Physician performance pay: Experimental evidence*

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

PRELIMINARY VERSION

PLEASE DO NOT CITE OR CIRCULATE!

Abstract

We present causal evidence on the effect of performance pay on medical service provision from an artefactual field experiment with a representative sample of German resident primary care physicians. In the experiment, we introduce performance pay, which is adjusted according to patients' severities of illnesses, to complement capitation. Performance pay is granted if a health care quality threshold is met. In line with standard theory, we find that performance pay significantly reduces underprovision of medical services, and, on average, it increases the patients' health benefit. The magnitude of these effects depends, however, on patients' characteristics. Findings are robust towards variations in levels of performance pay. Beyond standard theory, we find evidence for a crowding-out of altruistic behavior when physicians receive performance pay. Physicians' characteristics such as gender and practice location significantly affect crowding-out of altruistic behavior.

Keywords: capitation, pay for performance, artefactual field experiment, representative physician sample, crowding-out

JEL-Classification: C91, I11

*Brosig-Koch: University of Duisburg-Essen, Faculty of Economics and Business Administration, and CINCH—Health Economics Research Center Essen, Berliner Platz 6-8, 45127 Essen, Germany, e-mail: jeannette.brosig@uni-due.de; Kairies-Schwarz: University of Duisburg-Essen, Faculty of Economics and Business Administration, and CINCH Essen, Berliner Platz 6-8, 45127 Essen, Germany, e-mail: nadja.kairies@ibes.uni-due.de; Hennig-Schmidt: Laboratory for Experimental Economics, Department of Economics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, and Institute of Health and Society, University of Oslo, Norway, e-mail: hschmidt@uni-bonn.de; Kokot: Zi–*Zentralinstitut fuer die kassenärztliche Versorgung* of the Federal Republic of Germany, Berlin (Germany); Wiesen: Behavioral Service Management, Department of Health Care Management, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany, and Institute of Health and Society, University of Oslo, Norway, e-mail: wiesen@wiso.uni-koeln.de. We are grateful for valuable comments and suggestions from John Cawley, Jim Cox, Randy Ellis, Glenn W. Harrison, Albert Ma, Henry Mak, Anne Sophie Oxholm, Daniele Paserman, Meredith Rosenthal, Tony Scott, and by participants of seminars at BU, Cornell, iHEA Boston, IUPUI, U Lucerne and the 5th BEH Workshop in Atlanta. Financial support by the German Research Foundation (DFG, grant: BR 2346/2-1/2) is gratefully acknowledged.

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 and capitation to underserve patients.

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, however—with identification of the causal impact of physician incentives being the main challenge. Also, evidence from field studies is quite mixed on whether an effect of performance pay with respect to improving the quality of care exists. If at all, rather moderate effects of performance pay are reported (e.g., Mullen et al., 2010; 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 because of incomplete performance measures or 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).

Pay for performance may not only have positive effects on physicians' medical service provision. It is 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). This is a particularly important challenge as other-regarding motivations are a fundamental determinant of public service provision (e.g.,

[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 also 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).

Besley and Ghatak, 2005).

There is some empirical evidence for motivation crowding-out, for instance, from real work settings (e.g., Huffman and Bognanno, forthcoming), and blood donation (Mellström and Johannesson, 2008). To the best of our knowledge, empirical evidence is lacking, however, on whether P4P actually does crowd out physicians' altruistic (patient-regarding) motivation and therefore the quality of medical service provision. This is surprising as it is often argued 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, like pride in excellent work (e.g., Woolhandler et al.,2012). If crowding out exists, it is also important to know whether the level of the incentive has an effect on physicians' behavior, as has been found in real-effort experiments (e.g., Gneezy and Rustichini, 2000; Ariely et al., 2009).

This paper contributes to providing empirical evidence on the challenges described above. We run an artefactual field experiment (according to the taxonomy of experiments of Harrison and List, 2004) with real doctors as participants. Our sample is based on a representative sample of resident physicians in Germany. Our study is the first field experiment that examines the causal effect of introducing performance pay on doctors' behavior. Our unique data set also allows us to link their behavioral data to survey data on physicians' individual characteristics attitudes towards altruism and competition according to the World Value Survey, and risk attitudes according to the German Socio-Economic Panel, as well as objective data on physicians' gender, age, residence by regional type according to the Federal Institute for Research on Building, Urban Affairs and Spatial Development, and years in practice.

In our field experiment, all physicians decide on the provision of medical services. We randomly assign physicians to different payment conditions. We exogenously change physicians' remuneration at a within-subject level: each physician makes his or her decisions under two consecutive payment systems. First, physicians are incentivized by a lump-sum capitation system (CAP), which serves as the baseline payment. We then introduce performance pay in form of a bonus in addition to the baseline payments. Our between-subject condition variation includes two levels of performance pay: 5%-Bonus and 20%-Bonus.

The physicians' task in the experiment is to choose the quantity of 'medical treatment for patients' with different severities of illness. Quantity choices determine the physicians' own profit and the patients' health benefits. Real patients' health is affected by these 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 (similar to Hennig-Schmidt et al., 2011; Kesternich et al., 2015; Brosig-Koch et al., 2016, 2017). The controlled environment allows us to implement a 'clean' measure of individual physicians' performance tied to the quality of medical care, which is the difference between the chosen quantity and the patient-optimal quantity of medical services, the latter implying the highest health benefit for a patient. The larger this difference is, the lower is the provided quality. Performance pay is linked to this measure of individual physicians' performance and is granted if a health care quality threshold is met. Bonus rates are adjusted for patients' severities of illness.[2]

Our parsimonious experimental design abstracts from several effects that are associated with performance pay and are unintended. These are, for instance, substitution of effort on incentivized tasks 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). In our experiment, the bonus is paid whenever a quality threshold is reached.

Our study contributes to several strands of the health economics literature. First, we contribute to the growing number of research topics in health economics addressed by controlled experiments. Controlled behavioral experiments analyze, for example, health care markets (e.g., Kessler and Roth, 2012; Kessler and Roth, 2014), non-monetary incentives (e.g., Kesternich et al., 2015, Godager et al., 2016), provider behavior under fee-for-service and capitation payment (e.g., Hennig-Schmidt et al., 2011, Hennig-Schmidt and Wiesen, 2014, Brosig-Koch et al., 2016), mixed payment systems (Brosig-Koch et al., 2017), and health care financing (e.g., Buckley et al., 2012); for a comprehensive summary see Galizzi and Wiesen (forthcoming).

Our study substantially contributes to the empirical literature on how introducing performance pay influences individual physicians' medical service provision. Based on our unique data set that is built on a representative sample of resident physicians in Germany, we are the first to investigate the following research questions in a controlled field experiment: (i) how does introducing performance pay affect individual physicians' provision behavior? (ii) does the level of the bonus (5% versus 20% in addition to a lump-sum payment) affect individual physicians' provision behavior? (iii) does performance pay crowd-out individual physicians' altruistic (patient-regarding) behavior? (iv) how do individual physicians' characteristics relate to behavioral responses when pay-for-performance is introduced, in particular to crowding-out

[2]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 high severity of illness, for example, face the highest 'risk' of being undertreated under capitation—a behavioral pattern which has been indicated by recent experimental findings (e.g., Hennig-Schmidt et al., 2011; Brosig-Koch et al., 2017).

of altruistic behavior?

Behavioral data from the experiment show that: (i) undertreatment under capitation varies with patients' severity of illness, (ii) introducing performance pay significantly reduces under-provision under capitation and the reduction in non-optimal service provision is significantly affected by the severity of illness, (iii) performance pay induces crowding-out of patient-regarding behavior for a substantial share of patients who have been treated optimally prior to the introduction of performance pay, with the extent of crowding-out varying with the patients' severities of illness; (iv) physicians' gender and location significantly relate to crowding-out of altruistic behavior.

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 crowding-out of patient-regarding behavior reduces the health benefit for a considerable proportion of patients.

This paper proceeds as follows. In Section 2, we provide some background on the setting and our sample. We present our experimental design in Section 3. Section 4 presents our behavioral results and Section 5 concludes.

2 Background and sample characteristics

In Germany, the vast majority of health care is provided under the statutory health insurance scheme. About 72 million people, around 88% of the German population, are enrolled in the statutory health insurance. Health insurance is mandatory in Germany. Around 32,000 resident self-employed primary care physicians (PCP) contract with the statutory health insurance. About 3 percent of all PCPs are included in the 'Physician practice panel' (*Praxis-Panel*, ZiPP) of the *Zentralinstitut für die Kassenärztliche Versorgung* (Zi) of the Federal Republic of Germany. ZiPP is a representative sample of resident physicians in Germany and is run annually with about 5,000 resident physicians across all specializations. It is a unique data base originally designed to analyze the cost structure of health care in Germany.[3]

In the ZiPP, the subsample of resident PCPs is randomly recruited from all resident PCPs in Germany and stratified according to three regional areas (city, outer conurbation, and rural). The sample represents the general population of resident PCPs in Germany measured by the

[3]See Zi-Praxis-Panel (2017): *Jahresbericht 2015, Wirtschaftliche Situation und Rahmenbedingungen in der vertragsärztlichen Versorgung der Jahre 2011 bis 2014, 6. Jahrgang, Berlin, Juli 2017, p.19, https://www.zi-pp.de/pdf/ZiPPJahresbericht2015.pdf; retrieved December 2017*

number of medical treatments per physician, the remuneration per physician, the remuneration per medical treatment, and the ratio between remuneration and medical treatments required.

For our experiment, the Zi randomly selected a subsample of 662 physicians from all resident

Table 1: Sample characteristics

	Our PCP sample	PCPs in ZiPP	PCPs in Germany
Age (in years)	54	54	53
Share of females	35%	39%	44%
Location			
City	37%	34%	–
Outer conurbation	44%	37%	–
Rural	19%	29%	–

Notes: This table shows characteristics of our primary care physician (PCP) sample ($N = 104$) and compares it to the ZiPP sample of PCPs as well as to primary care physicians in Germany. Sources are the annual reports of the Zi and the Bundesärtzekammer for 2015.

PCPs in the ZiPP who were invited to participate in our online experiment conducted in 2016. A total of 104 resident PCPs participated in the experiment, i.e., about 10% of all PCPs enrolled in the ZiPP form the subsample we analyzed in our experiment.

Comparing the characteristics of our sample (see Table 1) with those of the ZiPP sample of resident PCPs and the population of all German PCPs shows a rather good approximation regarding average age; regarding the share of female PCPs and regional areas, there are slight differences.

3 Experimental design

3.1 Decision situation

In our experiment, all physicians decide on the provision of medical services. We employ a within-subject design to analyze the effect of introducing physicians' performance pay on medical service provision. Each physician decides under two different payment systems. First, physicians' decisions on the quantity of medical services are incentivized by a lump-sum capitation payment (CAP), which serves as the baseline scheme (part I of the experiment). Second, physicians decide under a payment system comprising performance pay in addition to the baseline capitation (CAP+P4P in part II). We randomly assign physicians to two different conditions: either 5 percent or 20 percent performance bonus in addition to the lump-sum payment.

Physician i decides on the quantity of medical services $q \in [0, 10]$ for nine different patients ($j = 1, \ldots, 9$) in all payment systems. Patients differ in illnesses $k \in \{A, B, C\}$ and in severities of illness $l \in \{x, y, z\}$. Patients are fully insured, which seems to be a natural assumption given that patients insured under the German statutory health insurance scheme typically do not make co-payments. Moreover, full insurance is commonly assumed in the health economics literature (e.g., McGuire, 2000). Patients' illness and severity of illness are the same in all payment conditions. This design feature implies that behavioral changes between payment conditions are not confounded by variations in the patient population.

With each decision, a physician determines his or her own profit and a patients' health benefit. While all physicians decide for abstract patients in the experiment, real patients' health outside the lab is affected by their choices. Physicians 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 3.3. This mechanism ensures that the patients' health benefit is made salient.

3.2 Payment systems, profits, and patient health benefit

We consider a threshold-based performance-pay system designed to mitigate inherent incentives in CAP to provide too few services. Physicians receive a bonus payment b that maximizes their profit whenever their treatment reaches a quality threshold. We determine the profit-maximizing quantities in such a way that they are 'closer' to the patient-optimal quantities than in CAP, but do not coincide with them.[4] In particular, the bonus is paid if the chosen quantity does not differ by more than one unit from the patient-optimal treatment. Thus, the bonus reduces the trade-off between profit maximization and patient health benefit optimization. Quality thresholds are quite common in practice; for example, in the Quality and Outcomes Framework in the UK (e.g., Roland, 2004).

Due to the fact that profit-maximizing and patient-optimal quantities do not coincide—neither under CAP nor under CAP+P4P—we are able to identify crowding-out of patient-regarding altruistic behavior. In particular, this design feature allows us to test whether introducing a performance payment induces physicians that have chosen the patient-optimal treatment under CAP to deviate and choose their profit maximum instead.

More formally, physician i's payment is $R(q) = \Lambda + b_l^\bullet I_{b_l}$, with Λ being the lump-sum payment 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 by more than one unit from the patient optimal treatment, and

[4] This design feature accounts for the fact that service quality is not fully contractible due to information asymmetry.

0 otherwise. In CAP, $b_l^{\mathrm{CAP}} = 0$.

Physician i's profit is given as

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

with $\Lambda b_l^{\bullet} > 0$, $c'(q) > 0$ and $c''(q) > 0$. In the experiment, $c(q) = q^2/4$ for both 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 $\theta > 0$.[5]

Patient health benefits in our experiment are systematically varied with 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.[6] 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.

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 A.2.

While all physicians make decisions for abstract patients in the experiment, real patients' health is affected by their choices. Participants 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 3.3.

Table 2 provides an overview of the payment systems employed in our experiment. Physicians are paid by CAP in part I and decide under one of the two performance-pay systems

[5] 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 $B_1 = 4 + 2 \cdot 1 \cdot 3 = 10$.

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

CAP+P4P-5% or CAP+P4P-20% in part *II*. Beyond the within-subject comparison on the effect of introducing performance pay (part *I* vs. part *II*), the experimental design also allows us to make a between-subject comparison between CAP+P4P-5% and CAP+P4P-20% in part *II*. The complete set of parameter values is shown in Table A1 in Appendix A.1.

Table 2: Payment systems in the main experimental conditions

Exp.	Part *I* of the experiment		Part *II* of the experiment				# of physicians
Condition	Payment	Λ	Payment	Severity l	Λ	$b_l^\bullet$	
5%-Bonus	CAP	10	CAP+P4P-5%	x	10	0.9	51
				y	10	2.1	
				z	10	4.1	
20%-Bonus	CAP	10	CAP+P4P-20%	x	10	2.4	53
				y	10	3.6	
				z	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 if their quantity choice fulfills the quality requirement $|q - q^*| \leq 1$, otherwise $b_l^\bullet = 0$.

In CAP, physicians receive a lump-sum payment of $\Lambda = 10$ per patient, independent of the quantity of medical services. Physicians' profit per patient is $\pi(q) = 10 - c(q)$ with the maximum attainable profit being 10 in CAP, when choosing $\hat{q}_j^{\text{CAP}} = 0$. Note that all payments, costs, profits and patient benefits are multiplied by a factor of 2.5 to arrive at the amount to be paid in Euro. To facilitate calculations for the physicians all monetary amounts are presented in Euro. The maximal profit is thus EUR 25. The benefit-maximizing choices yield profits of EUR 22.75 (18.75, 12.75) for x (y, z, respectively).

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, $b_x^{\text{P4P-5\%}} = 0.9$, $b_y^{\text{P4P-5\%}} = 2.1$, $b_z^{\text{P4P-5\%}} = 4.1$ in CAP+P4P-5%, and $b_x^{\text{P4P-20\%}} = 2.4$, $b_y^{\text{P4P-20\%}} = 3.6$, $b_z^{\text{P4P-20\%}} = 5.6$ in CAP+P4P-20%, respectively. The bonus implies an increase in the maximum attainable profit $\pi(\hat{q}_j)$ by 5 or 20 percent. For each severity, choosing $\hat{q}_j$ equal to 2 (4), 4 (6), or 6 (8) in CAP+P4P-5% and CAP+P4P-20% thus yields a maximal profit for the physician of 10.5 and 12, which amounts to EUR 26.25 and EUR 30, respectively. The benefit-maximizing choices in CAP+P4P-5% yield profits of EUR 25 (24, 23) for x (y, z, respectively). In CAP+P4P-20% they are EUR 28.75 (27.75, 26.75).

3.3 Experimental protocol

We employ a double-blind procedure according to the data protection guidelines of the ZiPP. All studies using the ZiPP have to implement the following double blind procedure, and participants are informed accordingly. Invitations to physicians including log-in data and IDs are sent out via a trustee at the Zi. All decisions in the online experiment are made using these IDs to which we can relate decisions only. Payment of participants is made via a notary authorized by the Zi, who receives a list with participants' names and IDs from the trustee and a list with IDs and payoffs from the IT department of the Zi, but is not informed about participants' decisions. The notary transfers the money to the banking accounts of the participants.

The online experiment was programmed with the software SoPHIE (www.sophielabs.com) and conducted in Germany in 2016. The experimental procedure was as follows: Physicians locked in with their ID and were alternately assigned to one of the two conditions Bonus-5% or Bonus-20%; i.e., the physician locked in first was assigned to Bonus-5%, the second one to Bonus-20%, the third one again to Bonus-5%, and so forth. This procedure ensured that we had a random assignment of an equal number of physicians to the two conditions. Physicians then received instructions for part I onscreen. They were advised to print out the instructions by clicking on a button that directly linked to a .pdf file. This link was provided also on every subsequent screen physicians saw during the experiment. Physicians 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 each physician's understanding of the decision task, he or she had to answer a set of control questions. The experiment did not start unless the physician had answered all control questions correctly (instructions and control questions are included in Appendices A.2 and A.3).

In each of the two parts of the experiment, physicians 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 physicians and all conditions: Bx; Cx; Az; By; Bz; Ay; Cz; Ax; Cy. Before making their decision for a specific patient, physicians 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 were given in EUR. The procedure was exactly the same in part II of the experiment.

After having finished the second part of the experiment, we asked physicians to complete a questionnaire on social demographics (age and gender), on risk preferences (based on respective questions included in the German Socio Economic Panel; see Dohmen et al., 2011), on the social traits altruism and competitiveness (based on respective questions included in the World

Value Survey), and on their general attitude regarding pay for performance. In addition, the Zi matched physicians' IDs to data on their location and other specifics of the physician (practice owner, yearly working hours) available for the ZiPP panelists.

At the end of the experiment, we randomly determined one decision in each part of the experiment to be relevant for a physician's actual payoff and the patient benefit. This was done to rule out income effects. Physicians were paid according to these two randomly determined decisions. As mentioned above payment was made double blind via the notary of the Zi. The notary of the Zi also transferred the sum of patient benefits resulting from the two randomly determined decisions to the Christoffel Blindenmission, which used the money to support surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists from the charity.

The experiment lasted for about 25 minutes. Physicians earned, on average, EUR 45.93. In total, EUR 5,002.50 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 166 patients to be treated.

4 Results

4.1 Medical service provision under capitation

To begin with, we investigate the quantities of medical services physicians provided under capitation payments (part I of the experiment) in both treatments 20%-Bonus and 5%-Bonus. The average quantity of medical services is 4.27 in both conditions, see Table 3. Recall that the *average* patient in our experiment is treated optimally with five medical services. Hence, our results indicate that physicians underprovide patients on average. Applying non-parametric statistics, we find that physicians significantly underprovide medical services under capitation ($p \leq 0.007$, Wilcoxon signed-rank test). Comparing average medical services per subject and severity across the identical CAP-parts I of our two conditions shows no significant differences ($p \geq 0.7000$, Mann-Whitney U-test).[7]

Given our design of different patient types the aggregate quantities may, however, not provide the full picture. We therefore investigate the quantity choices disaggregated by the severity of illness. Figures 1 show physicians' average quantities of medical services per severity in con-

[7]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 three quantity decisions per severity of illness in part I (part II) of the experiment. We apply non-parametric tests to the individual subjects' averages per severity of illness. For between-subject analyses, we employ Mann-Whitney U-tests, for within-subject analyses we use the Wilcoxon matched-pairs signed rank-test.

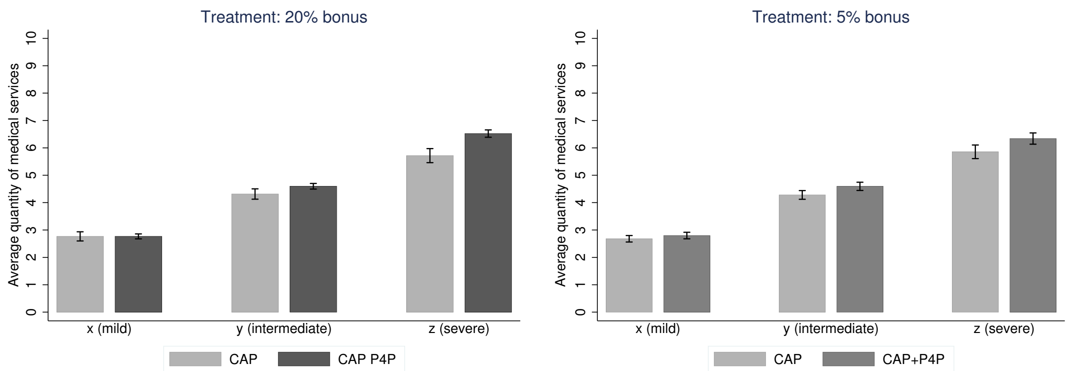
Table 3: Quantities of medical services by payment system

	Capitation (part I)		Capitation + P4P (part II)		
Condition	Mean	s.d.	Mean	s.d.	N
20%-Bonus	4.27	1.73	4.58	1.77	459
5%-Bonus	4.27	1.79	4.63	1.69	477

Notes: This table shows descriptive statistics on the quantity of medical services q at the payment system level. Note that 104 German resident physicians (51 in 20%-Bonus and 53 in 5%-Bonus) each decide for nine patients on the quantity of medical services.

ditions 20%-Bonus and 5%-Bonus. Looking at the light-grey shaded bars only, medical service increase in the patients' severity of illness. The severity of illness also significantly affects physicians' quantity choices ($p < 0.050$, Wilcoxon signed-rank test). In sum, we state the following result:

Figure 1: Quantity choices by payment system and severity of illness



Notes: This figure shows average quantities of medical services for the three severities of illness and for the payment conditions. 95% CI are also indicated. The left panel reports average deviations under payment systems CAP and CAP+P4P-20%. For each severity, 51 subjects made three decisions in each payment system. The right panel shows quantities of service provision for payment systems CAP and CAP+P4P+5%. 53 subjects made three decisions in each payment system for each severity.

Result 1. *Capitation induces physician to significantly underprovide medical services. The severity of illness significantly affects quantity choices.*

The behavioral results on the provision of medical services under CAP by a representative subsample of resident PCPs in Germany are in line with findings reported in earlier empirical studies (e.g., Gaynor and Gertler 1995; Cutler, 1995) and more recent and experimental studies Hennig-Schmidt et al. 2011). Regarding the severity of illness, our results complement findings of related lab experiments (e.g., Hennig-Schmidt et al., 2011; Brosig-Koch et al., 2017) 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 bonus payments for different severities of illness.

4.2 Effect of performance pay on physician provision behavior

We now analyze the effect of introducing performance pay. On aggregate, the quantity of medical services increases from 4.27 under CAP to 4.58 and 4.63 under CAP+P4P in conditions 20%-Bonus and 5%-Bonus, respectively; see Table 3. Comparing average quantities for each severity between the two parts of our experiment indicates that quantities increase when performance pay is introduced. Also the change in quantities increases with the severities of illness; see Figure 1.

To control for the impact of patient characteristics, in particular the severity of illness and illness, we run a series of regressions; see Table 4. We find that underprovision is highly significantly reduced under performance pay by 33.4 percentage points. Our regression results also show that the severity of illness highly significantly affects quantity choices while the illness does not; see model (1). The effect is robust when controlling for physicians' characteristics, such as gender, age, practice years, and location; see models (2-4) in Table Table 4. In sum, we state:

Result 2. *Introducing performance pay reduces underprovision of medical services compared to capitation. The reduction is significantly stronger for intermediately and severely ill patients.*

We now consider how the level of the bonus payment affects physicians' medical service provision. In a between-subject analysis, we investigate whether physicians' behavior differs significantly when comparing decisions in parts II between 20%-Bonus and 5%-Bonus. We find no significant differences in average medical services per subject and severity between the two conditions ($p \geq 0.4964$, Mann-Whitney U-test). Regression analyses provide further support that the level of the bonus did not affect physicians' medical service provision; see Table 4. In sum, we state the following result:

Result 3. *The level of the bonus payment—either 5% or 20% in addition to the lump-sum capitation payment—does not affect physicians' medical service provision.*

Table 4: Effect of performance pay on physicians' medical service provision

Model:	(1)	(2)	(3)	(4)
Dependent variable:	q	q	q	q
Performance pay (= 1 if part II)	0.334***	0.334***	0.334***	0.334***
	(0.071)	(0.071)	(0.071)	(0.071)
Level of bonus (= 1 if 20%-Bonus)	0.011	0.017	0.009	0.073
	(0.065)	(0.064)	(0.063)	(0.075)
Interm. severity (= 1 if $l = y$)	1.694***	1.694***	1.694***	1.694***
	(0.054)	(0.054)	(0.054)	(0.054)
High severity (= 1 if $l = z$)	3.356***	3.356***	3.356***	3.356***
	(0.097)	(0.098)	(0.098)	(0.098)
Illness	-0.017	-0.017	-0.017	-0.017
	(0.022)	(0.022)	(0.022)	(0.022)
Gender		0.081	0.142	0.051
		(0.157)	(0.166)	(0.140)
Age		0.010	0.010	0.013
		(0.008)	(0.016)	(0.015)
Constant	2.217***	1.508***	1.298**	0.592
	(0.410)	(0.531)	(0.572)	(0.694)
Practice years	No	No	Yes	Yes
Physicians' characteristics	No	No	No	Yes
R^2	0.621	0.623	0.626	0.641
Observations	1872	1872	1872	1872
Physicians	104	104	104	104

Notes: Ordinary Least Square (OLS) estimates are reported with robust standard errors clustered for subjects (in brackets). The reference category is the 'mild severity of illness', $l = x$. In addition to gender and age, physicians' characteristics comprise a question each for the attitude towards altruism and competition from the World Value Survey, risk attitudes according to the German Socio-Economic Panel, and regional type. Practice years include how many years physicians have been practicing employed in a practice, as self-employed resident physician, and in hospital. Estimates from Tobit regressions yield qualitatively very similar results. *** $p < 0.01$, ** $p < 0.05$, and *$p < 0.1$.

4.3 Crowding-out of altruistic behavior

We next analyze the dynamics of treatment patterns for individual patients between CAP and the performance pay systems. In particular, we analyze changes in treatment patterns focussing on the crowding-out of patient-regarding behavior by performance pay. For the following analysis, we consider 936 data points per payment system (104 subjects $\times$ 9 decisions).

We distinguish between three main treatment patterns: quantity choices that maximize physician profit (PM), quantity choices maximizing the patient benefit (BM), and trade-off choices (TO) which capture Pareto-optimal quantity choices but are neither PM nor BM. Category 'Other' comprises Pareto-inferior medical service provision. As we do not observe significant differences between conditions for both parts of the experiment, we pool the data for our classification of choices. We find the following treatment patterns under CAP: PM: 1.5%; BM: 54%; TO: 41%,Other: 2.5%. Under CAP+P4P we observe: PM: 30%; BM: 64%; and Other: 4%.[8]

When performance pay is introduced we observe that PM increases by 28.5 percentage points, BM by 10 and Other by 1.5 percentage points. Despite the rise in BM, we do find evidence for crowding-out of altruistic behavior. In total, crowding-out, i.e., a transition from BM to PM, amounts to 7% of all physicians' choices, which means 14% of all BM choices (choices that lead to crowding-out of patient-regarding behavior in the second part) under CAP, which allow to identify crowding-out of patient-regarding behavior in our design (i.e., all BM choices). 24% of choices transition from TO to PM. We also observe crowding-in (PM to BM) which amounts to 1%. Finally, 17% of choices transition from TO to BM.

We next analyze how crowding-out of physicians' altruistic behavior relates to physicians' characteristics. In particular, we explore the effect of physicians' age, gender, and location (rural, outer conurbation, and city) while controlling for physicians' practice years and other stated preferences. Table 5 shows estimation results (marginal effects) from logit regressions.

Crowding-out of altruistic behavior is significantly affected by physicians' gender and location. For male physicians crowding-out is less likely to be observed. Moreover, physicians in cities and outer conurbation areas are significantly less likely to show crowding-out behavior compared to the reference category 'rural'. Finally, crowding-out is less likely for patients characterized by a higher marginal health benefit, see model (1) in Table 5. These findings are robust when including controls for physicians' years in practice; see model (2) in Table 5.

[8] As the bonus is paid if the chosen quantity does not differ by more than one unit from the patient-optimal treatment, the classification TO does not exist in part II.

Table 5: Logit regressions on crowding-out of altruism, marginal effects

Model:	(1)	(2)
Dependent variable:	Crowding out	Crowding out
Condition (= 1 if 20%-Bonus)	0.0089	0.0105
	(0.0094)	(0.0092)
Low severity (= 1 if $l = x$)	0.0376	0.0364
	(0.0274)	(0.0264)
Interm. severity (= 1 if $l = y$)	0.0203	0.0195
	(0.0207)	(0.0200)
Marginal health benefit (= 1 if $\theta = 2$)	-0.0300**	-0.0289**
	(0.016)	(0.0142)
Age	-0.0005	-0.0034
	(0.0012)	(0.0023)
Gender (= 1 if male)	-0.0392*	-0.0430**
	(0.0220)	(0.0202)
City	-0.0447***	-0.4581***
	(0.1738)	(0.0162)
Outer conurbation	-0.0416**	-0.0432**
	(0.1915)	(0.1846)
Years in practice controls	No	Yes
Other characteristics	Yes	Yes
Observations	936	936
Subjects	104	104

Notes: The table shows marginal effects from logit regressions with robust standard errors clustered for subjects (in parentheses).The reference category is 'high severity', $l = z$. Marginal health benefit is a dummy equal to 1 if $\theta = 2$ for illness C and $= 0$ if $\theta = 1$ for illness A, B. 'City' is a dummy being 1 if the physician is practicing in a city with more than 100,00 inhabitants. 'Outer conurbation' is an area with more than 100 inhabitants per square kilometer excluding cities with 100,000 inhabitants. The reference category is 'rural', a dummy which equals one if a physician practices in an area with less than 100 inhabitants per square kilometer. 'Years in practice controls' includes how many years a physician has been practicing employed in a practice, as self-employed resident physician, and in a hospital. 'Other characteristics' comprises ordinarily-scaled variables for attitudes towards altruism and competition from the World Value Survey and risk attitudes according to the German Socio-Economic Panel. Probit regressions yield very similar estimation results. *** $p < 0.01$, ** $p < 0.05$, and *$p < 0.1$.

Result 4. *Physicians' gender and location of practice significantly relate to the likelihood of crowding-out of altruistic behavior. Male physicians and physicians in cities and outer conurbation areas are less prone to crowding-out compared to females and to rural areas.*

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 complements capitation. Under performance pay, subjects increase, on aggregate, the quality of health care provision compared to non-blended capitation. The intensity of a response to performance pay is, however, significantly affected by the severity of illness. We also observe a crowding-out of patient-regarding behavior for around 14 percent 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 (RCT) in the field.

In our experiment, we found performance pay—implying an increase in a physician's maximum attainable payoffs by 5% and 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 undertreatment of high-severity patients under capitation. Nonetheless, 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 with a representative sample of primary care physicians, although we left little room in physcisians' 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., 2017; Godager et al., 2016). 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.

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.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009): "Large Stakes and Big Mistakes," *Review of Economic Studies*, 76, 451–469.

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 (2016): "Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision," *Journal of Economic Behavior & Organization*, 131, Part B, 17 – 23.

——— (2017): "The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence," *Health Economics*, 26, 243–262.

Buckley, N. J., K. Cuff, J. Hurley, L. McLeod, S. Mestelman, and D. Cameron (2012): "An experimental investigation of mixed systems of public and private health care finance," *Journal of Economic Behavior Organization*, 84, 713 – 729.

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.

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

Cutler, D. M. (1995): "The Incidence of Adverse Medical Outcomes under Prospective Payment," *Econometrica*, 63, 29–50.

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.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): “Individual risk attitudes: Measurement, Determinants, and behavioral consequences,” *Journal of the European Economic Association*, 9, 522–550.

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.

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.

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

Galizzi, M. M. and D. Wiesen (forthcoming): “Behavioral Experiments in Health,” in *Oxford Research Encyclopedia of Economics and Finance*, ed. by J. Hamilton, Oxford University Press, Oxford, UK.

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 (2016): “Does performance disclosure influence physicians? medical decisions? An experimental study,” *Journal of Economic Behavior & Organization*, 131, Part B, 36 – 46.

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

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.

Harrison, G. W. and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009–1055.

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 (forthcoming): "High-Powered Performance Pay and the Crowding out of Worker Non-Monetary Motivations," *Management Science*.

Kessler, J. B. and A. E. Roth (2012): "Organ Allocation Policy and the Decision to Donate," *American Economic Review*, 102, 2018–2047.

——— (2014): "Getting More Organs for Transplantation," *American Economic Review*, 104, 425–30.

Kesternich, I., H. Schumacher, and J. Winter (2015): "Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus?" *Journal of Public Economics*, 131, 1 – 11.

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

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.

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.

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.

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 Appendix

A.1 Parameters of the experiment

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
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
CAP+P4P: 5%-Bonus											
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	0.9	0.9	0.9	0.0	0.0	0.0	0.0	0.0	0.0
b_y	0.0	0.0	0.0	0.0	2.1	2.1	2.1	0.0	0.0	0.0	0.0
b_z	0.0	0.0	0.0	0.0	0.0	0.0	4.1	4.1	4.1	0.0	0.0
π_x	10.0	9.9	10.5	10.0	9.3	7.5	6.4	5.1	3.6	1.9	0.0
π_y	10.0	9.9	9.6	9.1	10.5	9.6	8.5	5.1	3.6	1.9	0.0
π_z	10.0	9.9	9.6	9.1	8.4	7.5	10.5	9.2	7.7	1.9	0.0
CAP+P4P: 20%-Bonus											
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	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	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. Λ is the lump-sum payment in CAP, $b_l^\bullet$ is the bonus paid when the quality requirement is met in CAP+P4P, and π is the physician's profit.

A.2 Instructions of the experiment

Notice that the text in squared brackets [ ] denotes text on the computer screen not contained in the instructions.

[*Text on Computer Screen*: **Welcome to the Experiment!**

Thank you very much for your participation. During the study, you will be asked to make decisions for which you will receive an allowance. This allowance we call payoff in the following. Your payoff depends on the decisions you make. At the end of the study, your total payoff will be transferred to you by the notary of the Zentralinstitut für die kassenärztliche Versorgung in Deutschland. Thereby, anonymity of your decisions is guaranteed.

The experiment will take about 30 minutes and consists of two parts. Before each part, you will receive detailed instructions that you can download during the respective part of the study using the 'Link to Instructions'. If possible, please print the instructions for your assistance before the respective part of the study starts.

Pls. note that neither your decisions in part I nor in part II will have any influence on the respective other part of the study. The study ends by a small questionnaire. Pls. klick OK to proceed to the instructions of part I of the study.]

Instructions to part I

In part I of the study you will participate in nine decision rounds.

Description of decision rounds

In each round, you decide as a physician on the medical treatment for a patient. That means, in each round you have to determine the quantity of medical services you wish to provide to this 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 of the nine decision rounds, you will consecutively and in random order face one patient who is characterized by one of the nine possible combinations of illnesses and degrees of severity. Each of these nine patients you can provide with a quantity of $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, or 10 medical services. Providing the medical treatment for each patient is independent from that for the other patients.

Payoff

In each round, you receive a lump-sum remuneration for treating the patient irrespective of the amount of medical treatment you provide. You also incur costs for treating the patient, which depend on the quantity of services you provide. Your payoff in each decision round is calculated by subtracting these costs from the lump-sum remuneration for treating the patient. Your remuneration, your costs and your payoff will be stated in Euro.

Each quantity of medical services yields a particular health status—contingent on illness and severity—, i.e., a particular benefit for the patient. Hence, in choosing the medical services you provide, you determine not only your own payoff but also the patient's benefit. The benefit is stated in monetary units (Euro).

Before taking your decision, in each round you will be shown on your screen the illness (A, B, or C), the severity of the illness (x, y or z), and—for each possible amount of medical treatment—your lump-sum remuneration, your costs, your payoff, as well as the benefit for the patient. You, therefore, need not calculate these values yourself.

Payment

At the end of the study, one of the nine rounds of this part of the study will be chosen at random. Your payoff in that round together with your payoff from part II of the study will be transferred to you by the notary of the Zentralinstitut für die kassenärztliche Versorgung in Deutschland.

The benefit (in Euro) that a patient gets from your medical treatment in the chosen round, will be beneficial for a real patient. The amount will be transferred to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, which will use the money exclusively for enabling the treatment of patients with eye cataract. Transferring the money to the Christoffel Blindenmission Deutschland e.V. will also be carried by the notary of the Zentralinstitut für die kassenärztliche Versorgung in Deutschland.

[*Text on Computer Screen*: In the following, you are kindly asked to answer some comprehension questions. Pls. note, that the comprehension questions are not meant to recommend taking a specific decision in the study to follow. The questions are only intended to improve and sharpen your understanding of the decision situation you will be facing in the study.]

Comprehension questions

Prior to the decision rounds, we kindly ask you to answer a few comprehension questions. They are intended to help familiarize yourself with the decision situation. Having answered all questions correctly, part I of the study will begin immediately. Otherwise you are asked to answer the respective question again.

Instructions to part II

In part II of the study you will again participate in nine decision rounds.

Description of decision rounds

As in part I of the study, in each round, you decide as a physician on the medical treatment for a patient. That means, you have to determine in each round the quantity of medical services you wish to provide to this patient for a given illness and a given severity of this illness.

As in part I, you will in the nine decision rounds consecutively and in random order face one patient who is characterized by one of the nine patients who is characterized by one of the three illnesses (A, B, C), and by one of the three different degrees of severity (x, y, z). Each of these nine patients you can provide with a quantity of $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, or 10 medical services. Providing the medical treatment for each patient is independent from that for the other patients.

Payoff

In each round, you receive a lump-sum remuneration for treating the patient 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 depend on the quantity of services you provide. **Your payoff in each decision round is calculated by the sum of the lump-sum remuneration and the bonus payment minus the costs from treating the patient.** Your lump-sum remuneration, your costs, your bonus payment and your payoff will be stated in Euro.

As in part I, each quantity of medical service yields a particular health status—contingent on illness and severity—, i.e., a particular benefit for the patient. Hence, in choosing the medical services you provide, you determine not only your own payoff but also the patient's benefit. The benefit is stated in monetary units (Euro).

Before taking your decision, in each round you will be shown on your screen the illness (A, B, or C), the severity of the illness (x, y or z), and—for each possible amount of medical treatment—the amounts of your lump-sum remuneration and the bonus payment, your costs, your payoff, as well as the benefit for the patient. You, therefore, need not calculate these values yourself.

Payment

At the end of the study, one of the nine rounds of this part of the study will be chosen at random. Your payoff in that round together with your payoff from part I of the study will be transferred to you by the notary of the Zentralinstitut für die kassenärztliche Versorgung in Deutschland.

As in part I of the study, the benefit (in Euro) that a patient gets from your medical treatment in the chosen round, will be beneficial for a real patient. The amount together with the amount from part I will be transferred by the notary to the Christoffel Blindenmission Deutschland e.V., 64625 Bensheim, which will use the money exclusively for enabling the treatment of patients with eye cataract.

[*Text on Computer Screen*: In the following, you are kindly asked to answer some comprehension questions. Pls. note, that the comprehension questions are not meant to recommend taking a specific decision in the study to follow. The questions are only intended to improve and sharpen your understanding of the decision situation you will be facing in the study.]

Comprehension questions

Prior to the decision rounds, we again kindly ask you to answer a few comprehension questions. They are intended to help familiarize yourself with the decision situation. Having answered all questions correctly, part II of the study will begin immediately. Otherwise you are asked to answer the respective question again.

A.3 Comprehension questions

The questions were asked for different benefit functions and for both CAP and CAP+P4P.

- Assume you want to provide for the patient shown in the table the quantity of services that yields the **lowest benefit** for this patient. Which quantity of medical services you have to choose?
- Assume you want to provide for the patient shown in the table the quantity of services that yields the **highest payoff** for you. Which quantity of medical services you have to choose?
- Assume you want to provide for the patient shown in the table the quantity of services that yields the **highest benefit** for this patient. Which quantity of medical services you have to choose?
- Assume you want to provide for the patient shown in the table the quantity of services that yields the **lowest payoff** for you. Which quantity of medical services you have to choose?