distribution, the conditional mean

$$E(\theta \mid q_{t-1}) = \bar{\theta} + \frac{\eta_1\sigma^2}{\eta_1^2\sigma^2 + \sigma_\varepsilon^2/(1-\lambda^2)}(q_{t-1} - E(q_{t-1})).$$

To compute the value of $E(q_t \mid q_{t-1})$, start with (6), writing it as

$$q_t - \lambda q_{t-1} = \eta_0(1-\lambda) + \eta_1(1-\lambda)\theta + \varepsilon_t.$$

This implies that

$$\begin{aligned} E(q_t \mid q_{t-1}) &= \lambda q_{t-1} + \eta_0(1-\lambda) + \eta_1(1-\lambda)E(\theta \mid q_{t-1}) \\ &= \gamma_0 + \left(\lambda + \eta_1(1-\lambda)\frac{\eta_1\sigma^2}{\eta_1^2\sigma^2 + \sigma_\varepsilon^2/(1-\lambda^2)}\right) q_{t-1}, \end{aligned}$$

where γ_0 is a constant which does not depend on q_{t-1}. This expression can be simplified to the following:

$$E(q_t \mid q_{t-1}) = \gamma_0 + \left(1 - \frac{\sigma_\varepsilon^2/(1+\lambda)}{\eta_1^2\sigma^2 + \sigma_\varepsilon^2/(1-\lambda^2)}\right) q_{t-1}.$$

It follows from Proposition 6 that both the numerator and denominator of this fraction are positive, the numerator decreases in α and the denominator increases in α. Therefore, the mixed derivative in q_{t-1} and α is positive. This completes the proof. □

Appendix C Robustness to functional form assumptions

C.1 Hospital's utility function in general form

The model in Section 3 employs a linear demand for hospital services, a linear benefit from altruism and a linear marginal disutility of quality-inducing efforts. Here we relax these assumptions about the functional form. Specifically, we assume that the demand for hospital services $Q(q_t)$ is monotonically increasing and concave in hospital quality: $Q(q_t) > 0$, $Q'(q_t) > 0$, $Q''(q_t) \leq 0$; the benefit from altruism $u(q_t)$ is monotonically increasing and concave in hospital quality: $u'(q_t) > 0$, $u''(q_t) \leq 0$; and the disutility of efforts $C(e_t)$ is an increasing and convex function: $C'(e_t) > 0$, $C''(e_t) > 0$. The one-period utility of the hospital becomes

$$U_t(m_t, q_t, e_t) = \theta u(q_t) + Q(q_t)(R_t + p_t(1 + (\kappa m_t - 1)\alpha) - d_t) - C(e_t)$$

and the first order condition under constant prices and unit costs of healthcare services from Assumption 2 is

$$E_t[\theta u'(q_t) + Q'(q_t)(R - d + p(1 + (\kappa m_t - 1)\alpha)) - C'(q_t) + \beta Q(q_{t+1})p\kappa\alpha] = 0,$$

which leads to the following nonlinear difference equation for q_t:

$$\theta u'(q_t) + Q'(q_t)(R - d + p(1 + (\kappa(q_{t-1} + \varepsilon_t) - 1)\alpha)) - C'(q_t) + \beta p\kappa\alpha E_t[Q(q_{t+1})] = 0. \quad (27)$$

We cannot solve this equation directly, so we linearize it along a non-stochastic steady state.

C.2 Mean value

In order to calculate the mean value for the stationary solution of the linearized equation, we rewrite the first order condition (27) for the constant $q_t \equiv q$ and $\varepsilon \equiv 0$:

$$\theta u'(q) + Q'(q)(R - d + p(1 + (\kappa q - 1)\alpha)) - C'(q) + \beta p\kappa\alpha Q(q) = 0. \tag{28}$$

Solve equation (28) with respect to q, denote the solution as μ. Note that $\mu = \mu(\alpha, \theta)$ and the equation is nonlinear, so only a numerical solution is feasible in the general case.

Using the implicit function theorem, we can compute

$$\frac{\partial \mu}{\partial \theta} = -\frac{u'(\mu)}{\theta u''(\mu) + Q''(\mu)(R - d + p(1 + (\kappa\mu - 1)\alpha)) + Q'(\mu)p\kappa\alpha - C''(\mu) + \beta p\kappa\alpha Q'(\mu)} \tag{29}$$

and

$$\frac{\partial \mu}{\partial \alpha} = -\frac{Q'(\mu)p(\kappa\mu - 1) + \beta p\kappa Q(\mu)}{\theta u''(\mu) + Q''(\mu)(R - d + p(1 + (\kappa\mu - 1)\alpha)) + Q'(\mu)p\kappa\alpha - C''(\mu) + \beta p\kappa\alpha Q'(\mu)}. \tag{30}$$

The numerators of (29) and (30) are greater than zero. The denominators of (29) and (30) are less than zero for sufficiently high values of $C''(\mu)$. Therefore the mean optimal quality level increases in the level of altruism θ and in the size of incentive α.

The mixed partial derivative is

$$\begin{aligned}\frac{\partial^2 \mu}{\partial \alpha \partial \theta} &= \frac{u'(\mu)(Q''(\mu)p(\kappa\mu - 1) + (1+\beta)Q'(\mu)p\kappa) + u''(\mu)(Q'(\mu)p(\kappa\mu - 1) + \beta p\kappa Q(\mu))}{(\theta u''(\mu) + Q''(\mu)(R - d + p(1 + (\kappa\mu - 1)\alpha)) + Q'(\mu)p\kappa\alpha - C''(\mu) + \beta p\kappa\alpha Q'(\mu))^2} \\ &= \frac{(u''(\mu)Q'(\mu) + u'(\mu)Q''(\mu))p(\kappa\mu - 1) + u'(\mu)(1+\beta)Q'(\mu)p\kappa + u''(\mu)\beta p\kappa Q(\mu)}{(\theta u''(\mu) + Q''(\mu)(R - d + p(1 + (\kappa\mu - 1)\alpha)) + Q'(\mu)p\kappa\alpha - C''(\mu) + \beta p\kappa\alpha Q'(\mu))^2},\end{aligned}$$

which means that if the second derivative of $u(\mu)$ and/or $Q(\mu)$ is large then the effect of an increase in α can be negative for high values of θ. If $u(\mu)$ and $Q(\mu)$ have small second derivative then the sign of $\dfrac{\partial^2 \mu}{\partial \alpha \partial \theta}$ is determined by $u'(\mu)(1+\beta)Q'(\mu)p\kappa$, which is positive.

C.3 Cycle around the mean

Denote $\tilde{q}_t = q_t - \mu$. Then the linearized first order condition becomes

$$\begin{aligned}&Q'(\mu)p\kappa\alpha\tilde{q}_{t-1} + (\theta u''(\mu) + Q''(\mu)(R - d + p(1 + (\kappa\mu - 1)\alpha)) + Q'(\mu)p\kappa\alpha - C''(\mu))\tilde{q}_t \\ &\qquad + \beta p\kappa\alpha Q'(\mu)E_t\tilde{q}_{t+1} = -Q'(\mu)p\kappa\alpha\varepsilon_t.\end{aligned} \tag{31}$$

The value of the persistence parameter λ is determined by the characteristic equation of (31). Similarly to the main model, there are two real roots for the characteristic equation of (31) for sufficiently large values of $C''(\mu)$. One of the roots must be discarded because of the transversality condition. The other root determines how persistent the deviations of measured quality from the mean value are in the steady state.

Note that in contrast with Section 3, λ depends on θ in nonlinear case. This means that the regression analysis allows us to estimate the average value of the persistence parameter λ.

C.4 Numerical solution

The figures 3–5 are plotted for $u(q_t) \propto q_t^{0.75}$, $Q(q_t) \propto q_t^{0.85}$, $C(q_t) \propto q_t^2$ to demonstrate typical behavior of the mean value $\mu(\alpha, \theta)$, its derivative $\partial\mu/\partial\alpha$ (presented as an increment $\Delta\mu = \mu(\alpha) - \mu(\alpha - \Delta)$ for $\Delta = 0.001$) and the value of the persistence parameter $\lambda(\alpha, \theta)$. The figures

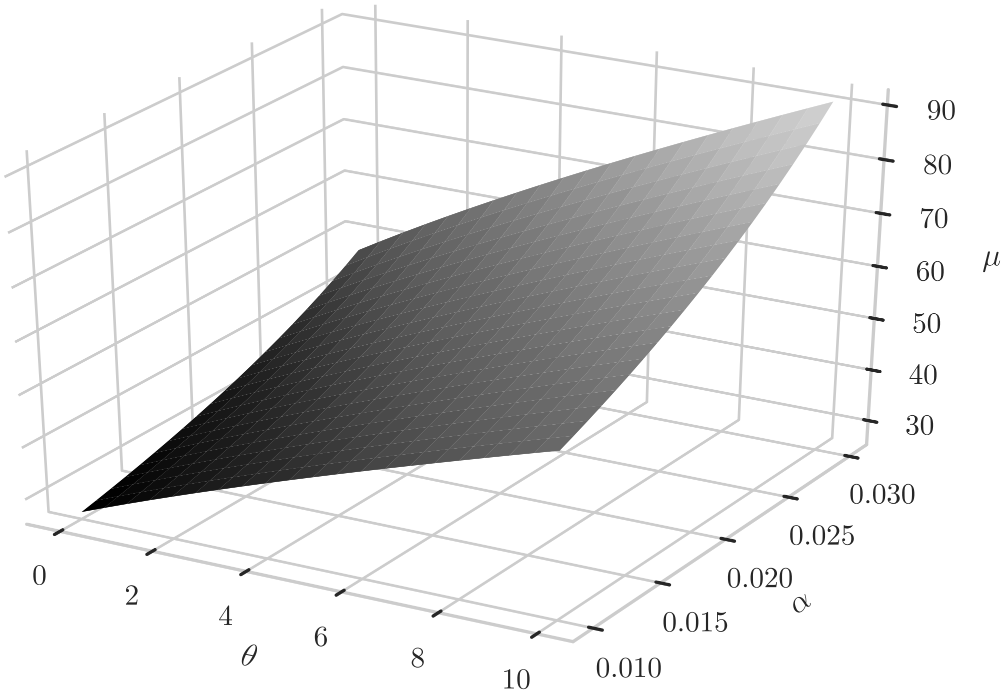

Figure 3: Mean value of the optimal quality μ as the function of θ and α

illustrate that the pay-for-performance effect, albeit non-linear, still has all the properties of the effect in the main model: it is positive, monotonically increases in α and in θ. The persistence parameter λ also increases in α.

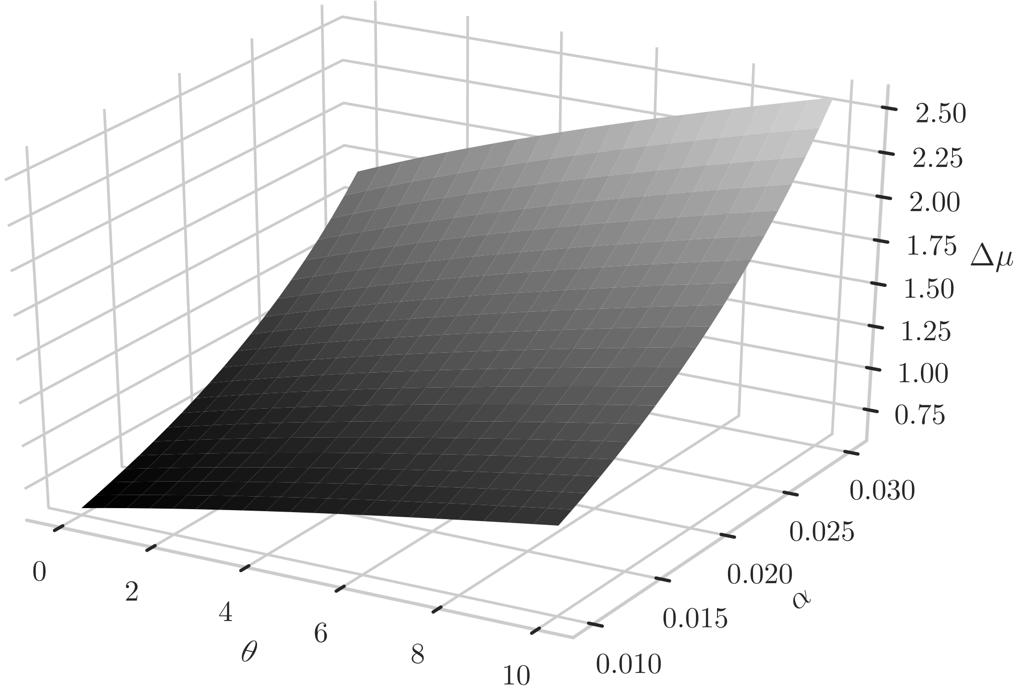

Figure 4: Increment of the mean value of the optimal quality to an increase in α on 0.001 $\Delta\mu$ as the function of θ and α

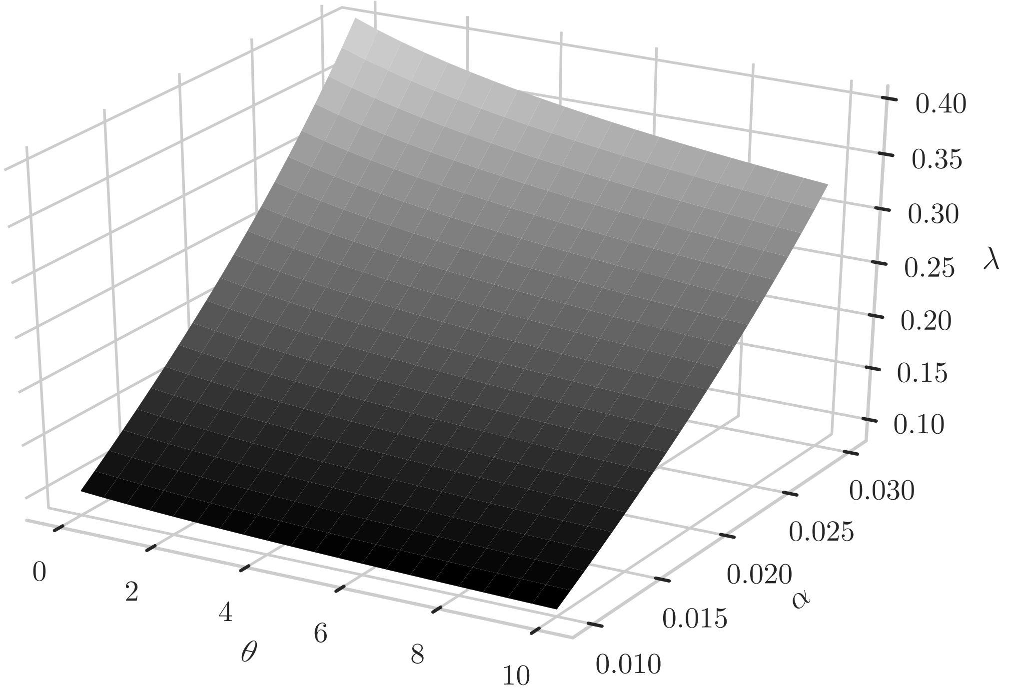

Figure 5: Value of the persistence parameter λ as the function of θ and α

Appendix D Estimation with the dynamic panel

Table 10: Explaining the total performance score in Medicare's hospitals

	TPS	
$L(TPS)$	0.319^{***}	(0.049)
$L^2(TPS)$	-0.221^{***}	(0.042)
$L(TPS) \cdot medicare\ share \cdot \alpha$	0.282^{***}	(0.077)
$L^2(TPS) \cdot medicare\ share \cdot \alpha$	0.443^{***}	(0.069)
$medicare\ share$	-15.078^{***}	(5.775)
$medicare\ share \cdot \alpha$	-42.817^{***}	(5.093)
$casemix$	6.783^{***}	(1.577)
$physicians/beds$	1.155^{***}	(0.266)
$physicians/beds \cdot medicare\ share \cdot \alpha$	-1.669^{***}	(0.338)
$nurses/beds$	0.048	(0.062)
$nurses/beds \cdot medicare\ share \cdot \alpha$	−0.152	(0.117)
dsh	−1.708	(3.629)
$\log(beds)$	-7.253^{***}	(0.689)
$\log(beds) \cdot medicare\ share \cdot \alpha$	5.061^{***}	(0.918)
$HRRP\ penalty$	0.318	(0.246)
$MUEHR$	1.625	(1.111)
$year = 2013$	2.625^{***}	(0.746)
$year = 2014$	2.023^{***}	(0.378)
$year = 2015$	-3.044^{***}	(0.161)
$year = 2016$	0.195	(0.264)
$year = 2017$	−0.650	(0.460)
$year = 2018$	-1.148^{***}	(0.468)
$constant$	58.955^{***}	(5.565)
Observations	15971	
Hospitals	2968	
Arellano–Bond test statistic	−1.541	

Notes: Coefficients for year dummies are normalized to zero sum.
Robust standard errors are in parentheses.
*, ** and *** show significance at levels of 0.1, 0.05 and 0.01, respectively.
The Sargan statistic is not applicable to the specification with robust standard errors.

Appendix E Data sources

The Hospital Compare data archive contains hospital-level files on aggregate quality measure (TPS_{it}) and measures for the clinical-process-of-care, patient-experience-of-care, outcome-of-care and efficiency domains. The same data source contains hospital characteristics such as ownership and acute-care status.

- The files on clinical process of care give values of 12–13 incentivized measures. Each measure is the percent of patient cases, for which the corresponding clinical requirement is satisfied (i.e., certain type of therapy or drug provided within a given time interval).
- Patient-experience-of-care files give data for eight measures, considered by value-based purchasing, as well as a few measures outside the incentives schedule. Each measure is the percent of discharged patients who gave the most positive ("top-box") response to the corresponding question (e.g. communication with doctors, nurses, medical staff, assessment of cleanliness and quietness of the hospital environment).
- Outcome-of-care data contains hazard rates for patient mortality after discharge. Hazard rates for each of three conditions (AMI, heart failure and pneumonia) were included in the value-based purchasing. The domain also reports complication rates, two of which are considered by the reform: a measure of the complication rate for selected conditions and a measure of central line associated blood stream infections.
- Safety data provides information about healthcare associated infections and several other safety measures. The safety domain appeared in 2015 and its measures were disentangled from the outcome-of-care domain, although measures of knee or hip surgery complications remained in the outcome-of-care domain.
- The efficiency-of-care domain has only one measure, which is Medicare spending per beneficiary.

The list of hospital quality measures used for computing total performance score is given in Table 11. Table 12 shows sources of hospital and patient data.

Specifically, Provider of Service data were downloaded from https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services (Table 13).

Impact Files data were downloaded from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS (Table 14).

Total performance scores and the individual performance measures were downloaded from https://data.medicare.gov/data/hospital-compare (Table 15).

The Hospitals Readmissions Reduction Program data were downloaded from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Archived-Supplemental-Data-Files (Table 16).

Table 11: List of measures for each domain of TPS

Variable	Definition
Patient experience of care domain	
Comp-1-*ap*	Nurses always communicated well
Comp-2-*ap*	Doctors always communicated well
Comp-3-*ap*	Patients always received help as soon as they wanted
Comp-4-*ap*	Pain was always well controlled
Comp-5-*ap*	Staff always gave explanation about medicines
Comp-6-*yp*	Yes, staff did give patients discharge information
Clean-hsp-ap	Room was always clean
Quiet-hsp-ap	Hospital always quiet at night
Recmnd-dy	Patients who would definitely recommend the hospital
CTM-3	3-item care transitions measure
Hsp-rating-910	Patients who gave hospital a rating of 9 or 10 (high)
Clinical process of care domain	
AMI-8*a*	Primary PCI received within 90 minutes of hospital arrival
AMI-10	Statin prescribed at discharge
HF-1	Discharge instructions
PN-3*b*	Blood cultures performed in the emergency department prior to initial antibiotic received in hospital
PN-6	Initial antibiotic selection for CAP in immunocompetent patient
IMM-2	Influenza immunization
*SCIP-Card*2	Surgery patients on beta-blocker therapy prior to arrival who received a beta-blocker during the perioperative period
*SCIP-Inf*1	Prophylactic antibiotic received within 1 hour prior to surgical incision
*SCIP-Inf*2	Prophylactic antibiotic selection for surgical patients
*SCIP-Inf*3	Prophylactic antibiotics discontinued within 24 hours after surgery end time
*SCIP-Inf*4	Cardiac surgery patients with controlled 6 A.M. postoperative blood glucose
*SCIP-Inf*9	Postoperative urinary catheter removal on post operative day 1 or 2
*SCIP-VTE*1	Surgery patients with recommended venous thromboembolism prophylaxis ordered
*SCIP-VTE*2	Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery
Outcome of care domain	
Mort-30-*AMI*	Hospital 30-day mortality rates from heart attack
Mort-30-*HF*	Hospital 30-day mortality rates from heart failure
Mort-30-*PN*	Hospital 30-day mortality rates from pneumonia
Comp-hip-knee	Complication rate following elective primary total hip arthroplasty and/or total knee arthroplasty
Safety domain	
PSI-90	Patient safety (weighted complication rate) for selected conditions
CLABSI	Central line associated blood stream infection
CAUTI	Catheter-associated urinary tract infection
SSI-Colon	Surgical site infection from colon surgery
SSI-Hysterect	Surgical site infection from abdominal hysterectomy
MRSA	Methicillin-resistant Staphylococcus aureus (MRSA) bacteremia
CDI	Clostridium difficile infection
PC-01	Elective delivery prior to 39 completed weeks gestation
Efficiency domain	
MSPB-1	Spending per hospital patient with Medicare (ratio to the median)

Table 12: Sources of the hospital and patient data

Variable	Source
TPS	Hospital Compare
casemix	Impact Files
dsh	Impact Files
medicare share	Impact Files
urban	Impact Files
public	Hospital Compare
emergency	Hospital Compare
physicians	Provider of Service Files
nurses	Provider of Service Files
teaching	Provider of Service Files
beds	Provider of Service Files
forprofit	Hospital Compare
regional dummies	Hospital Compare
individual performance measures	Hospital Compare
HRRP penalty	Hospital Readmissions Reduction Program Supplemental files to Acute Inpatient PPS Final Rules
MUEHR	Electronic Health Records Incentive Program Eligible Hospitals Public Use files

Table 13: Download links for the Provider of Services data

Year	Link
2011	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2011
2012	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2012
2013	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2013
2014	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2014
2015	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2015
2016	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2016
2017	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2017
2018	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2018

The EHR (Electronic Hospital Records) Incentive Program (renamed to the Promoting Interoperability (PI) Program) data were downloaded from https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/PUF (Table 17).

Table 14: Download links for the Impact Files data

Year	Link
2011	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_13_FR_Impact_File.zip
2012	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_14_FR_Impact_File.zip
2013	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Impact-File.zip
2014	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Impact.zip
2015	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Impact.zip
2016	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-CMS-1677-FR-Impact.zip
2017	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2019-CMS-1694-FR-Impact.zip
2018	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2020-FR-Impact-File.zip

Table 15: Download links for the TPS data

Year	Date	Link
2011	10/01/2013	https://medicare.gov/download/HospitalCompare/2013/October/HOSArchive_Revised_Flatfiles_20131001.zip
2012	10/23/2014	https://medicare.gov/download/HospitalCompare/2014/October/HOSArchive_Revised_Flatfiles_20141023.zip
2013	10/08/2015	https://medicare.gov/download/HospitalCompare/2015/October/HOSArchive_Revised_FlatFiles_20151008.zip
2014	11/10/2016	https://medicare.gov/download/HospitalCompare/2016/October/Hospital_Revised_FlatFiles_20161110.zip
2015	10/24/2017	https://medicare.gov/download/HospitalCompare/2017/October/HOSArchive_Revised_FlatFiles_20171024.zip
2016	10/31/2018	https://medicare.gov/download/HospitalCompare/2018/October/HOSArchive_Revised_FlatFiles_20181031.zip
2017	10/30/2019	https://medicare.gov/download/HospitalCompare/2019/October/HOSAchive_Revised_Flatfiles_20191030.zip
2018	01/29/2020	https://medicare.gov/download/HospitalCompare/2020/January/HOSArchive_Revised_Flatfiles_20200129.zip

Table 16: Download links for the HRRP data

Year	Link
2013	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY_2013_FR_Readmissions_File.zip
2014	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/Readmissions-Supplemental-Data-PUF.zip
2015	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2015-FR-Readmit-Supp-Data-File.zip
2016	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2016-CMS-1632-FR-Readmissions.zip
2017	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2017-CMS-1655-FR-Hospital-Readmissions.zip
2018	https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2018-FR-Hospital-Readmissions.zip

Table 17: Download links for the EHR Incentive Program data

Stage	Link
1	https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage1.zip
2	https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/EH_PUF_Q32018Stage2.zip