RUHR
ECONOMIC PAPERS

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

How Effective are Pay-for-Performance Incentives for Physicians?

A Laboratory Experiment

UNIVERSITÄT DUISBURG ESSEN

#413

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

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #413

Responsible Editor: Volker Clausen

ISSN 1864-4872 (online) – ISBN 978-3-86788-468-6

Ruhr Economic Papers #413

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

How Effective are Pay-for-Performance Incentives for Physicians?

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/10.4419/86788468
ISSN 1864-4872 (online)
ISBN 978-3-86788-468-6

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

How Effective are Pay-for-Performance Incentives for Physicians? – A Laboratory Experiment

Abstract

Recent reforms in health care have introduced a variety of pay-for-performance programs using financial incentives for physicians to improve the quality of care. Their effectiveness is, however, ambiguous as it is often difficult to disentangle the effect of financial incentives from the ones of various other simultaneous changes in the system. In this study we investigate the effects of introducing financial pay-for-performance incentives with the help of controlled laboratory experiments. In particular, we use fee-for-service and capitation as baseline payment schemes and test how additional pay-for-performance incentives affect the medical treatment of different patient types. Our results reveal that, on average, patients significantly benefit from introducing pay-for-performance, independently of hether it is combined with capitation or fee-for-service incentives. The magnitude of this effect is significantly influenced by the patient type, though. These results hold for medical and non-medical students. A cost-benefit analysis further demonstrates that, overall, the increase in patient benefits cannot overcompensate the additional costs associated with pay-for-performance. Moreover, our analysis of individual data reveals different types of responses to pay-for-performance incentives. We find some indication that pay-forperformance might crowd out the intrinsic motivation to care for patients. These insights help to understand the effects caused by introducing pay-for-performance schemes.

JEL Classification: C91, I11

Keywords: Physician incentive schemes; pay-for-performance; fee-for-service; capitation; laboratory experiment

April 2013

1 Jeannette Brosig-Koch, Nadja Kairies, Daniel Wiesen, University of Duisburg-Essen; Heike Hennig-Schmidt, University of Bonn. *– We thank the participants of the European Conference on Health Economics in Zurich, of the Economic Science Association International Conference in New York and the Economic Science Association European Conference in Cologne for their valuable comments and suggestions. Financial support provided by the Deutsche Forschungsgemeinschaft (German Research Foundation) is gratefully acknowledged.* *– All correspondence to Jeannette Brosig-Koch, University of Duisburg-Essen, Faculty of Economics and Business Administration, Universitätsstr. 12, 45117 Essen, Germany, E-Mail: jeannette.brosig-koch@ibes.uni-due.de.*

1. Introduction

For policy makers, improving quality in the health care market has been a major issue in recent years (McCellan, 2011). Thus, the reform of reimbursement for health care providers, i.e. physicians, has become seminal. As established capitation or fee-for-service schemes rather "pay" physicians for the quantity than for the quality of medical treatment, policy makers have started to implement monetary *pay-for-performance* (P4P) schemes to incentivize quality improvement. By now, payment schemes including P4P incentives have been introduced in many countries, e.g. the UK (Doran et al., 2006, Campbell et al., 2009), the USA (Rosenthal et al., 2004, and Rosenthal, 2008), Australia (Duckett et al. 2008, Scott 2008, Scott et al. 2009), France (Mousquès et al., 2012), Korea (Kim, 2010), New Zealand (Perkins and Seddon, 2006, and Buetow, 2008), or Spain (Gené-Badia et al. , 2007).

Existing P4P incentives vary to some extent, the key determinants are alike, however. First, P4P incentives are usually designed as a monetary bonus paid in addition to the basic payment, which usually is capitation (CAP), fee-for-service (FFS), or a combination of both.[1] P4P programs with a basic payment of CAP have been introduced, among others, in the UK within the Quality and Outcomes Framework (QOF), in the US within the California P4P Program, and in Spain. Countries that have implemented P4P programs with a basic FFS payment scheme include, e.g., the US (Medicare), Australia, New Zealand and France. Second, performance measures usually include a combination of structure, process and outcome measures (Donabedian, 2005). Third, the basis for the financial bonus is either an absolute measure such as targets or intervals (see, e.g., the QOF in the UK), based on improvements in the measure (see, e.g., the Medicare Physician Group Practice Demonstration in the US), or a relative ranking (see, e.g., the Value Incentive Program in Korea). Fourth, the size of the bonus is rather small, i.e., less than 5 percent of the entire remuneration (Borowitz et al., 2010). An exception is the QOF in the UK where the bonus amounts to about 20 percent (Doran et al., 2006).

Despite the increasing popularity of monetary P4P, it is yet unclear whether they actually enhance quality and efficiency of care (Borowitz et al., 2010). Reasons for this uncertainty include limited access to valid data and design problems leading to negative effects (see Maynard, 2012, for a survey). The latter comprise, among others, substituting away from the non-rewarded towards the rewarded aspects of quality (Mullen et al., 2010), improving quality to the performance target (Campbell et al., 2009), or gaming with quality indicators in terms of exception reporting (Gravelle et al., 2010). For the California P4P Program, Mullen et al. (2010), e.g., find that P4P did have a positive impact on some of the rewarded clinical measures. However, in the cases in which physicians substituted away from unrewarded aspects of medical treatment, the gain in quality of the P4P indicator could sometimes not be offset by the reduction in the other indicators. In the UK, incentives within

[1] Most P4P programs include incentives for primary care physicians, specialists, or hospitals. Since we will focus on the direct implications of introducing P4P incentives on physicians' behavior, hospitals are not considered as for their case P4P payments imply additional agency problems. Accordingly, specific payment forms used in hospitals such as budgets or DRGs are also not discussed in this study.

the QOF seemed to have resulted only in a short-term increase in quality due to quality thresholds (Campbell et al., 2009) and even triggered unintended gaming with the quality indicators, e.g., in terms of exception reporting (Gravelle et al., 2010).

While these examples underline problems of existing P4P schemes resulting from the specific designs of performance measures (in particular the difficulty of identifying quality), they do not tell much about the effectiveness of P4P incentives *per se*. Moreover, field research on P4P programs often faces the problem that relevant parameters are difficult to control like the implementation of additional incentives (e.g., public reporting), regional or institutional characteristics of study groups, or the health status of patients treated.[2] Hence, it is difficult to assess whether the quality metrics used are not reliable or whether the financial P4P incentives themselves fail to work. That introducing monetary rewards can have negative behavioral effects has been demonstrated in previous field and laboratory research. For example, running an IQ task and a donation collection task Gneezy and Rustichini (2000) observe that implementing a monetary bonus can significantly decrease work effort. Similar results have been obtained by Frey and Oberholzer-Gee (1997) on the willingness to host a nuclear waste repository and by Mellström and Johannesson (2008) on the supply of blood donors (see also Frey and Jegen, 2001, for a survey). Accordingly, it is an open question whether introducing P4P incentives aimed at achieving an optimal provision of medical care actually crowds out physicians' intrinsic motivation.

In order to prevent costly failures and to guarantee the success of P4P programs, a controlled analysis of the effects of financial P4P incentives on physician provision behavior is of great importance. The purpose of our paper is to provide such a controlled analysis of P4P incentives. In particular, our study contributes to previous research by investigating the effects of P4P incentives on the provision of medical care in a controlled laboratory experiment. Laboratory experiments are a new and emerging method in health economics (see, e.g., Hennig-Schmidt et al., 2011). It allows analyzing individual behavior in a controlled environment where only one parameter is changed at a time. External factors like patients' health status can be isolated and, if a change in behavior occurs, it can be attributed to the change of that one parameter, here the change of the payment scheme. Moreover, in the laboratory we can employ a reliable measure for patients' health status and, accordingly, the quality of care making it possible to isolate the effect of monetary P4P incentives from effects of deficient quality indicators. Laboratory experiments are a relatively inexpensive method to study behavioral responses to reforms. They are, thus, an effective additional tool to guide policy makers in designing appropriate incentive schemes.

In our experiment, participants act in the role of physicians and make decisions about the medical treatment of nine different patients. The treatment choice affects the patient benefit. The patients differ systematically in their illness and the degree of severity of these illnesses. We induce a sequential within-subject design to account for a payment reform given the same patient population before and after implementation. In part one of the

[2] Many of the P4P reimbursement schemes like the Quality and Outcomes Framework in the UK or Pacificare in California have been implemented along with public evaluation programs.

experiment, physicians decide about the quantity of medical treatment either under a capitation or a fee-for-service payment scheme. In part two, they are confronted with a P4P scheme based on capitation and fee-for-service payment, respectively. P4P is designed as a performance-based bonus paid in case physicians treat patients close to their optimum. The bonus applies to all patients; hence substitution away from unrewarded patients or exception reporting is impossible. The experimental design not only allows a controlled investigation of physicians' responses to P4P incentives, but also enables us to compare cost-benefit ratios between the different payment schemes. Moreover, due to our within-subject design we are able to identify whether and how patterns of behavior revealed in the basic payment schemes change with P4P incentives.

Our results demonstrate that introducing P4P incentives leads to treatment levels which are significantly closer to patient optimal levels than those observed with the baseline incentives of CAP and FFS. The degree to which patients benefit from P4P depends on their individual characteristics (i.e., the severity of their illness). These results hold for both, medical and non-medical student participants. Moreover, comparing patient benefits and physician remuneration between the different payment schemes reveals that the increase in benefits is outweighed by the additional cost of P4P incentives. Analyzing individual treatment behavior shows that there are different behavioral types and that P4P incentives differently affect these types. In particular, there is some indication that the financial incentives included in P4P might crowd out the intrinsic motivation to care.

The paper is organized as follows. In Section 2, we present design and procedure of our experiment with a special focus on how we designed the P4P incentives. In Section 3 we present our results and in Section 4 we conclude.

2. Experimental Design and Procedure

In the experiment, we analyze the impact of financial P4P incentives on the quantity of medical treatment and, thus, the patient health benefit. Capturing the character of a reform we use a sequential design consisting of two parts. In part 1 of the experiment, subjects are confronted with either one of the baseline payment schemes, CAP or FFS. In part 2, the only parameter we change is introducing a P4P incentive. This procedure allows a controlled ceteris paribus analysis.

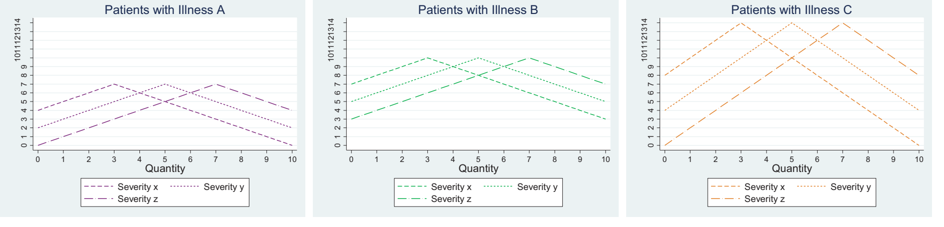
The basic design of the **decision situation** follows Brosig-Koch et al. (2013). Subjects are in the role of a physician i and decide on the quantity of medical services $q = \{0, 1, \dots, 10\}$ for nine different patients $j \in [0,1, \dots, 9]$ who vary in their illnesses $k \in [A, B, C]$, and the severities $l \in [x, y, z]$ of these illnesses. For each patient a physician receives a remuneration R depending on the experimental condition and bears costs $c_{kl}(q) = 0.1 \cdot q^2$ which are the same in all conditions.[3] With each decision, the physician determines her profit π^i_{kl} , i.e. $R - c_{kl}(q)$, as well as the patient's health benefit B_{kl}. Both, profits and patient benefits, are

[3] Convex cost functions are used in several theoretical models describing physician behavior (e.g., Ma, 1994 and Choné and Ma, 2011).

measured in monetary terms. Patients in the experiment are assumed to be passive and fully insured, accepting the quantity of medical treatment provided by the physician.

Patient benefit functions (see Figure 1) take on a global optimum on the quantity interval [0,10].[4] The optimal level of patient benefit depends on the illnesses, i.e. $B_{Al}(q^*) = 7$, $B_{Bl}(q^*) = 10$, and $B_{Cl}(q^*) = 14$, and is achieved when the patient optimal quantity q^* is provided. Moreover, each illness implies a certain slope of the benefit function, i.e., a certain change of benefit resulting from an additional unit of medical service. While the slope of the benefit function is the same for illnesses A and B, it is different for illness C. The patient optimal quantity q^* depends on the severity of an illness, i.e. $q^* = 3$ for severity *x*, $q^* = 5$ for severity *y*, and $q^* = 7$ for severity *z*. Taking q^* as a benchmark, we can identify the magnitude of overprovision and underprovision. The optimal amount of medical services is specified for each patient and is known to the physician.[5]

Figure 1: Patient Benefit Functions for Illnesses k and Severities l



Patients in the experiment are not present in the lab. The experimental participants are informed, however, that the monetary equivalent to the realized patient benefit is transferred to a charity (the *Christoffel Blindenmission*) caring for real patients with eye cataract (see also section 2.2). In the following, we explain the payment schemes used in the experimental conditions in more detail.

2.1 Experimental Conditions

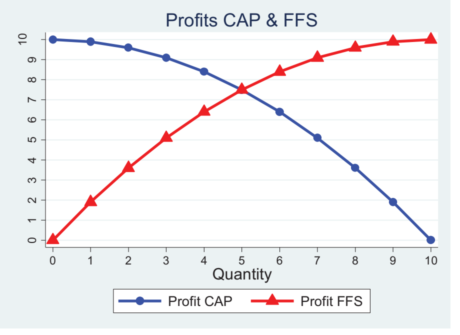
2.1.1. Baseline Incentive Schemes (Part 1)

The design of the baseline incentives CAP and FFS follows Brosig-Koch et al. (2013). Under CAP, a physician receives a lump-sum payment per patient, i.e. $R = LS = 10$. Thus, physician i's profit per patient is $\pi^i_{kl}(q) = 10 - c_{kl}(q)$. Under FFS, physicians are paid a fee of $p = 2$ per service they provide, i.e. $R(q) = 2q$. Accordingly, physician i's profit per patient is $\pi^i_{kl}(q) = 2q - c_{kl}(q)$ (see Figure 2).

[4] A concave patient benefit function has been widely assumed in theoretical papers; see e.g., Ellis and McGuire, 1986, Ma, 1994, Choné and Ma, 2011.

[5] Thus, there is no uncertainty about the impact of the chosen quantity of medical services on the patient's health benefit.

Figure 2: Profits of CAP and FFS Schemes in Part1 of the Experiment



The maximum profit a physician can achieve is identical for CAP and FFS ($\hat{\pi}_{kl}^{CAP} = \hat{\pi}_{kl}^{FFS} = 10$), as is the (absolute value of) marginal changes of profits. The profit maximizing quantity of medical services $\hat{q}$, however, differs and is 0 for CAP and 10 for FFS (see also Appendix A.3 for the complete set of parameter values). Given this parameterization, the profit functions for the two payment schemes are perfectly symmetric. Since patient benefit functions are also symmetric, we are able to fully compare behavior revealed under the two payment schemes.

2.1.2. P4P Incentive Scheme (Part 2)

P4P programs for primary care physicians and specialists use financial incentives to stimulate improvements in the quality of care and, in some cases, reductions in costs. Motivated by the composition of actual P4P schemes we chose our experimental parameters. We use the CAP and FFS payment schemes as a basis and provide an additional performance-based bonus in case the physician provides treatment within a predetermined performance interval. We link the performance measure to the individual patient's health benefit that can be interpreted as the health outcome of a certain medical procedure.

In particular, we design the P4P scheme as follows. P4P incentives apply to all patients; hence we exclude problems of multitasking or exception reporting. Physicians are rewarded for being "close" to the patient optimal quantity of medical care q^*, which is perfectly identifiable in our design. A physician is rewarded with a bonus for those quantity choices that do not differ by more than *one* unit from the patient optimal quantity q^*. Whenever the physician's quantity choice differs by more than one unit, she does not receive a bonus. The bonus is specified such that we provide higher incentives to treat those patients optimally that are in need of a high (low) quantity of medical services in CAP (FFS). More specifically, we designed the bonus in a way that maximum incentives are fully comparable across different patient types. This means, physicians receive the same maximum profit when treating the patients within the predetermined performance interval, irrespective of the severity of the illness.[6] We increased the physicians' profit maximum by 20 percent to $\hat{\pi}$=12 for all patient types to achieve a sufficient difference in incentives to the baseline schemes. This increase is in line with the Quality and Outcomes Framework in the UK which was intended to increase family physicians' income up to 25 percent depending on their performance (Doran et al., 2006). The remuneration from P4P is

$$\text{P4P} = \begin{cases} P_l & if\ |q - q^*| \leq 1 \\ 0 & otherwise \end{cases}$$

where P_l is the P4P bonus which depends on the patient's severity of illness l. Given that the physician's profit function in CAP is symmetric to the profit function in FFS, and given that the maximum profit with P4P incentives is equal to 12 for all patient types, we have symmetric profit functions also with P4P incentives. Table 1 shows the bonus for each severity in CAP and in FFS.

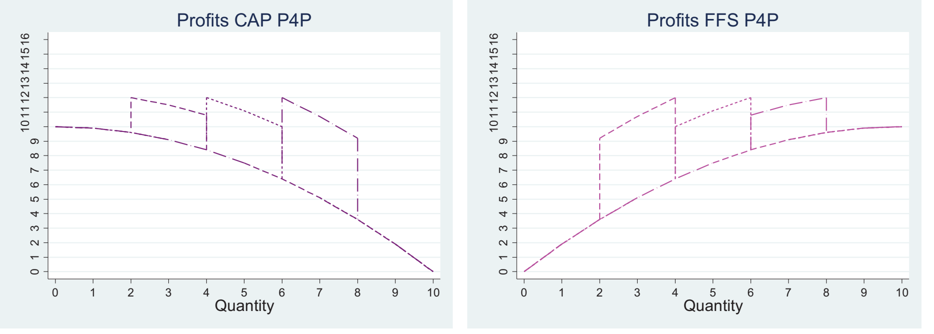
Table 1: Parameters for the P4P bonus in CAP and FFS

Severity	P_l^{CAP}	P_l^{FFS}
x (q*=3)	2.4	5.6
y (q*=5)	3.6	3.6
z (q*=7)	5.6	2.4

Physicians' total profits under the P4P payment schemes are calculated by $\pi_{kl}^i = R_{kl}^{CAP/FFS} P_k^{CAP/FFS} - 0.1q^2$ and are depicted in Figure 3 (see also Appendix A.3 for the complete set of parameter values). By paying the bonus if the chosen quantity is within an interval of quantities instead of if it exactly matches the patient optimal quantity, the physician's and the patient's interests are not perfectly aligned. The physician profit maximizing quantities are $\hat{q}_x^{CAP\ P4P}$=2, $\hat{q}_y^{CAP\ P4P}$=4, $\hat{q}_z^{CAP\ P4P}$=6 under CAP P4P and $\hat{q}_x^{FFS\ P4P}$=4, $\hat{q}_y^{FFS\ P4P}$=6, $\hat{q}_z^{FFS\ P4P}$=8 under FFS P4P, whereas the patient optimal quantities are q_x^*=3, q_y^*= 5 and q_z^*=7. Hence, the incentive from the baseline schemes are still inherent in the P4P schemes, but to a substantially lower extent.

[6] Recall that the severity of an illness determines the patient optimal quantity q*, which is related to a certain payment from the baseline schemes.

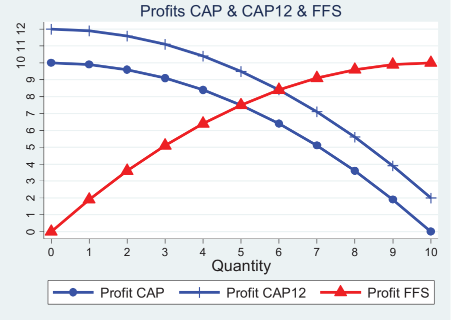
Figure 3: Profits of CAP P4P and FFS P4P Schemes in Part 2 of the Experiment



2.1.3. Constant Profit Maximum

Policy-makers who aim at reforming physician payment towards a P4P scheme are often constrained to a constant budget. Assuming selfish physicians, it appears reasonable to keep the profit maximum constant before and after reforming a payment scheme. To test the effect of a constant profit maximum, we model a CAP payment scheme with a profit maximum equal to the one in CAP P4P $\hat{\pi}$=12, i.e. CAP12. The lump-sum payment is assumed to be 12 resulting in physician *i*'s profit $\pi_{kl}^{i}(q) = 12 - c_{kl}(q)$ (see Figure 4).

Figure 4: Profits of CAP, FFS and CAP12 Schemes



2.2. Experimental Protocol

The computerized experiment was programmed with z-Tree (Fischbacher, 2007) and conducted at elfe, the Essen Laboratory for Experimental Economics of the University of Duisburg-Essen, Germany. Overall 44 medical students (who are supposed to become physicians in the future) and 56 non-medical students from the University of Duisburg-Essen participated in our experimental sessions. See Table 2 for an overview of experimental conditions. Subjects were recruited by the online recruiting system ORSEE (Greiner, 2004).[7]

Table 2: Number of Subjects per Condition and Type of Students

Condition	Payment Scheme Part 1 - Part 2	Type of Students		**Total**
		Non-Medical	Medical	
C1	CAP - CAP P4P	23	22	**45**
C2	FFS - FFS P4P	20	22	**44**
CAP12	CAP12 - CAP P4P	13	-	**13**
Total		**56**	**44**	**100**

The procedure follows Brosig-Koch et al. (2013). Prior to the actual experiment subjects were randomly assigned to their cubicles. They had enough time for reading the instructions for part 1 of the experiment (when deciding under one of the baseline payment schemes) and for privately asking the experimenter clarifying questions. After reading the instructions, subjects had to answer several control questions to make sure they had understood the decision task. Once everyone had answered these questions correctly, part 1 of the experiment started. The order of the nine patients to be treated was randomly determined and kept constant for all subjects in part 1 and part 2 of all conditions to make the data straightforwardly comparable across payment schemes (see Table 3).

Table 3: Randomized Order of Illnesses and Severities of Illness

Patient j	1	2	3	4	5	6	7	8	9
Illness k	B	C	A	B	B	A	C	A	C
Severity l	x	x	z	y	z	y	z	x	y

For each of the nine patients, subjects are informed on their decision screen about their remuneration, their cost and profit, as well as about the patient's benefit for each quantity of medical treatment. All monetary amounts are given in the experimental currency Taler the exchange rate being 1 Taler = €0.08.[8] In part 2 of the experiment, we applied exactly the same procedure except for subjects now being confronted with a P4P payment scheme. At the end of part 2, one decision for each part of the experiment was randomly chosen to be relevant for the subject's actual payoff and the patient benefit. We used this procedure to

[7] Students who registered in ORSEE to participate in laboratory experiments at the Essen Laboratory for Experimental Economics were invited via automatically generated e-mails and registered for a special session. We can thus say that subjects were randomly allocated to the experimental conditions. Moreover, subjects were not informed about the content of the experimental conditions unless they participated in a session.

[8] Instructions, control questions, and an example of the decision screen are included in Appendix A.2.

avoid wealth and averaging effects.[9] After the experiment, subjects privately received their payment and were dismissed.

The monetary value of patient benefits for the two payoff-relevant decisions aggregated over all subjects was transferred to the *Christoffel Blindenmission*. To verify this transfer, we applied a procedure similar to the one used in Hennig-Schmidt et al. (2011) and Eckel and Grossman (1996). After the experiment, a previously randomly determined subject acted as our monitor and verified that a correct transfer order on the aggregated benefit in the respective session was written to the university's financial department. The monitor and experimenter then walked together to the nearest mailbox and deposited the order in a sealed envelope. The monitor was paid an additional 5€.

Sessions lasted for about 70 minutes. Subjects earned, on average, 16.58€. The average benefit per patient was 13.38€. In total, 1,351.78€ were transferred to the *Christoffel Blindenmission*. The money supported surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists of the *Christoffel Blindenmission*. Average costs for such an operation amounted to about 30€. Thus, the money from our experiment allowed treating 45 patients.[10]

3. Results

In this section, we first present the average quantity of treatment physicians provide for each decision under the baseline payment schemes CAP and FFS. Second, we analyze whether patients actually benefit from P4P payment incentives and present a within-subject comparison of treatment behavior before and after a reform towards P4P payment. For this, we restrict the analysis to non-medical students. Third, we compare provision behavior between non-medical students and prospective physicians. Fourth, we check whether our results hold when keeping the profit maximum constant. Finally, we classify subjects according to their behavior in part 1 and describe type-specific changes of behavior in part 2.

3.1. Provision Behavior under Baseline Payment Schemes

Analyzing physician provision behavior under the baseline payment schemes, we find that subjects are not purely selfish in the sense that many of them do not provide their profit-maximizing quantity of treatment $\hat{q}_{kl}^{CAP}$=0 or $\hat{q}_{kl}^{FFS}$=10. However, in CAP patients are significantly underserved, i.e. the average quantities provided per subject per severity are significantly lower than the patient optimal quantities (p=0.0000 Fisher Pitman Permutation test for paired replicates, two-sided; FPPP in the following). In FFS, patients are significantly overserved, i.e. the average quantities provided per subject and severity are significantly higher than the patient optimum (p=0.0000, FPPP two-sided).

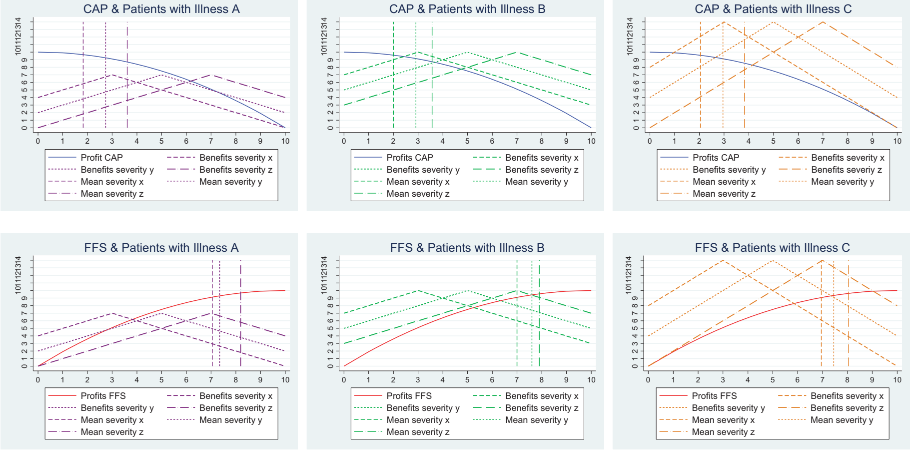
[9] Various studies confirm that the random payment technique does not affect the power of the monetary incentive for non-complex choice tasks (Starmer and Sugden, 1991, Cubitt et al., 1998, Laury, 2006, Baltussen et al., 2012).

[10] Subjects were not informed that the money would be transferred to a developing country in order to avoid motives like compassion.

The symmetry of physicians' profit functions and patients' benefit functions allows comparing the extent of deviations from the patient optimum between CAP and FFS. From a psychological point of view, people might regard choosing a higher quantity of medical treatment for the patient (which is incentivized in FFS) less severe than not providing this quantity to the patient (which is incentivized in CAP). We do not find such an effect, however. In both payment schemes, we observe a similar degree of over- or underprovision, respectively (p=0.5911, Fisher Pitman Permutation test for independent replicates, two-sided; FPPI in the following).

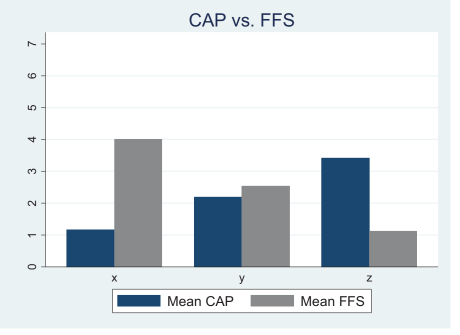
Figure 5 shows the average quantities chosen per patient for each severity (depicted by the vertical lines) as well as physician's profit and patient's benefit for each quantity of medical treatment in CAP and FFS.

Figure 5: Average Treatment Quantity per Severity in CAP and FFS



Testing the influence of patient characteristics on provision behavior yields a significant and systematic effect of the severity of illness (p<0.0746, FPPP, two-sided), but almost no significant effect of the illness itself (p>0.2722, except for 1 out of the 6 comparisons where p=0.0501, FPPP, two-sided). Thus, neither the maximum level of patient health benefit nor the change of benefit that is associated with an additional unit of medical treatment (both are implied by an illness) systematically affect provision behavior. Only the quantity yielding the maximum health benefit (which is implied by the severity of an illness) influences the choice of medical treatment. The more the patient optimal quantity deviates from the physician's profit maximizing quantity (i.e., the more severe the illness is in CAP and the less severe the illness is in FFS), the more does the quantity choice deviate from the patient optimum (p<0.0004, FPPP, two-sided), see Figure 6. These results are fully in line with the observations made by Brosig-Koch et al. (2013).

Figure 6: Distance between Patient Optimum and Chosen Quantity per Severity



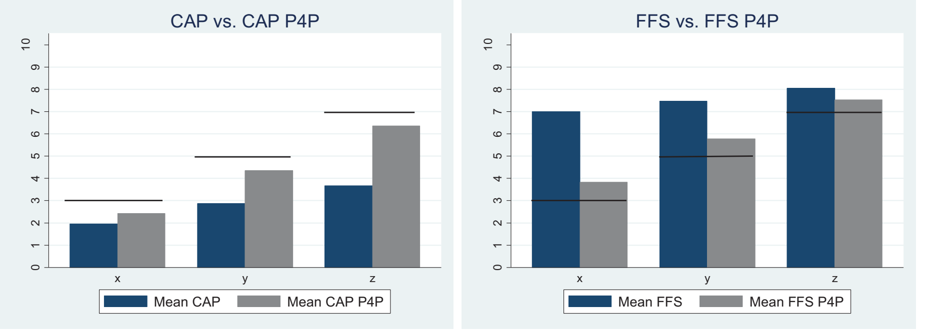
Result 1: Under CAP (FFS) physicians significantly underserve (overserve) patients. Deviations from the patient optimal medical treatment are significantly influenced by the severity of a patient's illness.

3.2. (Change of) Provision Behavior under P4P Incentives

Next, we compare subjects' provision behavior before and after introducing P4P incentives. As a patient's illness has almost no significant (and no systematic) effect on the quantity of treatment provided under P4P incentives ($p>0.6234$, except for 2 out of the 6 comparisons where $p<0.0232$, FPPP, two-sided), we pool decisions over illnesses in all subsequent analyses.

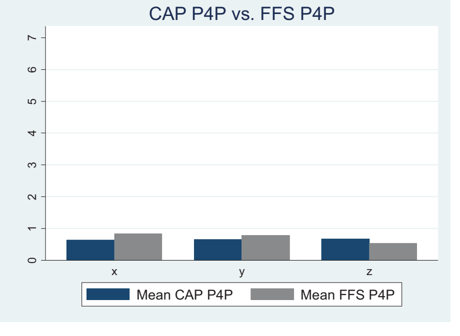
Figure 7 depicts the average treatment quantities for each severity without and with P4P incentives. The black horizontal lines indicate the patient optimal quantities. We find that implementing a bonus payment significantly reduces the underprovision observed under CAP ($p<0.0314$, FPPP, two-sided) and the overprovision observed under FFS ($p=0.0000$, FPPP, two-sided). Moreover, patients significantly benefit from introducing P4P incentives in CAP and in FFS as subjects in the role of physicians reduce their deviation from the patient optimal quantity to a similar extent ($p>0.5926$, FPPI, two-sided). The reduction of deviation is significantly affected by the severity of illness, though ($p<0.0002$, FPPP, two-sided). In particular, we observe the highest reduction for the most severe illness in CAP and the least severe illness in FFS, respectively.

Figure 7: Average Treatment Quantity per Severity without and with P4P



With P4P incentives, we still observe a significant underprovision under a CAP baseline payment (p=0.0000, FPPP, two-sided) and a significant overprovision under a FFS baseline payment (p=0.0000, FPPP, two-sided).[11] But, in contrast to the baseline incentives, the distances to the patient optimal level now only significantly differ across some severities in FFS P4P, i.e. between severities *x* and *z* (p=0.0019) and between severities *y* and *z* (p=0.0074 FPPP, two-sided; see Figure 8). The results are summarized in observation 2.

Figure 8: Distance between Patient Optimum and Chosen Quantity with P4P Incentives



Result 2: Patients benefit from introducing P4P incentives as these incentives significantly mitigate the underprovision (overprovision) observed with the baseline scheme CAP (FFS). Though, the deviations from the patient optimal treatment do not completely vanish with P4P.

3.3. Medical versus non-medical students

Our third research question analyzes whether differences in provision behavior exist between non-medical students – who are the typical subjects in laboratory experiments – and prospective physicians. In particular, we test the supposition that making a decision in a

[11] This result is supported by an OLS regression of the P4P incentive on the quantity of medical treatment, see Appendix A.1.

medical context might be induced by the different professional background and, accordingly leads to different decisions made by the two subject groups (see Ahlert et al., 2012).

Figure 9: Average Treatment Quantity per Severity for Medical and Non-medical Students

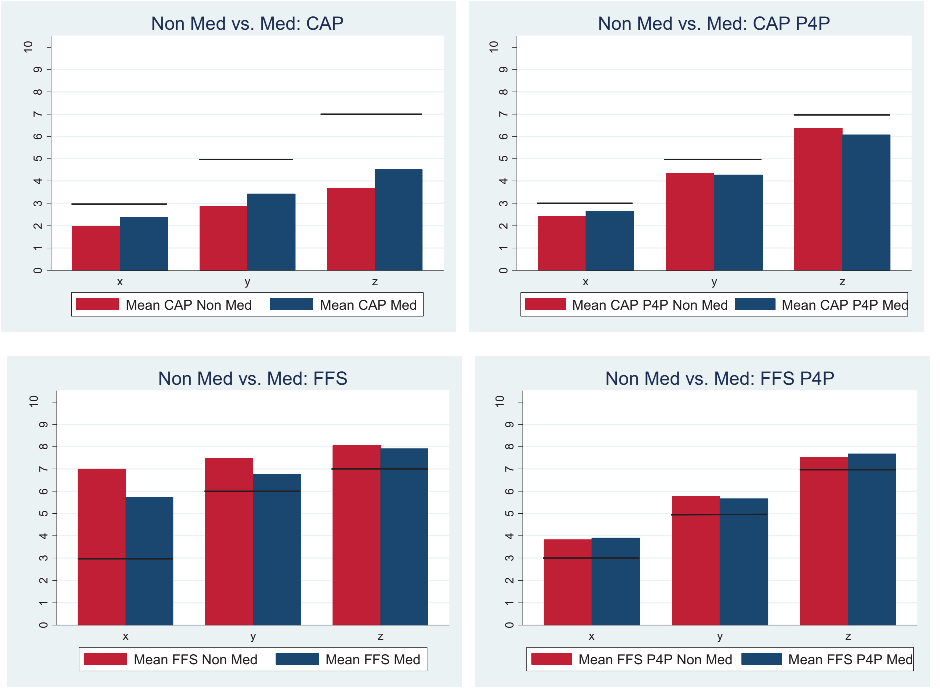


Figure 9 shows the average of treatment levels per severity for non-medical and medical students. There are no significant differences between the two subject groups in all payment schemes (p>0.1599, FPPI, two sided), except a weak one for severity *x* in FFS (p=0.0505, FPPT, two-sided). Only in the latter case we find that medical students are somewhat more patient-oriented than non-medical students. Moreover, all previous observations on the baseline payment schemes and the effect of P4P incentives also hold for prospective physicians.

Result 3: There are almost no significant differences in provision behavior between medical and non-medical students.

3.4. Robustness Check: Constant Profit Maximum

Introducing P4P incentives in our experiment is associated with an increase in physicians' maximum profit. We, therefore, test whether similar behavioral effects can be observed when keeping the profit maximum constant between the baseline and the P4P schemes.[12] In condition CAP12, we designed a CAP payment scheme in such a way that the physician profit

[12] As we find no differences between medical and other students, we pool the data from here on.

maximum is identical with and without P4P incentives, i.e. $\hat{\pi}$=12. We find no significant difference in average treatment quantities per severity provided between a CAP scheme with $\hat{\pi}$=12 (part 1 in condition CAP12) and a CAP scheme with $\hat{\pi}$=10 (part 1 in condition C1; p≥0.2397 FPPI, two-sided). Similarly, we find no significant difference between the CAP P4P schemes in part 2 of conditions CAP12 and C1, which have the same profit maximum of $\hat{\pi}$=12 (p>0.4405 FPPI, two-sided). Accordingly, patients benefit from introducing P4P incentives also when this payment reform is not associated with an increase of physicians' profit maximum.

Result 4: It appears that no increase in the maximum profit level is necessary to achieve the same patient health benefits.

This result raises the question on how patient benefits and the expenditure for physicians' payment are related to each other in the different conditions.

3.5. Ratio of Patient Benefit/Physician Remuneration

In this section, we analyze how patient benefits and the expenditure for physicians' remuneration are related in the different conditions. For this, we calculate the individual ratios *r* of the sum of benefits over the nine patients and the sum of respective remunerations and compare them between the different payment schemes.[13] As already noted, patient benefits significantly increase after introducing P4P incentives. But, expenses for physicians' remuneration also significantly increase with P4P incentives, particularly in conditions C1 and C2 (p<0.0140, FPPP, two-sided). Even in CAP12, where the profit maximum is the same in the baseline and in the P4P payment schemes, we find a significant increase in expenses (p=0.0002, FPPP, two-sided). We, thus, observe both, an increase in patient benefits and in physicians' remuneration. These results support recent field studies reporting that P4P schemes increase the quality of care, but come at increased costs (Mullen et al., 2010). Compared to these field studies, our controlled experimental environment allows clearly identifying the overall effect of P4P on the benefits per monetary unit spent (see Table 4).

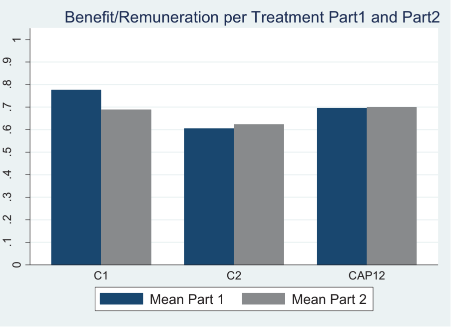
Table 4: Patient Benefit and Expenditure on Physician Remuneration per Condition

Condition	Part 1			Part 2		
	Patient Benefit	Expenditure on remuneration	Avg. *r*	Patient Benefit	Expenditure on remuneration	Avg. *r*
C1	7.7605	10.0000	0.7761	9.4074	13.7807	0.6876
C2	7.4656	14.2751	0.6056	9.3624	15.3016	0.6241
CAP12	8.3504	12.0000	0.6959	9.5983	13.8667	0.6991

[13] The minimum and maximum possible ratios for the four payment schemes given the nine patient types are: CAP (CAP P4P) r_{min}=0 (0), r_{max}=1.4 (1.13) and FFS (FFS P4P) r_{min}= 0 (0), r_{max}=5 (3.45). Note that the maximum levels for FFS (FFS P4P) are given for a quantity of one as for a quantity of zero the expenditure would be zero and thus the ratio undefined. However, we can exclude these cases as none of the subjects chose zero given FFS (FFS P4P).

Comparing the individual ratios *r* between the baseline and P4P schemes, we find that they do not differ significantly for C2 and CAP12 (p>0.6136, FPPT, two-sided), whereas they significantly decrease for C1 (p=0.0040, FPPP, two-sided). Thus, the patient benefits per monetary unit spent might even decrease with P4P incentives. Figure 10 illustrates the average benefit-expenditure ratios per condition before and after the P4P reform.

Figure 10: Average Benefit/Remuneration per Condition before and after the Reform



Result 5: The increase in patient benefits cannot overcompensate the additional expenditures associated with pay-for-performance incentives.

3.6. Individual Behavior

In this section, we focus on individual behavior and its changes between the two parts of the experiment. In particular, we identify four behavioral types: profit maximizing subjects (PMs) choosing their profit maximum quantity, benefit maximizing subjects (BMs) choosing the patient optimal quantity, trade-off types (TOs) choosing a quantity between the profit maximum and the patient optimal quantity, and others (Os) whose quantity choices cannot be explained by these types.[14] We define a subject to match one of these four behavioral types if the majority of his/her decisions in each part of the experiment is consistent with the classification criterion. If this is not the case we classify the subject as non-consistent (NCs).[15] The number of subjects per type in each condition and part of the experiment is given in Table 5. We pool our data over medical and non-medical students as the distribution of behavioral types across these subject groups is not significantly different (baseline: p=0.834, P4P incentives: p=0.970, two-sample Kolmogorov-Smirnov test; supporting our results in section 3.3). Similarly, we pool our data over (the symmetric) CAP and FFS schemes as we find also no significant effect across these schemes (baseline:

[14] Our design does not allow identifying social welfare optimizers, i.e. subjects who choose the quantity yielding the maximum sum of patient benefit and physician profit. The reason is that we cannot differentiate clearly between subjects who maximize social welfare and those who maximize patient benefits except for two of the nine patients for whom social welfare-maximizing quantities differ from patient optimal ones.
[15] This class comprises subjects that choose Pareto-inefficient quantities.

p=0.900, P4P incentives: p=0.970, two-sample Kolmogorov-Smirnov test; supporting our results in sections 3.1 and 3.2).

Table 5: Distribution of Patient Types per Part

	Baseline Payment	With P4P Incentives
PM	10	63
BM	26	33
TO	56	0
O	1	1
NC	7	3
Total	**100**	**100**

Our classification reveals that, without P4P incentives, 56 percent of subjects can be classified as TOs, 26 percent as BMs, and only 10 percent as PMs. After introducing P4P incentives, 63 percent of subjects can be classified as PMs. Also, the fraction of BMs is increased to 33 percent. Obviously, as the loss of profits associated with an optimal medical treatment decreases with P4P incentives, more subjects are willing to bear this lower loss to consistently treat their patients in an optimal way. Since the distance between the profit maximum quantity and the benefit maximum quantity is only one unit in part 2, it is not possible to identify TOs in this part. While, in part 1, there are 7 percent of subjects whose behavior is not consistent with one of the four types, in part 2 the percentage of NCs is decreased to 3 percent. Only 1 subject (1 percent of overall decisions) does not follow the behavioral pattern of PM, BM, or TO.

The analysis of individual changes of behavior between the two parts reveals a more detailed picture of how P4P incentives work (see Table 6). We find that only 29 percent of subjects do not change their type. These are particularly PMs (90 percent of them do not change their type) and BMs (77 percent of them do not change their type). For all other subjects (71 percent) the type classification changes with P4P incentives. The majority of them are TOs (56 out of 71) who mainly switch to either PM (45) or, to a minor degree, BM (8). The six BMs who change their behavior, switch to PM. The latter finding indicates that the monetary P4P incentives might crowd out the subjects' intrinsic motivation to maximize patient benefits, at least to some degree.

Table 6: Individual Type Changes from Part 1 to Part 2 per Patient

Type Changes	Number of Subjects
PM-PM	9
PM-O	1
BM-BM	20
BM-PM	6
TO-PM	45
TO-BM	8
TO-NC	3
O-BM	1
NC-PM	4
NC-BM	3
Total	**100**

Another way to classify subjects is to use a measure averaging the treatment behavior over all patients. We base our classifications on the individual absolute distance between the chosen quantity and the patient optimal quantity averaged over all nine patients per part (d_i).[16] The classifications are derived as follows.

First, we consider the classification of BMs. The patient optimal quantity averaged over all nine patients is $q^*_{kl} = 5$ irrespective of the payment scheme. Accordingly, patient optimal treatment implies that the chosen quantities averaged over all nine patients do not deviate from $q^*_{kl} = 5$, i.e. the patient optimal distance over all nine patients is $d^*_{kl} = 0$. To be classified as a BM we allow for an individual deviation of less than 0.5 from the patient optimal distance, i.e, $d_i < 0.5$.

Second, we derive the classification of PMs. With baseline incentives in part 1, the profit maximizing quantity for all nine patients is $\hat{q}_{kl} = 0$ for CAP and $\hat{q}_{kl} = 10$ for FFS. Thus, the profit maximizing distance to the patient optimal quantity ($q^*_{kl} = 5$) is $\hat{d}_{kl} = 5$ under both baseline schemes. With P4P incentives in part 2, the profit maximizing quantity for all nine patients is $\hat{q}_{kl} = 4$ for CAP P4P and $\hat{q}_{kl} = 6$ for FFS P4P. Hence, the profit maximizing distance to the patient optimal quantity ($q^*_{kl} = 5$) is $\hat{d}_{kl} = 1$ for both P4P schemes, We keep the classification of PMs comparable between the two schemes insofar as the thresholds, i.e. the allowed deviations from the optimal distances, are the same in both schemes. With baseline incentives in part 1, subjects are classified as PMs, if $d_i > 4.5$. With P4P incentives in part 2, subjects are classified as PMs, if $d_i > 0.5$.

Third, we consider TOs who trade-off patient benefits with own profits. Their individual distances are in between the ones of BMs and PMs. Hence, with baseline incentives in part 1 a subject is classified as TO, if $0.5 \leq d_i \leq 4.5$, and with P4P incentives in part 2 a subject is classified as TO, if $d_i = 0.5$.[17]

[16] Due to the symmetric design of the benefit and payment functions in part 1 and part 2, respectively, this measure does not need to be adjusted between CAP (CAP P4P) and FFS (FFS P4P).
[17] Note that given these classifications social welfare optimizers are identical to TOs.

Table 7: Distribution of Patient Types per Part

	Baseline Payment	With P4P Incentives
PM	10	67
BM	16	33
TO	74	-
Total	**100**	**100**

Given this classification, we find that, in part 1, 74 percent of subjects are TOs, 16 percent are BMs, and only 10 percent are PMs (see Table 7). This distribution of types is very similar to the one based on individual decisions. The percentages change when a P4P bonus is introduced in part 2 of the experiment. With P4P incentives, 67 percent of subjects can be classified as PMs and 33 percent as BMs. The shift toward profit maximizing behavior, again, underlines the effectiveness of financial incentives. At the same time we see that the lower loss of profits associated with an optimal medical treatment of patients induces more subjects to behave in this way.

Again, the comparison of the individual classification between the two parts reveals a more detailed picture of the functioning of P4P incentives (see Table 8). We find that, with this second classification, a similar percentage of subjects (25 percent) who do not change their type. Again, these are particularly PMs (100 percent of them do not change their type) and BMs (94 percent of them do not change their type). All other subjects (75 percent) do change their type classification with P4P incentives. The majority of them are TOs (74 out of 75) who mainly switch to either PM (56) or, to a minor degree, BM (18). The one BM who changes her behavior, switches to PM. That is, the insights generated with our first classification also hold for the second one. Note that with this second classification we find almost no hint for a crowding out of the intrinsic motivation to treat the patients in an optimal way.

Table 8: Switching behavior of Types between Part 1 and Part 2 over all Patients

Type Changes	Number of Subjects
BM-BM	15
BM-PM	1
PM-PM	10
TO-BM	18
TO-PM	56
Total	**100**

Result 6: Profit maximizing subjects and benefit maximizing subjects reveal a rather stable behavior between the two parts of the experiment. Particularly those, who trade off the own profit with the patient benefit, change their behavior with P4P incentives. The majority of them switches to profit maximization (which comes along with an increase in patient benefits).

4. Conclusion

The aim of this study is to analyze P4P incentives within a laboratory experiment. In contrast to a large part of the existing research, the laboratory environment allows isolating the effects of financial P4P incentives from other influencing factors and, thus, drawing conclusions on how P4P schemes work. In particular, the experimental setting allows introducing a clear-cut performance measure directly linked to the patient's benefit which is different from many field studies. Our data reveal that physician provision behavior under P4P schemes differs significantly from behavior under more traditional fee-for-service and capitation schemes. In particular, being paid a bonus for good performance, physicians provide patients with medical services much closer to their optimal treatment levels, hence improving the quality of care. The observed effects hold for both non-medical students (the pre-dominant population analyzed in laboratory experiments) and prospective physicians.

Our results lead us to conclude that in case quality metrics are well designed in order to prevent other problems observed in the field such as substitution from non-incentivized to incentivized aspects of medical care (Eggleston, 2005, Kaarboe and Siciliani, 2011), patient exclusion (Gravelle et al., 2010), or reliability of evidence-based outcome measures (Maynard, 2012), P4P incentives may actually lead to better patient outcomes.

However, comparing the benefit/expenditure ratio of payment schemes without and with P4P incentives, we observe in almost all treatments that this ratio is not significantly improved when P4P is introduced. In one case it even slightly decreases. From a welfare economic perspective, the increase in patient health benefits, thus, cannot overcompensate the additional expenditure due to P4P incentives in our parameter setting. This result is in line with findings of previous field studies (see e.g. Mullen et al. 2010). Hence, in case policy makers aim at improving the quality of care in terms of better treatment of patients, P4P incentives can be a successful policy means. However, if policy makers also want to improve the patient benefit per monetary unit spent, P4P incentives alone may not be sufficient. This insight holds even if the P4P scheme is designed in a way that the profit maximum is held constant compared to the baseline payment scheme. This may be an important indicator for policy makers in terms of money needed to provide P4P incentives.

Finally, our analysis of individual switching behavior supports the effectiveness of financial P4P incentives. While the majority of physicians trades off patient benefits against their own payoff under the basic payment schemes in part 1 of our experiment, most of them become profit maximizers in part 2 which goes along with the targeted increase in patient benefits. That is, aligning financial incentives for physicians with patient health benefits is a successful means to bring medical treatment closer to the patient optimum. Though, in some individual cases financial incentives might also crowd out the intrinsic motivation to treat patients well.

Literature

Ahlert, M., Felder, S., and B. Vogt (2012): Which Patients Do I Treat?: An Experimental Study with Economists and Physicians, *Health Economics Review*, 2(1), 1–11.

Baltussen, G., Post, G. T., Van Den Assem, M. J., and P.P. Wakker, (2012): Random incentive systems in a dynamic choice experiment. *Experimental Economics*, *15*(3), 418-443.

Borowitz, M., Scheffler, R.M. and B.M., Fulton (2010): Improving value for money in health by paying for performance. In: OECD Health Policy Studies, Value for Money in Health Spending; Paris: Organisation for Economic Co-operation and Development.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies, D. Wiesen (2013): How to improve patient care? An analysis of capitation, fee-for-service, and mixed incentive schemes for physicians, *Ruhr Economic Papers No*. 412 .

Buetow, S. (2008): Pay-for-Performance in New Zealand Primary Health Care, *Journal of Health Organisation and Management*, 22, 36-47.

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, 368-78.

Choné, P. and C.-t.A. Ma (2010): Asymmetric Information from Physician Agency: Optimal Payment and Health Care Quantity, Working Paper, University of Boston.

Cubitt, R. P., Starmer, C., and R. Sugden (1998): On the validity of the random lottery incentive system, *Experimental Economics*, 1(2), 115-131.

Donabedian, A. (2005): Evaluating the Quality of Medical Care, *The Milbank Quarterly*, 83 (4), 691-729.

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.

Duckett, S., Daniels, S., Kamp, M., Stockwell, A., and G. Walker (2008): Pay for Performance in Australia: Queensland's new Clinical Practice Improvement Payment, *Journal of Health Services Research and Policy*. 13, 174-177.

Eckel, C. and P. Grossman, (1996): Altruism in anonymous dictator games, *Games and Economic Behavior*, 16, 181.

Eggleston, K. (2005): Multitasking and Mixed Payment Systems for Provider Payment, *Journal of Health Economics* 24, 211-223.

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

Fischbacher, U. (2007): Z-tree: Zurich Toolbox for Readymade Economic Experiments, *Experimental Economics*, 10, 171-178.

Frey, B. S. and F. Oberholzer-Gee (1997): The cost of price incentives: An empirical analysis of motivation crowding-out, *The American economic review*, *87*(4), 746-755.

Frey, B. S. and R. Jegen (2001): Motivation crowding theory, *Journal of economic surveys*, 15(5), 589-611.

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

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

Gravelle, H., Sutton, M., and A. Ma (2010): Doctor behaviour under a pay for performance contract: treating, cheating and case finding?, *The Economic Journal* 120, 129–156.

Greiner, B. (2004): An Online Recruitment System for Economic Experiments, in V. M. Kurt Kremer (ed.), Forschung und Wissenschaftliches Rechnen 2003. GWDG Bericht 63, Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, 79-93.

Hennig-Schmidt, H., Selten, R., and D. Wiesen (2011): How Payment Systems Affect Physicians' Provision Behavior? An Experimental Investigation, *Journal of Health Economics*, 30(4), 637-646.

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

Kim, B.Y. (2010): Quality Assessment and Pay for Performance in Korea, HIRA, Korean Health Insurance Review Agency.

Laury, S.K. (2006): Pay One or Pay All: Random Selection of One Choice for Payment, Experimental Economics Center Working Paper Series, No 2006-24, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.

Ma, C.-t.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 (1), 3-12.

McClellan, M. (2011): Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality?, *Journal of Economic Perspectives*, 25, 69-92.

Mellström, C. and M. Johannesson (2008). Crowding out in blood donation: was Titmuss right?, *Journal of the European Economic Association*, *6*(4), 845-863.

Mousquès, J., Afritte, A., and Y. Bourgueil (2012): Efficiency of Multidisciplinary Group Practices in Primary Care – An assessment of a French Experiment of Pay for Performance for Integrated Primary Care Organization, mimeo, ECHE.

Mullen, K., Frank, R., and M. Rosenthal (2010): Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers, *The RAND Journal of Economics*, 41(1), 64-91.

Perkins, R. and M. Seddon, on behalf of EPIQ, (2006): Quality improvement in New Zealand healthcare. Part 5: measurement for monitoring and controlling performance–the quest for external accountability, New Zealand Medical Journal, 119.

Rosenthal, M.B., Fernandopulle, R., Song, H.R., and B. Landon (2004): Paying for Quality: Providers' Incentives for Quality Improvement, *Health Affairs*, 23 (2), 127-41.

Rosenthal, M.B., (2008): Beyond Pay for performance – Emerging Models of Provider Payment Reform, *New England Journal of Medicine*, 359, 1197-1200.

Scott, I. (2008): Pay for Performance Programs in Australia: A Need for Guiding Principles, *Australian Health Review*, 32, 740–9.

Scott, A., Schurer, S., Jensen P.H, and P. Sivey (2009): The Effects of an Incentive Program of Quality of Care in Diabetes Management, Health Economics, 9, 1091–108.

Starmer, C. and R. Sugden (1991): Does the random-lottery incentive system elicit true preferences? An experimental investigation, *The American Economic Review*, 81(4), 971-978.

APPENDICES

A.1. OLS Regression: P4P incentive on the quantity of medical treatment clustered by subject

VARIABLES	C1	C2
P4P incentive	1.546***	-1.789***
	(0.366)	(0.283)
Constant	1.285	9.294***
	(0.798)	(0.620)
Observations	414	360
R-squared	0.124	0.216
N_clust	23	20

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A.2 Instructions and Control Questions

Welcome to the Experiment!

Preface

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 equals 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 One

Please read the following 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 one of the experiment consists of 9 rounds of decision situations.

Decision Situations

In each round you take on 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.

Every 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.

Profit

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. 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.

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 - your costs, profit as well as the benefit for the patient with the corresponding illness and severity.

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

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 serverity (in Taler)

Which quantity of medical treatment do you want to provide?

Your decision:

OK

Payment

At the end of the experiment one of the 9 rounds of part one will be chosen at random. Your profit in this 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, an organization which funds 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 University of Duisburg-Essen's finance department. The form is then sealed in a postpaid envelope and posted 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 which 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 one of the experiment will begin once all participants have answered the comprehension questions correctly.

Part II

Please read the following 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 two of the experiment also consists of 9 rounds of decision situations.

Decision Situations

As in part one 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.

Every 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.

Profit

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 one, 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.

Patient 1 with illness

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 and severity (in Taler)

Which quantity of medical treatment do you provide?

Your decision:

OK

Patient 1 with illness

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 and severity (in Taler)

Which quantity of medical treatment do you want to provide?

Your decision:

OK

<u>Payment</u>

At the end of the experiment one of the 9 rounds of part two 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 one 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 one, no patients are physically present in the laboratory for part two 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, an organization which funds the treatment of patients with eye cataract.

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

<u>Comprehension Questions</u>

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 two of the experiment will begin once all participants have answered the 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 collaboration!

Comprehension Questions Part I: CAP (FFS)

Questions Tables 1-4:

1-4 a) What is the capitation (fee-for-service)?

1-4 b) What are the costs?

1-4 c) What is the profit?

1-4 d) What is the patient benefit?

Quantity of medical treatment	Capitation (Fee-for-service) (in Taler)	Costs (in Taler)	Profit (in Taler)	Benefit of the patient with illness F and severity y (in Taler)
0	20.00 (0.00)	0.00	20.00 (0.00)	15.00
1	20.00 (4.00)	0.20	19.80 (3.80)	16.00
2	20.00 (8.00)	0.80	19.20 (7.20)	17.00
3	20.00 (12.00)	1.80	18.20 (10.20)	18.00
4	20.00 (16.00)	3.20	16.80 (12.80)	19.00
5	20.00 (20.00)	5.00	15.00 (15.00)	20.00
6	20.00 (24.00)	7.20	12.80 (16.80)	19.00
7	20.00 (28.00)	9.80	10.20 (18.20)	18.00
8	20.00 (32.00)	12.80	7.20 (19.20)	17.00
9	20.00 (36.00)	16.20	3.80 (19.80)	16.00
10	20.00 (40.00)	20.00	0.00 (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	Capitation (Fee-for-service) (in Taler)	Costs (in Taler)	Profit (in Taler)	Benefit of the patient with illness G and severity z (in Taler)
0	20.00 (0.00)	0.00	20.00 (0.00)	10.00
1	20.00 (4.00)	0.20	19.80 (3.80)	12.00
2	20.00 (8.00)	0.80	19.20 (7.20)	14.00
3	20.00 (12.00)	1.80	18.20 (10.20)	16.00
4	20.00 (16.00)	3.20	16.80 (12.80)	18.00
5	20.00 (20.00)	5.00	15.00 (15.00)	20.00
6	20.00 (24.00)	7.20	12.80 (16.80)	22.00
7	20.00 (28.00)	9.80	10.20 (18.20)	24.00
8	20.00 (32.00)	12.80	7.20 (19.20)	22.00
9	20.00 (36.00)	16.20	3.80 (19.80)	20.00
10	20.00 (40.00)	20.00	0.00 (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.

Comprehension Questions Part II: CAP(FFS) P4P

1-4 a) What is the capitation (fee-for-service)?

1-4 a) What is the bonus payment?

1-4 b) What are the costs?

1-4 c) What is the profit?

1-4 d) What is the patient benefit?

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

1. Assume that a physician wants to provide 1 quantities 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	Capitation (Fee-for-service (in Taler)	Bonus payment (in Taler)	Costs (in Taler)	Profit (in Taler)	Benefit of the patient with illness G and severity z (in Taler)
0	20.00 (0.00)	0.00 (0.00)	0.00	20.00 (0.00)	10.00
1	20.00 (4.00)	0.00 (0.00)	0.20	19.80 (3.80)	12.00
2	20.00 (8.00)	0.00 (0.00)	0.80	19.20 (7.20)	14.00
3	20.00 (12.00)	0.00 (0.00)	1.80	18.20 (10.20)	16.00
4	20.00 (16.00)	0.00 (0.00)	3.20	16.80 (12.80)	18.00
5	20.00 (20.00)	0.00 (0.00)	5.00	15.00 (15.00)	20.00
6	20.00 (24.00)	11.20 (4.80)	7.20	24.00 (21.60)	22.00
7	20.00 (28.00)	11.20 (4.80)	9.80	21.40 (23.00)	24.00
8	20.00 (32.00)	11.20 (4.80)	12.80	18.40 (24.00)	22.00
9	20.00 (36.00)	0.00 (0.00)	16.20	3.80 (19.80)	20.00
10	20.00 (40.00)	0.00 (0.00)	20.00	0.00 (20.00)	18.00

3. Assume that a physician wants to provide 1 quantities 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.

A.3 Parameter Tables

Treatment	Variable	Quantity (q) 0	1	2	3	4	5	6	7	8	9	10
C1	R_{kl}^{I}	10	10	10	10	10	10	10	10	10	10	10
	R_{kl}^{II}	10	10	10	10	10	10	10	10	10	10	10
	$P_x^{II\ CAP}$	0	0	2.4	2.4	2.4	0	0	0	0	0	0
	$P_y^{II\ CAP}$	0	0	0	0	3.6	3.6	3.6	0	0	0	0
	$P_z^{II\ CAP}$	0	0	0	0	0	0	5.6	5.6	5.6	0	0
C2	R_{kl}^{I}	0	2	4	6	8	10	12	14	16	18	20
	$R_{kl}^{II\ FFS}$	0	2	4	6	8	10	12	14	16	18	20
	$P_x^{II\ FFS}$	0	0	5.6	5.6	5.6	0	0	0	0	0	0
	$P_y^{II\ FFS}$	0	0	0	0	3.6	3.6	3.6	0	0	0	0
	$P_z^{II\ FFS}$	0	0	0	0	0	0	2.4	2.4	2.4	0	0
CAP12	R_{kl}^{I}	12	12	12	12	12	12	12	12	12	12	12
	R_{kl}^{II}	10	10	10	10	10	10	10	10	10	10	10
	$P_x^{II\ CAP}$	0	0	2.4	2.4	2.4	0	0	0	0	0	0
	$P_y^{II\ CAP}$	0	0	0	0	3.6	3.6	3.6	0	0	0	0
	$P_z^{II\ CAP}$	0	0	0	0	0	0	5.6	5.6	5.6	0	0
all	c_{kl}	0	0.1	0.4	0.9	1.6	2.5	3.6	4.9	6.4	8.1	10
C1	π_{kl}^{I}	10	9.9	9.6	9.1	8.4	7.5	6.4	5.1	3.6	1.9	0
	π_x^{II}	10	9.9	12	11.5	10.8	7.5	6.4	5.1	3.6	1.9	0
	π_x^{II}	10	9.9	9.6	9.1	12	11.1	10	5.1	3.6	1.9	0
	π_x^{II}	10	9.9	9.6	9.1	8.4	7.5	12	10.7	9.2	1.9	0
C2	π_{kl}^{I}	0	1.9	3.6	5.1	6.4	7.5	8.4	9.1	9.6	9.9	10
	π_x^{II}	0	1.9	9.2	10.7	12	7.5	8.4	9.1	9.6	9.9	10
	π_x^{II}	0	1.9	3.6	5.1	10	11.1	12	9.1	9.6	9.9	10
	π_x^{II}	0	1.9	3.6	5.1	6.4	7.5	10.8	11.5	12	9.9	10
CAP12	π_{kl}^{I}	12	11.9	11.6	11.1	10.4	9.5	8.4	7.1	5.6	3.9	2
	π_x^{II}	10	9.9	12	11.5	10.8	7.5	6.4	5.1	3.6	1.9	0
	π_x^{II}	10	9.9	9.6	9.1	12	11.1	10	5.1	3.6	1.9	0
	π_x^{II}	10	9.9	9.6	9.1	8.4	7.5	12	10.7	9.2	1.9	0
all	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