CINCH competence in competition + health series

#2019/03

Galina Besstremyannaya and Sergei Golovan

Physicians' altruism in incentives contracts:
Medicare's quality race

Imprint

CINCH SERIES

CINCH – Health Economics Research Center
Weststadttürme, Berliner Platz 6-8
45127 Essen

www.cinch.uni-due.de
cinchseries@cinch-essen.de
Phone +49 (0) 201 183 - 3679
Fax +49 (0) 201 183 - 3716

ISSN 2199-8744 (online)

CINCH
competence in competition + health
series

#2019/03

Galina Besstremyannaya and Sergei Golovan

Physician's altruism in incentive contracts: Medicare's quality race

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Bundesministerium für Bildung und Forschung

rwi

Galina Besstremyannaya[*] and Sergei Golovan [†]

Physician's altruism in incentive contracts: Medicare's quality race

Abstract

The paper analyzes the impact of physicians' altruism and motivation on the outcomes of pay-for-performance schemes in healthcare, where a fixed price contract on quantity is supplemented with a relative performance contract on quality. Our theoretical model forecasts crowding out of most altruistic types. In an empirical application to the Medicare's nationwide natural experiment with a relative performance contract on quality for acute inpatient care since 2013, we observe the proof of this prediction. Namely, the quality dimensions, which are linked to patient's benefit, demonstrate higher deterioration among top-performing hospitals than other incentivized dimensions.

Keywords: incentives contracts, altruism, dynamic panels, healthcare

JEL Classification Codes: C22, C23, D21, D22, I18

We thank Jay Bhattacharya, Katharina Blankart, Ruben Enikolopov, Sergei Izmalkov, Martin Karlsson, Sergey Kokovin, Anna Obizhalova, Vladimir Smirnov, Alexei Zakharov, participants of the European Economic Association 2015 conference and seminars at New Economic School (2015), Higher School of Economics/ New Economic School Political Economy (2016) and CINCH Health Economics Research Center at University of Duisburg-Essen (2019) for helpful comments.

[*] *Lead Research Fellow, Centre for Economic and Financial Research at New Economic School (Moscow, Russia) and Visiting Researcher, CINCH Health Economics Research Center at University of Duisburg-Essen (Essen, Germany), Moscow, Russia, Nakhimovsky pr.47, gbesstre@cefir.ru*

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

1 Introduction

Public contracting with firms under asymmetric information about their technology provides a classic example of an agency problem, where government as a principal can achieve social optimum in terms of product's quantity and agent's efforts through nonlinear prices (Laffont and Tirole, 1993).[1] However, firms face a number of competing objectives, and this context of multi-tasking may result in trade-offs between quality, quantity and efforts, especially if demand does not respond to quality (Hölmstrom and Milgrom, 1991). A solution has been found in incentives contracts on quality, stemming from pay-for-performance in managerial economics. Overall, incentives contracts based on piece-rate compensation may be viewed as a most natural application of performance-pay. However, the desirable performance level may be unknown to the principal or subject to a common shock. Additionally, under an asymmetric information about the link between efforts and outcomes, agents may "game" on performance (Baker, 1992). So contests with empirically-based absolute or relative performance standards are widely exploited as a means to solve inefficiency of incentives contracts, particularly on the markets with a large number of agents and imprecisely known distribution of their outcomes (Green and Stokey, 1983; Hölmstrom, 1982; Lazear and Rosen, 1981). The payment mechanism has become widespread in various public sector industries: civil service, healthcare, education, and social work.

Incentives contracts are targeted at increasing aggregate performance, which is observed in the experimental and empirical literature (Eijkenaar et al., 2013; Houle et al., 2012). However, the mean effect hinders heterogeneity in the responses of agents who differ in their abilities. In particular, both theoretical literature and natural experiments point to deteriorating performance of the front-runners. Theoretical explanations include a crowding out of motivation, which may be explained by intrinsic behavior (Bénabou and Tirole, 2006; Kreps, 1997) or conformism (Murdock, 2002). Additionally, there is an issue of slacking efforts in tournaments (Prendergast, 1999), and especially, in the dynamic context (Casas-Arce and Martínez-Jerez, 2009; Radner, 1985). However, altruistic agents would be interested in a social value of their performance per se. There is limited theoretical literature on altruism in public good games and piece-rate incentives contracts (Makris and Siciliani, 2014, 2013; Buurman and Dur, 2012; Makris, 2009), but little is known about the influence of altruism on the outcomes of tournaments.

This paper analyzes the impact of agent's altruism on the effects of pay-for-performance remuneration in healthcare, where a fixed price contract on quantity is supplemented with a relative performance contract on quality.[2] Our theoretical model predicts quality con-

[1] A particularly notable implementation is yardstick competition – setting the cost of comparable firms as the benchmark for a given firm (Shleifer, 1985).

[2] Various forms of pay-for-performance are particularly prevalent in healthcare, which is a classic case of

vergence, although, altruism may lead to quality decrease among subgroups of the high-performing agents. The model exploits the Siciliani (2009) approach of adding the Bénabou and Tirole (2006) motivation to the agent's utility. Different from Siciliani (2009), we model a relative performance remuneration instead of a piece-rate contract and regard quality as a continuous variable.

In an empirical application to the U.S. Medicare's nationwide natural experiment with value-based purchasing for inpatient acute care since 2013, we show that quality dimensions, which are linked to patient's benefit, demonstrate deterioration among top-performing hospitals. In other words, altruism may become a reason for crowding out on the healthcare market.

The novelty of our empirical approach is severalfold. Firstly, we use dynamic panel data estimations to account for "habit-formation". The analysis excludes "regression-to-the-mean" effect by modelling the time-dependent long-term mean as a function of hospital characteristics. Secondly, while previous studies exploited the data for prototypes of value-based purchasing and concentrated on composite measures, we use longitudinal datasets on each quality measure of all acute-care Medicare's hospitals before and after the reform (fiscal years 2004–2017). The data are supplemented with patient case-mix, ownership, share of Medicare population and various hospital control variables, coming from: Medicare's Impact Files, Final Rules, Provider of Service Data, and Provider Utilization and Payment Data.

The remainder of the paper is structured as follows. Section 2 reviews the literature on incentives contracts and altruism in healthcare. Section 3 explains incentives regulation for the inpatient care, as implemented in the U.S. within Medicare's value-based purchasing. Section 4 provides theoretical model, predicting heterogeneous effects of uneven tournaments with altruism, and describes econometric methodology for estimating dynamic panel data models with endogeneity. Section 5 outlines the data. Section 6 presents the results of the estimations, while Section 7 is concerned with the discussion of our estimates. The economic and behavioral evidence on existence of altruism in healthcare, the price-setting within Medicare's value-based purchasing, and history of incentives contracts on the U.S. healthcare market are given in the Appendices.

an industry with an asymmetric information and physician-agency problem.

2 Related literature

2.1 Incentives contracts

The origins of incentives regulation under asymmetric information may be found in the approach by Baron and Myerson (1982) and the yardstick competition model by Shleifer (1985), which aim to set a benchmark for evaluating the potential for a regulated monopolistic firm. The model establishes the fixed price contract for each firm dependent on the costs of similar firms and independent of the firm's own price. Laffont and Tirole (1986) extend the approach to the case when a firm's cost-reducing efforts are not observed. The authors propose a cost-sharing contract: a lump-sum transfer and a fraction of actually incurred costs. The contract allows for risk sharing under uncertainty about the firm's costs.

If quality and output are independent objectives, quality may be treated as an additional output in the framework of the multi-product firm (Laffont and Tirole, 1990) and the same model may be used. However, quality and output are likely to be dependent. Therefore, the incentives for quality enhancement and cost-reducing efforts should be analyzed in their totality as interrelated objectives of a multi-task agency problem (Hölmstrom and Milgrom, 1991). In this regard, Laffont and Tirole (1993) investigate the influence of quality on the power of incentive schemes and discover differential results depending on whether quality and quantity are net complements or net substitutes. Ma (1994) shows that a fixed price contract leads to efficient levels of costs and quality when these are the only two objectives of a hospital, while incentive trade-offs arise in the presence of other objectives.

A solution for dealing with such trade-offs may be discovered in incentives regulation, related to performance pay to employees in labor contracts. At the level of pricing the whole firm, it dates back to the early 1980s when various performance targets, often related to piece-rate pay, were employed for enhancing the quality of natural monopolies and telecommunications (Kridel et al., 1996; Joskow and Schmalensee, 1986).

The main purpose of an incentives contract is maximization of the aggregate output (or equivalently, aggregate effort) of several agents. Here, the seminal paper by Lazear and Rosen (1981) offers an alternative to a piece-rate or an absolute standard incentives contract with two players if there is a large variance in a common shock to agents' productivity. The result about the effectiveness of rank-order contracts (e.g. in the form of tournaments) holds in a multi-agent setting, where the use of rank-order or relative performance incentives scheme is explained by nonstationary environment and (precisely) unknown distribution of an outcome (Green and Stokey, 1983; Hölmstrom, 1982). Experimental literature supports the theoretical expectations about an increasing aggregate effort in relative performance contracts, yet it comes at the cost of high variance (Eriksson and Villeval, 2008; Bull et al., 1987). Accordingly, some agents may demonstrate deteriorating performance.

If heterogeneous abilities are known to the principal, competitive handicaps or winner-discriminating prizes allow for an efficient outcome of a relative performance incentives contract (Gürtler and Kräkel, 2010; O'Keeffe et al., 1984; Lazear and Rosen, 1981). For instance, in a real world implementation, a jockey's weight handicap decreases variation in the finishing time of horse races (Lynch, 2005). However, abilities are likely to be unknown, and this explains the unintended consequences of this remuneration mechanism. Front-runners decrease their efforts owing to overconfidence in success while underdogs get demotivated due to inevitable defeat and may engage in a risky behavior (Eriksson et al., 2009; Prendergast, 1999; Gibbs, 1991; Dixit, 1987; Rosen, 1986).

2.2 Altruism in incentives contracts

Pure altruism may be defined as a concern about monetary and non-monetary benefits of others, and is modeled through entering welfare of others into agent's utility function (Becker, 1976). An extension of impure ("warm-glow") altruism deals with the act of giving by itself, rather than the consideration of the welfare gain of another agent (Andreoni, 1989).

Theoretical analyses of the effects of a piece-rate incentives contract on street-level bureaucrats, general practitioners or hospitals reveal that pure altruism[3] causes motivation crowding in for the best-performers (Buurman and Dur, 2012; Siciliani, 2009). It should be noted that incentives contracts with organizations are often linked to public reporting and enhanced competition,[4] so quality may increase through extra stimuli to low quality/low altruism agents, who are forced to raise quality to catch up with agents of a high quality/high altruism type (Kairies-Schwarz, 2014). Front-runner crowding in may offset negative effects of potential crowding out owing to extrinsic motivation (Siciliani, 2009).

2.3 Application to healthcare markets

Major actors

The market for healthcare is centered on consumers, who incur a risk of getting sick and receive a benefit B from healthcare in case of illness (Ellis and McGuire, 1986). Benefit may be viewed as the overall increase in patient's health (e.g. in terms of quality-adjusted life years), distributed with a given pdf and independent from hospital's efforts (De Fraja,

[3]Towards the benefits of clients and equivalently, towards the performance of the agent per se

[4]Incentive contract on quality may not even be directly monetized, so quality advancement happens owing to behavioral aspects related to reputation, signaling and non-price competition for clients. For example, "Advancing Quality Initiative" at hospitals in the North West of England consists solely of disclosing information on various quality dimensions, and awards are essentially a recognition of hospital's/physician's achievement (National Health Service, 2015).

2000).[5] Similarly, benefit may be taken equivalently to an amount of health or quantity of healthcare, produced by physicians in the course of healthcare treatment (Bardey and Lesur, 2006; Lee, 1995). Benefit is implicitly monetized and enters the utility or value functions in the same units as does the cost function. Bardey and Lesur (2006) explicitly phrase that benefit is a financial equivalent of the healthcare supplied by the physician. In a medical sense, a physician's concern about the patient's benefit as well as physicians being perfect agents to patients implies that the supply of healthcare is in the best interest of a patient's health and well-being, despite economic or other matters of physician or patient (Lee, 1995; Emanuel and Emanuel, 1992).

The suppliers are hospitals, also called healthcare providers. Hospitals are viewed as profit maximizers even if they have a nonprofit status.[6] Healthcare services are offered by physicians, who are employed by hospitals,[7] are interested in hospital's profits but also, according to the plausible assumptions by Arrow (1963) and Newhouse (1970),[8] show at least some degree of ethical behavior (altruism)[9] and motivation (pride).[10]

Social preferences on the healthcare market are commonly analyzed in terms of altruism towards patients: physician's concern about patient's benefit, i.e. health and well-being (McGuire, 2000; Lee, 1995; Emanuel and Emanuel, 1992; Ellis and McGuire, 1986). The type of altruism is commonly agent's private information, although it may be revealed owing to a high altruism-high quality link (Kairies-Schwarz, 2014; Brekke et al., 2012) or through a menu of linear contracts which are characterized by the degree of cost-sharing (Jack, 2005). Consequently, the benefit to patient B or the overall consumer utility (Siciliani et al., 2013) come into physicians' objective function[11] with a certain factor $\theta \in [0, 1]$. The altruistic component may be added to the objective function of a hospital if it is analyzed as a single agent. It should be noted that the quantity of healthcare by ethical suppliers exceeds a certain benchmark (Ma and McGuire, 1997) and the quality must be above a certain malpractice threshold set by the regulator (Siciliani et al., 2013). A review on the existence of altruism in healthcare is given in Appendix A.

[5] Here efforts are treated as intrinsic efficiency.

[6] Nonprofit hospitals may add financial surplus to their reserves, or regard it as non-pecuniary compensation (Brekke et al., 2012; Siciliani, 2006; Jack, 2005; Chalkley and Malcomson, 1998; Dranove and White, 1994).

[7] Hospitals here include physician groups and individual practices.

[8] See review on medical ethics and physician objectives in McGuire (2000).

[9] Kessel (1958) focuses on physician's price discrimination in the interest of the poor, which is similar to the presence of the altruistic behavior for price-taking physicians.

[10] May be viewed synonymous, e.g. Siciliani et al. (2013), or may be separated into altruism *per se* and motivation, aimed at sustaining the level of reputation.

[11] Ellis and McGuire (1990) show that benefit to patient in physician's objective function may be viewed as a Roth-Nash solution to a physician-patient cooperative game on the quantity of healthcare.

Contracts

A simplified framework delegates suppliers to perform all decisions about the quantity and quality, and fully insured consumers turn for healthcare once they get sick[12] and accept all the treatment (Ellis and McGuire, 1986). The contract is essentially a price schedule targeted at stimulating optimal quantity/quality of services and/or agent's cost-reducing efforts. Additional issues may include hospital's competition for physicians[13] or patients,[14] and price-adjustment owing to asymmetry of patient types (Miraldo et al., 2011; Chalkley and Khalil, 2005).[15]

Incentives regulation on the healthcare markets requires the identification of products and determination of a reasonable cost for each product. This is accomplished with the help of a restricted number of medically justified groups (diagnosis-related groups, DRGs), with a statistically stable distribution of resource consumption within each group (Thompson et al., 1979). This classification is a core part of a prospective payment system (PPS) – a method of reimbursement 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, this approach has been adopted in most healthcare systems. It may be noted that such average cost pricing is a version of yardstick competition, when lump-sum transfers are unavailable. The contract has been expanded to a more efficient cost-sharing principle, which allows incorporating outlier cases: a fixed price and a fraction of actually incurred costs (Laffont and Tirole, 1993).

Regarding incentives contracts, a piece-rate pay-for-performance (also called "payment by results") started with quality measures for family practices in the U.K. These targets were established in 2004 and covered patient experience of care, management of chronic diseases and practice organization (Campbell et al., 2009). Rank-order tournaments have been later applied for acute-inpatient care: for instance, length-of-stay relative performance in Japan (since 2003) and benchmarking with inpatient quality measures in the U.S. (piloted in 2006). Overall, healthcare attracts attention in terms of incentives contracts, owing to the large share of public expenditures and the presence of welfare issues, demanding regulation (Chalkley and Malcomson, 2000).

Incentive contracts generally enhance the aggregate level of performance in healthcare (Eijkenaar et al., 2013; Houle et al., 2012; Moreno-Serra and Wagstaff, 2010), although the

[12]See RAND health insurance experiment-related literature on demand for healthcare and its major determinants.

[13]Where physician has a certain degree of altruism and an inclination to provide a certain quantity of services, e.g. Ellis and McGuire (1986)

[14]The Hotelling (1929) quality competition, e.g. Brekke et al. (2012) or dumping certain diagnoses/cream-skimming as in Barros (2003).

[15]Heterogeneity in health status due to severity of their illnesses, potentially even within homogeneously defined diagnosis groups.

observed mean effect hinders differential responses by under-performing and over-performing agents. In particular, hospitals already above the target may not have enough incentives for improvement (Besstremyannaya, 2016; Nawata and Kawabuchi, 2013; Ryan et al., 2012; Werner et al., 2011; Miraldo et al., 2011; Grabowski et al., 2011; Mannion et al., 2008; Doran et al., 2008; Lindenauer et al., 2007; Rosenthal et al., 2005).

Effect of altruism

Heterogeneous altruism is expected to cause differential dynamics under piece-rate contracts: higher piece-rate increases performance of agents with low and high altruism, while agents with medium altruism are subject to motivation crowding out (Siciliani, 2009). Kaarboe and Siciliani (2011) show that only intermediate values of altruism allow implementing the first best under piece-rate incentives contracts and multitasking, whereas the price for verifiable dimensions of quality is related to consumer's marginal benefit. Agent's heterogeneity in altruism is also related to socially optimal price-setting in piece-rate contracts (Makris and Siciliani, 2014).

Laboratory experiments have mixed results about altruism and outcomes of incentives contracts on the healthcare markets. For instance, Green (2014) demonstrates a certain crowding out effect under piece-rate incentives contracts, which supplement cost-based or fixed-price pay. The experimental study by Keser et al. (2014) however, does not discover adverse selection owing to piece-rate incentives contracts combined with cost-based remuneration. The difference in results may relate to varying setups. Keser et al. (2014) use the Hennig-Schmidt et al. (2011) experiment design, which is specific to healthcare and is directly linked to benefits for real patients. At the same time, Green (2014) defines healthcare treatment as physicians' correction of mistakes in the course of proofreading patients' essays, where correct or wrong edits are interpreted as beneficial or harmful care (i.e. leading to monetary gains or losses by subjects, who represent patients).

3 Medicare's incentive contract

Value-based purchasing, started in fiscal year 2013, applies to discharges within the inpatient prospective payment system for acute-care Medicare's hospitals (with an exception of two states: Puerto Rico and Maryland). The reform decreases Medicare's DRG-based payment to each hospital by a factor α and redistributes the accumulated fund. The adjustment coefficient γ is calculated as:

$$\gamma_i = 1 + \left(s\frac{tps_i}{100} - 1 \right) \cdot \alpha, \tag{1}$$

where i is the index for hospital, tps_i is hospital's *total performance score* ($0 \leq tps_i \leq 100$) and s is the slope of a linear exchange function ($s = 1.936$ for 2013, $s = 2.096$ for 2014 and $s = 2.580$ for 2015). Hospitals with $\gamma_i < 1$ suffer a loss, while $\gamma_i > 1$ implies that hospitals are rewarded under the reform. The factor α is uniform across hospitals: $\alpha = 0.01$ in 2013 and is annually increased by 0.0025 in 2014–2017.

In our analysis we regard the participation in the reform as a predetermined variable.[16] Indeed, there are eligibility criteria: 1) at least 100 surveys on the patient experience of healthcare; 2) data on at least 4 measures of clinical process of care, with at least 10 respondents on each measure.

Additionally, there is an issue of non-compliance: out of all hospitals eligible for the reform, 96% submitted data in the pre-reform year, 89% joined in the first year (2013), and 92% joined in the second year (2014). Non-compliance is not directly related to hospital size or other observable characteristics but rather linked to hospital's managerial practices and desire to exert quality improving efforts (Centers for Medicare and Medicaid Services, 2007).

The *total performance score* is a weighted average of the scores for several domains:

1. Clinical process of care (12–13 measures).

   Each measure is the percentage of patient cases for which the corresponding clinical requirement is satisfied (i.e. certain type of therapy provided within a given time interval).

2. Patient experience of care (8 measures).

   Each measure is the percentage 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 cleanness and quietness of hospital environment).

3. Outcome of care (3–5 measures, domain added in 2014).

   Measures of outcome of care are 30-day mortality rates (hazard rates) for patients with each of the three conditions: AMI, heart failure and pneumonia.

4. Efficiency (1 measure – Medicare spending per beneficiary, domain added in 2014).

The domain score is the sum of the scores for each measure. The score for each measure m is computed as the maximum of points for hospital's improvement and achievement.

[16] Although participation (and eligibility) for the reform are weakly correlated with the hospital size in FY2013, hospital size would be an insufficient proxy under a potential use of an instrumental variable approach. Indeed, there is no dependency between hospital size and compliance in FY2014. Moreover, we do not observe any "scale" economy in terms of quality performance and hospital size. In fact, compliers have slightly *lower* scores for most quality measures, but differences between compliers and non-compliers are statistically insignificant.

Achievements points a_i^m $(0 \leq a_i^m \leq 10)$ are assigned with a stepwise function, which positions a hospital within the empirical distribution of a quality measure:

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

where y_i^m is the value of measure m for hospital i in the current period, m_b is benchmark, m_a is achievement threshold for measure m. The benchmark and achievement threshold are respectively set as the mean of the top decile (or 95th percentile) and the median in the empirical distribution of y^m, according to the survey in the baseline period.

The pricing schedule includes *improvement points*, yet they are irrelevant for the front-runner hospitals, i.e. hospitals with at least 9 achievement points out of 10 (see Appendix B for details about price-setting in value-based purchasing and Appendix D for the history of incentives contracts on the U.S. healthcare market).

4 Model

4.1 Overview

Under the principal-agent approach on the healthcare market, a principal (a government or a social planner) contracts agents (physicians in our model) on behalf of consumers (patients). Note that pay-for-performance incentives schemes, for instance, as implemented in Medicare's value-based purchasing, target hospitals rather than physicians. However, our model assumes that the reform impacts the physician utility function. Here we rely on the changes in the management of Medicare's hospitals, which bridge the gap between the incentives of hospitals and physicians. For instance, hospital boards pay attention to public reporting of quality performance, review the relative assessment of their hospital and stimulate quality improvement by physicians (Centers for Medicare and Medicaid Services, 2007).

In our model, agents have a type-specific altruism θ with pdf $f(\theta)$. The parameter θ would only be homogeneous in the industries with strict social norms (Makris and Siciliani, 2013), so we follow the approach which differentiates between agents with high/low altruism or consider a continuous case of $\theta \in [\underline{\theta}, \hat{\theta}]$ (Liu and Ma, 2013; Makris and Siciliani, 2013; Siciliani, 2009, 2007; Jack, 2005). Agents have an additively separable welfare function with respect to patient's health (i.e. benefit B) and wealth (in terms of hospital's net profit) – this assumption, shared by the theoretical literature, comes from Blomqvist (1997) and is explained by avoiding extra complexity of analysis due to cross-effects of health and welfare

(Bardey and Lesur, 2006). Agent's program in the presence of the altruism is amended by reservation utility or limited liability constraint: physicians agree to a contract if their utility is above a certain reservation utility level $U_0 \geq 0$ (Makris and Siciliani, 2013; Jack, 2005). Indeed, unlimited altruism may lead to the bankruptcy of a hospital.[17]

We model a relative performance contract by introducing a linear pricing rule, which rewards/punishes the agents for quality q above/below the standard q_0. Since q_0 and performance of each agent are publicly reported, we exploit the Siciliani (2009) approach of adding the Bénabou and Tirole (2006) motivation to the agent's utility. However, in contrast to Siciliani (2009), we focus on a tournament and distinguish between quantity and quality. This accounts for a combination of a yardstick competition with quality taxation, as implemented in Medicare's value-based purchasing. Note that patients may distinguish between the standard q_0 and their own perception of the reputation-related quality $\tilde{q}$, so the two values do not need to coincide.

The continuous treatment variable t enables analyzing an agent's behavior under anticipation of the reform (i.e. with $0 < t < 1$). For simplicity the model centers upon a fixed price contract. However, it can be generalized to a cost-sharing contract by replacing a fixed price A by a two-part tariff $A + bC(q)$, where $C(q)$ is the cost function and $0 \leq b \leq 1$.

4.2 Altruism and motivation

The approach in the discrete version of the model is based on the Siciliani (2009) formulation of the Bénabou and Tirole (2006) intrinsic motivation approach: an agent with quantity above a benchmark $\tilde{q}$ derives a fixed utility of good reputation under an incentives contract and has a linear disutility, proportional to monetary reward for quality. The major differences in our study are as follows. While Siciliani (2009) analyzes piece-rate pay for absolute performance, we model contracts on relative performance. Additionally, Siciliani (2009) focuses on the change in piece price for quality within the pay-for-performance contract, while our objective is comparative statics under a changeover from a prospective pay to pay-for-performance.

Hospital's profit is defined as

$$\pi = A + At(-\alpha + p(q - q_0)) - C(q) \tag{2}$$

where $t \in [0, +\infty)$ is a treatment variable (incentives contract), A is fix pay, $q_0 \geq 0$ is the absolute standard: agents with quality above/below the absolute standard are subject to a

[17] A limited liability form of the minimum profit constraint Makris and Siciliani (2013); Bardey and Lesur (2006) equates $\pi_0 = 0$. It should be noted that owing to intrinsic character of medical profession, π_0 maybe a certain *negative* number, not very large in the absolute value (Choné and Ma, 2011). Liu and Ma (2013) use the term "minimum income" constraint for the Choné and Ma (2011) formulation.

reward/penalty under incentives contract, α is the maximal value of loss and p is the piece-rate for quality q, $C(q)$ is the cost function with $C_q > 0$, $C_{qq} > 0$, and $C(0) = C_q(0) = 0$.

Hospital's objectives are own profit and altruistic concern about patient's benefit, so $U = \pi + \theta B(q)$. Following Ellis and McGuire (1986), we assume that $B_q > 0$, $B_{qq} \leq 0$. The third derivatives equal zero (Laffont and Tirole, 1993).

The key feature of the model in this paper is the participation constraint $\pi \geq 0$: hospital chooses quality level q so that the hospital's profits are non-negative. Given that the profit function is strictly concave, the set $\mathcal{Q}$ of q, which satisfy the participation constraint, is either empty or constitutes a segment $[\underline{q}(t), \bar{q}(t)]$.

Note that the inequality $\pi \geq 0$ simplifies to $A - C(q) \geq 0$ under $t = 0$, which means that the set $\mathcal{Q}$ is non-empty. So, using the continuity argument, we conclude that the solution of hospital's maximization problem exists for $t \in [0, \bar{t}]$ (where $\bar{t}$ can be infinite).

The interior solution for the problem of $\max_q \mathcal{U}$ in *absence* of motivation is given by

$$\begin{aligned} &\max_q A + At(-\alpha + p(q - q_0)) - C(q) + \theta B(q) \\ &s.t. \pi \geq 0 \end{aligned} \tag{3}$$

$q^*(\theta, t)$ comes from the FOC:

$$Atp + \theta B_q = C_q \tag{4}$$

The zero-profit equation may have two solutions: $\underline{q}(t)$ and $\bar{q}(t)$. Comparing (4) with the condition for the profit maximization ($Atp = C_q$) and using the fact that $B(q)$ is an increasing concave function, we conclude that $q^*(\theta, t)$ is greater than the profits maximizing value for $\theta > 0$. So the left limit $\underline{q}(t)$ is not binding.

Consider the zero-profit equation:

$$A + At(-\alpha + p(q - q_0)) - C(q) = 0. \tag{5}$$

Owing to the implicit function theorem

$$\frac{\partial \bar{q}(t)}{\partial t} = \frac{A(p(\bar{q} - q_0) - \alpha)}{C_q - Atp}.$$

$\bar{q}(t)$ lies to the right of the value of q that maximizes the profits π, therefore the denominator in (5) is positive. Subsequently, we can see that: (i) $\bar{q}(t)$ decreases in t if the values of the absolute standard q_0 or α are sufficiently large, (ii) otherwise $\bar{q}(t)$ increases in t. For the rest of the paper, we assume that the case (i) holds.

The right limit $\bar{q}(t)$ can be binding for sufficiently high values of θ. Define $\check{\theta}(t)$ as the value for which $q^*(\theta, t) = \bar{q}(t)$ and assume that $\check{\theta}(t) < \bar{\theta}$.

The second order condition implies that $\theta B_{qq} - C_{qq} < 0$.[18] Using the implicit function theorem to differentiate q^* in t and θ, we obtain comparative statics:[19]

$$\frac{\partial q^*}{\partial t} = -\frac{Ap}{\theta B_{qq} - C_{qq}} \geq 0 \tag{6}$$

$$\frac{\partial^2 q^*}{\partial t \partial \theta} = \frac{ApB_{qq}}{(\theta B_{qq} - C_{qq})^2} \leq 0 \tag{7}$$

Result 1: The partial derivative of the optimal quality in t is non-negative because of the second order condition. So, given an interior solution, there is a quality increase under incentives contract t.

Result 2: Altruism decelerates quality increase, since the mixed partial derivative of the optimal quality in t and θ is non-positive.

Note that since

$$\frac{\partial q^*}{\partial \theta} = -\frac{B_q}{\theta B_{qq} - C_{qq}} \geq 0,$$

the optimum quality is higher for higher altruism.

Now we follow Siciliani (2009) to add the discrete motivation component to the linearly separable utility function of an agent under the assumptions: 1) the reputation is associated with quality above a certain benchmark value $\tilde{q} \geq 0$, 2) the optimal value of quality $q^*(\bar{\theta}, t) \geq \tilde{q}$, so group with $q^*(\theta, t) \geq \tilde{q}$ is nonempty, 3) there exists $\hat{\theta}(t)$, so that $U(\hat{\theta}, q^*(\hat{\theta}, t)) - U(\hat{\theta}, \tilde{q}) = (V - wp)t$, where $q^*(\tilde{\theta}, t) = \tilde{q}$. In our application, the reputation is observed under non-zero values of the reform parameter t, so the full utility function becomes:

$$\mathcal{U} = \pi + \theta B(q) + 1(q > \tilde{q})(V - wp)t \tag{8}$$

where $0 \leq \underline{\theta} \leq \theta \leq \bar{\theta}$ is the parameter of altruism with pdf $f(\theta)$, V is a constant, reflecting utility of high reputation, w is marginal disutility of monetary reward (coming from the Bénabou and Tirole (2006) model).

Compute total amount of quality:

$$Q = \int_{\underline{\theta}}^{\hat{\theta}(t)} q^*(\theta, t) f(\theta) d\theta + \int_{\hat{\theta}(t)}^{\tilde{\theta}(t)} \tilde{q} f(\theta) d\theta + \int_{\tilde{\theta}(t)}^{\check{\theta}(t)} q^*(\theta, t) f(\theta) d\theta + \int_{\check{\theta}(t)}^{\bar{\theta}} \bar{q}(t) f(\theta) d\theta \tag{9}$$

Here we use the fact that $q^* < \tilde{q}$ is chosen by hospitals with $\theta(t) \in [\underline{\theta}, \hat{\theta}(t))$, $q^* = \tilde{q}$ is

[18] The SOC holds automatically if $\theta \geq 0$, since $C_{qq} > 0$ and $B_{qq} \leq 0$.

[19] The third derivatives are zero, so terms $\frac{\partial q^*}{\partial \theta} \theta B_{qqq}$ and $-\frac{\partial q^*}{\partial \theta} C_{qqq}$ are not added to B_{qq} as factors of Ap in the numerator of (7).

provided by hospitals with $\theta(t) \in [\hat{\theta}(t), \tilde{\theta}(t)]$, while hospitals with $\theta(t) \in (\tilde{\theta}(t), \check{\theta}(t))$ have $q^* > \tilde{q}$ (Siciliani, 2009). The remaining hospitals have $\theta(t) > \check{\theta}(t)$ and choose $\bar{q}(t)$ owing to the participation constraint. (They would like to choose $q^*(\theta(t), t) > \bar{q}(t)$ but this leads to negative profits).

Using Leibnitz integral rule and differentiating Q in t we obtain:

$$
\begin{aligned}
\frac{dQ}{dt} = & \int_{\underline{\theta}}^{\hat{\theta}(t)} \frac{\partial q^*(\theta, t)}{\partial t} f(\theta) d\theta + q^*(\hat{\theta}(t), t) f(\hat{\theta}(t)) \frac{\partial \hat{\theta}(t)}{\partial t} \\
& + \tilde{q} f(\tilde{\theta}(t)) \frac{\partial \tilde{\theta}(t)}{\partial t} - \tilde{q} f(\hat{\theta}(t)) \frac{\partial \hat{\theta}(t)}{\partial t} \\
& + \int_{\tilde{\theta}(t)}^{\check{\theta}(t)} \frac{\partial q^*(\theta, t)}{\partial t} f(\theta) d\theta - q^*\big(\tilde{\theta}(t), t\big) f(\tilde{\theta}(t)) \frac{\partial \tilde{\theta}(t)}{\partial t} + q^*(\check{\theta}(t), t) f(\check{\theta}(t)) \frac{\partial \check{\theta}(t)}{\partial t} \\
& + \int_{\check{\theta}(t)}^{\bar{\theta}} \frac{\partial \bar{q}(t)}{\partial t} f(\theta) d\theta - \bar{q}(t) f(\check{\theta}(t)) \frac{\partial \check{\theta}(t)}{\partial t}
\end{aligned}
\tag{10}
$$

where each line corresponds to each summand in equation (9).

Terms $q^*(\tilde{\theta}(t), t)) f(\tilde{\theta}(t)) \dfrac{\partial \tilde{\theta}(t)}{\partial t}$ and $\tilde{q} f(\tilde{\theta}(t)) \dfrac{\partial \tilde{\theta}(t)}{\partial t}$ cancel out by definition of $\tilde{\theta}(t)$. Similarly, terms $q^*(\check{\theta}(t), t)) f(\check{\theta}(t)) \dfrac{\partial \check{\theta}(t)}{\partial t}$ and $\bar{q} f(\check{\theta}(t)) \dfrac{\partial \check{\theta}(t)}{\partial t}$ cancel out by definition of $\check{\theta}(t)$. So

$$
\begin{aligned}
\frac{dQ}{dt} = & \int_{\underline{\theta}}^{\hat{\theta}(t)} \frac{\partial q^*(\theta, t)}{\partial t} f(\theta) d\theta + \int_{\tilde{\theta}(t)}^{\check{\theta}(t)} \frac{\partial q^*(\theta, t)}{\partial t} f(\theta) d\theta + \int_{\check{\theta}(t)}^{\bar{\theta}} \frac{\partial \bar{q}(t)}{\partial t} f(\theta) d\theta \\
& + [q^*(\hat{\theta}(t), t) - \tilde{q}] f(\hat{\theta}(t)) \frac{\partial \hat{\theta}(t)}{\partial t}
\end{aligned}
\tag{11}
$$

Applying the implicit function theorem to

$$U(\hat{\theta}, q^*(\hat{\theta}, t)) - U(\hat{\theta}, \tilde{q}) - (V - wp)t = 0 \tag{12}$$

we get

$$\frac{\partial \hat{\theta}}{\partial t} = \frac{Ap(\tilde{q} - q^*(\hat{\theta}(t), t)) + (V - wp)}{B(q^*(\hat{\theta}(t), t)) - B(\tilde{q})} \tag{13}$$

Therefore, the change in quality simplifies to

$$\frac{dQ}{dt} = \int_{\underline{\theta}}^{\hat{\theta}(t)} \underbrace{\frac{\partial q^*(\theta,t)}{\partial t}}_{+} f(\theta)d\theta + \int_{\tilde{\theta}(t)}^{\check{\theta}(t)} \underbrace{\frac{\partial q^*(\theta,t)}{\partial t}}_{+} f(\theta)d\theta + \int_{\check{\theta}(t)}^{\bar{\theta}} \underbrace{\frac{\partial \bar{q}(t)}{\partial t}}_{-} f(\theta)d\theta$$
$$+ \left(\underbrace{\tilde{q} - q^*(\hat{\theta}(t),t)}_{+}\right) \underbrace{\frac{f(\hat{\theta}(t))}{B(\tilde{q}) - B(q^*(\hat{\theta}(t),t)}}_{+} \left[\underbrace{Ap\left(\tilde{q} - q^*(\hat{\theta}(t),t)\right)}_{+} + (\underbrace{V - wp}_{-})\right]$$

Result 3: Hospitals with $\theta \in [\underline{\theta}, \hat{\theta}(t)]$ and $\theta \in [\tilde{\theta}(t), \check{\theta}(t)]$ increase their quality and hospitals with $\theta \in (\check{\theta}(t), \bar{\theta}]$ decrease their quality. Additionally, there is a change in the number of hospitals on the interval $(\hat{\theta}(t), \tilde{\theta}(t))$, as its limit points depend on t. For instance, $\hat{\theta}(0) < \hat{\theta}(1)$, so some of the former hospitals with $\tilde{q}$ now provide $q^* < \tilde{q}$. At the same time, $\tilde{\theta}(0) > \tilde{\theta}(1)$ and some hospitals with $\tilde{q}$ now provide $q^* > \tilde{q}$. In other words, there is a certain convergence: quality goes up at the lowest tail of the distribution of altruism where $\theta \in [\underline{\theta}, \hat{\theta}(t')]$ and a part of the highest tail with $\theta \in [\tilde{\theta}(t'), \check{\theta}(t')]$. Quality may decrease for medium types of altruism, where $\theta \in [\hat{\theta}(t), \hat{\theta}(t')]$. Finally, owing to participation constraint, quality goes down for the highest types: $\theta \in [\tilde{\tilde{\theta}}(t'), \bar{\theta}]$.

Regarding the net aggregate effect, if $[Ap(\tilde{q} - q^*(\hat{\theta}(t),t)) + (V - wp)]$ is negative and rather large in absolute terms, the third and the forth summand in (14) may be larger than the first two positive summands. So the total effect of the incentives contract for all hospitals would be negative. Figure 1 below illustrates this result.

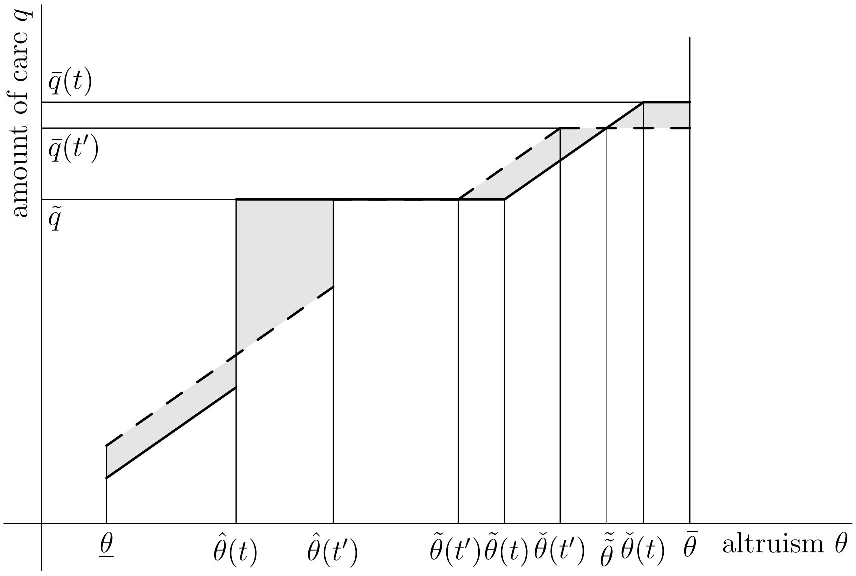

Figure 1: **Optimal amount of care under the changeover from policy t (solid line) to t' (dotted line)**

4.3 Empirical approach

4.3.1 Dynamic panel data model

The empirical analysis exploits dynamic panel data models for hospital quality measures with a binary treatment variable for the tournament-type incentives contracts. The use of the dynamic models is explained by the conjecture that a hospital strongly adheres to its practice patterns. Consequently, the value of each quality measure depends on its values in the previous periods.

The data is likely to exhibit mean reversion as there are strong peer effects leading to a dissemination of the quality-enhancing practices (Damberg et al., 2014). To separate the reform effect and the potential mean reversion, we distinguish the pre-treatment and post-treatment long-term means in the autocorrelation models. Furthermore, the long-term means are explicitly modeled as functions of a hospital's characteristics.

We extend the autocorrelation specification of Hamilton (1994), assuming there are two distinct long-term means in the pre-reform and post-reform periods, and allowing each long-term mean to be a function of hospital variables $\mathbf{x}$. Denote $\tilde{\mathbf{x}}_{it} = [\mathbf{x}_{it}1(t \geq t_0), \mathbf{x}_{it}1(t < t_0)]$, where t_0 is the year of incentivizing the given quality measure y through a contract and $\mathbf{x}$ is a vector of hospital control variables, which does not include a constant.

We study the dynamics of each quality measure separately and focus on the treatment effect of the reform for the groups of hospitals. In particular, the top decile and percentiles 95–100 approximate the groups of front-runner agents.

The analysis is based on the second-order[20] autocorrelation model:

$$\begin{aligned}
y_{it} - \tilde{\mathbf{x}}_t\boldsymbol{\beta} - \mu = {} & \alpha_1(y_{i,t-1} - \tilde{\mathbf{x}}_{i,t-1}\boldsymbol{\beta} - \mu) + \alpha_2(y_{i,t-1} - \tilde{\mathbf{x}}_{i,t-1}\boldsymbol{\beta} - \mu)r_{it} \\
& + \alpha_3(y_{i,t-2} - \tilde{\mathbf{x}}_{i,t-2}\boldsymbol{\beta} - \mu) + \alpha_4(y_{i,t-2} - \tilde{\mathbf{x}}_{i,t-2}\boldsymbol{\beta} - \mu)r_{it} \\
& + \tilde{\mathbf{x}}_{it}\gamma + \tilde{\mathbf{x}}_{i,t-1}r_{it}\delta + \theta r_{it} + \tilde{\mathbf{x}}_{i,t-2}r_{it}\lambda + \tilde{\mathbf{x}}_{i,t-2}\kappa \\
& + \nu_i + \epsilon_{it}
\end{aligned} \tag{14}$$

The dependent variable, y_{it}, is quality measure, r_{it} is the reform dummy which equals unity if hospital i participates in value-based purchasing in year t, ν_i are hospital fixed effects, ϵ_{it} are i.i.d. with zero mean. The terms in the third line are included for identification. Interaction terms $(y_{i,t-s} - \tilde{\mathbf{x}}_{i,t-s}\boldsymbol{\beta} - \mu)r_{it}$, $s = 1, 2$ capture the effect of the reform conditional on the pre-reform value of the dependent variable.

It should be noted that we separate mean reversion owing to the incentives contract per se (through variable r) and owing to the non-financial impact of incentives contract, such as

[20]According to the results of the Arellano-Bond test, the first order lag did not allow excluding serial correlation.

peer-effects, public reporting etc. (through time-variable t_0). When our sensitivity analysis relaxes the assumptions about two types of mean reversion, we clearly find regression-to-the-mean owing to t_0, or both t_0 and r.

For convenience, we collect terms in (14) and rewrite:

$$
\begin{aligned}
y_{it} &= c_0 + c_1 y_{i,t-1} + c_2 y_{i,t-1} r_{it} + c_3 r_{it} + \tilde{\mathbf{x}}_{it}\mathbf{c}_4 + \tilde{\mathbf{x}}_{i,t-1}\mathbf{c}_5 + \tilde{\mathbf{x}}_{i,t-1}\mathbf{c}_6 r_{it} \\
&\quad + c_7 y_{i,t-2} + c_8 y_{i,t-2} r_t + \tilde{\mathbf{x}}_{i,t-2}\mathbf{c}_9 + \tilde{\mathbf{x}}_{i,t-2}\mathbf{c}_{10} r_{it} + \nu_i + \epsilon_{it} \qquad (15)
\end{aligned}
$$

Equation (15) is estimated using Arellano and Bover (1995)/Blundell and Bond (1998) estimator, with robust variance-covariance matrix (Windmeijer, 2005). The reform and its interaction terms with $\tilde{\mathbf{x}}$ and $y_{i,t-s}$ are treated as predetermined variables, which means that lagged levels and lagged differences of y_{it}, r_{it}, $\tilde{\mathbf{x}}_{i,t-s} r_{it}$ and $y_{i,t-s} r_{it}$ ($s = 1, 2$) are used as instruments for the difference equation.

Equating coefficients in (14) and (15) we obtain: $\boldsymbol{\beta} = -\mathbf{c}_5/c_1$ and $\mu = c_0/(1 - c_1 - c_7)$.

4.3.2 Hypotheses

We assume that altruism is heterogeneous across hospitals and the values of altruism in each hospital are higher for quality measures which are strongly associated with patient's benefit. The data show that quality measures of the clinical process of care domain are not linked to patient's health outcomes, and so physicians would not have altruistic concerns for performance on these measures. On the other hand, patient experience of care measures would reveal more altruism. The effect of the incentives contract is analyzed across different percentile groups (deciles and percentiles 95–100). Indeed, the Medicare's pricing schedule stimulates hospitals above the median performance in a step-wise manner, so deciles in our approach approximate different levels of quality. The mean of top decile is regarded by the pricing schedule as the best performance, so we approximate hospital group with the highest quality as percentiles 95–100.

Our model demonstrates that a changeover to an incentives contract causes quality increase in the group of agents with high performance in absence of altruism. However, quality may not change or fall for subgroups of altruistic agents.

The average treatment effect of the reform (at group means) is estimated by setting $r = 1$ in (15):

$$
\tau = c_3 + c_2 \bar{y}_{t-1} + c_8 \bar{y}_{t-2} + \bar{\tilde{\mathbf{x}}}_{t-1}\mathbf{c}_6 + \bar{\tilde{\mathbf{x}}}_{t-2}\mathbf{c}_{10} \qquad (16)
$$

We conjecture that the average treatment effect at top percentile groups would be negative for quality measures that are strongly related to patient's benefit.

Hypothesis I: $\tau \geq 0$ for the measures of clinical process of care.

Hypothesis II: $\tau < 0$ for some measures of patient experience of care at the hospitals with the highest quality.

5 Data

5.1 Hospital quality measures

The data for quality measures and the reform participation come from the *Hospital Compare data archive* by the Centers for Medicare & Medicaid Services (October 2018 update with data for value-based purchasing in 2013–2017). Our analysis focuses on Medicare's acute-care hospitals, as the incentives contract applies exclusively to this subgroup. The dichotomous variable for reform participation equals unity in fiscal years from 2013 onwards if a hospital is listed as a value-based purchasing hospital in the corresponding year (Table 1). We regard the reform participation as a predetermined variable. It is not directly linked to hospital size, but is associated with hospital's desire to invest in the resources for data collection and data validation, along with the overall intent to focus on quality improvement (Centers for Medicare and Medicaid Services, 2007).

The unit for the time period in our analysis is a fiscal year. The data for most quality measures in *Hospital Compare* are reported on a 12-month basis. The values of patient experience of care (Hospital Consumer Assessment of Healthcare Providers and Systems, HCAHPS) measures are updated quarterly, adding the data for the most recent quarter and excluding the data for the earliest quarter. However, clinical process of care measures are only updated annually. Similarly, hospital control variables are reported on a fiscal year basis.

Medicare's acute care hospitals are incentivized to report the measures of clinical process of care since 2004 and patient experience of care since July 2007. Namely, eligible hospitals who have a sufficient number of cases, but do not submit the data for the corresponding clinical process of care or patient experience of care measure, suffer 0.4 percent or 2 percent financial losses, respectively.[21] Our analysis uses the data for the clinical process of care between 2004 and 2015 (measures are not collected in later years) and for patient experience of care between 2007 and 2017.

Measures in outcome of care/efficiency domain were incentivized only in 2014, so we lack post-reform data to fit second order autocorrelation model with interaction terms.

[21] Enactment of the Deficit Reduction Act of 2005, HCAHPS Fact sheet, 2012; section 501(b) of the Medicare Prescription Drug, Improvement and Modernization Act of 2003, Hospital Quality Initiative Overview, 2008).

5.2 Hospital control variables

Hospital characteristics are taken from the hospital files by *Hospital Compare*, which contain variables on hospital location and ownership. The number of hospital beds, share of Medicare's discharges, resident-to-bed ratio, ownership and the dichotomous variables for urban location come from Medicare's *Impact Files*. *Provider of Service Data* are exploited for the variables on the history of Medicare affiliation, number of Medicare certified beds, numbers of doctors, nurses and residents.

5.3 Patient control variables and risk adjustment

We take the casemix variable from the *Impact Files*, which supplement the *Final Rules* on Medicare's payments to each hospital for a given fiscal year. The variable reflects the relative weight of each DRG in financial terms and is adjusted for transfers of patients across hospitals.[22] Accordingly, it enables to control for the composition of patient cases in view of an objective link between the severity of illness and required hospital's resources.

The disproportionate share index, coming from the *Impact Files*, accounts for the share of Medicare and the low-income Medicaid patients. The use of the variable allows to proxy patients' income.

Other patient control variables are included in the original Medicare's risk-adjustment models as follows. The patient experience of care measures, reported by the *Hospital Compare*, are risk adjusted within a multi-step procedure: 1) at the hospital level with coefficients from the linear regression of the response on the major covariates, which account for patient-mix: education, self-assessed health, time after discharge, second language, age, type of care (maternity, surgical or medical), 2) rescaled by the overall national mean of patient-mix across all hospitals, and 3) adjusted for mode of the survey: mail, telephone, mail combined with telephone follow-up, and active interactive voice response, see Elliott et al. (2009).

Note that the clinical process of care measures relate exclusively to the clinical condition and therefore do not require risk-adjustment (Medicare.gov, 2017).

The lists of hospital and patient control variables are motivated by results of the Centers for Medicare and Medicaid Services (2007) interviews with hospitals and hospital associations, as well as the findings in the literature on the pilot programs (Damberg et al., 2014).

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

Table 1: **Descriptive statistics for Medicare's acute-care hospitals in 2004–2017**

Variable	Definition	Obs	Mean	St.Dev	Min	Max
Patient experience of care measures (2007–2017)						
Comp-1-*ap*	Nurses always communicated well	34880	76.574	6.279	16	100
Comp-2-*ap*	Doctors always communicated well	34880	80.081	5.295	23	100
Comp-3-*ap*	Patients always received help as soon as they wanted	34876	64.013	8.896	8	100
Comp-4-*ap*	Pain was always well controlled	31714	69.266	5.747	0	100
Comp-5-*ap*	Staff always gave explanation about medicines	34854	61.478	6.855	2	100
Comp-6-*yp*	Yes, staff did give patients discharge information	34868	83.664	5.362	27	100
Clean-hsp-ap	Room was always clean	34880	70.656	7.444	7	100
Quiet-hsp-ap	Hospital always quiet at night	32652	58.873	10.568	0	100
Hrecomddy	Patients who would definitely recommend the hospital	32652	69.762	10.156	2	100
Hsp-rating-910	Patients who gave hospital a rating of 9 or 10 (high)	34877	68.017	9.453	10	100
Clinical process of care measures (2004–2015)						
AMI-8*a*	Primary PCI received within 90 minutes of hospital arrival (AMI)	8163	66.186	24.594	0	100
HF-1	Discharge instructions (heart failure)	34285	77.075	25.148	0	100
PN-3*b*	Blood cultures performed in the emergency department prior to initial antibiotic received in hospital (pneumonia)	18857	88.219	9.955	0	100
PN-6	Initial antibiotic selection for CAP in immunocompetent patient (pneumonia)	35019	88.203	11.780	0	100
*SCIP-Card*2	Surgery patients on beta-blocker therapy prior to arrival who received a beta-blocker during the perioperative period	30993	90.459	15.108	0	100
*SCIP-Inf*1	Prophylactic antibiotic received within 1 hour prior to surgical incision	25739	95.718	8.900	0	100
*SCIP-Inf*2	Prophylactic antibiotic selection for surgical patients	30882	87.838	16.424	0	100
*SCIP-Inf*3	Prophylactic antibiotics discontinued within 24 hours after surgery end time	7471	91.376	14.712	0	100
*SCIP-Inf*4	Cardiac surgery patients with controlled 6 A.M. postoperative blood glucose	25854	90.338	14.885	0	100
*SCIP-VTE*2	Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery	18408	93.313	12.082	0	100
Reform dummy (2004–2017)						
r	=1 in 2013 onwards if was a value-based purchasing hospital in the corresponding fiscal year	48523	0.419	0.493	0	1
Hospital characteristics (2004–2017)						
public	=1 if managed by federal, state or local government, or hospital district or authority	48523	0.142	0.349	0	1
emergency	=1 if emergency hospital	48523	0.932	0.252	0	1
urban	=1 if urban hospital	48523	0.735	0.441	0	1
resbed	=1 resident-to-bed ratio	46205	0.060	0.154	0	1.996
beds	Number of beds	40769	184.814	171.444	1	1928
medicare_share	Share of Medicare cases	44956	0.455	0.156	0.001	1
Patient characteristics (2004–2017)						
casemix	Transfer-adjusted casemix index	46205	1.416	0.321	0.356	4.363
dsh	Disproportionate share index, reflecting the prevalence of low-income patients	46205	0.271	0.174	0	1.492

Notes: Clean-hsp-ap and Quiet-hsp-ap albeit measured separately, are regarded as one measure "Cleanliness and quietness of hospital environment" in The Final Rule. Hrecomddy is not listed in The Final Rule, yet, we analyze it since it relates to overall rating of hospital. Section 401 hospitals are treated as rural hospitals.

6 Results

6.1 Identification

We focus on all measures of patient experience of care: nine incentivized measures (cleanliness and quietness of hospital environment are studied separately) and a measure on recommending hospital *Hrecomdd*.

The analysis uses nine out of twelve measures of clinical process of care. The measure *SCIP-Inf*4 (Cardiac surgery patients with controlled 6 A.M. postoperative blood glucose) was incentivized only in 2014. Accordingly, they have only two post-reform periods, which is sufficient only for identification of the model with the first order autocorrelation.[23]

The data for measure *HF*-1 (Discharge instructions (heart failure)) is available only till 2014, so the order two model is not identifiable for it as well. Hence we use the order one model for this measure similarly to *SCIP-Inf*4.

We do not study the dynamics of two clinical process of care measures: *AMI*-7*a* (Fibrinolytic therapy received within 30 minutes of hospital arrival), owing to availability of post-2012 data only for a few hospitals, and *SCIP-VTE*1 (Surgery patients with recommended venous thromboembolism prophylaxis ordered), for which the data collection was discontinued in 2013.

The results are presented for seven out of ten patient experience of care measures and for six out of nine clinical process of care measures because the Arellano and Bond (1991) test rejects the hypothesis about the absence of order two serial correlation in the first differenced errors for other three patient experience of care measures and for other three clinical process of care measures. Owing to unavailability of longer time-series for post-reform data, we cannot estimate higher order lags and limit our analysis to models with the above-mentioned 13 measures. The stationary conditions for the AR(2) process ($c_1 + c_7 < 1$, $c_7 - c_1 < 1$ and $|c_7| < 1$) are satisfied for each of these 13 models. The results of the estimation and coefficients for lagged dependent variables and other explanatory variables are presented in tables 6 and 7 in Appendix C.

6.2 Effect of altruism

Concerning patient experience of care, value-based purchasing has a negative effect for a few measures.

The reform decreases quality for front-runner hospitals: percentiles 95–100 of measures on nurse communication (*Comp*-1-*ap*), the ability for patients to receive help quickly

[23] The Arellano and Bond (1991) test show no order two serial correlation in the first differenced errors, so the order one model is applicable for *SCIP-Inf*4.

(*Comp*-3-*ap*) and discharge instructions (*Comp*-6-*yp*). Quality measures associated with clean room (*Clean-hsp-ap*) and overall rating of hospital (*Hsp-rating*-910) do not decrease across hospitals in top percentiles (Table 2).[24]

None of clinical process of care measures deteriorates across hospitals in top percentiles and any other percentile groups. The measures on prophylactic antibiotics, selection for surgical patients (*SCIP-Inf*2) and on surgery patients on beta-blocker therapy prior to arrival who received a beta-blocker during the perioperative period (*SCIP-Card*2) increase in all percentiles owing to the reform, and the absolute value of the effect is higher in the lowest percentiles 0–10 and the highest percentiles 90–100 and 95–100.

Accordingly, we may not reject *Hypothesis* I or *Hypothesis* II, and conjecture that social preferences are revealed in the behavior of hospitals with the highest quality.

6.3 Internal validity

The analysis exploits longitudinal data, which may lack observations in certain years. Overall, the panels are unbalanced but 85–93% of hospitals would have observations in each year. As robustness check, we conducted analysis with balanced panels and discovered similar distribution of the dependent variables and negligible difference in the values for the coefficients for the explanatory variables.

To assess internal validity of the estimates in terms of financial dependence on Medicare's beneficiaries, we use the subsample of hospitals with a share of Medicare revenues above 50% and observe higher values for the adverse effects for patient experience of care measures.

The major limitation of our analyses is the lack of patient-level data. Nonetheless, the quality measures exploited in the empirical part of the paper are risk-adjusted according to major patient control variables, for instance age, education and co-morbidities. Yet, our estimates do not fully control for individual socio-demographic characteristics, which may influence treatment patterns.

[24]The only measure, where the effect is positively significant – "definitely recommend this hospital, *Hrecomddy*" – is not financially stimulated.

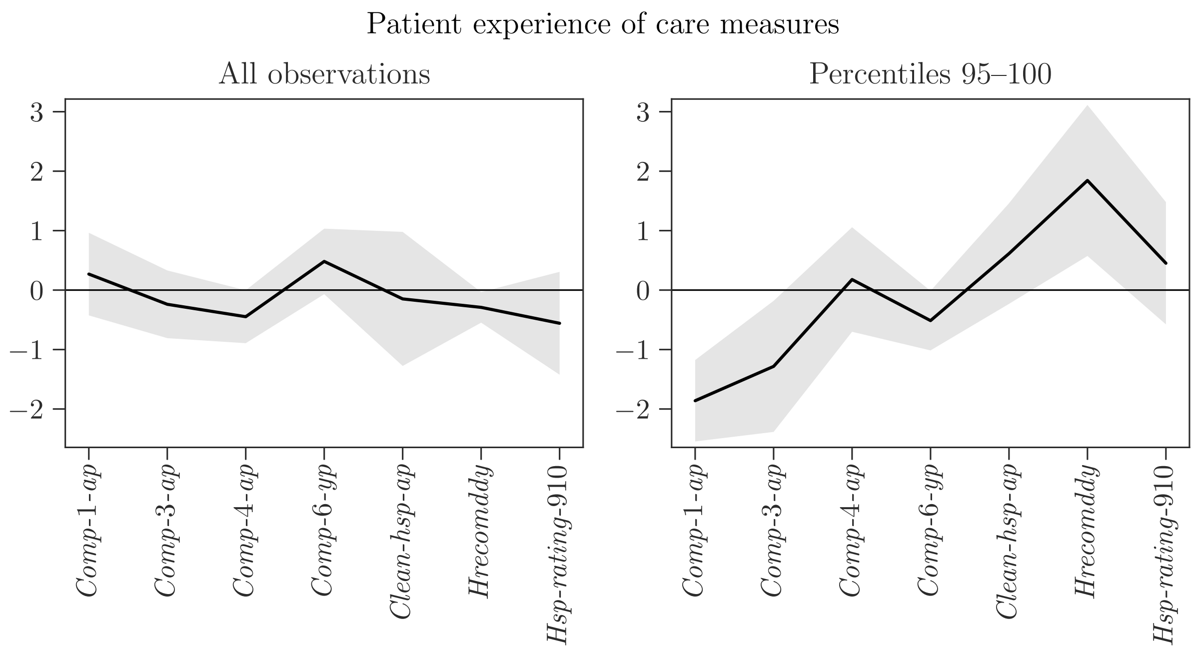

Figure 2: The average treatment effect of the reform for quality measures with the 95% confidence intervals

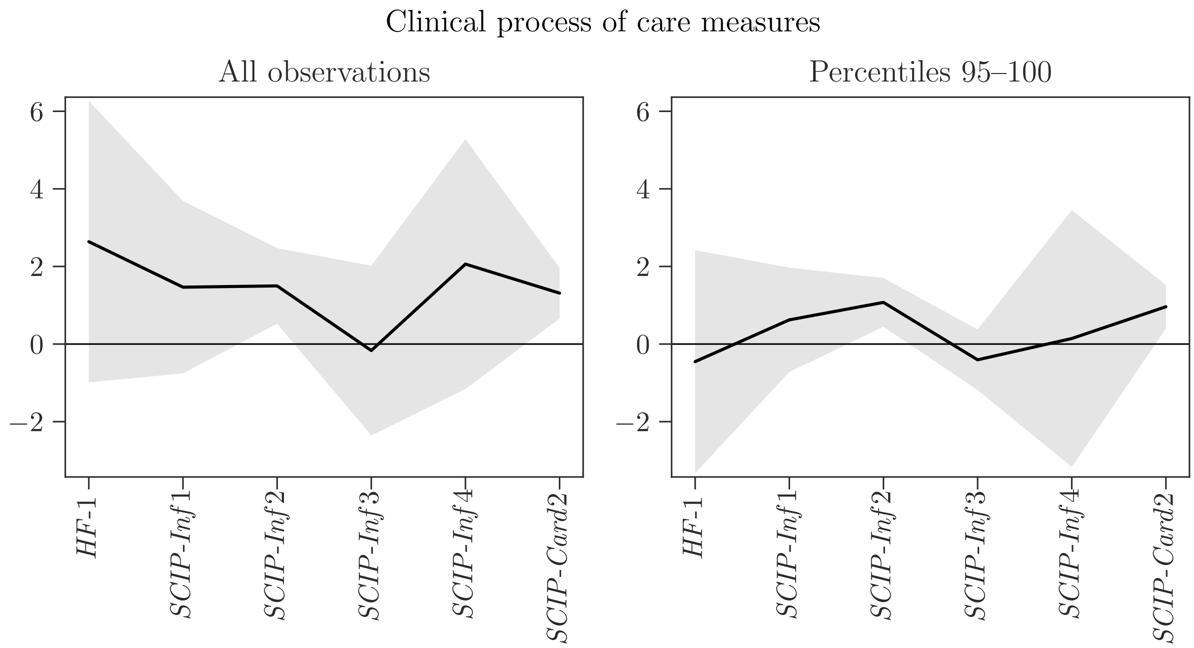

Figure 3: The average treatment effect of the reform for quality measures with the 95% confidence intervals

Table 2: **Average treatment effect of incentives contract in percentile groups for each quality measure in dynamic panel data regression (Patient experience of care)**

	Comp-1-*ap*	*Comp*-3-*ap*	*Comp*-4-*ap*	*Comp*-6-*yp*	*Clean-hsp-ap*	*Hrecomddy*	*Hsp-rating*-910
τ_{0-10}	2.064***	0.510	−0.839	1.506***	−0.718	−2.076***	−1.307**
	(0.558)	(0.644)	(0.526)	(0.413)	(0.940)	(0.583)	(0.596)
τ_{10-20}	1.199***	0.205	−0.654*	1.008***	−0.458	−1.338***	−0.957*
	(0.468)	(0.505)	(0.362)	(0.361)	(0.842)	(0.358)	(0.503)
τ_{20-30}	0.854*	0.061	−0.582*	0.775***	−0.366	−0.974***	−0.799
	(0.453)	(0.430)	(0.308)	(0.329)	(0.769)	(0.261)	(0.484)
τ_{30-40}	0.568	−0.065	−0.547**	0.636**	−0.255	−0.706***	−0.693
	(0.429)	(0.373)	(0.279)	(0.320)	(0.721)	(0.189)	(0.478)
τ_{40-50}	0.365	−0.175	−0.491*	0.499	−0.176	−0.427***	−0.602
	(0.403)	(0.332)	(0.254)	(0.320)	(0.679)	(0.141)	(0.469)
τ_{50-60}	0.155	−0.271	−0.430*	0.373	−0.110	−0.137	−0.506
	(0.378)	(0.286)	(0.221)	(0.301)	(0.596)	(0.133)	(0.494)
τ_{60-70}	−0.052	−0.371	−0.432**	0.253	−0.055	0.119	−0.418
	(0.364)	(0.262)	(0.219)	(0.299)	(0.524)	(0.168)	(0.490)
τ_{70-80}	−0.248	−0.492**	−0.355*	0.126	0.035	0.438*	−0.303
	(0.343)	(0.233)	(0.199)	(0.279)	(0.450)	(0.236)	(0.511)
τ_{80-90}	−0.561*	−0.668***	−0.255	−0.018	0.165	0.792***	−0.170
	(0.299)	(0.240)	(0.207)	(0.267)	(0.382)	(0.333)	(0.506)
τ_{90-100}	−1.431***	−1.070***	0.032	−0.364	0.462	1.522***	0.216
	(0.297)	(0.431)	(0.342)	(0.252)	(0.380)	(0.546)	(0.475)
τ_{95-100}	−1.861***	−1.283**	0.178	−0.513**	0.614	1.843***	0.452
	(0.349)	(0.562)	(0.448)	(0.255)	(0.431)	(0.648)	(0.525)
All observations	0.269	−0.239	−0.448**	0.482*	−0.148	−0.291**	−0.558
	(0.354)	(0.290)	(0.227)	(0.281)	(0.575)	(0.131)	(0.441)

Notes: Robust standard errors in parentheses. The deciles of hospitals approximate different levels of quality with respect to each measure. Mean of top decile is regarded by the Medicare's pricing schedule as the best performance, so we approximate hospital group with the highest quality as percentiles 95–100.

Table 3: **Average treatment effect of incentives contract in percentile groups for each quality measure in dynamic panel data regression (Clinical process of care)**

	HF-1	*SCIP-Inf*1	*SCIP-Inf*2	*SCIP-Inf*3	*SCIP-Inf*4	*SCIP-Card*2
$\tau_{0\text{–}10}$	8.016***	4.868	5.260**	0.172	7.849	2.864**
	(2.721)	(3.519)	(2.339)	(2.981)	(4.876)	(1.384)
$\tau_{10\text{–}20}$	5.037**	2.560	2.174***	−0.116	3.339	1.782***
	(2.240)	(1.873)	(0.816)	(1.853)	(2.766)	(0.614)
$\tau_{20\text{–}30}$	3.952*	1.892	1.535***	−0.095	2.559	1.465***
	(2.208)	(1.419)	(0.507)	(1.524)	(2.532)	(0.418)
$\tau_{30\text{–}40}$	3.081	1.547	1.230***	−0.170	1.726	1.273***
	(2.122)	(1.169)	(0.359)	(1.270)	(2.320)	(0.356)
$\tau_{40\text{–}50}$	2.445	1.252	1.049***	−0.243	1.325	1.141***
	(2.063)	(0.979)	(0.275)	(1.148)	(2.087)	(0.345)
$\tau_{50\text{–}60}$	1.932	1.007	0.883***	−0.254	0.938	1.074***
	(2.045)	(0.829)	(0.207)	(0.958)	(2.097)	(0.347)
$\tau_{60\text{–}70}$	1.320	0.812	0.757***	−0.291	0.440	0.996***
	(1.933)	(0.679)	(0.174)	(0.848)	(2.108)	(0.368)
$\tau_{70\text{–}80}$	0.851	0.625	0.663***	−0.348	0.198	0.885***
	(1.881)	(0.580)	(0.149)	(0.717)	(2.061)	(0.367)
$\tau_{80\text{–}90}$	0.243	0.493	0.689***	−0.420	−0.105	0.891***
	(1.708)	(0.518)	(0.158)	(0.581)	(2.083)	(0.279)
$\tau_{90\text{–}100}$	−0.309	0.515	1.009***	−0.462	0.128	0.941***
	(1.539)	(0.600)	(0.288)	(0.431)	(1.651)	(0.284)
$\tau_{95\text{–}100}$	−0.453	0.624	1.074***	−0.406	0.142	0.962***
	(1.466)	(0.686)	(0.320)	(0.398)	(1.688)	(0.286)
All observations	2.638	1.466	1.499***	−0.167	2.061	1.312***
	(1.850)	(1.134)	(0.496)	(1.116)	(1.643)	(0.332)

Notes: Robust standard errors in parentheses. The deciles of hospitals approximate different levels of quality with respect to each measure. Mean of top decile is regarded by the Medicare's pricing schedule as the best performance, so we approximate hospital group with the highest quality as percentiles 95–100.

7 Discussion

The knowledge about altruism may be exploited by a social planner for extracting rents/ providing subsidies based on an altruistic type. However, identification of altruism requires careful setup and consideration of various factors related to potential noise. Laboratory experiments in healthcare applications control for both physician and patient effects, and compare physician's desire to sacrifice profits for an increase in a monetized equivalent of patient's benefit (Hennig-Schmidt et al., 2011; Ellis and McGuire, 1986). Namely, lab subjects act as physicians and choose among potential healthcare services, given prices and costs (and hence, physician's profit) and benefits to a patient with a given severity (Godager et al., 2015; Godager and Wiesen, 2013; Hennig-Schmidt et al., 2011). Concerning empirical work, we are not aware of any literature which would explicitly quantify heterogeneity in agents' altruism. The common approach would be a revelation of a mean relative weight of the altruistic component in the physicians' utility function, which may be accomplished through field experiments or analysis of prescription records (Galizzi et al., 2015). An alternative approach might focus on a potential correlation of altruism with observable quality level, for instance through agents' response to various payment schemes (Kristensen et al., 2016).

It should be noted that altruism is likely to be a function of agents' and clients' characteristics (e.g. may be linked to reciprocity towards clients in greater economic or health need, or of certain socio-ethnic background) and may have a profession-specific variation over time (Buurman and Dur, 2012).[25] Therefore, extending the theoretical approach to a dynamic relative performance contract and incorporating time-dependent altruism becomes one of several possible directions of future work.

The importance of focusing on a dynamic setting may be also justified by the results on relative performance contracts in the theoretical literature, for instance with regard to loss of motivation due to abilities gap, which exacerbates over time (Ederer, 2010; Casas-Arce and Martínez-Jerez, 2009; Gershkov and Perry, 2009; Rosen, 1986). Indeed, agents bid on their current performance given their earlier rank (Prendergast, 1999). Here, the information revelation principle of Milgrom and Weber (1982) does not necessarily hold in multi-period setups, and efficiency of interim public reporting for agents with heterogeneous abilities may depend on the functional form of their disutility in effort (Aoyagi, 2010). Other examples of unintended effects of dynamic contracts on relative performance can be found in sports elimination contests: the replacement of a team's best-performing players by substitutes at the end of the game when the team is evidently winning or losing (Prendergast, 1999; Gibbs, 1991) and negative incentive effects for favorites in tennis tournaments (Sunde, 2009).

[25] In particular, while street-bureaucrats and public sector employees may get more cynical and less altruistic with longer tenure, physicians might become more altruistic with experience and medical expertise (Roland and Dudley, 2015; Godager et al., 2015; Buurman et al., 2012; Buurman and Dur, 2012).

Finally, altruism towards the social value of a product in our approach captures only certain aspects of the "other-regarding behavior" and may be interpreted as heterogeneous personal norms associated with an action which is desirable by a principal (Fischer and Huddart, 2008; Hoffman et al., 1994). However, social norms that introduce behavioral and physiological effects to pure economic motives of incentives contracts are much broader. Examples include intrinsic, ethical, cooperative, reciprocal and other social issues (Arce, 2013; Bolton et al., 2008).

8 Conclusion

Incentives contracts cause unintended effects for most capable agents: their performance deteriorates owing to intrinsic behavior, conformism and slacking efforts. Altruistic agents, however, would be interested in the social value of their performance per se. There is limited theoretical literature on social preferences in public good games and piece-rate incentives contracts, but little is known about the effect of altruism on the outcomes of relative performance reimbursement. Yet, incentives contracts based on relative performance become increasingly widespread in public industries, where the number of agents is large and the distribution of their outcomes is precisely unknown.

This paper has demonstrated differential behavioral response to incentives regulations in healthcare when a fixed-price contract on quantity is combined with reimbursement based on relative quality performance. We propose a theoretical model, forecasting the adverse effects for hospitals with highest quality, where motivation and altruism towards the social value of performance cause quality deterioration.

The predictions of the model are tested in the empirical part of the paper, which uses Medicare's hospital level administrative panel data on a recent changeover to a relative quality performance incentive contract for acute care hospitals. The analysis exploits the dynamic panel data approach with habit-formation and self-selection. The models account for potential mean reversion due to the knowledge about the reform and the actual participation in the reform.

We focus on hospitals with top performance and show that quality dimensions, which may be linked to patient's benefit, demonstrate higher deterioration than other incentivized dimensions. In other words, relative performance contracts may be associated with motivation crowding out owing to the existence of altruism on the healthcare market.

The unintended effects of altruism may be adjusted by the social planner through designing a revelation mechanism and subsidizing the altruistic types.

References

Andreoni, J. (1989). Giving with impure altruism: applications to charity and Ricardian equivalence. *Journal of Political Economy*, 97(6):1447–1458.

Aoyagi, M. (2010). Information feedback in a dynamic tournament. *Games and Economic Behavior*, 70(2):242–260.

Arce, D. G. (2013). Principals' preferences for agents with social preferences. *Journal of Economic Behavior & Organization*, 90:154–163.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297.

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

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5):941–973.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614.

Bardey, D. and Lesur, R. (2006). Optimal regulation of health system with induced demand and ex post moral hazard. *Annales d'Économie et de Statistique*, pages 279–293.

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

Barros, P. P. (2003). Cream-skimming, incentives for efficiency and payment system. *Journal of Health Economics*, 22(3):419–443.

Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature*, 14(3):817–826.

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

Besstremyannaya, G. (2016). Differential effects of declining rates in a per diem payment system. *Health Economics*, 25(12):1599–1618.

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

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

Bolton, G. E., Brandts, J., Katok, E., Ockenfels, A., Zwick, R., et al. (2008). Testing theories of other-regarding behavior: A sequence of four laboratory studies. *Handbook of Experimental Economics Results*, 1(C):488.

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

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2015a). Supplementary data. http://www.sciencedirect.com/science/article/pii/S0167268115001158.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2015b). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization*. http://dx.doi.org/10.1016/j.jebo.2015.04.011.

Bull, C., Schotter, A., and Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1):1–33.

Buurman, M., Delfgaauw, J., Dur, R., and Van den Bossche, S. (2012). Public sector employees: Risk averse and altruistic? *Journal of Economic Behavior & Organization*, 83(3):279–291.

Buurman, M. and Dur, R. (2012). Incentives and the sorting of altruistic agents into street-level bureaucracies. *Scandinavian Journal of Economics*, 114(4):1318–1345.

Campbell, S. M., Reeves, D., Kontopantelis, E., Sibbald, B., and Roland, M. (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 Martínez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, 55(8):1306–1320.

Centers for Medicare and Medicaid Services (2007). Report to Congress: Plan to implement a Medicare hospital value-based purchasing program. *Washington, DC: US Department of Health and Human Services, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/downloads/hospitalVBPplanRTCFinalsubmitted2007.pdf*.

Chalkley, M. and Khalil, F. (2005). Third party purchasing of health services: Patient choice and agency. *Journal of Health Economics*, 24(6):1132–1153.

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

Chalkley, M. and Malcomson, J. M. (2000). Government purchasing of health services. *Handbook of Health Economics*, 1:847–890.

Chalkley, M. and Malcomson, J. M. (2002). Cost sharing in health service provision: An empirical assessment of cost savings. *Journal of Public Economics*, 84(2):219–249.

Choné, P. and Ma, C.-t. A. (2011). Optimal health care contract under physician agency. *Annales d'Économie et de Statistique*, 101/102:229–256.

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

Damberg, C. L., Raube, K., Williams, T., and Shortell, S. M. (2005). Paying for performance: implementing a statewide project in California. *Quality Management in Healthcare*, 14(2):66–79.

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

De Fraja, G. (2000). Contracts for health care and asymmetric information. *Journal of Health Economics*, 19(5):663–677.

Devlin, R. A. and Sarma, S. (2008). Do physician remuneration schemes matter? The case of Canadian family physicians. *Journal of Health Economics*, 27(5):1168–1181.

Dixit, A. (1987). Strategic behavior in contests. *American Economic Review*, 77(5):891–898.

Doran, T., Fullwood, C., Kontopantelis, E., and Reeves, D. (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.

Dranove, D. and White, W. D. (1994). Recent theory and evidence on competition in hospital markets. *Journal of Economics & Management Strategy*, 3(1):169–209.

Dumont, E., Fortin, B., Jacquemet, N., and Shearer, B. (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics*, 27(6):1436–1450.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*, 19(3):733–769.

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

Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Beckett, M. K., and Giordano, L. (2009). Effects of survey mode, patient mix, and nonresponse on cahps hospital survey scores. *Health Services Research*, 44(2p1):501–518.

Ellis, R. and McGuire, T. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5(2):129–151.

Ellis, R. P. and McGuire, T. G. (1990). Optimal payment systems for health services. *Journal of Health Economics*, 9(4):375–396.

Emanuel, E. J. and Emanuel, L. L. (1992). Four models of the physician-patient relationship. *Journal of American Medical Association*, 267(16):2221–2226.

Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688.

Eriksson, T. and Villeval, M. C. (2008). Performance-pay, sorting and social motivation. *Journal of Economic Behavior & Organization*, 68(2):412–421.

Fischer, P. and Huddart, S. (2008). Optimal contracting with endogenous social norms. *American Economic Review*, 98(4):1459–1475.

Galizzi, M., Tammi, T., Godager, G., Linnosmaa, I., and Wiesen, D. (2015). Provider altruism in health economics. Tampere, Finland.

Gershkov, A. and Perry, M. (2009). Tournaments with midterm reviews. *Games and Economic Behavior*, 66(1):162–190.

Gibbs, M. J. (1991). An economic approach to process in pay and performance appraisals. Manuscript, available at http://faculty.chicagobooth.edu/michael.gibbs/research/papers/Process.pdf.

Godager, G., Hennig-Schmidt, H., and Iversen, T. (2015). Does performance disclosure influence physicians' medical decisions? An experimental study. *Journal of Economic Behavior & Organization*.

Godager, G. and Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32(6):1105–1116.

Grabowski, D., Afendulis, C., and McGuire, T. (2011). Medicare prospective payment system and the volume and intensity of skilled nursing facility services. *Journal of Health Economics*, 30:675–684.

Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization*, 106:367–378.

Green, J. and Stokey, N. (1983). A comparison of tournaments and contracts. *Journal of Political Economy*, 91(3):349–364.

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

Gürtler, O. and Kräkel, M. (2010). Optimal tournament contracts for heterogeneous workers. *Journal of Economic Behavior & Organization*, 75(2):180–191.

Hamilton, J. (1994). *Time Series Analysis*, volume 2. Princeton University Press, Princeton.

HCAHPS online (2013). HCAHPS fact sheet (HCAHPS Hospital Survey) August 2013. http://www.hcahpsonline.org/files/August_2013_HCAHPS_Fact_Sheet3.pdf.

Hennig-Schmidt, H., Selten, R., and Wiesen, D. (2011). How payment systems affect physicians' provision behaviour – an experimental investigation. *Journal of Health Economics*, 30(4):637–646.

Hennig-Schmidt, H. and Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science & Medicine*, 108:156–165.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3):346–380.

Hölmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, 13(2):324–340.

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

Hotelling, H. (1929). Stability in competition. *Economic Journal*, 39(153):41–57.

Houle, S. K., McAlister, F. A., Jackevicius, C. A., Chuck, A. W., and Tsuyuki, R. T. (2012). Does performance-based remuneration for individual health care practitioners affect patient care? A systematic review. *Annals of Internal Medicine*, 157(12):889–899.

Jack, W. (2005). Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, 24(1):73–93.

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

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

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

Kairies-Schwarz, N. (2014). Altruism heterogeneity and quality competition among healthcare providers. RUHR Economic Papers, available at http://www.cinch-essen.de/files/REP_14_507.pdf.

Keser, C., Peterle, E., and Schnitzler, C. (2014). Money talks – paying physicians for performance. Center for European, Governance and Economic Development Research, Discussion Paper 173, available at http://wwwuser.gwdg.de/~cege/Diskussionspapiere/DP173.pdf.

Kessel, R. (1958). Price discrimination in medicine. *Journal of Law and Economics*, 1(Oct):20–53.

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

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

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

Laffont, J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation*. MIT Press.

Laffont, J.-J. and Tirole, J. (1986). Using cost observation to regulate firms. *Journal of Political Economy*, pages 614–641.

Laffont, J.-J. and Tirole, J. (1990). The regulation of multiproduct firms: Part I: Theory. *Journal of Public Economics*, 43(1):1–36.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864.

Lee, C. (1995). Optimal medical treatment under asymmetric information. *Journal of Health Economics*, 14(4):419–441.

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

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

Liu, T. and Ma, C.-t. A. (2013). Health insurance, treatment plan, and delegation to altruistic physician. *Journal of Economic Behavior & Organization*, 85:79–96.

Lynch, J. G. (2005). The effort effects of prizes in the second half of tournaments. *Journal of Economic Behavior & Organization*, 57(1):115–129.

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 McGuire, T. G. (1997). Optimal health insurance and provider payment. *American Economic Review*, 87(4):685–704.

Makris, M. (2009). Incentives for motivated agents under an administrative constraint. *Journal of Economic Behavior & Organization*, 71(2):428–440.

Makris, M. and Siciliani, L. (2013). Optimal incentive schemes for altruistic providers. *Journal of Public Economic Theory*, 15(5):675–699.

Makris, M. and Siciliani, L. (2014). Should incentive schemes be high-powered or low-powered in the presence of motivated agents? http://www.bris.ac.uk/media-library/sites/cmpo/migrated/documents/luigisiciliani.pdf.

Mannion, R., Marini, G., and Street, A. (2008). Implementing payment by results in the English NHS: Changing incentives and the role of information. *Journal of Health Organization and Management*, 22(1):79–88.

McGuire, T. G. (2000). Physician agency. *Handbook of Health Economics*, 1:461–536.

Medicare.gov (2017). Timely & effective care measures. https://www.medicare.gov/hospitalcompare/Data/Measures.html.

Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, 50(5):1089–1122.

Miraldo, M., Siciliani, L., and Street, A. (2011). Price adjustment in the hospital sector. *Journal of Health Economics*, 30(1):112–125.

Moreno-Serra, R. and Wagstaff, A. (2010). System-wide impacts of hospital payment reforms: evidence from Central and Eastern Europe and Central Asia. *Journal of Health Economics*, 29(4):585–602.

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

National Health Service (2015). Advancing Quality Aliance. http://www.advancingqualitynw.nhs.uk/.

Nawata, K. and Kawabuchi, K. (2013). Evaluation of the dpc-based inclusive payment system in japan for cataract operations by a new model. *Mathematics and Computers in Simulation*, 93:76–85.

Newhouse, J. P. (1970). Toward a theory of nonprofit institutions: An economic model of a hospital. *American Economic Review*, pages 64–74.

O'Keeffe, M., Viscusi, W. K., and Zeckhauser, R. J. (1984). Economic contests: Comparative reward schemes. *Journal of Labor Economics*, 100(3):27–56.

Pearson, S. D., Schneider, E. C., Kleinman, K. P., Coltin, K. L., and Singer, J. A. (2008). The impact of pay-for-performance on health care quality in Massachusetts, 2001–2003. *Health Affairs*, 27(4):1167–1176.

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.

Roland, M. and Dudley, R. A. (2015). How financial and reputational incentives can be used to improve medical care. *Health Services Research*, 50(S2):2090–2115.

Rosen, S. (1986). Prizes and incentives in elimination tournaments. *American Economic Review*, 76(4):701–715.

Rosenthal, M. B. (2008). Beyond pay for performance–emerging models of provider-payment reform. *New England Journal of Medicine*, 359(12):1197–1200.

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

Rosenthal, M. B., Frank, R. G., Li, Z., and Epstein, A. M. (2005). Early experience with pay-for-performance: from concept to practice. *JAMA*, 294(14):1788–1793.

Ryan, A. M. and Blustein, J. (2011). The effect of the MassHealth hospital pay-for-performance program on quality. *Health Services Research*, 46(3):712–728.

Ryan, A. M., Blustein, J., and Casalino, L. P. (2012). 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.

Sarma, S., Devlin, R. A., and Hogg, W. (2010). Physician's production of primary care in Ontario, Canada. *Health Economics*, 19(1):14–30.

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

Siciliani, L. (2006). Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics*, 25(3):479–499.

Siciliani, L. (2007). Paying for performance with altruistic or motivated providers.

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

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

Sunde, U. (2009). Heterogeneity and performance in tournaments: a test for incentive effects using professional tennis data. *Applied Economics*, 41(25):3199–3208.

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

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

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 On the existence of altruism

A.1 Experimental literature

Laboratory experiments use a straightforward application of the agency test, comparing physician's desire to sacrifice profits for an increase in a monetized equivalent of patient's benefit. For example, subjects choose among potential healthcare services within a *fixed-price* contract, given prices and costs (and hence, physician's profit) and benefits to patient (Godager and Wiesen, 2013). Subjects learn that benefits are converted into local currency and transferred to an actual charitable institution to support real patients (Godager et al., 2015; Godager and Wiesen, 2013; Hennig-Schmidt et al., 2011). Interestingly, disclosure of the subjects' choices to their peers increases the prevalence of altruistic behavior, but has no influence on the probability of the fully selfish choices (Godager et al., 2015).

Alternatively, healthcare services and resulting benefits to patients may be contrasted across *different* contracts: i.e. cost-based and fixed budget (per capita), as in (Hennig-Schmidt et al., 2011). The results of experiment, where subjects were medical students, proved the Ellis and McGuire (1986) assumption about the excess supply under cost-based contracts[26] and insufficient provision under the fixed-price contracts. Additionally, the experiment showed heterogeneity in patient's benefit, which depends not only on the payment mechanism but also on patient's type (state of health).

An extension of the same experiment contrasts the altruism of medical and non-medical[27] students (Hennig-Schmidt and Wiesen, 2014) and reveals that medical students are more altruistic under each of the analyzed contracts. Verbal responses from all but one medical student in this experiment showed concern about patient's benefit, while non-medical students generally cared about their own profits.

A further extension of the experiment compares altruism of physicians, medical and non-medical students (Brosig-Koch et al., 2015b) and finds that physicians appear to be more altruistic than medical students under each contract. An explanation may be linked to an increase of altruism in the course of actual practice and medical expertise (Roland and Dudley, 2015). The experiment assumes that benefit is a concave function of quantity as in Ellis and McGuire (1986) and an increase in patient's severity shifts the benefit function to the right – the same quantity leads to lower benefit for a more severe patient on the zero to optimal quantity interval and to higher benefit on the optimal quantity to infinity interval (Brosig-Koch et al., 2015a).

[26] The finding is similar to the Gruber and Owings (1996) conclusions about cesarean section deliveries in the U.S. in 1970–1982.

[27] Almost half of non-medical students come from the faculty of economics

Green (2014) conducts a laboratory experiment with a cost-based, fixed price and incentives (pay-for-performance) contracts and an agency framework: the subjects are physicians and patients. Physicians can offer harmful and beneficial services to patients. The results demonstrated oversupply of beneficial services under cost-based contracts if compared to fixed price/salaried contracts. Additionally, there is a certain crowding-out effect owing to the incentives contracts, supplementing cost-based or fixed price pay. The experimental study by Keser et al. (2014), however, does not discover adverse selection owing to incentives contracts combined with cost-based remuneration. The difference in results may relate to varying setups: while Keser et al. (2014) use the Hennig-Schmidt et al. (2011) experiment design with issues specific to healthcare and benefits for real patients, Keser et al. (2014) exploits a usual motivation scheme: he defines healthcare treatment as physicians' correction of mistakes in the course of proofreading patients' essays, where correct or wrong edits are interpreted as beneficial or harmful care (i.e. leading to monetary gains or losses by subjects, who represented patients).

A.2 Empirical literature

Empirical literature commonly focuses on physicians' responses to different contracts, with most popular analysis of (the difference in) output under the cost-based and/or fixed-price/pay-for-performance schedules[28] Although there is a certain response in the volume of services or other output indicators, the presence of multitasking (Dumont et al., 2008) may not enable quantification of the degree of altruism.

Chalkley and Malcomson (2002) consider the Laffont and Tirole (1993) menu of contracts in case of *perfect* agency, derive optimal efforts as a function of patient type and cost parameters, and calibrate the model using the Medicare payments data. The calculation allows quantifying cost savings owing to a changeover from a fixed price to an optimal payment schedule.

[28]See earlier works, such as Gruber and Owings (1996) and recent literature, e.g. on the natural experiment on the fee-for-service and capitation payments in Canada: Li et al. (2014); Sarma et al. (2010); Devlin and Sarma (2008).

Appendix B Price-setting in the U.S. Medicare's value-based purchasing

The aggregation of scores within domains is conducted as follows. For each hospital i and each measure m in clinical process of care, patient experience of care, outcome of care and efficiency domains *achievements 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 benchmark, m_a is 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 clinical process of care, patient experience of care along with survival rate measures of outcome of care. The means of the *bottom* deciles are employed for complication/infection measures of the outcome of care 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 a value of the measure in the current period and the baseline period, normalized by a hospital's distance from the benchmark in the baseline period:

$$p_i^m = \text{Round}\left[10\frac{y_i^m - y_{i0}^m}{m_b - y_{i0}^m} - 0.5\right],$$

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: $\max\{a_i^m, p_i^m\}$.

Additionally, *consistency points* c_i for patient experience of care domain are calculated as the lowest of the 8 dimension scores d_i^m:

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

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 $m = 1, \ldots, 8$.

The scores for clinical process of care and outcome of care domains are the sum of the values for all quality measures within domain, divided by the total potential score and translated into percentage points: $d_{i1} = \dfrac{\sum_{m=1}^{12} \max\{a_i^m, p_i^m\}}{120} \cdot 100$ for clinical process of care and $d_{i3} = \dfrac{\sum_{m=1}^{3} \max\{a_i^m, p_i^m\}}{30} \cdot 100$ for outcome of care.

In case of patient experience of care, the domain score is the sum of the values for each measure plus consistency points, divided by total potential score for quality measures plus maximum value of consistency points (percentage points): $d_{i2} = c_i + \sum_{m=1}^{8} \max\{a_i^m, p_i^m\}$.

Finally, 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 by the regulator[29] uniform across all hospitals (Table 4).

Table 4: **Domain weights**

Domain	2013	2014	2015	2016
Clinical process of care	0.7	0.2	0.2	0.10
Patient experience of care	0.3	0.3	0.3	0.25
Outcome of care	–	0.3	0.3	0.40
Efficiency	–	0.2	0.2	0.25

The values of threshold, floor and benchmark are annually re-estimated, based on the empirical distribution of hospital-level quality measures (Table 5).

[29]The lower weight for patient experience of care domain is explained by subjective character of the measures. With respect to decreasing the weight of the clinical process of care domain, the regulator explains it by the fact that most measures of the domain are already "topped-up", i.e. have reached high threshold and benchmark values (no statistical difference between 75th and 90th percentiles). At the same time, medical practitioners believe that some clinical process of care measures are not strongly correlated with patients' adverse outcomes. Accordingly, giving more weight to outcome of care domain (with survival rates and complication/infection rates) becomes an attempt of a more reasonable approximation of medical quality.

Table 5: **Incentivized quality measures within each domain in 2013–2015**

		2013			2014			2015		
Measure	Definition	Threshold	Benchmark	Floor	Threshold	Benchmark	Floor	Threshold	Benchmark	Floor
Clinical process of care domain										
AMI-7*a*	Fibrinolytic therapy received within 30 minutes of hospital arrival	65.48	91.91	–	80.66	96.30	–	72.73	100.00	–
AMI-8*a*	Primary percutaneous coronary intervention received within 90 minutes of hospital arrival	91.86	100.00	–	93.44	100.00	–	92.86	100.00	–
AMI-10	Statin prescribed at discharge	–	–	–	–	–	–	90.47	100.00	–
HF-1	Discharge instructions	90.77	100.00	–	92.66	100.00	–	92.09	100.00	–
PN-3*b*	Blood cultures performed in the emergency department prior to initial antibiotic received in hospital	96.43	100.00	–	97.30	100.00	–	97.13	100.00	–
PN-6	Initial antibiotic selection for CAP in immunocompetent patient	92.77	99.58	–	94.46	100.00	–	93.67	99.83	–
*SCIP-Inf*1	Prophylactic antibiotic received within 1 hour prior to surgical incision	97.66	100.00	–	98.07	100.00	–	97.87	100.00	–
*SCIP-Inf*2	Prophylactic antibiotic selection for surgical patients	95.07	99.68	–	98.13	100.00	–	97.88	100.00	–
*SCIP-Inf*3	Prophylactic antibiotics discontinued within 24 hours after surgery end time	94.28	99.63	–	96.63	99.96	–	96.15	99.91	–
*SCIP-Inf*4	Cardiac surgery patients with controlled 6 A.M. postoperative blood glucose	95.00	100.00	–	96.34	100.00	–	95.80	99.77	–
*SCIP-Inf*9	Postoperative urinary catheter removal on post operative day 1 or 2	–	–	–	92.86	99.89	–	93.33	100.00	–
*SCIP-Card*2	Surgery patients on beta-blocker therapy prior to arrival who received a beta-blocker during the perioperative period	97.35	99.58	–	95.65	100.00	–	95.12	100.00	–
*SCIP-VTE*1	Surgery patients with recommended venous thromboembolism prophylaxis ordered	93.07	99.85	–	94.62	100.00	–	–	–	–
*SCIP-VTE*2	Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery	93.99	100.00	–	94.92	99.83	–	94.89	99.99	–
Patient experience of care domain										
Clean-hsp-ap/*Quiet-hsp-ap*	Room was always clean/Hospital always quiet at night	62.80	77.64	36.88	63.54	78.10	38.52	64.07	78.90	41.94
Comp-1-*ap*	Nurses always communicated well	75.18	84.70	38.98	75.79	84.99	42.84	76.56	85.70	47.77
Comp-2-*ap*	Doctors always communicated well	79.42	88.95	51.51	79.57	88.45	55.49	79.88	88.79	55.62
Comp-3-*ap*	Patients always received help as soon as they wanted	61.82	77.69	30.25	62.21	78.08	32.15	63.17	79.06	35.10
Comp-4-*ap*	Pain was always well controlled	68.75	77.90	34.76	68.99	77.92	40.79	69.46	78.17	43.58
Comp-5-*ap*	Staff always gave explanation about medicines	59.28	70.42	29.27	59.85	71.54	36.01	60.89	71.85	35.48
Comp-6-*yp*	Yes, staff did give patients discharge information	81.93	89.09	50.47	82.72	89.24	54.73	83.54	89.72	57.67
Hsp-rating-910	Patients who gave hospital a rating of 9 or 10 (high)	66.02	82.52	29.32	67.33	82.55	30.91	67.96	83.44	32.82

Table 5: **Incentivized quality measures within each domain in 2013–2015**

		2013			2014			2015		
Measure	Definition	Threshold	Benchmark	Floor	Threshold	Benchmark	Floor	Threshold	Benchmark	Floor
Outcome of care domain										
Mort-30-*AMI*	Acute myocardial infarction 30-day survival rate	–	–	–	84.77	86.73	–	84.74	86.24	–
Mort-30-*HF*	Heart failure 30-day survival rate	–	–	–	88.61	90.42	–	88.15	90.03	–
Mort-30-*PN*	Pneumonia 30-day survival rate	–	–	–	88.18	90.21	–	88.27	90.42	–
PSI-90	Patient safety (weighted complication rate) for selected conditions	–	–	–	–	–	–	62.29	45.18	–
Clabsi	Central line associated blood stream infection	–	–	–	–	–	–	43.70	0.00	–
Efficiency domain										
Mspb	Medicare spending per beneficiary	–	–	–	Median	Mean of lowest decile	–	Median	Mean of lowest decile	–

Note: Threshold is the percentage point score at 50th percentile, benchmark is score at the mean of top decile, floor is the minimum score based on survey in the baseline periods (E.g. of 3211 hospitals in the baseline period in Jul 2009-Mar 2010 for FY2013). *PSI*-90 and *Clabsi* are respectively, complication and infection rates, so threshold is lower than benchmark. Source: FY 2013–2015 final rules. Federal Register, Vol.76, No.88, May 6, 2011, Tables 4 and 9; Vol.76., No.230, Nov 30, 2011, pp.74538–74540; Vol.77, No.170, Aug.31, 2012, pp.53600–53602.

Appendix C Identification in dynamic panel data models

Table 6: **Coefficients for explanatory variables in dynamic panel data regression with excluded mean reversion (Patient experience of care)**

	Comp-1-*ap*	*Comp*-3-*ap*	*Comp*-4-*ap*	*Comp*-6-*yp*	*Clean-hsp-ap*	*Hrecomddy*	*Hsp-rating*-910
$L(y)$	0.779^{***}	0.684^{***}	0.496^{***}	0.642^{***}	0.608^{***}	0.614^{***}	0.636^{***}
	(0.024)	(0.022)	(0.028)	(0.025)	(0.025)	(0.030)	(0.021)
$L^2(y)$	0.184^{***}	0.216^{***}	0.135^{***}	0.236^{***}	0.111^{***}	0.129^{***}	0.151^{***}
	(0.014)	(0.016)	(0.019)	(0.020)	(0.015)	(0.019)	(0.015)
VBP	13.444^{***}	2.975	-4.560	12.310^{***}	-6.750^{**}	-8.690^{***}	-4.258^{**}
	(2.731)	(2.496)	(3.857)	(3.113)	(2.936)	(2.440)	(1.891)
$L(beds1) \cdot VBP$	0.002	0.001^{***}	0.000^{*}	0.000	0.002	0.001^{***}	0.003
	(0.002)	(0.000)	(0.000)	(0.001)	(0.003)	(0.000)	(0.002)
$L(beds2) \cdot VBP$	0.001	–	–	-0.000	0.002	–	0.001
	(0.002)			(0.001)	(0.003)		(0.002)
$L^2(beds1) \cdot VBP$	0.001^{***}	0.000	-0.002	0.001^{***}	0.002^{***}	0.001^{**}	-0.004^{**}
	(0.000)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.002)
$L^2(beds1) \cdot VBP$	–	-0.001	-0.003^{*}	–	–	–	-0.004^{**}
		(0.002)	(0.002)				(0.002)
$L(public1) \cdot VBP$	-0.162	1.806^{**}	1.010	–	–	0.889	–
	(0.468)	(0.900)	(0.809)			(0.665)	
$L(public2) \cdot VBP$	–	–	–	0.346	0.184	–	-1.019
				(0.542)	(0.782)		(0.694)
$L^2(public1) \cdot VBP$	–	0.489	–	0.287^{*}	–	0.295	–
		(0.301)		(0.167)		(0.258)	
$L^2(public2) \cdot VBP$	-0.472^{**}	–	-0.044	–	-0.305	–	-0.320
	(0.193)		(0.245)		(0.258)		(0.275)
$L(y) \cdot VBP$	-0.278^{***}	-0.077^{**}	-0.054	-0.119^{***}	0.029	-0.021	0.002
	(0.037)	(0.033)	(0.043)	(0.041)	(0.036)	(0.042)	(0.032)
$L^2(y) \cdot VBP$	0.102^{***}	0.022	0.115^{***}	-0.024	0.058^{**}	0.139^{***}	0.053^{**}
	(0.028)	(0.027)	(0.032)	(0.030)	(0.024)	(0.032)	(0.026)
emergency	0.073	0.288	0.284	-0.144	0.800^{**}	0.197	0.847^{**}
	(0.304)	(0.467)	(0.326)	(0.358)	(0.330)	(0.453)	(0.393)
urban	-0.700^{**}	-1.269^{***}	-1.472^{***}	-0.867^{***}	-1.168^{**}	0.790^{**}	-0.253
	(0.339)	(0.417)	(0.411)	(0.284)	(0.461)	(0.385)	(0.441)
medicare share	-0.054	-2.197^{*}	-2.403^{**}	-1.697^{**}	-0.309	-1.050	-2.573^{**}
	(0.869)	(1.197)	(0.946)	(0.794)	(1.134)	(1.030)	(1.135)
casemix	0.118	2.038^{***}	1.602^{**}	1.138^{**}	1.892^{***}	1.777^{***}	2.342^{***}
	(0.529)	(0.735)	(0.690)	(0.563)	(0.710)	(0.613)	(0.908)
resbed	1.828	-4.134	-1.053	0.233	0.629	-1.091	1.204
	(1.337)	(2.573)	(2.132)	(1.333)	(1.871)	(2.298)	(1.334)

Table 6: **Coefficients for explanatory variables in dynamic panel data regression with excluded mean reversion (Patient experience of care)**

	Comp-1-*ap*	*Comp*-3-*ap*	*Comp*-4-*ap*	*Comp*-6-*yp*	*Clean-hsp-ap*	*Hrecomddy*	*Hsp-rating*-910
dsh	−0.585	2.806	0.167	−0.720	−0.412	−4.483***	−2.662*
	(0.879)	(1.795)	(1.049)	(1.055)	(1.304)	(1.434)	(1.374)
*beds*1	−0.002	−0.004**	−0.003**	−0.003**	−0.003*	−0.001	−0.006***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
*beds*2	0.000	−0.003	−0.001	0.000	0.000	0.000	−0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
$L(beds1)$	0.004**	0.006**	0.002	0.004**	0.005*	0.002	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
$L(beds2)$	–	0.001	0.002	–	–	−0.000	–
		(0.003)	(0.002)			(0.002)	
$L^2(beds1)$	−0.002	−0.001	0.000	−0.001	−0.005**	−0.001	−0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$L^2(beds2)$	−0.003*	–	–	−0.002	−0.006***	−0.001	–
	(0.002)			(0.001)	(0.002)	(0.002)	
*public*1	−0.217	0.102	−0.095	0.035	−0.421*	0.024	−0.291
	(0.169)	(0.258)	(0.231)	(0.166)	(0.221)	(0.284)	(0.241)
*public*2	1.019**	0.701	−0.142	0.001	0.144	0.491	1.340*
	(0.516)	(0.688)	(0.694)	(0.632)	(0.701)	(0.698)	(0.771)
$L(public1)$	−0.447**	−0.596**	−0.384	−0.313	−0.235	0.133	−0.290
	(0.201)	(0.294)	(0.240)	(0.209)	(0.251)	(0.262)	(0.291)
$L(public2)$	−1.237**	−2.638***	−1.015	−0.241	−0.152	−1.034	−1.378*
	(0.550)	(0.934)	(0.895)	(0.613)	(0.711)	(0.891)	(0.770)
$L^2(public1)$	−0.197	−0.607**	−0.533*	−0.607**	−0.495*	−0.033	0.088
	(0.206)	(0.306)	(0.285)	(0.239)	(0.301)	(0.311)	(0.291)
$L^2(public2)$	0.267	1.567*	1.003	−0.444	0.002	0.916	1.285*
	(0.453)	(0.890)	(0.778)	(0.557)	(0.755)	(0.650)	(0.662)
constant	4.265*	5.820**	25.965***	11.545***	18.826***	16.879***	13.870***
	(2.248)	(2.550)	(3.050)	(2.250)	(2.716)	(2.627)	(2.096)
Observations	21066	21062	21056	21056	21066	18908	21064
Hospitals	3340	3340	3339	3338	3340	3315	3339
Arellano–Bond test statistic	−1.238	−1.635	−1.993	−1.615	−1.948	−0.956	0.685

Note: Robust standard errors in parentheses. Arellano-Bond test statistics for absence of order 2 serial correlation in the first-differenced errors.

Table 7: **Coefficients for explanatory variables in dynamic panel data regression with excluded mean reversion (Clinical process of care)**

	HF-1	*SCIP-Inf*1	*SCIP-Inf*2	*SCIP-Inf*3	*SCIP-Inf*4	*SCIP-Card*2
$L(y)$	0.726^{***}	0.608^{***}	0.424^{***}	0.655^{***}	0.502^{***}	0.426^{***}
	(0.014)	(0.021)	(0.079)	(0.019)	(0.073)	(0.048)
$L^2(y)$	−0.008	0.043^{***}	0.008	−0.006	0.059^{**}	0.071^{***}
	(0.009)	(0.013)	(0.019)	(0.011)	(0.026)	(0.019)
VBP	14.514^{***}	17.233	40.442^{**}	−1.434	41.387	12.258
	(5.052)	(13.801)	(19.721)	(11.185)	(29.417)	(9.090)
$L(beds1) \cdot VBP$	0.000	−0.000	0.001^{***}	0.003	−0.001	0.000
	(0.009)	(0.000)	(0.000)	(0.002)	(0.007)	(0.000)
$L(beds2) \cdot VBP$	−0.001	–	–	0.002	–	–
	(0.009)			(0.002)		
$L^2(beds1) \cdot VBP$	–	−0.000	−0.000	-0.002^{*}	–	0.000
		(0.000)	(0.000)	(0.001)		(0.002)
$L^2(beds1) \cdot VBP$	–	–	–	−0.002	–	0.001
				(0.002)		(0.002)
$L(public1) \cdot VBP$	–	0.067	−1.189	0.631	−0.519	–
		(0.648)	(0.956)	(0.655)	(1.415)	
$L(public2) \cdot VBP$	0.652	–	–	–	–	1.012
	(2.283)					(0.978)
$L^2(public1) \cdot VBP$	−0.700	0.227	0.720	−0.259	–	0.038
	(2.245)	(0.356)	(0.452)	(0.457)		(0.703)
$L^2(public2) \cdot VBP$	–	–	–	–	–	–
$L(y) \cdot VBP$	-0.235^{***}	-0.225^{**}	-0.348^{**}	−0.122	-0.414^{**}	−0.122
	(0.048)	(0.113)	(0.166)	(0.083)	(0.209)	(0.086)
$L^2(y) \cdot VBP$	0.081^{**}	0.050	−0.059	0.135^{***}	−0.015	0.004
	(0.035)	(0.059)	(0.057)	(0.051)	(0.120)	(0.039)
emergency	2.401^{***}	0.921	0.306	0.620	1.461	−0.723
	(0.888)	(0.576)	(0.488)	(0.594)	(1.111)	(1.252)
urban	0.044	0.397	0.921	0.199	1.349	0.864
	(1.367)	(0.520)	(0.640)	(0.733)	(1.237)	(0.546)
medicare share	-6.717^{**}	4.290^{**}	−1.176	0.596	-7.898^{**}	1.699
	(3.186)	(1.805)	(1.278)	(1.757)	(3.818)	(2.321)
casemix	3.420^{*}	0.359	1.650^{**}	1.473^{*}	0.622	0.683
	(1.791)	(0.722)	(0.663)	(0.839)	(1.542)	(1.412)
resbed	5.492	2.601	1.029	2.007	-11.666^{*}	1.862
	(4.233)	(2.024)	(1.120)	(2.580)	(6.862)	(2.591)
dsh	−5.705	−0.992	−0.835	−0.220	1.854	−2.847
	(3.510)	(1.627)	(1.785)	(1.850)	(4.408)	(2.479)
*beds*1	−0.001	−0.002	-0.002^{*}	−0.002	0.003	0.003

Table 7: **Coefficients for explanatory variables in dynamic panel data regression with excluded mean reversion (Clinical process of care)**

	HF-1	*SCIP-Inf*1	*SCIP-Inf*2	*SCIP-Inf*3	*SCIP-Inf*4	*SCIP-Card*2
	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
*beds*2	−0.003	0.002	−0.000	−0.001	0.001	−0.000
	(0.007)	(0.001)	(0.001)	(0.002)	(0.007)	(0.002)
$L(beds1)$	−0.002	−0.002	0.001	−0.001	0.001	−0.002
	(0.003)	(0.002)	(0.001)	(0.002)	(0.003)	(0.004)
$L(beds2)$	–	−0.002	−0.001	–	–	0.001
		(0.002)	(0.002)			(0.003)
$L^2(beds1)$	−0.000	0.003**	0.000	0.002	−0.001	0.001
	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
$L^2(beds2)$	−0.002	0.000	−0.000	–	0.002	–
	(0.005)	(0.001)	(0.001)		(0.003)	
*public*1	−1.580***	−0.094	0.667*	−0.651**	0.839	−0.395
	(0.526)	(0.199)	(0.400)	(0.268)	(0.910)	(0.405)
*public*2	0.948	0.453	−0.053	−0.816*	1.089	−0.021
	(1.561)	(0.552)	(0.504)	(0.496)	(0.973)	(0.731)
$L(public1)$	1.162*	0.651	−0.787**	0.434	0.546	−2.061
	(0.682)	(0.400)	(0.318)	(0.374)	(0.587)	(1.460)
$L(public2)$	−1.380	−1.018	0.857	−0.152	–	−0.770
	(1.550)	(0.783)	(0.747)	(0.748)		(0.772)
$L^2(public1)$	−0.510	0.353	0.522	0.282	0.470	0.226
	(0.707)	(0.298)	(0.517)	(0.355)	(0.976)	(0.510)
$L^2(public2)$	–	0.227	−1.639	0.956	–	−0.337
		(0.699)	(1.124)	(0.654)		(1.118)
constant	25.059***	31.334***	53.147***	31.709***	41.925***	47.440***
	(3.307)	(2.285)	(7.806)	(2.032)	(7.809)	(5.490)
Observations	26486	23230	18324	23136	4510	11475
Hospitals	3352	3323	3251	3321	1180	3069
Arellano–Bond test statistic	−0.616	−0.888	0.384	−0.312	0.977	−0.987

Note: Robust standard errors in parentheses. Arellano-Bond test statistics for absence of order 2 serial correlation in the first-differenced errors.

Appendix D History of incentives contracts on quality for the inpatient care in the U.S.

Numerous private and public programs linking quality and payment in healthcare existed in the U.S. in the early 2000s, mostly at employer or state level (Ryan and Blustein, 2011; Damberg et al., 2009; Pearson et al., 2008; Rosenthal, 2008; Damberg et al., 2005; Rosenthal et al., 2004). The nationwide quality-performance reimbursement started with the Hospital Quality Incentive Demonstration (HQID), when 33 quality measures for five clinical conditions (heart failure, acute myocardial infarction, community-acquired pneumonia, coronary-artery bypass grafting, an hip and knee replacement) were accumulated from voluntarily participating hospitals.[30] The project provided respectively 2% and 1% bonus payments for hospitals in the top and second top deciles of each quality measure. On the other hand, hospitals in the bottom two deciles (as of the end of the third year of the project) were to receive 1-2% penalties. It should be noted that HQID redistributed funds between top and bottom hospitals, while value-based purchasing applies deductions or rewards to all hospitals. Therefore, the potential impact of value-based purchasing might be expected to be higher than that of HQID (Kahn et al., 2006).

Overall, the financial incentives helped improve the quality of the participant hospitals, but the improvement was inversely related to baseline performance (Lindenauer et al., 2007). Moreover, low-quality hospitals required most investment in quality increase, yet, they were not financially stimulated (Rosenthal et al., 2004). This outcome might have been the reason for an extension of reimbursement rules within value-based purchasing into achievement and improvement points.

The accumulation of the measures within the Hospital Quality Incentive was followed by the launch of the Surgical Care Improvement Project (SCIP) and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). HCAHPS was started in 2007 as the first national standardized survey with public reporting on various dimensions of patient experience of care (HCAHPS online, 2013), and its measures constitute the patient experience of care domain in value-based purchasing. The measures of the clinical process of care domain are collected within Hospital Inpatient Quality Reporting (IQR) program. These are measures for acute clinical conditions stemming from the Hospital Quality Incentive (i.e. AMI, heart failure, pneumonia), as well as measures from the Surgical Care Improvement Project and Healthcare Associated Infections. Note that since 2009 Medicare's hospitals are financially punished for not submitting the IQR quality measures.

[30] The Centers for Medicare and Medicaid Services (CMS) - Premier database.

The success of the HQID pilot in terms of average enhancement of hospital quality has resulted in the nationwide introduction of pay-for-performance within a prospective payment system – a value-based purchasing reform, started in FY2013.

CINCH working paper series

1 Halla, Martin and Martina Zweimüller**. Parental Responses to Early Human Capital Shocks:** Evidence from the Chernobyl Accident. CINCH 2014.

2 Aparicio, Ainhoa and Libertad González. **Newborn Health and the Business Cycle:** Is it Good to be born in Bad Times? CINCH 2014.

3 Robinson, Joshua J. **Sound Body, Sound Mind?:** Asymmetric and Symmetric Fetal Growth Restriction and Human Capital Development. CINCH 2014.

4 Bhalotra, Sonia, Martin Karlsson and Therese Nilsson. **Life Expectancy and Mother-Baby Interventions:** Evidence from A Historical Trial. CINCH 2014.

5 Goebel, Jan, Christian Krekel, Tim Tiefenbach and Nicolas R. Ziebarth. **Natural Disaster, Environmental Concerns, Well-Being and Policy Action:** The Case of Fukushima. CINCH 2014.

6 Avdic, Daniel, **A matter of life and death? Hospital Distance and Quality of Care:** Evidence from Emergency Hospital Closures and Myocardial Infarctions. CINCH 2015.

7 Costa-Font, Joan, Martin Karlsson and Henning Øien. **Informal Care and the Great Recession.** CINCH 2015.

8 Titus J. Galama and Hans van Kippersluis. **A Theory of Education and Health.** CINCH 2015.

9 Dahmann, Sarah. **How Does Education Improve Cognitive Skills?:** Instructional Time versus Timing of Instruction. CINCH 2015.

10 Dahmann, Sarah and Silke Anger. **The Impact of Education on Personality:** Evidence from a German High School Reform. CINCH 2015.

11 Carbone, Jared C. and Snorre Kverndokk. **Individual Investments in Education and Health.** CINCH 2015.

12 Zilic, Ivan. **Effect of forced displacement on health.** CINCH 2015.

13 De la Mata, Dolores and Carlos Felipe Gaviria. **Losing Health Insurance When Young:** Impacts on Usage of Medical Services and Health. CINCH 2015.

14 Tequame, Miron and Nyasha Tirivayi. **Higher education and fertility:** Evidence from a natural experiment in Ethiopia. CINCH 2015.

15 Aoki, Yu and Lualhati Santiago. **Fertility, Health and Education of UK Immigrants:** The Role of English Language Skills. CINCH 2015.

16 Rawlings, Samantha B., **Parental education and child health:** Evidence from an education reform in China. CINCH 2015.

17 Kamhöfer, Daniel A., Hendrik Schmitz and Matthias Westphal. **Heterogeneity in Marginal Non-monetary Returns to Higher Education.** CINCH 2015.

18 Ardila Brenøe, Anne and Ramona Molitor. **Birth Order and Health of Newborns:** What Can We Learn from Danish Registry Data? CINCH 2015.

19 Rossi, Pauline. **Strategic Choices in Polygamous Households:** Theory and Evidence from Senegal. CINCH 2016.

20 Clarke, Damian and Hanna Mühlrad. **The Impact of Abortion Legalization on Fertility and Maternal Mortality:** New Evidence from Mexico. CINCH 2016.

21 Jones, Lauren E. and Nicolas R. Ziebarth. **US Child Safety Seat Laws:** Are they Effective, and Who Complies? CINCH 2016.

22 Koppensteiner, Martin Foureaux and Jesse Matheson**. Access to Education and Teenage Pregnancy.** CINCH 2016.

23 Hofmann, Sarah M. and Andrea M. Mühlenweg**. Gatekeeping in German Primary Health Care –** Impacts on Coordination of Care, Quality Indicators and Ambulatory Costs. CINCH 2016.

24 Sandner, Malte. **Effects of Early Childhood Intervention on Fertility and Maternal Employment:** Evidence from a Randomized Controlled Trial. CINCH 2016.

25 Baird, Matthew, Lindsay Daugherty, and Krishna Kumar. **Improving Estimation of Labor Market Disequilibrium through Inclusion of Shortage Indicators.** CINCH 2017.

26 Bertoni, Marco, Giorgio Brunello and Gianluca Mazzarella. **Does postponing minimum retirement age improve healthy behaviors**

before retirement? Evidence from middle-aged Italian workers. CINCH 2017.

27 Berniell, Inés and Jan Bietenbeck. **The Effect of Working Hours on Health.** CINCH 2017.

28 Cronin, Christopher, Matthew Forsstrom, and Nicholas Papageorge. **Mental Health, Human Capital and Labor Market Outcomes.** CINCH 2017.

29 Kamhöfer, Daniel and Matthias Westphal. **Fertility Effects of College Education:** Evidence from the German Educationl Expansion. CINCH 2017.

30 Jones, John Bailey and Yue Li. **The Effects of Collecting Income Taxes on Social Security Benefits.** CINCH 2017.

31 Hofmann, Sarah and Andrea Mühlenweg. **Learning Intensity Effects in Students' Mental and Physical Health –** Evidence from a Large Scale Natural Experiment in Germany. CINCH 2017.

32 Vollmer, Sebastian and Juditha Wójcik. **The Long-term Consequences of the Global 1918 Influenza Pandemic:** A Systematic Analysis of 117 IPUMS International Census Data Sets. CINCH 2017.

33 Thamarapani, Dhanushka, Rockmore, Marc, and Willa Friedman. **The Educational and Fertility Effects of Sibling Deaths**. CINCH 2018.

34 Lemmon, Elizabeth. **Utilisation of personal care services in Scotland:** the influence of unpaid carers. CINCH 2018.

35 Avdic, Daniel, Büyükdurmus, Tugba, Moscelli, Giuseppe, Pilny, Adam, and Ieva Sriubaite**. Subjective and objective quality reporting and choice of hospital:** Evidence from maternal care services in Germany. CINCH 2018.

36 Hentschker, Corinna and Ansgar Wübker. **Quasi-experimental evidence on the effectiveness of heart attack treatment in Germany.** CINCH 2018.

37 Pasha, Mochamad, Rockmore, Marc, and Chih Ming Tan. **Early Life Exposure to Above Average Rainfall and Adult Mental Health**. CINCH 2018

38 Elsner, Benjamin and Florian Wozny. **The Human Capital Cost of Radiation:** Long-run Evidence from Exposure outside the Womb. CINCH 2019

39 de la Mata, Dolores and Carlos Felipe Gaviria Garcés. **Exposure to Pollution and Infant Health:** Evidence from Colombia. CINCH 2019

40 Besstremyannaya, Galina and Sergei Golovan. **Physicians' altruism in incentives contracts:** Medicare's quality race. CINCH 2019