RUHR
ECONOMIC PAPERS

Jeannette Brosig-Koch
Heike Hennig-Schmidt
Nadja Kairies-Schwarz
Daniel Wiesen

Physician Performance Pay: Evidence from a Laboratory Experiment

UNIVERSITÄT DUISBURG ESSEN

#658

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

RWI Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #658

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-764-9

Ruhr Economic Papers #658

Jeannette Brosig-Koch, Heike Hennig-Schmidt,
Nadja Kairies-Schwarz, and Daniel Wiesen

Physician Performance Pay: Evidence from a Laboratory Experiment

UNIVERSITÄT DUISBURG ESSEN

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788764
ISSN 1864-4872 (online)
ISBN 978-3-86788-764-9

Jeannette Brosig-Koch, Heike Hennig-Schmidt,
Nadja Kairies-Schwarz, and Daniel Wiesen[1]

Physician Performance Pay: Evidence from a Laboratory Experiment

Abstract

We present causal evidence from a controlled experiment on the effect of pay for performance on physicians' behavior and patients' health benefits. At a within-subject level, we introduce performance pay to complement either fee-for-service or capitation. Performance pay is granted if a health care quality threshold is reached, and varies with the patients' severity of illness. We find that performance pay significantly reduces overprovision of medical services due to fee-for-service incentives, and underprovision due to capitation; on average, it increases the patients' health benefit. The magnitude of these effects depends, however, on the patients' characteristics. We also find evidence for a crowding-out of patient-regarding behavior due to performance pay. Health policy implications are discussed.

JEL Classification: C91, I11

Keywords: Fee-for-service; capitation; pay for performance; laboratory experiment; crowding-out

October 2016

1 Jeannette Brosig-Koch, UDE and CINCH; Heike Hennig-Schmidt, University of Bonn and University of Oslo; Nadja Kairies-Schwarz, UDE and CINCH; Daniel Wiesen, University of Cologne and University of Oslo. *– We are grateful for valuable comments and suggestions from Josse Delfgaauw, Robert Dur, Geir Godager, Oliver Gürtler, Glenn W. Harrison, Tor Iversen, Matthias Kräkel, and Tamar Kugler. We particularly thank Ludwig Kuntz who pointed us to a related literature in decision analysis. This paper also benefited from suggestions by participants of seminars at Berlin, Hamburg, iBMG Rotterdam, Mainz, Oslo, Tutzingen, Wuppertal, and at the 3rd Workshop on Behavioral and Experimental Health Economics Essen, COPE Vienna, ESA Europe Cologne, ESA New York, GfeW Hamburg, EuHEA Hamburg, and CAS Workshop Munich. We thank Emanuel Castillo and Heiko A. Schmidt for their excellent programming assistance, Johanna Kokot for helping to conduct the experiments, and Anne Classen for her research assistance. Financial support by the German Research Foundation (DFG, grant: BR 2346/2-1/2) is gratefully acknowledged. Heike Hennig-Schmidt acknowledges a fellowship from the Käte Hamburger Kolleg/ Centre for Global Cooperation Research. This paper is a completely revised version of the Ruhr Economic Paper No. 413 ("How Effective are Pay-for-Performance Incentives for Physicians? A Laboratory Experiment").* *– All correspondence to: Jeannette Brosig-Koch, University of Duisburg-Essen, Berliner Platz 6-8, 45127 Essen, Germany, e-mail: jeannette.brosig@uni-due.de*

1 Introduction

A fundamental question in health policy around the world is that of how to incentivize health care providers to improve the quality of care. While the traditional approaches to pay physicians have focussed on fee-for-service and capitation, there has been growing interest in directly measuring and incentivizing physicians' performance based on patients' health outcomes. Performance pay is typically granted conditional on achieving a performance threshold. The idea of paying physicians (at least partially) on the basis of direct performance measures has attracted particular attention as fee-for-service incentivizes physicians to overserve patients and capitation to underserve patients.

However, while the idea of using performance pay for physicians as a way of improving health care outcomes is increasingly making its way into policy,[1] the empirical evidence on the effectiveness of such policies is quite limited—with identification of the causal impact of physician incentives being the main challenge. Also, evidence from field studies on whether an effect of performance pay with respect to improving the quality of care exists is quite mixed. If at all, rather moderate effects of performance pay are reported (e.g., Li et al., 2014). Establishing the causal link between performance pay and individual physicians' provision behavior is particularly difficult due to the likely endogeneity of institutions (e.g., Baicker and Goldman, 2011), biases due to incomplete performance measures (e.g., Mullen et al., 2010), measurement errors (e.g., Campbell et al., 2009), gaming of performance indicators (e.g., Gravelle et al., 2010), the limited availability of data (e.g., Maynard, 2012), and the frequent introduction of performance pay accompanied by other interventions (e.g., Lindenauer et al., 2007).

Moreover, it is often argued in the economics and psychological literature that performance pay induces crowding-out of intrinsic motivation (e.g., Deci and Ryan, 1975; Frey, 1997; Kreps, 1997) among public service workers (e.g., Besley and Ghatak, 2005; Delfgaauw and Dur, 2008) and, more specifically, among physicians (e.g., Siciliani, 2009). There is some evidence that performance pay crowds out intrinsic motivation from questionnaire studies (e.g., Frey et al., 1996; Frey and Oberholzer-Gee, 1997) and controlled experiments (e.g., Gneezy and Rustichini, 2000; Mellström and Johannesson, 2008; Huffman and Bognanno, 2014). Evidence, however, on whether performance pay actually does crowd out individuals' patient-regarding behavior[2]—and therefore the quality of medical service provision—is, to the best of our knowledge, still lacking.

To address these issues, we introduce a new experiment that investigates the behavioral

[1] While performance pay contracts have been commonplace in many private sector industries for decades (see, e.g., Prendergast, 1999; Lazear, 2000), they have attracted growing attention in the health care sector. Performance pay for physicians has now been widely introduced, for example, in the UK (see, e.g., Roland, 2004; Doran et al., 2006; Roland and Campbell, 2014), the US (e.g., Rosenthal et al., 2006), and in several other OECD countries; see Cashin et al. (2014) for an overview. For a meta-study on the effectiveness of pay for performance initiatives, see Eijkenaar et al. (2013).

[2] Woolhandler et al. (2012) argue that offering performance pay to physicians, rather than enhancing their intrinsic motivation, may reduce the physicians' desire to perform an activity for its inherent rewards (e.g., pride in excellent work).

effects of performance pay for physicians under controlled conditions. In a medically framed decision situation, subjects in the lab decide, in the role of 'physicians', on the quantity of medical services for 'patients' with different severities of illness. Quantity choices determine the physicians' own profit and the patients' health benefits. Real patients' health outside the lab is affected by subjects' decisions as total health benefits (measured in monetary terms in the experiment) are transferred to a charity and are exclusively dedicated to the medical treatment of cataract patients (analogous to Hennig-Schmidt et al., 2011; Brosig-Koch et al., forthcoming a, forthcoming b). In our experiment, we implement a 'clean' measure of individual physicians' performance: the difference between the chosen quantity and the patient-optimal quantity of medical services, in which the latter implies the highest health benefit for a patient. A physician 'performs' best (according to this measure) whenever she is delivering the patient-optimal quantity of medical services (provides the highest quality of care). The larger the difference to the patient-optimal quantity, the lower is the quality.

To establish the causal link between performance pay and the quality of medical service provision, we exogenously change physicians' remuneration at a within-subject level from non-blended fee-for-service or capitation to blended performance pay systems. Performance pay is linked to the patients' optimal health benefit and bonus rates are adjusted for patients' severities of illness.[3]

In our parsimonious experimental design, we abstract from several unintended effects associated with performance pay such as substitution of effort on incentivized tasks (the quantity choice is one-dimensional) and gaming of the incentive scheme. We allow, however, for non-perfect contractibility of a physician's quality of medical service provision which, for instance, is implied by information asymmetry on the physician's treatment quality between physicians and payers (a common assumption in principal-agent theory; see, for example, Holmström and Milgrom, 1991). Performance pay is granted whenever a quality threshold is reached.

We derive hypotheses for physicians' behavior in an illustrative model related to Ellis and McGuire (1986), in which we allow for different degrees of a physician's altruism (see also Godager and Wiesen, 2013). We address the following research questions: First, we investigate whether fee-for-service leads to overprovision of medical service and capitation to underprovision of medical services differentiated for patients' severities of illness.[4] We then test how the introduction of performance pay, which relates to the patients' optimal health benefit and complements fee-for-service or capitation, affects medical service provision. Finally, we analyze at a within-subject level whether performance pay crowds out patient-regarding behavior. We derive a crowding-out hypothesis by allowing for a shift in physicians' weights on the patients' health benefit due to the introduction of performance pay (similar to the commonly known 'splitting

[3] Notice that the adjustment of the bonus rates based on the severity of illness can be interpreted as a kind of risk adjustment (for a definition, see, for example, Glazer and McGuire, 2000; van de Ven and Ellis, 2000). Patients with a mild severity of illness, for example, face the highest 'risk' of being overtreated under fee-for-service—a behavioral pattern which has been indicated by recent empirical and experimental findings (e.g., Hennig-Schmidt et al., 2011; Clemens and Gottlieb, 2014; Brosig-Koch et al., forthcoming b).

[4] Only such results would make the introduction of performance pay meaningful.

effect' in decision theory; see, for example, Weber et al., 1988).

In line with theory, behavioral data from the experiments show that: (i) overtreatment of patients under fee-for-service and undertreatment under capitation varies with patients' severity of illness, (ii) introducing performance pay significantly reduces overprovision of medical services under fee-for-service and underprovision under capitation and the reduction in non-optimal service provision is significantly affected by the severity of illness, (iii) performance pay induces a crowding-out of patient-regarding behavior for around one third of the patients who have been treated optimally prior to the introduction of performance pay, and the extent of a crowding-out varies with the patients' severities of illness.

Taken together, our behavioral results indicate that performance pay linked to a patient's health outcome, on aggregate, reduces non-optimal service provision and therefore enhances the patients' health benefit. However, the increase in quality varies with the patients' severities of illness and a crowding-out of patient-regarding behavior reduces the health benefit for a considerable proportion of patients. When considering a healthcare policy maker's costs due to additional pay for an increase in patients' health benefit, behavioral differences for patients' severities in our experimental setup indicate that it would be most cost effective to introduce performance pay for mildly ill patients when fee-for-service is the baseline payment and for severely ill patients when the baseline payment is capitation.

This paper proceeds as follows: We present our experimental design in Section 2 and derive behavioral hypotheses in Section 3. Section 4 presents our behavioral results and discusses implications for costs and benefits. Section 5 concludes.

2 Experimental design

2.1 Decision situation

In our experiment, all subjects decide in the role of physicians on the provision of medical services. We randomly assign subjects to different payment conditions. We employ a within-subject design to analyze the effect of physicians' performance pay on the provision of medical services. To this end, each subject makes his or her decisions under two consecutive payment systems. First, subjects are incentivized either by fee-for-service (FFS) or by capitation (CAP), which serve as baseline payments. Second, we introduce performance pay in addition to the respective baseline payments.[5]

In all payment systems, physician i decides on the quantity of medical services $q \in [0, 10]$ for nine different patients ($j = 1, \ldots, 9$). Patients differ in illnesses $k \in \{A, B, C\}$ and in severities of illness $l \in \{x, y, z\}$. Patients are assumed to be passive and fully insured, accepting each

[5]Notice that the general decision situation of our experiment is similar to Hennig-Schmidt et al. (2011), Hennig-Schmidt and Wiesen (2014) and Brosig-Koch et al. (forthcoming a, forthcoming b). In the latter two studies, incentives under FFS (CAP) are the same as in the present paper.

quantity of medical services provided by the physician. This is a common assumption in the theoretical health economics literature; for a comprehensive review, for example, see McGuire (2000). In our experiment, patients' characteristics are the same in all payment conditions. The patient population for which a physician chooses services thus remains constant.

Physician i's payment is $R(q) = \Lambda + pq + b_l^{\bullet} I_{b_l}$, with Λ being the lump-sum payment, p the fee per service rendered to a patient, and $b_l^{\bullet}$ the bonus payment; I_{b_l} denotes an indicator variable which equals 1, if the physician's chosen quantity does not differ more than one unit from the patient optimal treatment, and 0 otherwise. In FFS, $\Lambda = 0$ and $b_l^{\text{FFS}} = 0$ and in CAP $p = 0$, and $b_l^{\text{CAP}} = 0$.

Physician i's profit is given as

$$\pi(q) = \Lambda + pq + b_l^{\bullet} I_{b_l} - c(q), \tag{1}$$

with $\Lambda, p, b_l^{\bullet} > 0$, $c'(q) > 0$ and $c''(q) > 0$. In the experiment, $c(q) = q^2/10$ for all payment systems.

When deciding on q, physician i simultaneously determines her own profit $\pi(q)$ and the patient's health benefit $B(q)$ for patient j. Common to all patient health benefit functions is a global optimum at q^* on $q \in (0, 10)$. The patient health benefit function employed in our experiment is

$$B(q) = \begin{cases} B_0 + \theta q & \text{if } q \leq q^* \\ B_1 - \theta q & \text{if } q \geq q^*, \end{cases} \tag{2}$$

with $B_0, B_1 \geq 0$ and $\theta > 0$.[6]

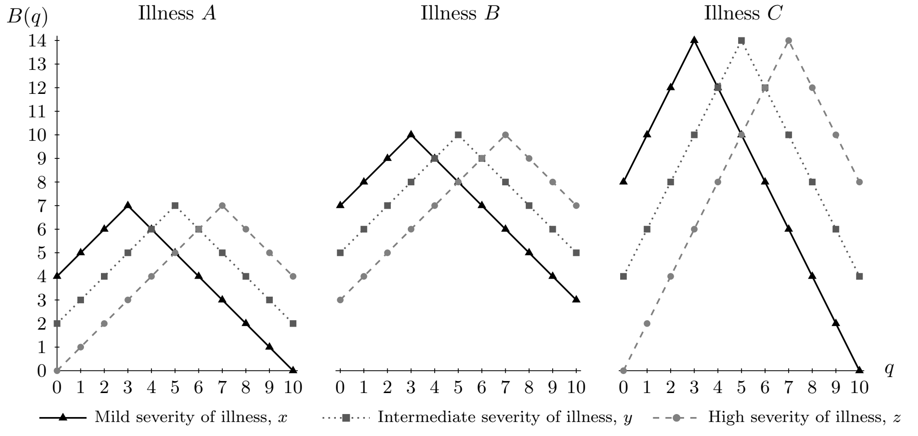
Figure 1 illustrates the patient health benefits in our experiment, which is varied systematically for the patients' illness k and severity of illness l. In particular, for illnesses A and B $\theta = 1$ and for illness C $\theta = 2$. For illnesses A, B, and C, the maximum health benefit is $B_{Al}(q^*) = 7$, $B_{Bl}(q^*) = 10$, and $B_{Cl}(q^*) = 14$, respectively. The patient-optimal quantity q^* varies with severities of illness l. For mild (x), intermediate (y), and severe (z) illnesses, the patient-optimal quantities are $q_x^* = 3$, $q_y^* = 5$, and $q_z^* = 7$, respectively.[7] In the experiment, the patient-optimal quantity q^* is known for all patients. We are therefore able, first, to analyze overprovision and underprovision of medical services and, second, to introduce a 'clean' performance measure of a physician's quality related to the patient-optimal treatment (see Subsection 2.2). Moreover, the symmetric design of patient health benefits implies that the marginal effects of over- and underprovision are equivalent.

All parameters of the experiment are common knowledge. When making their quantity choices, physicians are aware of cost, payment, profit, and the patient's health benefit for each quantity; for an illustration of the decision situation, see the instructions in Appendix B.

[6] Note that $B_1 = B_0 + 2\theta q^*$. B_0 and B_1 are allowed to be different, which reflects the patient health benefit parameters in the experiment. For example, for illness A (with $\theta = 1$) and severity x (with $q^* = 3$), $B_0 = 4$ and $B_1 = 10$, as $B_1 = 4 + 2 \cdot 1 \cdot 3 = 10$.

[7] Varying patients' characteristics in our lab experiment are motivated by the recent theoretical literature (see, e.g., Allard et al. 2011), which assumes that patient characteristics affect the physicians' behavior.

Figure 1: Patient health benefits by illness and severity of illness



Notes: This figure illustrates patient health benefit parameters $B(q)$ for illnesses $k = A, B, C$ and severities of illness $l = x, y, z$ on the quantity interval from 0 to 10. The left panel shows patient benefits for illness A, the middle panel for illness B, and the right panel for illness C. The black solid line indicates severity of illness x, the grey dotted line severity of illness y, and the grey dashed line severity of illness z. For illness A and B, $\theta = 1$ and for illness C, $\theta = 2$. Notice that the patient health benefits are kept constant for all payment conditions.

While all subjects in the experiment make decisions in the role of physicians for abstract patients in the lab, real patients' health outside the lab is affected by their choices. Subjects are informed that the monetary equivalent of the patient health benefit resulting from their decisions is transferred to a charity that uses the money for surgical treatments of cataract patients; for procedural details, see Subsection 2.3.

2.2 Payment systems

Recall that each subject decides in the role of a physician on the provision of medical services under two different payment systems. In part I of the experiment, subjects decide either under FFS or CAP. In part II, they receive performance pay.

Table 1 provides an overview of payment systems employed in our experiment. Subjects paid by FFS (CAP) in part I decide under the performance-pay system FFS+P4P (CAP+P4P) in part II. Beyond the within-subject comparison (part I vs. part II), the experimental design also allows us to compare between-subject behavior in FFS and CAP in part I. The profit functions of FFS and FFS+P4P systems mirror those of the respective CAP and CAP+P4P systems. While varying the components of the payment systems, we keep maximum profit levels and marginal profits constant. The profit parameters are illustrated in Figure 2, and the complete set of parameter values is shown in Table A1 in Appendix A.

In FFS, physicians are paid a fee of $p = 2$ per service. Accordingly, profit is $\pi(q) = 2q - c(q)$. In CAP, physicians receive a lump-sum payment of $\Lambda = 10$ per patient, independent

Table 1: Payment systems in the main experimental conditions

Exp.	Part I of the experiment				Part II of the experiment						# of subjects
cond.	Payment	Λ	p	R	Payment	Severity l	Λ	p	$b_l^\bullet$	R	(medical students)
1	FFS	–	2	$2q$	FFS+P4P	x	–	2	5.6	$2q + 5.6$	42 (22)
						y	–	2	3.6	$2q + 3.6$	
						z	–	2	2.4	$2q + 2.4$	
2	CAP	10	–	10	CAP+P4P	x	10	–	2.4	$10 + 2.4$	45 (22)
						y	10	–	3.6	$10 + 3.6$	
						z	10	–	5.6	$10 + 5.6$	

Notes. This table shows the parameters of the payment systems and the number of participants in our main experimental conditions. Note that the performance pay $b_l^\bullet$ is only granted to subjects if their quantity choice fulfills the quality requirement $|q - q^*| \leq 1$, otherwise the performance pay equals zero. In each condition, 22 medical students participated. Additionally, in a control condition (N=13), we lifted the lump-sum payment in CAP to 12 (part I of the experiment), while keeping the CAP+P4P system constant (part II of the experiment) to rule out income effects when introducing performance incentives.

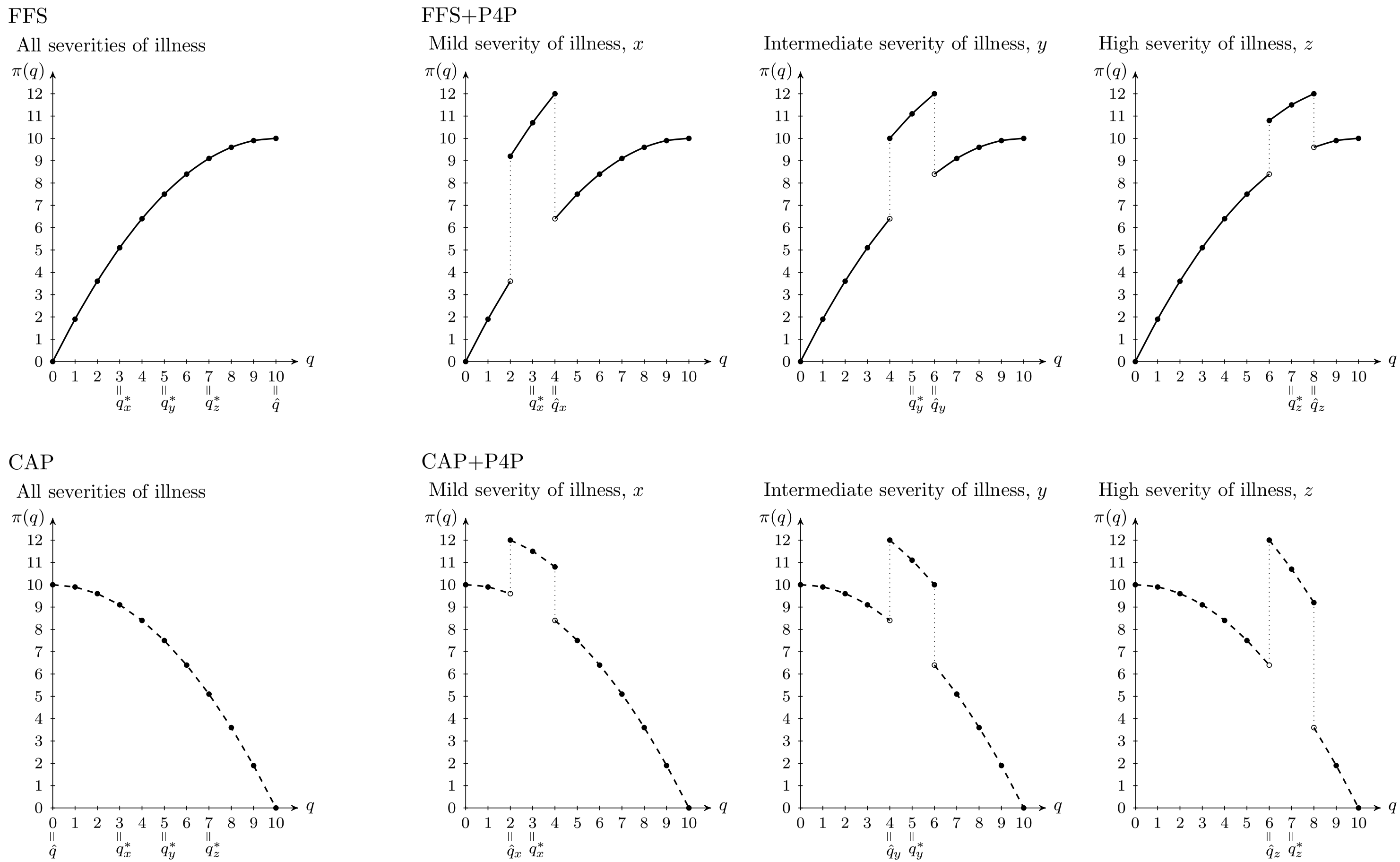
of the quantity of medical services. Physicians' profit per patient is thus $\pi(q) = 10 - c(q)$ with the maximum attainable profit being 10 in both payment systems FFS and CAP. The profit-maximizing quantity of medical services for each of the nine patients is $\hat{q}_j^{\text{FFS}} = 10$ and $\hat{q}_j^{\text{CAP}} = 0$ in FFS and CAP, respectively. Also, the absolute values of marginal profits are equal in CAP and FFS.

Our performance measure is linked to a patient's health outcome—in particular, to the optimal patient health benefit. The performance pay is granted if the quantity a physician has chosen does not deviate by more than one unit from the patient-optimal quantity q^*, i.e., $I_{b_l} = 1$, whenever $|q - q^*| \leq 1$, and $I_{b_l} = 0$ otherwise. We set bonus rates such that incentives are comparable across conditions. For severities x, y, and z, $b_x^{\text{FFS}} = 5.6$, $b_y^{\text{FFS}} = 3.6$, $b_z^{\text{FFS}} = 2.4$ in FFS+P4P, and $b_x^{\text{CAP}} = 2.4$, $b_y^{\text{CAP}} = 3.6$, $b_z^{\text{CAP}} = 5.6$ in CAP+P4P, respectively. The bonus implies an increase in the maximum attainable profit $\pi(\hat{q}_j)$ by 20 percent. For each severity, choosing $\hat{q}_j$ equal to 4, 6, or 8 (2, 4, or 6) in FFS+P4P (CAP+P4P) thus yields a profit of 12 for the physician.

In sum, performance-pay systems are designed to mitigate inherent incentives in FFS (CAP) to provide too many (few) services. However, we still assume that the quality is not fully contractible due to information asymmetry. Moreover, we determine the profit-maximizing quantities in such a way that profit-maximizing quantities are 'closer' to the patient-optimal quantities than in FFS or CAP, but do not coincide with them. Thus, we reduce the trade-offs between profit maximization and patient health benefit optimization.

We also run an experimental control condition that is described in the notes of Table 1.

Figure 2: Profit parameters in FFS/FFS+P4P and CAP/CAP+P4P



Notes: The upper panel of the figure illustrates profits in FFS and FFS+P4P, the lower panel analogously illustrates profits in CAP and CAP+P4P. Under basic payments, profits increase (in FFS) and decrease (in CAP) continuously, regardless of the severity of illness on the quantity interval. In the pay for performance conditions, a bonus payment is granted if the performance threshold $|q - q^*| \leq 1$ is reached. As the patient-optimal quantity q^* depends on the severity of illness, the performance thresholds differ accordingly. In the basic payment condition, the profit-maximizing quantity is $\hat{q}$=10 in FFS and $\hat{q}$=0 in CAP, respectively. In the pay for performance condition, $\hat{q}$ changes depending on the severity of illness.

2.3 Experimental protocol

The computerized experiment was programmed with z-Tree (Fischbacher, 2007) and conducted at elfe, the Essen Laboratory for Experimental Economics at the University of Duisburg-Essen. Overall, 100 students participated, of whom 87 students took part in our main conditions—42 subjects in condition 1 (FFS/FFS+P4P) and 45 in condition 2 (CAP/CAP+P4P); see Table 1. Among those were 44 medical students, of whom 22 each took part in conditions 1 and 2, and 13 subjects participated in the control condition. All subjects were recruited through the online recruiting system ORSEE (Greiner, 2015).

The procedure was as follows: Upon arrival, subjects were randomly allocated to cubicles. They then were given ample time to read the instructions for part I and to ask clarifying questions which were answered in private. Subjects were informed that the experiment consisted of two parts, but received detailed instructions for part II only after having finished part I of the experiment. To check for the subjects' understanding of the decision task, they had to answer a set of control questions. Instructions and comprehension questions are to be found in Appendices B and C. The experiment did not start unless all subjects had answered the control questions correctly. In each of the two parts of the experiment, subjects then subsequently decided on the quantity of medical services for each of the nine patients, i.e., for each possible combination of illnesses and severities. The order of patients was randomly determined and kept constant for all subjects and all conditions: $Bx, Cx, Az, By, Bz, Ay, Cz, Ax, Cy$.

Before making their decision for a specific patient, subjects were informed about their payment, their cost and profit, as well as about the patient benefit for each quantity from 0 to 10. All monetary amounts are given in Taler, our experimental currency, the exchange rate being 1 Taler = 0.08 EUR. The procedure was exactly the same in part II of the experiment. After finishing part II, we asked the subjects to complete a questionnaire on social demographics (e.g., age and gender) and on personality traits elicited by a ten-item personality inventory, which comprises five personality dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness (Gosling et al., 2003; Rammstedt and John, 2007). At the end of the experiment, when all subjects had made their decisions, we randomly determined one decision in each part of the experiment to be relevant for a subject's actual payoff and the patient benefit. This was done to rule out income effects. Subjects were paid in private according to these two randomly determined decisions.

To verify that the money corresponding to the sum of patient benefits in a session was actually transferred, we applied a procedure similar to Hennig-Schmidt et al. (2011) and Brosig-Koch et al. (forthcoming a, forthcoming b). One of the participants was randomly chosen to be the monitor. After the experiment, the monitor verified that a payment order on the aggregated benefit in the respective session was written to the financial department of the University of Duisburg-Essen to transfer the money to the Christoffel Blindenmission, which used the monetary transfers exclusively to support surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists from the charity. Notice that we did not inform the subjects that the money was assigned to a developing country.[8] The order was sealed in an envelope and the monitor and experimenter then walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional 5 EUR.

Sessions lasted for about 70 minutes. Subjects earned, on average, EUR 16.58. The average benefit per patient was EUR 13.38. In total, EUR 1,351.78 were transferred to the Christoffel Blindenmission. The average cost for a cataract operation amounts, according to the Christoffel Blindenmission, to about EUR 30. Thus, our experiment allowed 45 cataract patients to be treated.

3 Behavioral predictions

To organize our thoughts, to describe the physicians' behavior formally, and to derive behavioral predictions for our experiment, we introduce an illustrative model. Similar, for example, to Ellis and McGuire (1986), we assume that a physician derives utility from increasing his or her own profit and patients' health benefit. The weight the physician attaches to the patients' health benefit is interpreted as physician altruism. The assumption of an altruistic physician has become common in the health economics literature since Arrow (1963) coined the importance of physicians' patient-regarding motivation in the delivery of medical services. In addition to the importance for the design of optimal payment schemes (e.g., Ellis and McGuire, 1986, 1990; Ma, 1994; Chalkley and Malcomson, 1998; Jack, 2005; Choné and Ma, 2011), assumptions on a physician's altruism are essential, for example, to analyzing referral decisions (Allard et al., 2011) and the delegation of treatment decisions (Liu and Ma, 2013).

Let physician i choose the quantity of medical services q in order to maximize her utility

$$U_i(q) = \alpha_i B(q) + \beta_i \pi(q), \tag{3}$$

with $\alpha_i + \beta_i = 1$, $\alpha_i \geq 0$, and $\beta_i \geq 0$. α_i is a measure for physician i's altruism. For a purely profit-maximizing physician, for example, $\alpha_i = 0$ and $\beta_i = 1$. A profit-maximizing physician therefore obtains the highest utility, in our experiment, when choosing 10 medical services in FFS and when choosing 0 medical services in CAP.

First, we consider physician i's behavior under FFS and CAP. For profits and patient benefits given in our experiment, and the given altruism of physician i, we state the following lemma:[9]

Lemma 1. *Physician i overprovides medical services ($q > q^*$) if $p > q^*/5 + (\alpha_i/\beta_i)\,\theta$, and she underprovides medical services ($q < q^*$) if $p < q^*/5 - (\alpha_i/\beta_i)\,\theta$. Otherwise, physician i chooses the patient optimal quantity ($q = q^*$).*

[8] We wanted to avoid motives like compassion for people in developing countries that are independent of being in need of ophthalmic surgery. Feedback from the subjects in a pre-experimental pilot session in Hennig-Schmidt et al. (2011) actually raised this issue.

[9] Notice that Lemma 1 is a special case of Proposition 1 in Brosig-Koch et al. (forthcoming b). They consider a physician's behavior under mixed payment systems with a weight on a fee-for-service component and a lump-sum capitation.

Proof. Physician i's objective function $U_i(q) = \beta_i \pi(q) + \alpha_i B(q)$ is concave. Payment $R(q) = \Lambda + pq$ is linear and $-c(q)$ is concave as $c(q)$ is convex, thus $\pi(q)$ is a concave function. As $B(q)$ is also a concave function (as defined in equation 2) and $\alpha_i \geq 0$, it follows that $U_i(q)$ is concave.

Note that as $B(q)$ is not differentiable at $q = q^*$, with $q^* \in (0, 10)$. For $q < q^*$, the first-order condition $U_i'(q) = \beta_i \left[p - \frac{q}{5}\right] + \alpha_i \theta$. For $q > q^*$, the first-order condition $U'(q) = \beta_i \left[p - \frac{q}{5}\right] - \alpha_i \theta$. For $q > q^*$, consider $\lim_{q \to q^*} U_i'(q) = \beta_i \left[p - \frac{q^*}{5}\right] - \alpha_i \theta$. If $p < q^*/5 - (\alpha_i/\beta_i)\,\theta$, $\lim_{q \to q^*} U_i'(q)$ is positive. Also, because $U_i(q)$ is concave, $U_i'(q) > 0 \ \forall \ q < q^*$. Therefore any q such that $q \leq q^*$ cannot be optimal, i.e., physician i chooses $q > q^*$.

Analogously for $q < q^*$, consider $\lim_{q \to q^*} U_i'(q) = \beta_i \left[p - \frac{q^*}{5}\right] + \alpha_i \theta$. If $p < q^*/5 + (\alpha_i/\beta_i)\,\theta$, $\lim_{q \to q^*} U_i'(q)$ is negative. Also because $U_i(q)$ is concave, $U_i'(q) < 0 \ \forall \ q > q^*$. Therefore any q such that $q \geq q^*$ cannot be optimal, i.e., physician i chooses $q < q^*$. □

It directly follows from Lemma 1 that physician i's provision behavior depends on the severity of illness (i.e., the patient-optimal quantity q^* varying with severity of illness l), the fee for a medical service p, the marginal patient health benefit θ, and α_i, the physician i's degree of altruism. Intuitively, the higher physician i's altruism is towards her patient, the lower the degree of non-optimal service provision is. Based on Lemma 1, we hypothesize for the physicians' behavior in FFS and CAP:

Hypothesis 1. *Fee-for-service and capitation induce overprovision and underprovision of medical services, respectively. Overprovision under fee-for-service decreases and underprovision under capitation increases in the severity of a patient's illness. Also, the overprovision and underprovision of medical services decrease in patients' marginal health benefit.*

The second part of Hypothesis 1, the impact of the patients' severity of illness, adds to findings from recent behavioral experiments (e.g., Hennig-Schmidt et al., 2011; Brosig-Koch et al., forthcoming b) and empirical studies (e.g., Clemens and Gottlieb, 2014). A behavioral effect of the severity of illness is particularly relevant as, in our experiment, the levels of performance pay depend on the severity of illness.[10]

We now focus on the effect of introducing performance pay on physicians' behavior. Recall that a performance pay $b_l^\bullet$ is granted if a physician's quantity decision fulfills the quality requirement (i.e., if $|q - q^*| \leq 1$ for a patient with a severity of illness l). By linking performance pay to the patient's health benefit, the interests of physician and patient become (more) aligned. This means that less altruistic physicians are incentivized also to provide medical services 'close' to the patient-optimal quantity. Comparing physician i's provision behavior between FFS (CAP) and FFS+P4P (CAP+P4P), we state the following proposition:

[10] In our experiment, the performance pay for mild-severity patients, b_x^{FFS}, is highest compared to b_y^{FFS} and b_z^{FFS}, as the 'risk' that a mild-severity patient is overserved and therefore suffers disutility is highest under FFS. In CAP, the incentive to undertreat patients increases in the severity of illness; therefore, $b_z^{\mathrm{CAP}} > b_y^{\mathrm{CAP}} > b_x^{\mathrm{CAP}}$ (see Table 1).

Proposition 1. *Performance pay linked to the optimal patient's health benefit reduces physicians' overprovision of medical services in fee-for-service and underprovision in capitation.*

Proof. Let $q^{\text{Opt.}}$ be a physician's utility-maximizing choice for a patient j under FFS or CAP. Depending on a physician's quantity choice, we distinguish three cases. First, we consider $q^{\text{Opt.}} \in [q^* - 1, q^* + 1]$. As $b_l^{\bullet} > 0$, it follows that a physician with $\alpha_i \in [0, 1)$ does not change her behavior since $b_l^{\bullet} > 0$ is a constant. Second, we consider $q^{\text{Opt.}} > q^* + 1$. Here, the physician chooses q according to $\max\{U^{II}(q^{\text{Opt.}}), U(q^* + 1) + b_l^{\bullet}\}$. That means a physician either does not change her behavior or chooses $q^* + 1$ when performance pay has been introduced. Analogously for $q^{\text{Opt.}} < q^* - 1$, the same logic applies. □

Intuitively, a physician's medical service provision might meet the quality requirement ($|q - q^*| \leq 1$) for a patient or it might not meet it. This depends, according to Lemma 1, on physician i's degree of altruism towards the patient counterbalancing the incentive effects in FFS and CAP. For a physician's given altruism with $\alpha_i \in [0, 1)$, introducing performance pay therefore reduces non-optimal service provision under FFS and CAP. Based on Proposition 1, we state the following hypothesis for the effect of performance pay on physicians' behavior:

Hypothesis 2. *Introducing performance pay, which relates to the patient's optimal health benefit, reduces overprovision of medical services under fee-for-service and underprovision under capitation, respectively.*

We next consider the adverse effect of a crowding-out of patient-regarding behavior, often assumed to be inherent in performance pay for physicians (e.g., Siciliani, 2009; Maynard, 2012).[11] We say that a crowding-out under performance pay occurs if a physician deviates from her previous treatment decision such that her profit increases and the patient health benefit is reduced at the same time. Depending on the parameters of the choice environment, crowding-out can occur when treatment behavior in non-blended payment systems changes to choices under performance pay that raise profit and lower patient health benefit simultaneously. Given the parameters of our experiment, however, crowding-out can only occur whenever subjects change their medical service provision from choosing the patient-optimal quantity q^* in FFS or CAP to $\hat{q}$ in FFS+P4P or CAP+P4P.

To model the crowding-out of patient-regarding behavior, we make use of a multi-attribute utility theory framework (e.g., Keeney and Raiffa, 1976). In particular, we introduce a performance pay 'attribute' to a physician's utility function, and rewrite equation (3) as follows:

$$\tilde{U}_i = \tilde{\alpha}_i B(q) + \beta_{1i} \pi(q) + \beta_{2i} b_l^{\bullet}, \tag{4}$$

with $\beta_{1i} + \beta_{2i} + \tilde{\alpha}_i = 1$ and $\beta_{1i}, \beta_{2i}, \tilde{\alpha}_i \geq 0$.

Intuitively, the introduction of performance pay could lead physicians to place more em-

[11] Taking a more general perspective, the crowding-out of intrinsic *motivation* is a widely discussed phenomenon in health and labor contexts. Also, some evidence exists that performance pay crowds out the individuals' intrinsic motivation (e.g., Gneezy and Rustichini, 2000; Mellström and Johannesson, 2008).

phasis on profit and could therefore lead to an increase in the overall weight on profit. This phenomenon, well-known in decision theory, is called a 'splitting effect' (e.g., Weber et al., 1988). We make use of the splitting effect which, in our model, implies that $\beta_{1i} + \beta_{2i} > \beta_i$. Under the additional assumption that $\frac{\tilde{\alpha}_i}{\beta_{1i}} < \frac{\alpha_i}{\beta_i} = \frac{\alpha_i}{1-\alpha_i}$ and $\text{argmax}_q U_i = q^*$, this leads us to the following crowding-out proposition:[12]

Proposition 2. *For physicians who choose the patient-optimal treatment under fee-for-service or capitation, performance pay might lead to a crowding-out of altruistic, patient-regarding behavior. For a payment system with fee-for-service and performance pay linked to the patient's optimal health benefit, a physician deviates from choosing the patient-optimal quantity of medical service* ($\text{argmax}_{q\in[q^*-1,q^*+1]}\tilde{U} > q^*$), *if* $p \in \left[\frac{q^*}{5} + \frac{\tilde{\alpha}_i}{\beta_{1i}}\theta, \frac{q^*}{5} + \frac{\alpha_i}{1-\alpha_i}\theta\right]$. Analogously, for a system with capitation and performance pay, $\text{argmax}_{q\in[q^*-1,q^*+1]}\tilde{U}_i < q^*$ if $p \in \left[\frac{q^*}{5} - \frac{\tilde{\alpha}_i}{\beta_{1i}}\theta, \frac{q^*}{5} - \frac{\alpha_i}{1-\alpha_i}\theta\right]$.

Proof. Without loss of generality, we write for a payment system with fee-for-service and performance pay $\text{argmax}_{q\in[q^*-1,q^*+1]}\tilde{U}_i = \text{argmax}_{q\in[q^*-1,q^*+1]}(\beta_{1i}\pi(q) + \beta_{2i}b_l^\bullet + \tilde{\alpha}_i B(q))$ $= \text{argmax}_{q\in[q^*-1,q^*+1]}(\beta_{1i}\pi(q) + \tilde{\alpha}_i B(q))$. Using the assumption that $\frac{\tilde{\alpha}_i}{\beta_{1i}} < \frac{\alpha_i}{\beta_i} = \frac{\alpha_i}{1-\alpha_i}$, then the result follows immediately from Lemma 1 for U_i and $\tilde{U}_i$. □

Based on Proposition 2, we state the following hypothesis regarding the crowding-out of physicians' patient-regarding (altruistic) behavior:

Hypothesis 3. *Performance pay might induce crowding-out of patient-regarding behavior for some physicians who choose the patient-optimal treatment under non-blended payment systems.*

Intuitively, one could argue that, for physicians who choose the patient-optimal treatment q^* under capitation or fee-for-service, the quality of care provided to a patient is already very important. The introduction of the performance pay component now suggests to these physicians that treating the patient (perfectly) optimally is not of great importance anymore.

4 Results

4.1 Fee-for-service, capitation, and severities of illness

In the following, we analyze medical service provision according to Hypothesis 1, focusing on the deviation of choices from the optimal treatment of patients. We apply between-subject

[12]Note that Proposition 1 addresses the effect of performance pay on non patient-optimal behaviors under FFS and CAP. Proposition 2, however, particularly focuses on the effect of performance pay on patient-optimal choices under FFS and CAP. These are the quantity choices that allow for crowding-out in our experiment. We, therefore, deliberately extend our illustrative model (equation 3) as it so far does not consider the crowding-out of patient-regarding behavior for a given α_i—being constant in FFS, CAP, and under performance pay systems. Using the multi-attribute utility function (equation 4), we would arrive at a proposition similar to Proposition 1, however, as Proposition 1 addresses behavior outside the quantity interval meeting the quality requirement.

comparisons of the baseline payment systems FFS and CAP. In particular, we focus on the impact of the severity of illness on provision behavior under the two payment systems. First, we investigate Hypothesis 1 by analyzing provision behavior at an aggregate level and employ non-parametric statistical testing.[13] Second, we run a series of regressions to analyze the impact of the severity of illness, the marginal health benefit, and subjects' characteristics (i.e., medical background, subjects' demographics, and personality traits).

We find that subjects in our experiment respond to the incentives inherent in the respective

Table 2: Quantities of medical services and patient health benefits by payment system

	Quantity q		Benefit $B(q)$		
Payment system	Mean	s.d.	Mean	s.d.	N
FFS	7.14	1.88	7.47	3.39	378
FFS+P4P	5.74	1.61	9.36	2.70	378
CAP	3.13	2.15	7.76	3.60	405
CAP+P4P	4.36	1.71	9.41	2.84	405

Notes: This table shows descriptive statistics on the quantity of medical services q and patients' health benefit $B(q)$ at payment system level. Note that 42 subjects (22 medical and 20 non-medical students) each decide for nine patients on the quantity of medical services under FFS and FFS+P4P, while 45 subjects (22 medical and 23 non-medical students) each decide for nine patients under CAP and CAP+P4P. In our CAP-control condition with $\Lambda = 12$, we found very similar quantity choices compared to CAP.

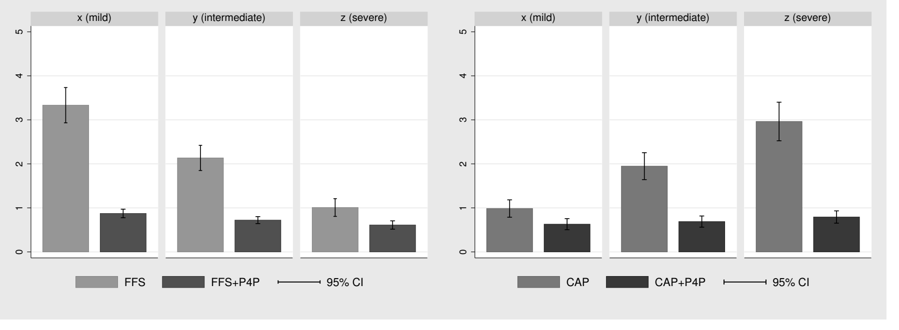
payment scheme FFS or CAP. Table 2 shows descriptive statistics on the quantity of medical services for all payment systems. Overall, about 56% fewer medical services are provided in CAP than in FFS. This difference is highly significant ($p < 0.001$, two-sided Mann-Whitney U-Test).

Behavioral data support Hypothesis 1, as we find overprovision of medical services in FFS and underprovision in CAP. Overall, the deviation of the chosen quantity from the patient-optimal quantity is 2.14 (s.d. 2.00) in FFS and −1.87 (s.d. 2.18) in CAP. Non-parametric tests

[13]Throughout the paper, p-values are reported from two-sided tests. If not reported otherwise, we cluster individual data by averaging for each individual subject over his or her nine quantity decisions in part I (part II) of the experiment. We apply non-parametric tests to the individual subjects' averages. For between-subject analyses, we employ Mann-Whitney U-tests; for within-subject analyses, we use the Wilcoxon matched-pairs signed rank-test. We also employ the Fisher-Pitman permutation test for unpaired samples and Fisher-Pitman permutation test for paired samples for between- and within-subject comparisons, respectively. Notice that test results from the Fisher-Pitman tests yield very similar results. Another alternative approach for the within-subject analyses is to apply a non-parametric test to individual quantity decisions in part I and in part II of the experiment, and to cluster the data per individual instead of applying the test to averages. We checked for the robustness of our results using this clustering approach by applying Somers' D. Both approaches yield qualitatively very similar findings.

show that deviations averaged over all subjects and computed for each of the nine patients are significant in both payment systems ($p < 0.001$, Wilcoxon signed-rank test). The extent of non-optimal medical service provision $|q - q^*|$ is not significantly different in FFS compared to CAP ($p = 0.316$, Mann Whitney U-Test). We also find no evidence for differences in patient health benefit across the basic payment systems ($p = 0.233$, Mann Whitney U-Test).

Figure 3: Differences to patient-optimal quantity of medical services by payment system



Notes: This figure shows the absolute average deviations from the patient optimal quantity $|q - q^*|$, for the three severities of illness. 95% CI are also indicated. The left panel reports average deviations under payment systems FFS and FFS+P4P. For each severity, 42 subjects made three decisions in each payment system ($N = 756$). The right panel shows average non-optimal service provision for payment systems CAP and CAP+P4P. 45 subjects made three decisions in each payment system for each severity ($N = 810$).

Figure 3 shows that the extent of non-optimal medical service provision varies with the severity of illness. In FFS, it decreases in a patient's severity of illness. It is highest for mild-severity patients with the average overprovision amounting to more than three medical services. Average overprovision is about two medical services for patients suffering from an intermediate severity and one medical service for patients with a high severity of illness. In CAP, the non-optimal service provision is in the reverse order and increases with a patient's severity of illness.

Regression analyses further emphasize that the severity of illness significantly influences subjects' medical service provision; see Table 3. Estimation results in model (1) and Wald-test results show that overprovision under FFS significantly decreases with increasing severity of illness. The illnesses characterized by the marginal health benefit θ does not affect optimal service provision. The effect is robust when including subjects' characteristics, such as being a medical student or not,[14] gender, age, and personality traits; see model (2). There is no significant relationship between overprovision and the marginal health benefit. In CAP, underprovision significantly increases with increasing severity of illness; see model (3) and Wald-test results.

[14]Notice that we also control for possible behavioral differences between medical and non-medical students. Recent experimental evidence (e.g., Hennig-Schmidt et al., 2011; Brosig-Koch et al., forthcoming a, forthcoming b) indicates that, although medical students and non-medical students respond to incentives in a similar way, their responses might be of different intensities.

Table 3: Impact of patients' characteristics on non-optimal medical service provision

	Fee-for-service, FFS		Capitation, CAP	
Model:	(1)	(2)	(3)	(4)
Dependent variable:	$\lvert q-q^*\rvert$	$\lvert q-q^*\rvert$	$\lvert q-q^*\rvert$	$\lvert q-q^*\rvert$
Intermediate severity (= 1 if $l = y$)	-1.198***	-1.198***	0.963***	0.963***
	(0.153)	(0.155)	(0.161)	(0.163)
High severity (= 1 if $l = z$)	-2.325***	-2.325***	1.978***	1.978***
	(0.237)	(0.239)	(0.288)	(0.290)
Illness, marginal health benefit (= 1 if $\theta = 2$)	-0.048	-0.048	-0.215**	-0.215**
	(0.065)	(0.066)	(0.099)	(0.100)
Subjects' characteristics	No	Yes	No	Yes
Constant	3.349***	0.878	1.057***	3.709*
	(0.337)	(2.371)	(0.156)	(2.228)
Wald test				
H_0 : inter. severity vs. high severity (p-value)	0.000	0.000	0.000	0.000
R^2	0.231	0.301	0.152	0.343
Observations	378	378	405	405
Subjects	42	42	45	45

Notes: OLS estimates are reported with robust standard errors clustered for subjects (in brackets). The reference category is the 'mild severity of illness', $l = x$. Illness is a dummy for indicating two different values of marginal health benefits (= 1 if $\theta = 2$ for illness C and = 0 if $\theta = 1$ for illness A, B). Subjects' characteristics comprise gender, age, a dummy for medical student, and BIG-five personality traits according to Gosling et al. (2003). Notice that Tobit regressions (censored at 10 for FFS and at zero for CAP) yield very similar estimation results. Wald tests indicate that the coefficients for intermediate and high severities are highly significantly different from each other ($p < 0.001$). *** $p < 0.01$, ** $p < 0.05$, and *$p < 0.1$.

Moreover, we find that underprovision of medical services in CAP slightly decreases with increasing marginal health benefits. The effect is robust when including subjects' characteristics; see model (4). In sum, we state the following result:

Result 1. *Fee-for-service induces overprovision of medical services, while under capitation patients are underserved. The severity of illness significantly affects deviations from the patient-optimal quantity of medical services. Overprovision under fee-for-service decreases with increasing severity of illness. Under capitation, underprovision increases when an illness becomes more severe. Under capitation, deviations from the patient-optimal quantity are also significantly affected by the marginal health benefit.*

The behavioral results on the provision of medical services under FFS and CAP are in line with findings reported in earlier empirical and experimental studies (e.g., Gaynor and Gertler 1995; Hennig-Schmidt et al. 2011; Brosig-Koch et al. forthcoming a, forthcoming b).

Regarding the severity of illness, our results complement findings of related lab experiments (e.g., Hennig-Schmidt et al., 2011; Brosig-Koch et al., forthcoming a) and of empirical studies emphasizing the interaction between the physicians' responses to incentives and the patients' severity of illness (e.g., Clemens and Gottlieb, 2014). Moreover, these results provide a rationale for adjusting the performance pay for different severities of illness.

4.2 Provision behavior under performance pay

We now study the effect of introducing performance pay. On aggregate, quantities of medical services decrease by about 20% under FFS+P4P compared to FFS. In CAP+P4P, the quantity of medical services increases by about 40% compared to CAP; see Table 2. Accordingly, the aggregate patients' health benefits are significantly higher when introducing performance pay ($p < 0.001$, Wilcoxon signed-rank test).

We now analyze whether performance pay affects subjects' behavior according to Hypothesis 2. To this end, we investigate the extent to which subjects deviate from the patient-optimal quantity of medical services. We find that implementing performance pay significantly reduces underprovision of medical services in CAP ($p < 0.001$, Wilcoxon signed-rank test). Likewise, overprovision of medical services in FFS significantly decreases ($p < 0.001$, Wilcoxon signed-rank test). Figure 3 indicates that the reduction in non-optimal medical service provision differs with the patients' severity of illness under performance pay also. The reduction in non-optimal service provision is highest for patients suffering from a mild illness ($l = x$) in FFS (about 74%) and for patients with a high severity of illness ($l = z$) in CAP (about 73%).

Table 4 shows estimation results from a series of OLS regressions. Estimation results of models (1) and (5) lend further support for Hypothesis 2, namely, that adding a performance pay tied to the patient-optimal quantity in either fee-for-service or capitation reduces the non-optimal service provision.

The main behavioral results are robust to including patients' and subjects' characteristics in the analysis. As already suggested by descriptive statistics, the estimation results in models (2) and (6) show that the deviation from patient-optimal service provision is affected by the severity of the patient's illness. In Condition 1 (FFS and FFS+P4P), non-optimal service provision is significantly lower for the intermediately and the severely ill patients compared to the reference category "mild illness" ($l = x$). In contrast, in Condition 2 (CAP and CAP+P4P), the non-optimal service provision is significantly higher for these patients compared to the reference category. In the latter condition, illness—the marginal health benefit θ—relates negatively to non-optimal service provision; see again model (6). When adding controls for subjects' medical major and further personal characteristics such as gender, age and personality traits, the estimation results remain (almost) unaffected; see models (3), (4), (7) and (8). In sum, we state the following result:

Result 2. *Introducing performance pay, on aggregate, reduces overprovision of medical services under fee-for-service and underprovision under capitation of medical services. The reduction in*

Table 4: Effect of performance pay on the quality of medical service provision

	A. Condition 1: FFS and FFS+P4P				B. Condition 2: CAP and CAP+P4P			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$	$\|q-q^*\|$
FFS+P4P	-1.423***	-1.423***	-1.423***	-1.423***				
	(0.195)	(0.195)	(0.195)	(0.196)				
CAP+P4P					-1.262***	-1.262***	-1.262***	-1.262***
					(0.206)	(0.207)	(0.207)	(0.208)
Intermediate severity (= 1 if $l = y$)		-0.675***	-0.675***	-0.675***		0.511***	0.511***	0.511***
		(0.091)	(0.091)	(0.092)		(0.0854)	(0.086)	(0.086)
High severity (= 1 if $l = y$)		-1.294***	-1.294***	-1.294***		1.070***	1.070***	1.070***
		(0.129)	(0.130)	(0.130)		(0.155)	(0.155)	(0.155)
Illness, marginal health benefit (= 1 if $\theta = 2$)		-0.046	-0.046	-0.460		-0.135***	-0.135***	-0.069**
		(0.060)	(0.061)	(0.061)		(0.047)	(0.047)	(0.048)
Med (= 1 if medical student)			-0.371	-0.414			0.047	0.250
			(0.258)	(0.328)			(0.328)	(0.275)
Sex (= 1 if female)			-0.107	-0.003			0.707*	0.538*
			(0.266)	(0.261)			(0.365)	(0.319)
Age			0.035	0.044			0.007	0.011
			(0.030)	(0.030)			(0.021)	(0.026)
Personality traits	No	No	No	Yes	No	No	No	Yes
Constant	7.100***	7.315***	6.606***	4.929***	3.227***	2.745***	1.672**	4.737***
	(0.805)	(0.880)	(1.190)	(1.514)	(0.445)	(0.389)	(0.816)	(1.371)
Observations	756	756	756	756	810	810	810	810
Subjects	42	42	42	42	45	45	45	45
R^2	0.195	0.314	0.323	0.339	0.139	0.206	0.243	0.302

Notes: OLS estimates are reported with robust standard errors clustered for subjects (in brackets). The reference categories in Models (1) and (5) are FFS and CAP for conditions 1 and 2, respectively. The reference category for severity of illness is $l = x$. Illness is a dummy for indicating two different values of marginal health benefits (= 1 if $\theta = 2$, for illness C, and = 0 if $\theta = 1$ for illness A, B). The subjects' characteristics comprise gender, age, a dummy for medical student major, and BIG-five personality traits according to Gosling et al. (2003). Notice that Tobit regressions (censored at zero) yield very similar estimation results. Wald tests indicate that the coefficients for intermediate and high severities are significantly different from each other ($p < 0.001$). *** $p < 0.01$, ** $p < 0.05$, and *$p < 0.1$.

non-optimal service provision is significantly affected by the severity of illness.

The aggregate effect of performance pay on the physicians' quality in our behavioral data relates to recent experimental studies on physician behavior under non-linear performance pay. In Keser et al.'s (2014) chosen effort experiment, which closely relates to Hennig-Schmidt et al. (2011), performance pay tied to the share of patients treated optimally leads to some increase in the physicians' quality, compared to a fee-for-service system. Lagarde and Blaauw (2015), who employ a real-effort experiment with a bonus and fine condition, report that performance pay enhances performance in the medically framed data entry task. In a hospital frame, Cox et al. (forthcoming) find that utilizing performance pay mechanisms incentivizes cost-effective reductions in hospital readmissions.

4.3 Individual patterns: Crowding-out of patient-regarding behavior

We next analyze the dynamics of treatment patterns for individual patients between the non-blended payment systems FFS and CAP and the blended performance pay systems FFS+P4P and CAP+P4P. By means of transition matrices, we analyze, for each physician and each of the nine patients, which quantities are chosen in both parts of the experiment and how the decisions change when the payment scheme is modified; for a related analysis, see Godager et al. (forthcoming). Investigating choice dynamics and behavioral patterns allows us to gain important insights into the driving forces of subjects' choices on medical service provision. According to Hypothesis 3, we analyze changes in treatment patterns while emphasizing the crowding-out of patient-regarding behavior by performance pay.

Under the non-blended payment systems FFS and CAP, on aggregate 228 of all 783 patients (about 29%) are treated optimally. In FFS, for 103 of 378 patients (27.25%) and, in CAP, for 125 of 405 patients (30.87%), subjects provide the patient-optimal quantity q^*; see Table 5. The profit-maximizing quantity of medical services $\hat{q}$ is provided for 125 of the 738 patients (15.96%): 52 (13.76)% under FFS and 73 (18.02)% under CAP. Two (five) subjects choose to provide $\hat{q}$ for each of the nine patients in FFS (CAP). A rather large fraction of subjects provide Pareto-efficient quantities of medical services, which differ from $\hat{q}$ and q^*. We label this treatment pattern 'Other Pareto-efficient' (OPE); here subjects trade off their own profit against the patient benefit when treating their patients. On aggregate, 415 of all 783 patients (53%) are treated according to OPE under the non-blended payment systems: in FFS, 220 of 378 patients (58.20%); and in CAP, 195 of 405 patients (48.15%), see Table 5.[15]

Under performance pay, the frequency of treatment choices, consistent with the patient-optimal quantity q^* and the profit-maximizing quantity $\hat{q}$, increases. For 28.57% (37.28%) of

[15] The choice of q^{OPE} includes choices which are characterized by, for example, deviating by one or two quantity units from q^*, or which involve choosing the social optimum (choice of q such that $\max(\pi + B)$) that can be distinguished from q^* under FFS (CAP) for patients 1, 8 (3, 5) only. We also observe a very small number of Pareto-inefficient choices which are neither patient-optimal, nor profit-maximal nor in the interval between the two quantities. (i.e., three out of 378 choices in FFS and 12 out of 405 choices in CAP).

Table 5: Treatment patterns for non-blended FFS, CAP and blended performance pay systems

Payment system	Treatment pattern	Mild illness ($q_x^* = 3$) abs. freq.	Mild illness rel. freq. (in %)	Interm. illness ($q_y^* = 5$) abs. freq.	Interm. illness rel. freq. (in %)	Severe illness ($q_z^* = 7$) abs. freq.	Severe illness rel. freq. (in %)	Aggregated abs. freq.	Share of total choices (in %)
FFS	Patient-optimal, q_l^*	17	4.50	26	6.88	60	15.87	103	27.25
	Profit-maximizing, $\hat{q}$	17	4.50	14	3.70	21	5.56	52	13.76
	Other Pareto-efficient, q^{OPE}	92	24.34	85	22.49	43	11.38	220	58.20
	Pareto-inefficient choices	0	0.00	1	0.26	2	0.53	3	0.79
FFS+P4P	Patient-optimal, q_l^*	22	5.82	35	9.26	51	13.49	108	28.57
	Profit-maximizing, $\hat{q}$	102	26.98	91	24.08	74	19.58	267	70.64
	Pareto-inefficient choices	2	0.53	0	0.00	1	0.26	3	0.79
	$\Delta\ q_l^*$ (FFS to FFS+P4P)	5	1.32	9	2.38	-9	-2.38	5	1.32
	$\Delta\ \hat{q}$ (FFS to FFS+P4P)	85	22.49	77	20.37	53	14.02	215	56.88
CAP	Patient-optimal, q_l^*	67	16.54	35	8.64	23	5.68	125	30.86
	Profit-maximizing, $\hat{q}$	22	5.43	23	5.68	28	6.92	73	18.03
	Other Pareto-efficient, q^{OPE}	41	10.12	72	17.78	82	20.25	195	48.15
	Pareto-inefficient choices	5	1.23	5	1.23	2	0.49	12	2.96
CAP+P4P	Patient-optimal, q_l^*	59	14.57	52	12.84	40	9.88	151	37.28
	Profit-maximizing, $\hat{q}$	71	17.53	80	19.75	91	22.47	242	59.75
	Pareto-inefficient choices	5	1.23	3	0.74	4	0.99	12	2.96
	$\Delta\ q_l^*$ (CAP to CAP+P4P)	-8	-1.98	17	4.20	17	4.20	26	6.42
	$\Delta\ \hat{q}$ (CAP to CAP+P4P)	49	12.10	57	4.07	63	15.56	169	41.73

Notes: This table shows absolute and relative frequencies of treatment patterns (i) patient-optimal, q_l^*, (ii) profit-maximizing, $\hat{q}$, (iii) other Pareto-efficient, q^{OPE}, and (iv) Pareto-inefficient choices for non-blended payment systems FFS, CAP, and blended performance pay systems FFS+P4P, CAP+P4P differentiated by the severity of illness. We define Pareto-efficient choices in our experiment as choices which do not allow to increase patient benefit and own profit simultaneously. Notice that, under performance pay, Pareto-efficient choices are either q_l^* or $\hat{q}$, while all other choice options are Pareto-inefficient.

all patients, the patient-optimal quantity is chosen in FFS+P4P (CAP+P4P), an increase by 1.32 (6.42) percentage points (see Table 5). Subjects choose the profit-maximizing quantity $\hat{q}$ to a much larger extent under performance pay than under the non-blended payment systems. On the aggregate, 70.64% of all choices in FFS+P4P and 59.75% in CAP+P4P coincide with $\hat{q}$, which constitutes an increase of 56.88 (41.73) percentage points from FFS to FFS+P4P (from CAP to CAP+P4P). The number of subjects choosing $\hat{q}$ for each of the nine patients triples by rising from three to 16 in FFS+P4P and from five to 14 in CAP+P4P. OPE choice alternatives do not exist in the pay-for-performance systems, as $\hat{q} = q^* + 1(-1)$ in FFS+P4P (CAP+P4P).

We now analyze the behavioral dynamics in treatment patterns when subjects face blended performance pay systems after having decided under the non-blended payment systems. We identified the following six transition patterns, which explain 96% of our behavioral data (see Table 6):

- ($\hat{q} \rightarrow q^*$): profit-maximizing quantity $\hat{q}$ in non-blended payment systems and patient-optimal quantity q^* under performance pay;
- ($q^* \rightarrow \hat{q}$): patient-optimal quantity q^* in non-blended payment systems and profit-maximizing quantity $\hat{q}$ under performance pay—crowding-out of patient-regarding behavior;
- ($q^{\text{OPE}} \rightarrow q^*$): quantity between the two polar quantities q^* and $\hat{q}$ in non-blended payment systems and transitioning to the patient-optimal quantity q^* under performance pay;
- ($q^{\text{OPE}} \rightarrow \hat{q}$): quantity between the two polar quantities q^* and $\hat{q}$ in the non-blended payment systems and transitioning to the profit-maximizing quantity $\hat{q}$ under performance pay;
- ($q^* \leftrightarrow q^*$): patient-optimal quantity q^* in both blended and non-blended payment systems;
- ($\hat{q} \leftrightarrow \hat{q}$): profit-maximizing quantity $\hat{q}$ in both blended and non-blended payment systems.

Other choice patterns include at least one Pareto-inefficient choice and are not captured by the above categories.

Table 6 shows the dynamics in treatment patterns aggregated and, separately, for the three severities of illness x, y, z. For an illustration of the dynamics in behavioral patterns at an individual patient level, see Figures A1 and A2 in Appendix A. The most frequent behavioral pattern we observe in our experimental data is ($q^{\text{OPE}} \rightarrow \hat{q}$), which amounts to 49.21% of all patients under FFS-payment systems and 34.81% of all patients under CAP-payment systems. For the FFS-systems, we observe that this pattern decreases with the severity of illness, while for CAP-systems the reverse holds. The second most common behavioral pattern is ($q^* \leftrightarrow q^*$) implying 18.78% in the FFS-payment systems and 21.73% in the CAP-systems. We observe the behavioral pattern ($\hat{q} \leftrightarrow \hat{q}$) for 12.70% and 15.56% of the choices in FFS-payment systems and in CAP-payment systems, respectively. Transitioning from q^{OPE} to q^* amounts to 8.20% and 12.59% in FFS-payment systems and in CAP-payment systems, respectively. Finally, in both payment conditions, the dynamic treatment pattern ($\hat{q} \rightarrow q^*$) occurs at a rather low frequency of around 1%. All the above dynamic patterns imply an increase in patient-regarding behavior under performance pay except for ($q^* \leftrightarrow q^*$), which provides the highest possible benefit already under the non-blended systems. That means introducing performance pay induced an increase in patient-regarding behavior for 71.16% of the patients under fee for service and for 63.70% under capitation. Intuitively, these behavioral dynamics are in line with a healthcare policy-maker's intended aim to increase patients' health benefit by introducing performance pay.

We now deal in more detail with the crowding-out of patient-regarding behavior ($q^* \to \hat{q}$) according to Hypothesis 3. The crowding-out of patient-regarding behavior occurs if introducing performance pay induces a subject to deviate from choosing q^* under the previous payment scheme such that the patient health benefit is reduced while the own profit is increased at the same time. More specifically, in our experiment, crowding-out for Pareto-efficient decisions can only occur when subjects change their quantity choice from q^* in FFS or CAP to $\hat{q}$ in FFS+P4P or CAP+P4P as, under performance pay, q^* and $\hat{q}$ differ only by one unit of medical services (i.e., $|\hat{q} - q^*| = 1$).

We do find evidence for a crowding-out of subjects' patient-regarding behavior in our experimental data. For ($q^* \to \hat{q}$), the percentage of all subjects' choices is 8.47 under FFS-payment systems and 8.89 under CAP-payment systems. The overall evidence for crowding-out does not seem to be very high in our experiment. However, when relating a crowding-out to q^*-choices under CAP and FFS only—in particular, to those choices for which a crowding-out can actually occur under performance pay in our experiment—31.07% and 28.80% of patients suffer from a crowding-out of patient-regarding behavior in FFS-systems and CAP-systems, respectively.

Finally, we find that the crowding-out of patient-regarding behavior varies with the severities of illness. For the FFS-systems, the crowding-out is highest for the high-severity patients (z) and lowest for low-severity patients (x). We observe the reverse for CAP-systems (see Table 6).[16]

Result 3. *While performance pay (linked to the patient-optimal health benefit) incentivizes subjects towards a more patient-regarding behavior it induces, at the same time, a crowding-out of patient-regarding behavior for around one third of the patients who have been treated optimally prior to the introduction of performance pay.*

4.4 Discussion: Benefits and costs of introducing performance pay

Pay for performance schemes have been increasingly used as a means to encourage physicians to improve their quality of care (for comprehensive reviews, see Emmert et al., 2012; Eijkenaar et al. 2013). Most research has focussed on the effect of performance pay initiatives on the targeted quality measures, thereby often neglecting the pertinent issue of their effect on health outcomes and costs (Meacock et al., 2014). In the following, we address this issue by analyzing,

[16]The numbers in FFS-systems are 5 out of 17 choices for illness x, 6 out of 26 choices for illness y, and 21 out of 60 choices for illness z. The numbers for CAP-systems are 23 out of 67 choices for illness x, 10 out of 35 choices for illness y, and 3 out of 23 choices for illness z, see Tables 5 and 6.

Table 6: Dynamics in treatment patterns between non-blended FFS, CAP and blended performance pay systems

Experimental	Dynamics in	Mild illness ($q_x^* = 3$)		Interm. illness ($q_y^* = 5$)		Severe illness ($q_z^* = 7$)		Aggregated	Share of total
condition	treatment patterns	abs. freq.	rel. freq. (in %)	abs. freq.	rel. freq. (in %)	abs. freq.	rel. freq. (in %)	abs. freq.	choices (in %)
FFS, FFS+P4P	$\hat{q} \to q^*$	0	0.00	0	0.00	4	1.06	4	1.06
	$q^* \to \hat{q}$, crowding-out	5	1.32	6	1.59	21	5.56	32	8.47
	$q^{\text{OPE}} \to q^*$	10	2.65	15	3.97	6	1.59	31	8.20
	$q^{\text{OPE}} \to \hat{q}$	80	21.16	70	18.52	36	9.52	186	49.21
	$q^* \leftrightarrow q^*$	12	3.17	20	5.29	39	10.32	71	18.78
	$\hat{q} \leftrightarrow \hat{q}$	17	4.50	14	3.70	17	4.50	48	12.70
	Other	2	0.53	1	0.26	3	0.79	6	1.59
	Total	126	33.33	126	33.33	126	33.33	378	100.00
CAP, CAP+P4P	$\hat{q} \to q^*$	2	0.49	1	0.25	0	0.00	3	0.74
	$q^* \to \hat{q}$, crowding-out	23	5.68	10	2.47	3	0.74	36	8.89
	$q^{\text{OPE}} \to q^*$	11	2.72	22	5.43	18	4.44	51	12.59
	$q^{\text{OPE}} \to \hat{q}$	30	7.41	49	12.10	62	15.31	141	34.81
	$q^* \leftrightarrow q^*$	43	10.62	25	6.17	20	4.94	88	21.73
	$\hat{q} \leftrightarrow \hat{q}$	17	4.20	20	4.94	26	6.42	63	15.56
	Other	9	2.22	8	1.98	6	1.48	23	5.68
	Total	135	33.33	135	33.33	135	33.33	405	100.00

Notes: This table shows absolute and relative frequencies for dynamic treatment patterns when subjects are sequentially confronted with blended and non-blended FFS and CAP-systems for the three severities of illness (x, y, z). Category 'Other' comprises Pareto-inefficient choices in either the non-blended or the blended payment system, or in both.

within the confines of our experiment, the effect of introducing performance pay on health outcomes (measured as patients' health benefit) and costs of incentive payments.[17]

We find that the patients' health benefit amounts, on aggregate, to 7.47 in FFS and to 7.76 in CAP (see Table 7). Under performance pay, the patient benefit significantly increases to 9.36 in FFS+P4P and to 9.41 in CAP+P4P ($p < 0.001$, Wilcoxon signed rank-test). The increase in the patients' health benefit is mainly driven by subjects' choices of $\hat{q}$ in FFS+P4P or CAP+P4P, which involves an increase in health benefit compared to the non-blended payment schemes. In FFS+P4P and CAP+P4P, patients' health benefit increases, on aggregate, for 63.23% and 55.06% of all patients, respectively. For 21.96% (33.33%) of the patients, the health benefit is the same under FFS (CAP) and FFS+P4P (CAP+P4P), and 14.81% (11.60%) of the patients suffer from reduced benefit under performance pay due to, for example, a crowding-out of patient-regarding behavior. Also, we observe a significant increase in the physicians' remuneration when introducing performance pay ($p < 0.014$, Wilcoxon Signed rank test).[18] This finding, which is in line with, for example, Mullen et al. (2010), does not come as a surprise, as subjects in our experiment do react to performance-pay incentives (recall the analyses in sections 4.2 and 4.3).

Moreover, it has often been argued that the key to an effective performance-pay system is the design of its elements (Epstein, 2012; Maynard, 2012; Kristensen et al., 2016). Therefore, we

Table 7: Patients' benefits, costs for physicians' payments, and marginal costs and benefits

	Mild illness (x)		Interm. illness (y)		Severe illness (z)		Aggregated	
A. Payment systems	$\overline{B}_{\bullet x}$	$\overline{R}_{\bullet x}$	$\overline{B}_{\bullet y}$	$\overline{R}_{\bullet y}$	$\overline{B}_{\bullet z}$	$\overline{R}_{\bullet z}$	$\overline{B}_{\bullet\bullet}$	$\overline{R}_{\bullet\bullet}$
FFS	5.94	12.67	7.47	14.21	8.99	15.95	7.76	14.28
FFS+P4P	9.16	13.26	9.38	15.04	9.55	17.60	9.39	15.30
CAP	9.04	10.00	7.77	10.00	6.47	10.00	7.47	10.00
CAP+P4P	9.49	12.35	9.42	13.52	9.31	15.48	9.41	13.78
B. Δ payment/Δ benefits								
FFS-systems	0.18		0.44		3.22		0.54	
CAP-systems	5.19		2.13		1.93		2.30	

Notes: In panel A, this table shows the average patients' health benefits $\overline{B}$ and remuneration $\overline{R}$ for FFS, CAP, FFS+P4P, and CAP+P4P, both aggregated and differentiated for the three severities of illness (x, y, z). Panel B shows the ratio of marginal payment to marginal patient health benefit, also both aggregated and separately for the three severities of illness.

[17]Notice that, for a cost effectiveness analysis of performance payment schemes, Meacock et al. (2014) propose the following 'cost categories': (i) set up/development costs (e.g., staff time, infrastructure investment); (ii) running costs (e.g., administration); (iii) incentive payments, (iv) costs to providers of participating in the scheme, (v) cost savings (e.g., reduced complications, LOS, readmissions) due to improving the quality of care resulting in superior health outcomes. In our parsimonious experimental design, we focus on the effect of additional *incentive payments* and therefore also restrict our analysis of costs to this category.

[18]Behavioral data from the experimental control condition with an increased lump-sum payment ($\Lambda = 12$) reveal a very similar pattern compared to the main experimental conditions ($\Lambda = 10$); see also the note of Table 2. This implies a smaller increase in physicians' remuneration payment in the control condition, while the absolute level of remuneration payment is higher.

take a closer look at cost and benefits for different severities of illness, as physicians' incentive payments are systematically varied for the three different severities of illness. We find that patients' health benefits increase (significantly) for all severities ($p < 0.038$, Wilcoxon Signed-rank test).[19] Under the FFS-systems, the increase in health benefit is highest for mildly ill patients (54.28%) while, under CAP-systems, the increase is highest for the severely ill patients (43.82%). Accordingly, the change in remuneration is 4.66% for the mildly ill patient in FFS-systems and 54.76% for the severely ill patients in CAP.

We now consider the 'change in payment' for an increase in patient health benefit between non-blended payment systems and blended performance pay systems—in particular, the ratio of marginal payment to marginal patient health benefit. On aggregate, for an increase in patient health benefit, a healthcare policy-maker's payment for physicians' remuneration needs to increase by 0.54 monetary units in FFS-systems and by 2.30 monetary units in CAP-systems. Under FFS-systems, the ratio is 0.18 for patients with a mild severity of illness, and for patients with an intermediate severity of illness the ratio is 0.44. This implies a fairly low increase in payment of fewer than 0.5 monetary units for an increase in patients' health benefit by one unit for patients with mild and intermediate severity of illness. Under CAP-systems, the ratio is lowest for severely ill patients (1.93), due to the large increase in patient health benefit. The ratio is highest for mildly ill patients (5.19) which is driven by the rather small increase of about 5% in patient health benefit.

Taken at face value, our results suggest that introducing performance pay is a good idea for health-care policy makers whose aim is only to enhance patients' health benefit, regardless of the costs due to additional payment to physicians. When taking into account the additional incentive payments that have to be paid for an increase in the patients' health benefit, it would be most effective to introduce performance pay for mildly ill patients in FFS-systems and for severely ill patients under CAP-regimes.

5 Conclusion

While the idea of using performance pay for physicians as a way of improving health care outcomes is increasingly making its way into health policy, the effects on physicians' provision behavior and patients' health benefits are not well understood. To this end, we introduced a controlled laboratory experiment to analyze the causal effect of pay for performance on the quality of medical service provision and on patients' health benefits. At a within-subject level, we implemented a performance pay system—with performance thresholds tied to the patient-optimal treatment and adjusted for the severities of illness—which either complements fee-for-service or capitation. Under performance pay, subjects increase, on aggregate, the quality of health care provision compared to non-blended fee-for-service and capitation. The intensity of a response to

[19]Notice that, when applying a Bonferroni correction, the difference is not significant for severity z in the FFS-systems and severity x in the CAP-systems.

performance pay is, however, significantly affected by the severity of illness. We also observe a crowding-out of patient-regarding behavior for around one third of the patients which had been treated optimally prior to the introduction of performance pay.

In our parsimonious experimental design, we reduced the complexity of a physician's treatment decisions, abstracted from multitasking, considered one-dimensional quality, and refrained from measurement issues of a physician's quality of treatment. In contrast, we focussed on *exogenously* introducing performance pay while keeping all other variables constant. We incentivized physicians for certain health outcomes—in particular, if a physician's treatment choice either renders the patient's health benefit or deviates only by one unit from the patient-optimal treatment—which did not generate uncertainty in physicians' payoffs, as all patient's outcomes are known. Taking a more general perspective, a controlled lab experiment could be regarded as a 'wind tunnel study', which allows testing for the behavioral effects of important design elements of performance pay prior to implementing these elements, for example, in a large-scale randomized controlled trial in the field.

In our experiment, we found performance pay—implying an increase in a physician's maximum attainable payoffs by 20%—to be effective in inducing a higher quality of medical services. This finding makes the case for having a sufficiently high-powered system. Also, our behavioral data showed that adjusting performance pay for the patient's severity of illness is reasonable to cope with strong overtreatment of low-severity patients under fee-for-service and with strong undertreatment of high-severity patients under capitation. Subjects were heterogenous in their responses to performance pay. We observed a considerable crowding-out of patient-regarding behavior. The unintended consequence of performance pay incentives referred to in the literature (e.g., Gneezy and Rustichini, 2000), thus, also exists in our experiment, although we left little room in subjects' choice set for a crowding-out. These effects should therefore be taken seriously given the evidence (also from other experiments) that a rather large share of subjects do provide patient-optimal treatment in the absence of performance pay (see, for example, Hennig-Schmidt et al., 2011; Godager and Wiesen, 2013; Brosig-Koch et al., forthcoming a, forthcoming b; Godager et al., forthcoming). Those patient-regarding subjects might be disposed to crowding-out under performance pay if they were given the opportunity. Moreover, as subjects do indeed respond to performance pay they may capitalize on the information asymmetry between physician and health policy-maker on the patient-optimal treatment.

Further, cost-effectiveness is another important issue for the economic evaluation of pay-for-performance systems (Meacock et al., 2014). It is often argued that pay-for-performance schemes can be considered cost-effective when quality increases are achieved with equal or lower costs or when the same quality is achieved with lower costs. When quality improvements are large, performance pay may be viewed as cost-effective even if it leads to cost increases. In general, the evidence on cost-effectiveness of performance pay seems rare and inconclusive, however. Most of the studies suggest quality improvements at the expense of cost (Emmert et al., 2012; Meacock et al., 2014). Calculations from our experiment reveal that the patients' health benefit does increase by introducing performance pay. On the other hand expenditures for physicians' incentive payments rise disproportionately indicating that (within the confines of our experimental setting and parameter choices) cost-effectiveness is not achieved when only considering incentive payments. However, there are other 'cost categories' which are might be affected by the introduction of performance pay. For example, 'cost savings' seem likely as performance pay induces care with superior health outcomes, which in turn has consequences for future health care costs. A healthcare policy-maker might also take these considerations into account, when evaluating the (cost-)effectiveness of performance pay schemes.

Finally, the observed heterogeneity in the physicians' responses to incentives calls for future work to understand better the self-selection into payment schemes (e.g., Dohmen and Falk, 2011). Knowledge on the underlying preferences that predict sorting are of great importance for researchers and health policy-makers alike. Therefore, evaluating the predictive power of preferences makes a strong case for further controlled experimental investigations.

References

Allard, M., I. Jelovac, and P. Léger (2011): "Treatment and Referral Decisions under Different Physician Payment Mechanisms," *Journal of Health Economics*, 30, 880–893.

Arrow, K. J. (1963): "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53, 941–969.

Baicker, K. and D. Goldman (2011): "Patient Cost-Sharing and Healthcare Spending Growth," *Journal of Economic Perspectives*, 25, 47–68.

Besley, T. and M. Ghatak (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95, 616–636.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz, and D. Wiesen (forthcoming a): "The Effect of Introducing Mixed Payment Systems for Physicians: Experimental Evidence," *Health Economics*.

——— (forthcoming b): "Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision," *Journal of Economic Behavior & Organization*.

Campbell, S. M., D. Reeves, E. Kontopantelis, B. Sibbald, and M. Roland (2009): "Effects of Pay for Performance on the Quality of Primary Care in England," *New England Journal of Medicine*, 361, 368–378.

Cashin, C., Y.-L. Chi, P. C. Smith, M. Borowitz, and S. Thomson (2014): *Paying for Performance in Health Care*, Open University Press, Buckingham, UK.

Chalkley, M. and J. M. Malcomson (1998): "Contracting for Health Services when Patient Demand Does not Reflect Quality," *Journal of Health Economics*, 17, 1–19.

Choné, P. and C. Ma (2011): "Optimal Health Care Contract under Physician Agency," *Annales d'Economie et de Statistique*, 101/202, 229–256.

Clemens, J. and J. D. Gottlieb (2014): "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104, 1320–1349.

Cox, J. C., V. Sadiraj, K. E. Schnier, and J. F. Sweeney (forthcoming): "Incentivizing cost-effective reductions in hospital readmission rates," *Journal of Economic Behavior Organization*.

Deci, E. L. and R. M. Ryan (1975): *Intrinsic motivation*, John Wiley & Sons.

Delfgaauw, J. and R. Dur (2008): "Incentives and Workers' Motivations in the Public Sector," *Economic Journal*, 118, 171–191.

Dohmen, T. and A. Falk (2011): "Performance pay and multidimensional sorting: Productivity, preferences, and gender," *American Economic Review*, 101, 556–590.

Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh, and M. Roland (2006): "Pay-for-Performance Programs in Family Practices in the United Kingdom," *New England Journal of Medicine*, 355, 375–384.

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

Ellis, R. P. and T. G. McGuire (1986): "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, 5, 129–151.

——— (1990): "Optimal Payment Systems for Health Services," *Journal of Health Economics*, 9, 375–396.

Emmert, M., F. Eijkenaar, H. Kemter, A. Esslinger, and O. Schöffski (2012): "Economic evaluation of pay-for-performance in health care: A systematic review," *European Journal of Health Economics*, 13, 755–767.

Epstein, A. M. (2012): "Will Pay for Performance Improve Quality of Care? The Answer Is in the Details," *New England Journal of Medicine*, 367, 1852–1853.

Fischbacher, U. (2007): "Z-tree: Zurich Toolboox for Readymade Economic Experiments – Experimenter's Manual," *Experimental Economics*, 10, 171–178.

Frey, B. S. (1997): "A constitution for knaves crowds out civic virtues," *Economic Journal*, 107, 1043–1053.

Frey, B. S. and F. Oberholzer-Gee (1997): "The cost of price incentives: An empirical analysis of motivation crowding-out," *American Economic Review*, 87, 746–755.

Frey, B. S., F. Oberholzer-Gee, and R. Eichenberger (1996): "The old lady visits your backyard: A tale of morals and markets," *Journal of Political Economy*, 1297–1313.

Gaynor, M. and P. Gertler (1995): "Moral Hazard and Risk Spreading in Partnerships," *Rand Journal of Economics*, 26, 591–613.

Glazer, J. and T. McGuire (2000): "Optimal risk adjustment of health insurance premiums: An application to Managed Care," *American Economic Review*, 90, 1055–1071.

Gneezy, U. and A. Rustichini (2000): "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115, 791–810.

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

Godager, G. and D. Wiesen (2013): "Profit or Patients' Health Benefit? Exploring the Heterogeneity in Physician Altruism," *Journal of Health Economics*, 32, 1105–1116.

Gosling, S. D., P. J. Rentfrow, and W. B. Swann Jr (2003): "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, 37, 504–528.

Gravelle, H., M. Sutton, and A. Ma (2010): "Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?" *Economic Journal*, 120, 129–156.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

Hennig-Schmidt, H., R. Selten, and D. Wiesen (2011): "How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation," *Journal of Health Economics*, 30, 637–646.

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

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

Huffman, D. and M. Bognanno (2014): "Does Performance Pay Crowd Out Worker Non-Monetary Motivations? Evidence from a Real Work Setting," Working paper, University of Pittsburgh.

Jack, W. (2005): "Purchasing Health Care Services from Providers with Unknown Altruism," *Journal of Health Economics*, 24, 73–93.

Keeney, R. and H. Raiffa (1976): *Decisions with multiple objectives: Preferences and value tradeoffs*, J. Wiley, New York.

Keser, C., E. Peterle, and C. Schnitzler (2014): "Money talks: Paying physicians for performance," University of Goettingen, cege Discussion Papers No. 218.

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

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

Lagarde, M. and D. Blaauw (2015): "Carrot, stick or competition? Experimental evidence of performance contracts in a multitasking setting," Working paper, London School of Hygiene and Tropical Medicine.

Lazear, E. P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90, 1346–1361.

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

Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler (2007): "Public Reporting and Pay for Performance in Hospital Quality Improvement," *New England Journal of Medicine*, 356, 486–496.

Liu, T. and C. Ma (2013): "Health Insurance, Treatment Plan, and Delegation to Altruistic Physician," *Journal of Economic Behavior & Organization*, 85, 79 – 96.

Ma, C. A. (1994): "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics and Management Strategy*, 3, 93–112.

Maynard, A. (2012): "The powers and pitfalls of payment for performance," *Health Economics*, 21, 3–12.

McGuire, T. G. (2000): "Physician Agency," in *Handbook of Health Economics, Vol. 1 A*, ed. by Cuyler and Newhouse, North-Holland, Amsterdam (The Netherlands), 461–536.

Meacock, R., S. R. Kristensen, and M. Sutton (2014): "The Cost-Effectiveness Of Using Financial Incentives To Improve Provider Quality: A Framework And Application," *Health economics*, 23, 1–13.

Mellström, C. and M. Johannesson (2008): "Crowding out in Blood Donation: Was Titmuss right?" *Journal of the European Economic Association*, 64, 845–863.

Mullen, K., R. Frank, and M. Rosenthal (2010): "Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers," *RAND Journal of Economics*, 41, 64–91.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

Rammstedt, B. and O. P. John (2007): "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 41, 203–212.

Roland, M. (2004): "Linking Physicians' Pay to the Quality of Care - A Major Experiment in the United Kingdom," *New England Journal of Medicine*, 351, 1448–1454.

Roland, M. and S. Campbell (2014): "Successes and Failures of Pay for Performance in the United Kingdom," *New England Journal of Medicine*, 370, 1944–1949.

Rosenthal, M. B., B. E. Landon, S.-L. T. Normand, R. G. Frank, and A. M. Epstein (2006): "Pay for Performance in Commercial HMOs," *New England Journal of Medicine*, 355, 1895–1902.

Siciliani, L. (2009): "Paying for Performance and Motivation Crowding Out," *Economics Letters*, 103, 68 – 71.

van de Ven, W. P. and R. P. Ellis (2000): "Risk Adjustment in Competitive Health Plan Markets," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, Elsevier, vol. 1, Part A, 755 – 845.

Weber, M., F. Eisenführ, and D. Von Winterfeldt (1988): "The effects of splitting attributes on weights in multiattribute utility measurement," *Management Science*, 34, 431–445.

Woolhandler, S., D. Ariely, and D. U. Himmelstein (2012): "Why pay for performance may be incompatible with quality improvement," *British Medical Journal*, 345, e5015.

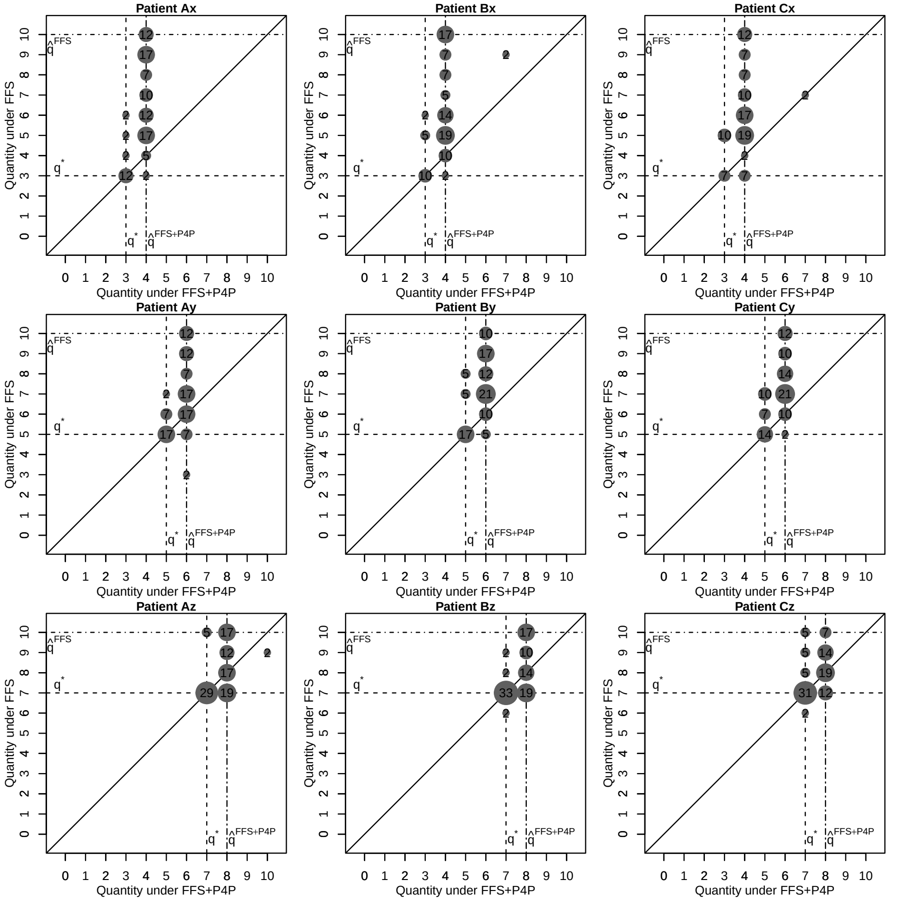
A Additional tables and figures

Table A1: Parameters of main experimental conditions

	Quantity (q)										
	0	1	2	3	4	5	6	7	8	9	10
Patient benefit											
B_{Ax}	4	5	6	7	6	5	4	3	2	1	0
B_{Ay}	2	3	4	5	6	7	6	5	4	3	2
B_{Az}	0	1	2	3	4	5	6	7	6	5	4
B_{Bx}	7	8	9	10	9	8	7	6	5	4	3
B_{By}	5	6	7	8	9	10	9	8	7	6	5
B_{Bz}	3	4	5	6	7	8	9	10	9	8	7
B_{Cx}	8	10	12	14	12	10	8	6	4	2	0
B_{Cy}	4	6	8	10	12	14	12	10	8	6	4
B_{Cz}	0	2	4	6	8	10	12	14	12	10	8
Costs											
c	0.0	0.1	0.4	0.9	1.6	2.5	3.6	4.9	6.4	8.1	10.0
FFS											
p	0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0
π	0.0	1.9	3.6	5.1	6.4	7.5	8.4	9.1	9.6	9.9	10.0
CAP											
Λ	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
π	10.0	9.9	9.6	9.1	8.4	7.5	6.4	5.1	3.6	1.9	0.0
FFS+P4P											
p	0.0	2.0	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0
b_x	0.0	0.0	2.4	2.4	2.4	0.0	0.0	0.0	0.0	0.0	0.0
b_y	0.0	0.0	0.0	0.0	3.6	3.6	3.6	0.0	0.0	0.0	0.0
b_z	0.0	0.0	0.0	0.0	0.0	0.0	5.6	5.6	5.6	0.0	0.0
π_x	0.0	1.9	9.2	10.7	12.0	7.5	8.4	9.1	9.6	9.9	10.0
π_y	0.0	1.9	3.6	5.1	10.0	11.1	12.0	9.1	9.6	9.9	10.0
π_z	0.0	1.9	3.6	5.1	6.4	7.5	10.8	11.5	12.0	9.9	10.0
CAP+P4P											
Λ	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
b_x	0.0	0.0	5.6	5.6	5.6	0.0	0.0	0.0	0.0	0.0	0.0
b_y	0.0	0.0	0.0	0.0	3.6	3.6	3.6	0.0	0.0	0.0	0.0
b_z	0.0	0.0	0.0	0.0	0.0	0.0	2.4	2.4	2.4	0.0	0.0
π_x	10.0	9.9	12.0	11.5	10.8	7.5	6.4	5.1	3.6	1.9	0.0
π_y	10.0	9.9	9.6	9.1	12.0	11.1	10.0	5.1	3.6	1.9	0.0
π_z	10.0	9.9	9.6	9.1	8.4	7.5	12.0	10.7	9.2	1.9	0.0

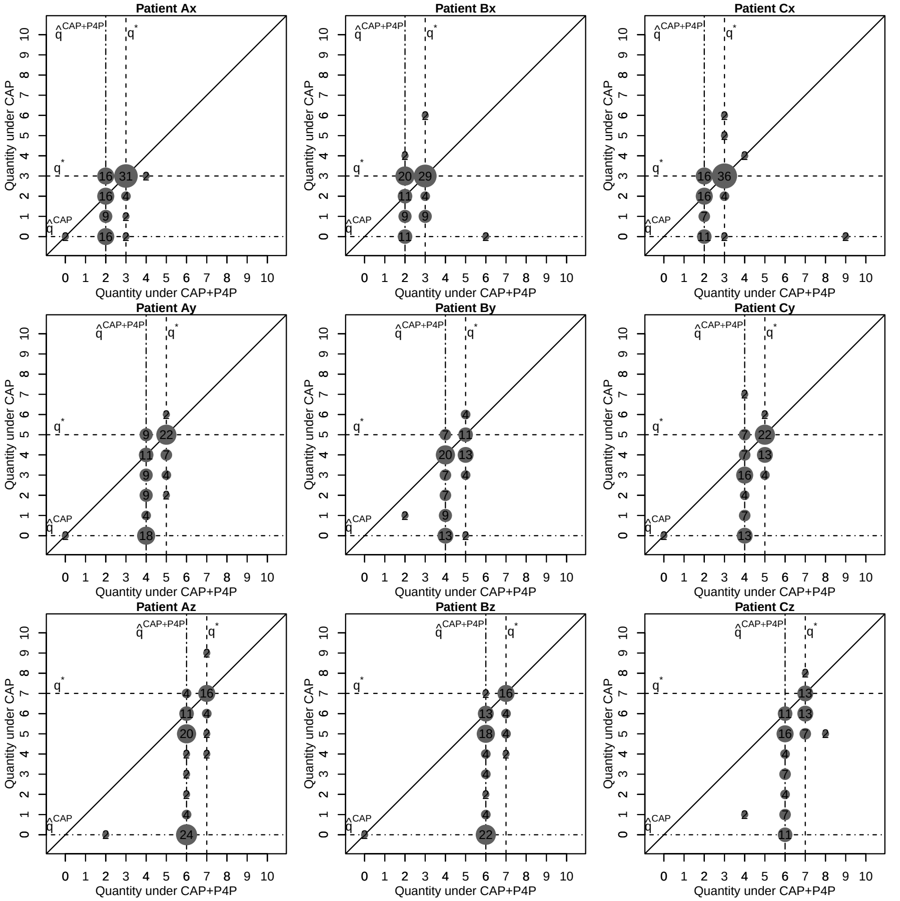
Notes: This table shows the parameters used in our experiment for all payment conditions. p is the fee per service rendered to a patient in FFS, Λ is the lump-sum payment in CAP, $b_l^\bullet$ is the bonus paid when the quality requirement is met in FFS+P4P (CAP+P4P), and π is the physician's profit. Notice that in the control CAP-condition, $\Lambda = 12$ which leads to an increase in profit by 2 monetary units irrespective of the quantity.

Figure A1: Quantity choices under FFS and FFS+P4P by patient



Notes: This figure illustrates relative frequencies of quantity choices (in %) of percentages of quantity choices of 42 subjects under FFS and FFS+P4P for each of the nine patients. A crowding-out of patient-regarding behavior occurs whenever $(q^*, \hat{q})$ is chosen.

Figure A2: Quantity choices under CAP and CAP+P4P by patient



Notes: This figure illustrates relative frequencies of quantity choices (in %) of 45 subjects under CAP and CAP+P4P for each of the nine patients. A crowding-out of patient-regarding behavior occurs whenever $(q^*, \hat{q})$ is chosen.

B Instructions of the experiment

Notice that the text in squared brackets denotes [Capitation, CAP] conditions.

Welcome to the Experiment!

You are participating in an economic experiment on decision behavior. You and the other participants will be asked to make decisions for which you can earn money. Your payoff depends on the decisions you make. At the end of the experiment, your payoff will be converted to Euro and paid to you in cash. During the experiment, all amounts are presented in the experimental currency Taler. 10 Taler equal 8 Euro. The experiment will take about 90 minutes and consists of two parts. You will receive detailed instructions before each part. Note that none of your decisions in either part have any influence on the other part of the experiment.

Part I of the experiment

Please read the instructions carefully. We will approach you in about five minutes to answer any questions you may have. If you have questions at any time during the experiment, please raise your hand and we will come to you. Part I of the experiment consists of 9 rounds of decision situations.

Decision situation

In each round, you are in the role of a physician and decide on medical treatment for a patient. That is, you determine the quantity of medical services you wish to provide to the patient for a given illness and a given severity of this illness. Each patient is characterized by one of three illnesses (A, B, C), each of which can occur in three different degrees of severity (x, y, z). In each consecutive decision round, you will face one patient who is characterized by one of the 9 possible combinations of illnesses and degrees of severity (in random order). Your decision is to provide each of these 9 patients with a quantity of $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, or 10 medical services.

Payment

In each round, you receive a fee-for-service [capitation] remuneration for treating the patient. Your remuneration increases with the amount of medical treatment [irrespective of the amount of medical treatment] you provide. You also incur costs for treating the patient, which likewise depend on the quantity of services you provide. Your profit for each decision is calculated by subtracting these costs from the fee-for-service [capitation] remuneration. Each quantity of medical service yields a particular benefit for the patient—contingent on his illness and severity. Hence, in choosing the medical services you provide, you determine not only your own profit but also the patient's benefit.

In each round you will receive detailed information on your screen (see below) for the respective patient, the illness, your amount of fee-for-service [capitation] remuneration—for each possible amount of medical treatment—your costs, profit, as well as the benefit for the patient with the corresponding illness and severity.

Fee-for-service, FFS:

Patient 1 with illness

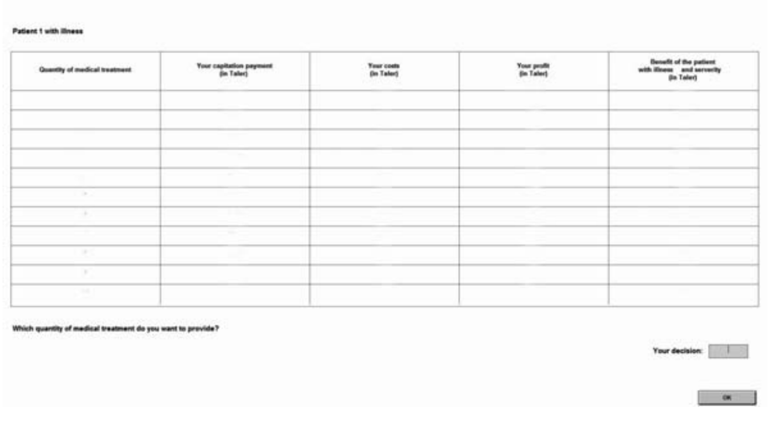
Quantity of medical treatment	Your fee-for-service payment (in Taler)	Your costs (in Taler)	Your profit (in Taler)	Benefit of the Patient with illness and severity (in Taler)

Which quantity of medical treatment do you want to provide?

Your decision:

OK

[Capitation, CAP:]



Patient 1 with illness

Quantity of medical treatment	Your capitation payment (in Taler)	Your costs (in Taler)	Your profit (in Taler)	Benefit of the patient with illness and severity (in Taler)

Which quantity of medical treatment do you want to provide?

Your decision:

OK

Payoff

At the end of the experiment, one of the 9 rounds in part I will be chosen at random. Your profit in that round will be paid to you in cash.

For this part of the experiment, no patients are physically present in the laboratory. Yet the patient benefit does accrue to a real patient: The amount resulting from your decision will be transferred to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, which will use the money for enabling the treatment of patients with eye cataract.

The transfer of money to the Christoffel Blindenmission Deutschland e.V. will be carried out after the experiment by the experimenter and one participant. The participant completes a money transfer form, filling in the total patient benefit (in Euro) resulting from the decisions made by all participants in the randomly chosen situation. This form prompts the payment of the designated amount to the Christoffel Blindenmission Deutschland e.V. by the finance department of the University of Duisburg-Essen. The form is then sealed in a stamped envelope and deposited in the nearest mailbox by the participant and the experimenter.

After the entire experiment is completed, one participant is chosen at random to oversee the money transfer to the Christoffel Blindenmission Deutschland e.V. The participant receives an additional compensation of 5 Euro for this task. The participant certifies that the process has been completed as described here by signing a statement that can be inspected by all participants at the office of the Chair of Quantitative Economic Policy. A receipt of the bank transfer to the Christoffel Blindenmission Deutschland e.V. may also be viewed here.

Comprehension questions

Prior to the decision rounds, we kindly ask you to answer a few comprehension questions. They are intended to help you familiarize yourself with the decision situations. If you have any questions about this, please raise your hand. Part *I* of the experiment will begin once all participants have answered all comprehension questions correctly.

Part *II* of the experiment

Please read the instructions carefully. We will approach you in about five minutes to answer any questions you may have. If you have questions at any time during the experiment, please raise your hand and we will come to you. Part *II* of the experiment also consists of 9 rounds of decision situations.

Decision situation

As in part *I* of the experiment, you take on the role of a physician in each round and decide on medical treatment for a patient. That is, you determine the quantity of medical services you wish to provide to the patient for a given illness and a given severity of this illness.

Each patient is characterized by one of three illnesses (A, B, C), each of which can occur in three different degrees of severity (x, y, z). In each consecutive decision round, you will face one patient who is characterized by one of the 9 possible combinations of illnesses and degrees of severity (in random order). Your decision is to provide each of these 9 patients with a quantity of $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, or 10 medical services.

Payment

In each round, you are remunerated for treating the patient. In each round, you receive a fee-for-service [capitation] remuneration for treating the patient. Your remuneration increases with the amount of medical treatment [is irrespective of the amount of medical treatment] you provide. In addition to this, in each round you receive a bonus payment, in case the quantity of medical services you provide is equal to the one that results in the highest benefit for the patient, or deviates by one quantity from the latter. You also incur costs for treating the patient, which likewise depend on the quantity of services you provide. Your profit for each decision is calculated by subtracting these costs from the sum of your fee-for-service [capitation] remuneration and bonus payment.

As in part I, every quantity of medical service yields a particular benefit for the patient contingent on his illness and severity. Hence, in choosing the medical services you provide, you determine not only your own profit, but also the patient's benefit.

In each round, you will receive detailed information on your screen (see below) for the respective patient, the illness, your amount of fee-for-service [capitation] remuneration—for each possible amount of medical treatment, the amount of your bonus payment, your costs, profit, as well as the benefit for the patient with the corresponding illness and severity.

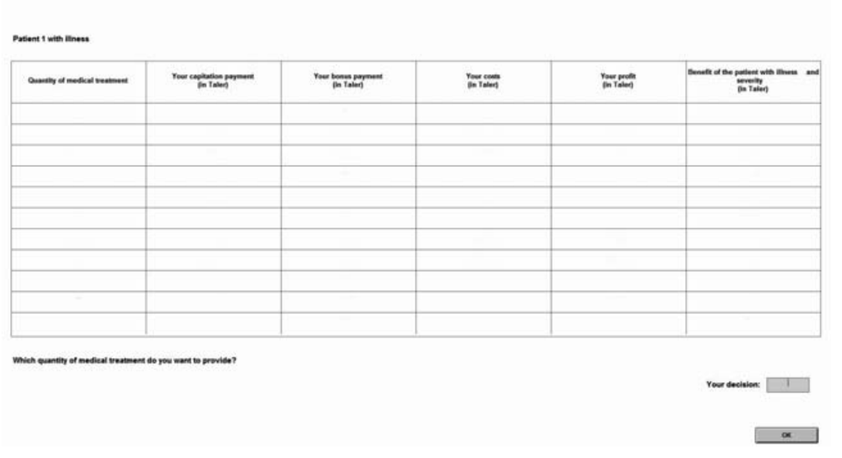
Payoff

At the end of the experiment, one of the 9 rounds of part II will be chosen at random. Your profit in this round will be paid to you in cash, in addition to your payment from the round chosen for part I of the experiment. After the experiment is over, please remain seated until the experimenter asks you to step forward. You will receive your payment at the front of the laboratory before exiting the room.

As in part I, no patients are physically present in the laboratory for part II of the experiment. Yet the patient benefit does accrue to a real patient: The amount resulting from your decision will be transferred to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, which will use the money for enabling the treatment of patients with

FFS+P4P:

[CAP+P4P:]



eye cataract.

The process for transferring the money to the Christoffel Blindenmission Deutschland e.V., as described for part *I* of the experiment, will be carried out by the experimenter and one participant.

Comprehension Questions

Prior to the decision rounds, we kindly ask you to answer a few comprehension questions. They are intended to help you familiarize yourself with the decision situations. If you have any questions about this, please raise your hand. Part *II* of the experiment will begin once all participants have answered all comprehension questions correctly.

Finally, we kindly ask you to not talk to anyone about the content of this session in order to prevent influencing other participants after you. Thank you for your cooperation!

C Comprehension questions

C.1 Experimental condition 1: FFS and FFS+P4P

FFS: (*For each of the situations 1. to 4. below, please answer the following questions.*)

- What is the fee-for-service payment?
- What are the costs?
- What is the profit?
- What is the patient benefit?

Quantity of medical treatment	**Your fee-for-service payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness F and severity y** (in Taler)
0	0.00	0.00	0.00	15.00
1	4.00	0.20	3.80	16.00
2	8.00	0.80	7.20	17.00
3	12.00	1.80	10.20	18.00
4	16.00	3.20	12.80	19.00
5	20.00	5.00	15.00	20.00
6	24.00	7.20	16.80	19.00
7	28.00	9.80	18.20	18.00
8	32.00	12.80	19.20	17.00
9	36.00	16.20	19.80	16.00
10	40.00	20.00	20.00	15.00

1. Assume that a physician wants to provide 2 quantities of medical treatment for the patient depicted above.
2. Assume that a physician wants to provide 9 quantities of medical treatment for the patient depicted above.

Quantity of medical treatment	**Your fee-for-service payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness G and severity z** (in Taler)
0	0.00	0.00	0.00	10.00
1	4.00	0.20	3.80	12.00
2	8.00	0.80	7.20	14.00
3	12.00	1.80	10.20	16.00
4	16.00	3.20	12.80	18.00
5	20.00	5.00	15.00	20.00
6	24.00	7.20	16.80	22.00
7	28.00	9.80	18.20	24.00
8	32.00	12.80	19.20	22.00
9	36.00	16.20	19.80	20.00
10	40.00	20.00	20.00	18.00

3. Assume that a physician wants to provide 2 quantities of medical treatment for the patient depicted above.
4. Assume that a physician wants to provide 9 quantities of medical treatment for the patient depicted above.

FFS+P4P: (*For each of the situations 1. to 4. below, please answer the following questions.*)

- What is the fee-for-service payment?
- What is the bonus payment?
- What are the costs?
- What is the profit?
- What is the patient benefit?

Quantity of medical treatment	**Your fee-for-service payment** (in Taler)	**Your bonus payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness F and severity y** (in Taler)
0	0.00	0.00	0.00	0.00	15.00
1	4.00	0.00	0.20	3.80	16.00
2	8.00	0.00	0.80	7.20	17.00
3	12.00	0.00	1.80	10.20	18.00
4	16.00	7.20	3.20	20.00	19.00
5	20.00	7.20	5.00	22.20	20.00
6	24.00	7.20	7.20	24.00	19.00
7	28.00	0.00	9.80	18.20	18.00
8	32.00	0.00	12.80	19.20	17.00
9	36.00	0.00	16.20	19.80	16.00
10	40.00	0.00	20.00	20.00	15.00

1. Assume that a physician wants to provide 1 quantity of medical treatment for the patient depicted above.
2. Assume that a physician wants to provide 8 quantities of medical treatment for the patient depicted above.

Quantity of medical treatment	**Your fee-for-service payment** (in Taler)	**Your bonus payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness G and severity z** (in Taler)
0	0.00	0.00	0.00	0.00	10.00
1	4.00	0.00	0.20	3.80	12.00
2	8.00	0.00	0.80	7.20	14.00
3	12.00	0.00	1.80	10.20	16.00
4	16.00	0.00	3.20	12.80	18.00
5	20.00	0.00	5.00	15.00	20.00
6	24.00	4.80	7.20	21.60	22.00
7	28.00	4.80	9.80	23.00	24.00
8	32.00	4.80	12.80	24.00	22.00
9	36.00	0.00	16.20	19.80	20.00
10	40.00	0.00	20.00	20.00	18.00

3. Assume that a physician wants to provide 1 quantity of medical treatment for the patient depicted above.
4. Assume that a physician wants to provide 8 quantities of medical treatment for the patient depicted above.

C.2 Experimental condition 2: CAP and CAP+P4P

CAP: (*For each of the situations 1. to 4. below, please answer the following questions.*)

- What is the capitation payment?
- What are the costs?
- What is the profit?
- What is the patient benefit?

Quantity of medical treatment	**Your capitation payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness F and severity y** (in Taler)
0	20.00	0.00	20.00	15.00
1	20.00	0.20	19.80	16.00
2	20.00	0.80	19.20	17.00
3	20.00	1.80	18.20	18.00
4	20.00	3.20	16.80	19.00
5	20.00	5.00	15.00	20.00
6	20.00	7.20	12.80	19.00
7	20.00	9.80	10.20	18.00
8	20.00	12.80	7.20	17.00
9	20.00	16.20	3.80	16.00
10	20.00	20.00	0.00	15.00

1. Assume that a physician wants to provide 2 quantities of medical treatment for the patient depicted above.
2. Assume that a physician wants to provide 9 quantities of medical treatment for the patient depicted above.

Quantity of medical treatment	**Your capitation payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness G and severity z** (in Taler)
0	20.00	0.00	20.00	10.00
1	20.00	0.20	19.80	12.00
2	20.00	0.80	19.20	14.00
3	20.00	1.80	18.20	16.00
4	20.00	3.20	16.80	18.00
5	20.00	5.00	15.00	20.00
6	20.00	7.20	12.80	22.00
7	20.00	9.80	10.20	24.00
8	20.00	12.80	7.20	22.00
9	20.00	16.20	3.80	20.00
10	20.00	20.00	0.00	18.00

3. Assume that a physician wants to provide 2 quantities of medical treatment for the patient depicted above.
4. Assume that a physician wants to provide 9 quantities of medical treatment for the patient depicted above.

CAP+P4P: (*For each of the situations 1. to 4. below, please answer the following questions.*)

- What is the capitation payment?
- What is the bonus payment?
- What are the costs?
- What is the profit?
- What is the patient benefit?

Quantity of medical treatment	**Your capitation payment** (in Taler)	**Your bonus payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness F and severity y** (in Taler)
0	20.00	0.00	0.00	20.00	15.00
1	20.00	0.00	0.20	19.80	16.00
2	20.00	0.00	0.80	19.20	17.00
3	20.00	0.00	1.80	18.20	18.00
4	20.00	7.20	3.20	24.00	19.00
5	20.00	7.20	5.00	22.20	20.00
6	20.00	7.20	7.20	20.00	19.00
7	20.00	0.00	9.80	10.20	18.00
8	20.00	0.00	12.80	7.20	17.00
9	20.00	0.00	16.20	3.80	16.00
10	20.00	0.00	20.00	0.00	15.00

1. Assume that a physician wants to provide 1 quantity of medical treatment for the patient depicted above.
2. Assume that a physician wants to provide 8 quantities of medical treatment for the patient depicted above.

Quantity of medical treatment	**Your capitation payment** (in Taler)	**Your bonus payment** (in Taler)	**Your costs** (in Taler)	**Your profit** (in Taler)	**Benefit of the patient with illness G and severity z** (in Taler)
0	20.00	0.00	0.00	20.00	10.00
1	20.00	0.00	0.20	19.80	12.00
2	20.00	0.00	0.80	19.20	14.00
3	20.00	0.00	1.80	18.20	16.00
4	20.00	0.00	3.20	16.80	18.00
5	20.00	0.00	5.00	15.00	20.00
6	20.00	11.20	7.20	24.00	22.00
7	20.00	11.20	9.80	21.40	24.00
8	20.00	11.20	12.80	18.40	22.00
9	20.00	0.00	16.20	3.80	20.00
10	20.00	0.00	20.00	0.00	18.00

3. Assume that a physician wants to provide 1 quantity of medical treatment for the patient depicted above.
4. Assume that a physician wants to provide 8 quantities of medical treatment for the patient depicted above.