Social Science & Medicine 278 (2021) 113939

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: http://www.elsevier.com/locate/socscimed

Allocation of health care under pay for performance: Winners and losers

Anne Sophie Oxholm [a,*], Sibilla Di Guida [b], Dorte Gyrd-Hansen [a]

[a] Danish Centre for Health Economics, Department of Public Health, University of Southern Denmark, J.B. Winsløws Vej 9B, 1st Floor, 5000, Odense C, Denmark
[b] Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230, Odense M, Denmark

ARTICLE INFO

JEL classification:
C91
I11

Keywords:
Pay for performance
Inequality
Access to care
Laboratory experiment
Denmark

ABSTRACT

Many physicians receive a payment for their performance (P4P). This performance is often linked to a health target that triggers a bonus when met. For some patients the target is easily met, while others require a significant amount of care to reach the target (if ever). This study contributes to the literature by providing evidence of how P4P affects allocation of care across patients with low and high responsiveness to treatment compared to a fixed payment, such as capitation and salary, under different degrees of resource constraint. Our evidence is based on a controlled laboratory experiment involving 143 medical students in Denmark in 2019. We find that patients who have the potential to reach the health target, gain care under P4P, whereas patients with no potential to reach it, may receive less care. Redistribution of care between patients under P4P arises when physicians are resource constrained. As many physicians are currently operating under tight resource constraints, policymakers should be careful to avoid unintended inequalities in patients' access to health care when introducing P4P. Risk-adjusting the performance target may potentially solve this issue.

1. Introduction

Physicians provide health care to improve their patients' health. However, similar to other workers, physicians may also be driven by profit or leisure (Ellis and McGuire, 1986; McGuire, 2000). If physicians' remuneration is independent of their effort, such as when they receive capitation payments, they face a trade-off between their altruistic preferences for delivering care to patients and their selfish interests in terms of monetary gain and/or increased leisure. Studies confirm this trade-off by showing that physicians supply fewer treatments when they are not paid for their activity (Hennig-Schmidt et al., 2011; Krasnik et al., 1990).

When treating patients, physicians may also face so-called patient opportunity costs, i.e. they may lack resources, such as time or equipment, to meet all their patients' needs. In fact, in many health care systems physicians are currently working close to their maximum capacity of care provision (see for example Emanuel et al. (2020); George and Gerada (2019)). In such cases, physicians face a trade-off between patients' care: if they provide more service to one patient, they have fewer opportunities to provide care to other patients. Studies find that patient opportunity costs affect physicians' treatment behaviour (see for example Di Guida et al. (2019); Oxholm et al. (2019)). More specifically, when physicians are not paid for their activity, it is especially the high-need patients that suffer from the lack of resources (Oxholm et al., 2019).

This paper focuses in particular on monetary and patient opportunity costs of providing care. These opportunity costs may lead physicians not to treat patients according to the payer's wishes. Payers may therefore try to incentivise physicians to change their behaviour. Pay for performance (P4P) is a popular incentive tool because it links physicians' payment directly to a health care target (Eijkenaar et al., 2013; Giuffrida et al., 1999). For some patients this target is easily met, while others require a significant amount of care before the target is met (if ever). If physicians' behaviour is driven by the bonus, they may select which patients to treat based on their responsiveness to treatment, and thereby create inequalities in access to care (Giuffrida et al., 1999).

Evidence is limited and mixed on how P4P affects inequalities in health care (Alshamsan et al., 2010; Eijkenaar et al., 2013; Milstein and Schreyoegg, 2016), and thus also on potential unintended consequences of this payment method in the form of cream-skimming behaviour. Many of the existing studies focus on geographical variation in care and suffer from weak evaluation designs. Our study contributes to this literature by providing causal evidence of how P4P affects physicians' allocation of care to different patient groups compared to a situation where they receive a fixed payment, such as salary or capitation, in the presence of different opportunity costs of treatment.

* Corresponding author. J.B. Winsløws Vej 9B, 1st floor, 5000, Odense C, Denmark.
E-mail addresses: asoxholm@sdu.dk (A.S. Oxholm), sidg@sam.sdu.dk (S. Di Guida), dgh@sdu.dk (D. Gyrd-Hansen).

https://doi.org/10.1016/j.socscimed.2021.113939
Received in revised form 6 April 2021; Accepted 11 April 2021
Available online 21 April 2021

As it is difficult to find suitable natural experiments, we make use of a laboratory setting with medical students. Laboratory experiments with medical students have been found to confirm findings from natural experiments with physicians (Galizzi and Wiesen, 2017; Lagarde and Blaauw, 2017) as well as from artefactual field experiments with physicians (Brosig-Koch et al., 2016b). Other studies have also tested the effect of P4P on physician behaviour in a laboratory setting. The effect of moving from capitation to P4P on the quality of care provided to different patient types has been shown with a representative sample of German general practitioners (Brosig-Koch et al., 2019) and with medical students (Brosig-Koch et al., 2013, 2016a). These studies conclude that P4P leads on average to higher quality of care provided across all patient types. This conclusion is, however, based on the assumption that physicians have sufficient resources to treat all patients optimally. In reality, physicians often face tight resource constraints forcing them to trade-off care between patients. We add to the literature by investigating whether P4P leads physicians to prioritise patients based on their profitability. Thus, we investigate possible inequalities arising from P4P when physicians face monetary and/or patient opportunity costs of care.

Inequality in health care can be measured in different ways. Some measure inequality in patients' health, whereas others measure inequality in their access to and utilisation of care (Culyer and Wagstaff, 1993). Studies show that the public often assesses equity in health care based on patients' health gains and outcomes (Ahlert and Schwettmann, 2017; Tsuchiya and Dolan, 2009). Our study therefore focuses on inequality in patients' health gains. Patients' health gains may differ due to many different factors, for example their compliance, comorbidities, age, and gender. Our experimental setting does not focus on these specific factors, but instead provides a generic framework in which some patients respond strongly to care, whilst others require more resources to obtain the same amount of gain. This approach follows other experimental studies on physicians' response to different payment schemes (e.g. Brosig-Koch et al. (2017a); Brosig-Koch et al. (2017b); Di Guida et al. (2019); Hennig-Schmidt et al. (2011); Martinsson and Persson (2019); Oxholm et al. (2019)).

Our findings suggest that patients who otherwise receive treatment below the performance target, but with a potential to reach it, gain care under P4P. On the other hand, patients who do not have the potential to receive enough treatment to reach the target, receive less care under P4P. This redistribution of care arises when physicians are resource constrained and thereby forced to trade-off care between their patients. As many physicians are currently lacking resources (see for example Emanuel et al. (2020); George and Gerada (2019)), policymakers may consider risk-adjusting the performance target to reduce unintended inequalities in healthcare under P4P.

Interestingly, we also find an unintended consequence of P4P when physicians are resource abundant. This finding differs from existing studies, which find that P4P leads to higher quality of care across all patient types when resources are abundant (Brosig-Koch et al., 2013, 2016a, 2019). Unlike the existing studies, we assume that patients gain from treatment above the performance target. We find that under P4P physicians reduce care (and thus the health gain) to patients who otherwise receive treatment above the performance target even though the foregone care does not help other patients reach the target. Thus, it is not a financial incentive driving physicians' behaviour, but potentially a consideration that care is less valuable above the performance target. From the payer's perspective, in these cases P4P is an additional cost, which results in less care provided to otherwise high-performing patients. We therefore conclude that when physicians are resource abundant, policymakers should consider risk-adjusting the performance target, such that health improving care is not reduced above target under P4P.

The paper is organised as follows. Section 2 presents a framework for understanding physicians' allocation of care under P4P. Section 3 explains the experimental setting. Section 4 present our results. Section 5 discusses our findings and concludes the paper.

2. A theoretical framework for understanding physicians' allocation of health care under P4P

The literature often uses agency theory to understand physician behaviour (McGuire, 2000; Scott, 2000). Following these studies (e.g. Ellis and McGuire (1986)), we assume that physicians choose the amount of effort (services) to exert into health care that maximises their utility. Thus, physician i's maximisation problem is:

$$\max_{e_i} u_i(e_i) = \alpha_i h(e_i) + r(e_i) - \gamma_i c(e_i), \tag{1}$$

$$s.t.\ e_i \leq \bar{e}$$

where u_i denotes the utility from providing effort, $e_i \geq 0$, into health care. The solution to this maximisation problem depends on the benefits and costs of providing effort. Several laboratory experiments (e.g. Brosig-Koch et al. (2016b); Godager and Wiesen (2013); Hennig-Schmidt et al. (2011)) have shown that the benefits are a function of the physicians' altruistic concerns, $\alpha_i h(e_i) > 0, \alpha_i h'(e_i) > 0$, $\alpha_i h''(e_i) < 0$, as well as their remuneration, $r(e_i) > 0$. We align our cost function with our experimental setting by assuming that the physicians' costs of providing care increases linearly in their effort, i.e. $\gamma_i c(e_i) \geq 0$, $\gamma_i c'(e_i) > 0$, $\gamma_i c''(e_i) = 0$. Our predictions are, however, similar in the case of a convex relationship between effort and costs. Also, aligned with our experiment, we capture the physicians' resource constraint, i.e. need for a minimum amount of leisure, by limiting their provision of effort to $\bar{e}$.

In line with the agency literature (e.g. Ellis and McGuire (1986)), we find that the interior solution to the maximisation problem is for physicians to exert effort until their marginal benefit is equal to their marginal cost of care:

$$\alpha_i h'(e_i) + r'(e_i) = \gamma_i c'(e_i) \tag{2}$$

From equation (2) it is clear that the remuneration scheme may play a role in physicians' provision of health care. As a baseline for our analysis, we consider a situation in which physicians receive only a fixed remuneration, $r(e_i) = s$, such as capitation and salary. As this remuneration is independent of the amount of effort exerted, $r'(e_i) = 0$, we find that physicians do not have a financial incentive to provide care to their patients. In fact, if the physicians themselves bear the treatment costs, $c'(e_i) > 0$, they are financially incentivised to reduce care (which is also shown in a laboratory setting by e.g. Brosig-Koch et al. (2016a); Brosig-Koch et al. (2013); Brosig-Koch et al. (2019); Hennig-Schmidt et al. (2011); Oxholm et al. (2019)).

The introduction of P4P may financially incentivise physicians to provide additional care to their patients. P4P schemes are designed in many different ways (Eijkenaar, 2013). Following our experimental setting, we consider a flat rate performance fee, implying that physicians receive a bonus, ρ, for each patient that reaches the performance target (see section 5 for a discussion of this choice). Physicians' maximisation problem thus becomes:

$$\max_{e_i} u_i(e_i) = \alpha_i h(e_i) + s + \rho p(e_i) - \gamma_i c(e_i), \tag{3}$$

$$s.t.\ e_i \leq \bar{e}$$

where $p(e_i) = \sum_{j=1}^{n} \theta_j(e_i)$ is the physicians' aggregated performance across their patient population, which depends on the binary performance measure, θ_j, linked to patient $j = 1 \ldots n$. The solution to this maximisation problem is not straight forward as the performance measure θ_j, is non-differentiable in effort. In the following, building on the work by Oxholm (2016), we argue that the solution depends on the amount of care required to reach the target as well as the existence of opportunity costs of care.

According to equation (2), physicians provide care until their

marginal benefit is equal to their marginal cost of effort. In our study, physicians may face both monetary as well as patient opportunity costs of effort. In cases, where physicians themselves bear the treatment costs, $c'(e_i) > 0$, they face monetary opportunity costs. Whereas if physicians are resource constrained, additional care to one patient implies forgone care to another patient, i.e. they face patient opportunity costs. P4P may therefore only incentivise physicians to provide extra care to patients compared to baseline (fixed payment) if the benefits (bonus, ρ, and additional health gain) offsets the physicians' opportunity costs of the additional care up to the performance target.

Physicians response to P4P also depends on whether patients otherwise are being provided enough care to reach the target and also their potential to reach it. Thus, physicians are only financially incentivised to increase effort to patients if they otherwise do not reach the target but have the potential to reach it. Some patients may be unable to be treated such that they reach target, whereas others receive more care than necessary to reach it even under no P4P. In such cases, physicians may have a financial incentive to substitute care away from these patients to help other patients reach the target. Lastly, as physicians do not receive a bonus for providing care above target, we hypothesize that patients who are provided just enough care to reach the target at baseline, receive the same amount of care under P4P.

Hypothesis 1. Patients who at baseline are not provided enough care to reach the performance target, but with a potential to reach it, receive just enough care to reach the target under P4P if the benefits offset the opportunity costs of additional care.

Hypothesis 2. Patients who at baseline are not provided enough care to reach the performance target, and without a potential to reach it, receive less care under P4P if the foregone care can help others reach the target.

Hypothesis 3. Patients who at baseline are provided just enough care to reach the target, receive the same amount of care under P4P.

Hypothesis 4. Patients who at baseline are provided more care than necessary to reach the target, receive less care under P4P if the foregone care can help others reach the target.

Physicians do not always face monetary or patient opportunity costs of treatment. An example could be salaried physicians with the resources to meet all patients' needs within their contractually fixed working hours. In such cases, we expect physicians at baseline to deliver care up to the target (if possible). Thus, we expect that the introduction of P4P does not incentivise these physicians to change their treatment patterns.

Hypothesis 5. In cases where physicians face no opportunity costs of care, patients receive the maximum available amount of care at both baseline and under P4P.

3. Experimental setting

We test our hypotheses using an incentivised computer-based experiment with 143 medical students (69% female, average age 23). An ex ante power calculation showed that 140 participants were sufficient to guarantee a power of 0.9 (see Appendix C1 for more information about this calculation). The experiment took place at the laboratory at University of Southern Denmark using a custom-made software developed using Z-tree (Fischbacher, 2007).

The experiment was developed following a within-subjects design. Each participant was asked to make 24 different decisions about the allocation of health care services. These decisions were framed such that each participant had to allocate health care services across 36 fictitious patients in 24 hypothetical workdays. The workdays were created systematically changing four variables: 1) the number of patients consulted (either one or two patients), 2) the maximum number of services that could be allocated (either 5 or 8 services), 3) the type of patient(s) consulted ("fast" patient, "slow" patient, "no reach" patient), and 4) the type of remuneration scheme (either "baseline" or "P4P").

All patients had the same initial health status of 60. To receive a bonus, the participant had to improve a patient's health enough to reach a predetermined performance target of 70 (see Table 1). The patient types differed in their responsiveness to care. As the types of patients differed in their responsiveness to care, they also differed in the number of services needed to reach the target. The first type ("fast") was very responsive to care and could reach the performance target with a relatively small number of services (3 services), the second type ("slow") was less responsive and therefore required a larger number of services to reach the target (5 services), while the last type ("no reach") was the least responsive and could never reach the target given the amount of services available. We did not allow patients to experience a decrease in health state above the performance target, i.e. they could not be overprovided (Di Guida et al., 2019). We opted for this design as our focus is mostly on situations of resource scarcity, where the case of overtreatment is unlikely, see section 2.

Each workday, participants could provide a predetermined maximum number of services (either 5 or 8 services) independently of the number of patients consulted. Following a well-established practice in experimental economics, we deliberately used the generic term "service", such that we did not prime participants. Service could therefore be interpreted as time dedicated to a patient as well as the use of a machinery that has to be shared by several patients (for example echographer, microscopes to analyse blood samples, and spirometers). The number of available services was specified each day as were the number of patients consulted by the physician. Some days the physicians consulted one patient whereas other days they consulted two patients, thereby facing a trade-off in patients' care. We varied the number of patients consulted per day to investigate the importance of patient opportunity costs of treatment. The physicians' degree of resource constraint depended on both the number of patients (either 1 or 2 patients) consulted in a day and on the number of services the physician could allocate (either 5 or 8 services).

Each participant treated patients under two remuneration schemes (within-subject design): a scheme where only a fixed payment was provided, and a scheme where a bonus was given for each patient that reached the performance target on top of the fixed payment (more details about the remuneration of the participants follow). The order of these two remuneration schemes was randomised to avoid order effects. Furthermore, within each remuneration scheme the order of the different workdays was randomised to avoid anchoring effects.

Pay for performance schemes may be designed in many different ways (Eijkenaar, 2013). We chose a flat rate fee of 6 experimental dollars for each patient reaching the performance target. The performance pay was linked to an absolute performance target of 70. Thus, the performance measure was independent of other participants' performances. Furthermore, the measure was based on patients' health state (an outcome measure). Section 5 discusses the expected implications of these payment characteristics for our findings.

To investigate the importance of monetary opportunity costs on physicians' response to P4P, we designed two experimental treatments involving two types of fixed payments: one which resembled a capitation-based payment scheme, where the participant bears the monetary treatment costs, another which resembled a fixed salary scheme, where the participant does not bear any monetary treatment costs. In both treatments, participants received a daily endowment of 10 experimental dollars. However, half of the participants faced monetary costs of 1 experimental dollar per service provided (72 participants), while the other half experienced no monetary costs (71 participants), see Fig. 1. In both treatments, participants were facing the same 24 workdays as described above. For simplicity, we henceforth denote the payment schemes "SAL" and "P4P SAL" for the physicians who did not face any monetary treatment costs, and "'CAP" and "P4P CAP" for physicians facing monetary costs. As our aim is to measure the within-subject

Table 1
Patient's health state for a given number of health care services provided.

Number of services	Patient type		
	Fast	Slow	No reach
0	60	60	60
1	64	62	61
2	68	64	62
3	72	66	63
4	76	68	64
5	80	70	65
6	84	72	66
7	88	74	67
8	92	76	68
9	96	78	69
10	100	80	70

Note: The shaded area indicates a health state at or above the P4P target.

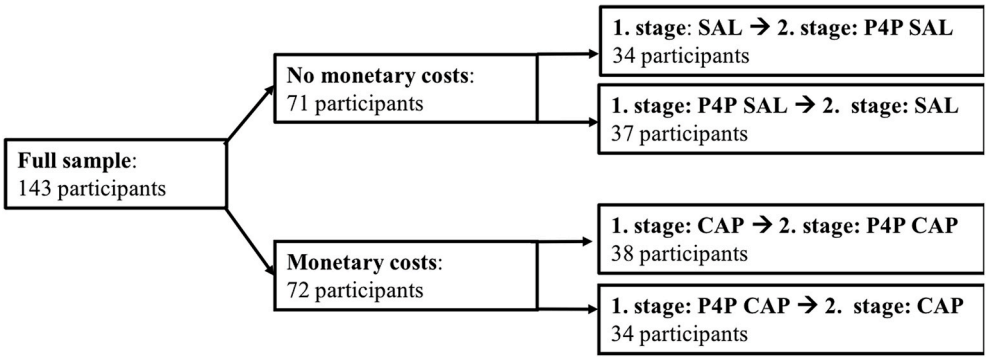

Fig. 1. Experimental design.

change in service delivery when P4P is introduced in the presence of various opportunity costs, we do not conduct a between-subject analyses across CAP and SAL scenarios.

Participation in the experiment was voluntary. The experiment was advertised during several classes in the medical programme at University of Southern Denmark roughly a week before each session took place. The different experimental sessions were run at different time slots during the day, and in different periods within the academic semester. Participants were recruited among students at different stages in their education, but who were yet to choose a medical specialisation. The gender ratio of our sample (69% female) is representative of the gender ratio of the students enrolled in a Danish medical program in 2019 (Ministry of Higher Education and Science, 2021). The use of medical students as experimental participants is widespread within the field of health economics (Di Guida et al., 2019; Hennig-Schmidt et al., 2011; Hennig-Schmidt and Wiesen, 2014; Oxholm et al., 2019). As shown by Brosig-Koch et al. (2016b), laboratory experiments with medical students yield qualitatively similar results to experiments with real physicians as participants. It is therefore reasonable to draw conclusions about physicians' behaviour based on data from medical students.

The experiment was incentivised. Participants were paid based on the outcome of two randomly drawn workdays. On average the experiment yielded the participants DKK 113 (USD 17), which included a show-up fee of DKK 40 (USD 6). As patients were fictional in our experiment, we incentivised participants to choose what they thought was best for the fictional patients by transforming their health gains into money and donating it to health-related charities which have an impact for real patients. Similar mechanisms have been employed in recent behavioural experiments in health analysing physician behaviour (e.g. Brosig-Koch et al. (2016a, 2016b); Brosig-Koch et al. (2019); Brosig-Koch et al. (2017a); Brosig-Koch et al. (2017b); Byambadalai et al. (2019); Di Guida et al. (2019); Godager et al. (2016); Godager et al. (2021); Hennig-Schmidt et al. (2011); Hennig-Schmidt and Wiesen (2014); Kesternich et al. (2015); Lagarde and Blaauw (2017); Martinsson and Persson (2019); Oxholm et al. (2019); Wang et al. (2020)). The names of the charities were communicated at the end of the experiment to avoid personal biases of the participants either against or in favour of specific charities. The selected charities had ongoing "patient support" activities, so the donations were beneficial for real patients.

Participants filled in a questionnaire proving whether their task was understood both at the start and end of the experiment. Participants in a pilot study found the task easy to understand and not tiring. Appendix C provides more details about the experimental protocol, design, instructions, and questionnaires. The experiment lasted roughly an hour, including the time needed to be allocated to a workstation, reading the instructions, and answering to pre- and post-experimental questionnaires. The choice task lasted between 10 min and 30 min.

Following other studies using laboratory experiments with medical students to analyse physicians' behaviour (Di Guida et al., 2019; Hennig-Schmidt et al., 2011; Hennig-Schmidt and Wiesen, 2014), we referred to the participants as "physicians", the fictional patients as "patients", and to each decision as a work "day" during the experiment.

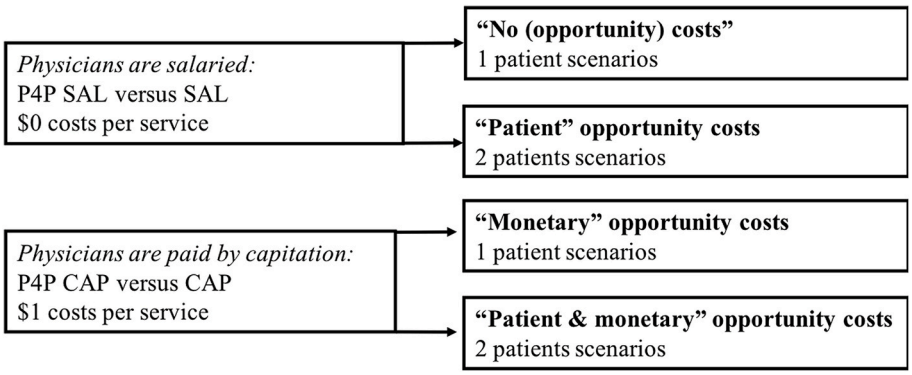

Fig. 2. Scenarios with or without opportunity costs of treatment.

While we called each decision a "day" to align with the literature (Di Guida et al., 2019), it does not have to be imagined necessarily as such a time interval. Instead of a day, it can represent a time slot within a longer working day, where the physician decides how much time to dedicate to patients.

4. Results

To analyse P4P's impact on physicians' behaviour, we compare their treatment under P4P with their corresponding treatment at baseline where they only receive a fixed payment (a within-subject analysis design). In Section 2, we hypothesize that the consequence of P4P for patients' care depends on both the care they receive at baseline (below, at or above the performance target) and the opportunity costs that the physicians face. Our results are therefore grouped based on both patients' care at baseline and the different opportunity costs. Fig. 2 illustrates the four different combinations of opportunity costs, which are characterised by the existence of monetary opportunity costs ($0 or $1 per service, SAL or CAP) and/or patient opportunity costs of treatment (1 or 2 patients per day).

We focus on our results for the scenarios with a high degree of resource constraints, i.e. a maximum of 5 services provided per day, and thus potentially large patient opportunity costs. However, appendix B shows that our conclusions are similar in cases of 8 services per day, and thus are robust to the specific number of services available. Appendix A and B provides detailed descriptive statistics and information about all our results. All presented test statistics are based on a two-sided *t*-test with standard errors obtained by bootstrapping (resampling clustered by physician, 10,000 replications). As we use a within-subject analysis design, we automatically control for individual physician differences, such as their gender, age etc., across the compared payment schemes (P4P versus fixed payment).

In sections 4.1 to 4.4 we consider situations where physicians face opportunity costs of treatment, therefore presenting data from the scenarios "SAL, two-patients", "CAP, one-patient", and "CAP, two-patients". In section 4.5 we consider situations where physicians face no opportunity costs of treatment and present data from the "SAL, one-patient" scenarios.

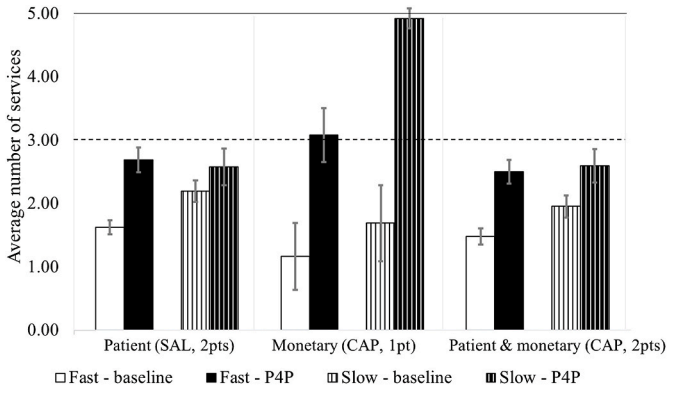

Fig. 3. This figure shows the average number of services (95%-confidence intervals are indicated) provided at baseline versus under P4P to patients who are provided below target at baseline but with a potential, i.e. "fast" and "slow" patients. The "fast" patients require 3 services to reach the target (dashed line), whereas the "slow" patients require 5 services (solid line). The physicians may provide up to 5 services per day. "Baseline" indicates a situation where physicians only receive a fixed payment, i.e. SAL or CAP, whereas "P4P" includes both the fixed payment and P4P. We present three different scenarios with opportunity costs: "Patient" opportunity cost (SAL, two-patients scenarios), "Monetary" opportunity cost (CAP, one-patient scenarios), "Patient & monetary" opportunity costs (CAP, two-patients scenarios).

4.1. Opportunity costs of treatment: patients provided below target at baseline with a potential

First, we focus on patients who are not provided enough care to reach the target at baseline but with a potential to reach it. As expected, we find that these patients receive more care under P4P, see Fig. 3. We observe the largest average increase in care if physicians only face monetary opportunity costs (slow patients: 3.2 services, $p < 0.001$; fast patients: 1.9 services, $p < 0.001$). In such cases, both patients with a

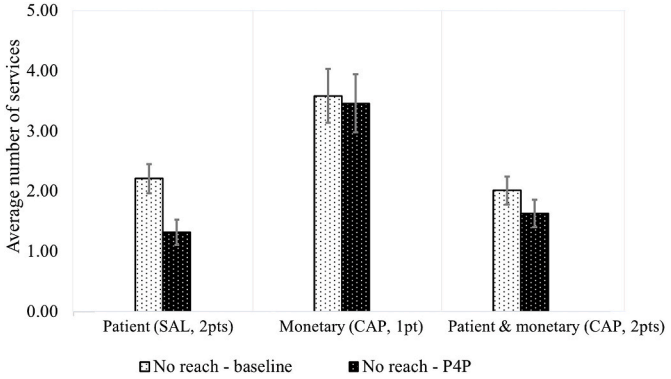

Fig. 4. This figure shows the average number of services (95%-confidence intervals are indicated) provided at baseline versus under P4P to patients who are provided below target at baseline without a potential to reach it, i.e. "no reach" patients. The physicians may provide up to 5 services per day. "Baseline" indicates a situation where physicians only receive a fixed payment, i.e. SAL or CAP, whereas "P4P" includes both the fixed payment and P4P. We present three different scenarios with opportunity costs: "Patient" opportunity cost (SAL, two-patients scenarios), "Monetary" opportunity cost (CAP, one-patient scenarios), "Patient & monetary" opportunity costs (CAP, two-patients scenarios).

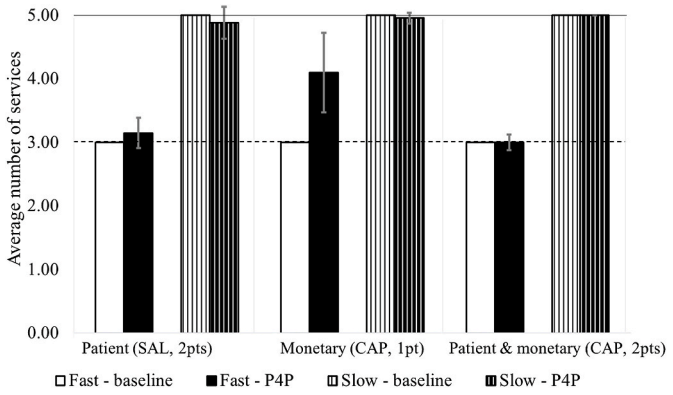

Fig. 5. This figure shows the average number of services (95%-confidence intervals are indicated) provided at baseline versus under P4P to patients who are provided at target at baseline, i.e. "fast" and 'slow' patients. The "fast" patients require 3 services to reach the target (dashed line), whereas the "slow" patients require 5 services (solid line). The physicians may provide up to 5 services per day. "Baseline" indicates a situation where physicians only receive a fixed payment, i.e. SAL or CAP, whereas "P4P" includes both the fixed payment and P4P. We present three different scenarios with opportunity costs: "Patient" opportunity cost (SAL, two-patients scenarios), "Monetary" opportunity cost (CAP, one-patient scenarios), "Patient & monetary" opportunity costs (CAP, two-patients scenarios).

slow and fast responsiveness to care on average reach the target. On the other hand, if physicians face patient opportunity costs, the average increase in care is more modest but still statistically significant (slow patients: 0.4–0.6 services, $p < 0.001$; fast patients: 1.0–1.1 services, $p < 0.001$). Thus, in such cases the average number of services are still below target, indicating that for some physicians the benefits do not outweigh the opportunity costs of additional care. However, overall, we confirm Hypothesis 1.

Result 1. *Patients who at baseline are not provided enough care to reach the performance target, but with a potential to reach it, receive additional care under P4P.*

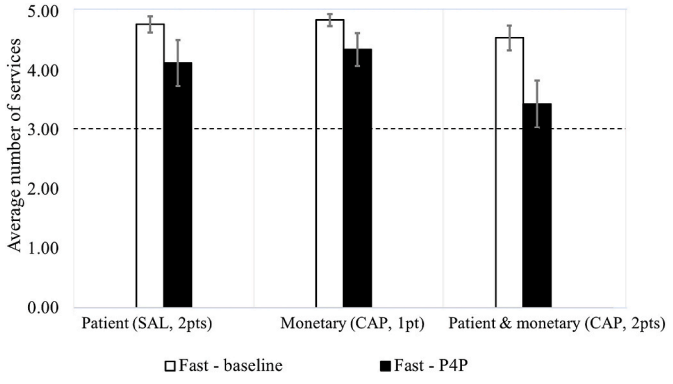

Fig. 6. This figure shows the average number of services (95%-confidence intervals are indicated) provided at baseline versus under P4P to patients who are provided above target at baseline, i.e. "fast" patients. The "fast" patients require 3 services to reach the target (dashed line). The physicians may provide up to 5 services per day. "Baseline" indicates a situation where physicians only receive a fixed payment, i.e. SAL or CAP, whereas "P4P" includes both the fixed payment and P4P. We present three different scenarios with opportunity costs: "Patient" opportunity cost (SAL, two-patients scenarios), "Monetary" opportunity cost (CAP, one-patient scenarios), "Patient & monetary" opportunity costs (CAP, two-patients scenarios).

4.2. Opportunity costs of treatment: patients provided below target at baseline without a potential

We now focus on patients who are not provided enough care to reach the target at baseline and without a potential to reach it. As hypothesized, the consequences of P4P for these patients' treatment depends on physicians' opportunity costs, see Fig. 4. If physicians face patient opportunity costs, we find a reduction in care (−0.9 to −0.4 services, $p < 0.001$) under P4P. However, we observe no statistically significant change in care (−0.1 services, $p = 0.327$) if physicians only face monetary opportunity cost. Thus, as expected, these patients only lose care under P4P if there are other patients that could gain from it. We thereby confirm Hypothesis 2.

Result 2. *Patients who at baseline are not provided enough care to reach the performance target, and without a potential to reach it, receive less care under P4P if the foregone care can help others reach the target.*

4.3. Opportunity costs of treatment: patients provided at the target at baseline

In this subsection we focus on patients who are provided just enough care to reach the target at baseline. By construction, "No reach" patients are not a part of this sample. Surprisingly, we find that care to these patients otherwise provided care at the target depends on physicians' opportunity costs of treatment, see Fig. 5. As expected, we see no change in care to these patients under P4P if physicians face patient opportunity costs. However, against our expectations, if physicians only face monetary opportunity costs, they become more generous towards these patients as P4P is introduced. Thus, in these cases we find that (if possible) these patients on average receive additional care under P4P (fast patients: 1.1 services, $p < 0.001$). We thereby only confirm Hypothesis 3 in cases where physicians face patient opportunity costs.

Result 3. *Patients who at baseline are provided just enough care to reach the target, receive the same amount of care under P4P if physicians face patient opportunity costs. If physicians only face monetary opportunity costs, then these patients receive additional care under P4P.*

4.4. Opportunity costs of treatment: patients provided above target at baseline

We focus next on patients who are provided more care than

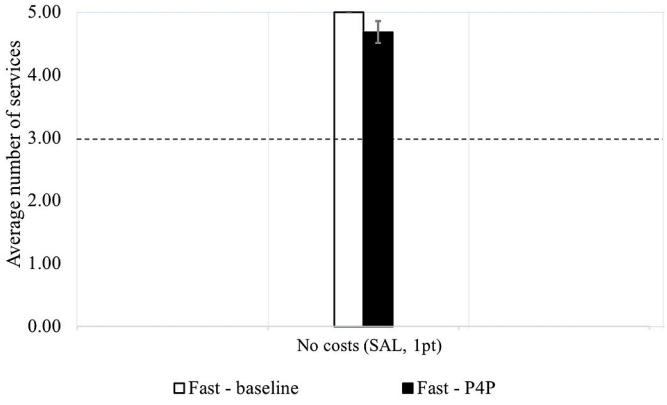

Fig. 7. This figure shows the average number of services (95%-confidence intervals are indicated) provided at baseline versus under P4P to patients who are provided above target at baseline, i.e. 'fast' patients. The 'fast' patients require 3 services to reach the target (dashed line). The physicians may provide up to 5 services per day. 'Baseline' indicates a situation where physicians only receive a fixed payment (SAL), whereas 'P4P' includes both SAL and P4P. We present one scenario without opportunity costs: 'No costs' (SAL, one-patients scenarios).

necessary to reach the target at baseline. By construction, when only 5 services can be allocated, this scenario is only feasible for "fast" patients. Across different types of opportunity costs, these patients receive fewer services under P4P, see Fig. 6. The most pronounced reduction in care arises when physicians face both monetary and patient opportunity costs. In such scenarios the fast patients on average receive 1.1 fewer services ($p < 0.001$) under P4P. Our finding that P4P leads to a reduction in care to patients provided care above target at baseline is surprising. According to Hypothesis 4, there is only a financial incentive to reduce care if the foregone care can help others reach the target. However, as only one patient is able to reach the target during a "day" in our 5-service setting, the foregone care cannot help others reach the target, and thus the bonus is not driving physicians' response.

Result 4. *Patients who at baseline are provided more care than necessary to reach the target, receive less care under P4P in the presence of monetary and/or patient opportunity costs of treatment.*

4.5. No opportunity costs of treatment: patients provided above target at baseline

Lastly, we consider situations where physicians face no opportunity costs of treatment. As expected, in this situation almost all fast patients (93%) receive more care than necessary to reach the target at baseline. Similar to situations where physicians face opportunity costs of treatment, we find a reduction in services (−0.3 services, $p < 0.010$) under P4P for patients provided care above target at baseline, see Fig. 7. This finding contradicts our Hypothesis 5, which states that patients receive the maximum available amount of care at both baseline and under P4P.

Result 5. *In cases where physicians face no opportunity costs of care, patients receive less care under P4P.*

Our Result 4 and 5 show that regardless of the opportunity costs, patients provided more care than necessary to reach the target at baseline on average receive less care under P4P. Thus, physicians on average respond to P4P over and above the financial incentive.

5. Discussion and concluding remarks

Our study provides causal evidence of the distributional consequences of paying physicians for their performance when faced with different opportunity costs. We make use of a laboratory setting with medical students to test physicians' behaviour. There are several examples in the literature of laboratory experiments with medical students confirming findings from natural experiments with physicians (Galizzi and Wiesen, 2017; Lagarde and Blaauw, 2017). These studies, including ours, make no pretence to mimic the real world perfectly. For example, in our study, we only represent treatment costs, $c(e_i)$, as monetary costs (and not also time costs, which would be present in real life). Thus, our results should only be interpreted in qualitative terms, i.e. they capture physicians' underlying decisional mechanisms that transcend our specific framework (but not the absolute size of patient benefits, service provision etc.).

To analyse P4P's impact on physicians' behaviour, we compare physicians' treatment under P4P with their corresponding treatment at baseline, i.e. when there is no bonus for reaching the target but only a fixed payment. We find that the implications of introducing P4P differs across patient groups and the opportunity costs physicians are facing. In cases where physicians are resource constrained and thus face patient opportunity costs, they may be financially incentivised to reallocate care between patients. More specifically, we find that patients who otherwise receive less care than needed to reach the target, but with a potential to reach it, gain care under P4P. Whereas patients who do not have the potential to reach it may receive less care under P4P.

Our study also shows that patients that receive more care than needed to reach the target at baseline receive less care under P4P. Interestingly, this finding holds when the foregone care does not help other patients to reach the target, indicating that physicians' response is not driven by a financial incentive. One explanation could be that physicians interpret the target as a standard for high-quality care (Cromwell and Smith, 2011), in which case care above this target is not considered as valuable. As a result, patients who exceed the target may receive less care under P4P, independently of the existence of opportunity costs of the treatment.

Our findings have important policy implications. Currently, physicians in many health care systems operate under tight resource constraints (see for example Emanuel et al. (2020); George and Gerada (2019)). Under such constraints, P4P may lead to underserving of patients who are unable to reach the performance target. If policymakers introduce P4P, they should therefore consider risk-adjusting the performance target such that patients have more equal opportunities for reaching the target (Casalino et al., 2007). Policymakers should also be careful setting targets when physicians are resource abundant, because less health improving care may be provided to patients otherwise performing above target. Thus, the payers risk paying more for less care to these patients. To solve this issue, policymakers may also consider risk-adjusting the performance target such that the requirements are raised for these high-performing patients.

There are some limitations to our study as our theoretical framework and experimental setting does not capture all factors that may affect patients' care. Following the standard health economics literature on physicians' behaviour (Ellis and McGuire, 1986; McGuire, 2000), our theoretical framework assumes that physicians are profit maximisers. However, to our surprise we find that physicians on average become more generous under P4P towards patients who at baseline receive care at the target. This finding only holds in cases where physicians are resource abundant and face monetary opportunity costs. One explanation for this finding could be that physicians may not always act as profit maximisers, but instead seek a target income (McGuire and Pauly, 1991). Thus, the income level may also play a role for physicians' decision-making.

In both our theoretical framework and experimental setting, patients' health improves as they receive more care than needed to reach the performance target. However, one could also imagine that it may deteriorate patients' health to receive care above target (Brosig-Koch et al., 2019). As our focus is primarily on situations of resource scarcity (opportunity costs) and the baseline payment is fixed (salary or capitation), the case of overtreatment with a consequently deterioration of health is not of interest. In such overtreatment cases, patients otherwise receiving care above target may improve their health as P4P is introduced (Brosig-Koch et al., 2019).

Similar to other laboratory experiments on physicians' behaviour under different payment schemes (e.g. Brosig-Koch et al. (2016a); Brosig-Koch et al. (2013); Brosig-Koch et al. (2019); Brosig-Koch et al. (2017a); Brosig-Koch et al. (2017b); Di Guida et al. (2019); Hennig-Schmidt et al. (2011); Hennig-Schmidt and Wiesen (2014); Oxholm et al. (2019)), our findings are independent of physicians' preferences for leisure. Introducing a labour-leisure choice would likely reduce physicians' effort across all payment schemes. To take into account that physicians' provision of labour to treat patients may vary and is limited, we vary the availability of resources across workdays. Thus, we uncover the physicians' preferences for providing care to patients under different payment schemes for a given limit on labour.

We also assume that physicians are certain about the relationship between their effort level and performance. However, in a real world setting it may be difficult for physicians to predict the amount of effort required for a patient to reach a specific performance target. Oxholm et al. (2018) show that this uncertainty affects physicians' response to P4P. Physicians may thus choose to provide care above target to ensure the bonus. Despite the presence of uncertainty, we argue that patients with a potential to reach the target will still be prioritised under P4P at the detriment of patients without a potential.

P4P schemes are designed in several different ways (Eijkenaar, 2013). The type of incentivised indicators may differ both across and within P4P schemes. We focus on performance targets related to patients' health outcomes. Physicians are paid based on outcome indicators in many health care system, for example the English Quality and Outcomes Framework (National Institute for Health and Care Excellence, 2020). Targets may, however, also be related to processes, for example whether the patients have received a spirometry test or whether a blood sample has been drawn. In such cases, we expect patients' ability to reach the target to be more similar. However, some patients may still require more effort to treat to the target (or even to treat) compared to others, which makes our findings relevant also for these performance measures.

The number of incentivised indicators may also vary significantly between P4P schemes. We introduce a simple P4P design with only a single performance indicator capturing patients' entire health state. However, in reality performance indicators are related to a specific health domain (disease area) and do not capture all dimensions of the patients' health state or care. P4P schemes therefore typically include several performance indicators with targets that should be met. Our experimental design disregards issues related to physicians' multitasking (Eggleston, 2005; Mehrotra et al., 2010), which makes the distributional consequences of P4P even more complex.

The basis for reward may also be designed in different ways. The basis for the reward is both related to the payment function that determines the size of the bonus and the performance measure that triggers the reward. Similar to other P4P schemes, for example the Australian Practice Incentives Programme (Cashin et al., 2014), our scheme is based on a flat rate fee, i.e. a bonus is paid per patient that reaches the target. Other schemes may differ in terms of the payment function, for example by only triggering a bonus if a certain percentage of patients are provided care to the target. However, as the underlying patient performance targets are similar, we expect that our findings are qualitatively the same for these schemes.

The performance measure that triggers the reward may be based on an absolute performance level (based on one time point) or improved performance level (based on two time points) (Scott et al., 2018). In our experiment the initial health state (baseline) is similar across patients, which implies that rewarding for absolute health level or improved health level yield the same result. Thus, our findings suggest that prioritisation of patients based on their profitability may take place for both types of performance measures. This behaviour is especially worrisome in cases where performance is measured at the absolute health level, because patients with the lowest health level will likely receive less care even though they may be most in need of care. In cases where improved performance is rewarded, P4P may on the other hand lead to a more equitable distribution if care. Thus, one can imagine that it may be easier to improve health for patients with a lower initial health state than patients with a higher initial state.

Finally, our analysis of P4P is based on a comparison with situations where physicians only receive a fixed payment, such as capitation or salary. As fixed payment schemes do not incentivise physicians to provide care, P4P are commonly added on top to induce activity, see for example the English Quality and Outcomes Framework (Cashin et al., 2014). However, physicians may also receive other types of payments such as fee-for-service. P4P may be introduced in this setting to incentivise physicians to focus on quality instead of quantity of care, see for example the French Public Health Objectives ROSP (Cashin et al., 2014). In such cases, we expect P4P to incentivise physicians to reduce care to patients otherwise being overprovided (Brosig-Koch et al., 2019). However, many physicians are working under tight resource constraints, forcing them to trade-off care between their patients despite being paid for their activity (Di Guida et al., 2019). As these physicians are facing patient opportunity costs of providing care, we expect the same distributional consequences of introducing P4P for their care as identified in a setting with a fixed payment.

This study provides an important step in understanding the distributional consequences of paying physicians for their performance. We find that patients with a higher responsiveness to care are prioritised at the detriment of patients with a low responsiveness under P4P if physicians are resource constrained. An avenue for future research could be to investigate how physicians' allocative preferences affect their responsiveness to the introduction of P4P.

Credit author statement

Anne Sophie Oxholm: Writing – original draft preparation, Conceptualisation, Methodology, Formal analysis, Funding acquisition Sibilla Di Guida: Writing – original draft preparation, Conceptualisation, Methodology, Software, Funding acquisition Dorte Gyrd-Hansen: Writing – original draft preparation, Conceptualisation, Methodology, Funding acquisition

Ethical approval

This study was approved by the Research Ethics Committee of the University of Southern Denmark (case no. 19/66789).

Acknowledgments

Financial support from the Novo Nordisk Foundation (Grant number: NNF18OC0033978), the Danish Council for Independent Research (Grant number: DFF-6109-00132), and the FAIRCARE research programme at University of Southern Denmark is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2021.113939.

References

Ahlert, M., Schwettmann, L., 2017. Allocating health care resources: a questionnaire experiment on the predictive success of rules. Int. J. Equity Health 16, 112.
Alshamsan, R., Majeed, A., Ashworth, M., Car, J., Millett, C., 2010. Impact of pay for performance on inequalities in health care: systematic review. J. Health Serv. Res. Pol. 15, 178–184.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016a. *Physician Performance Pay: Evidence From a Laboratory Experiment*: Ruhr Economic Papers.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016b. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. J. Econ. Behav. Organ. 131, 17–23.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2013. How Effective Are Pay-For-Performance Incentives for Physicians?-A Laboratory Experiment. A Laboratory Experiment, pp. 1–34. April 2, 2013).
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., Wiesen, D., 2019. Physician performance pay: experimental evidence. Available at SSRN 3467583 1–81.
Brosig-Koch, J., Hehenkamp, B., Kokot, J., 2017a. The effects of competition on medical service provision. Health Econ. 26, 6–20.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2017b. The effects of introducing mixed payment systems for physicians: experimental evidence. Health Econ. 26, 243–262.
Byambadalai, U., Ma, C.-t.A., Wiesen, D., 2019. Changing Preferences: an Experiment and Estimation of Market-Incentive Effects on Altruism. Available at: SSRN 3431002.
Casalino, L.P., Elster, A., Eisenberg, A., Lewis, E., Montgomery, J., Ramos, D., 2007. Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities? These rapidly proliferating programs do not appear to be devoting much attention to the possible impact on disparities in health care. Health Aff. 26, w405–w414.
Cashin, C., Chi, Y.-L., Smith, P.C., Borowitz, M., Thomson, S., 2014. Paying for Performance in Health Care: Implications for Health System Performance and Accountability. McGraw-Hill Education, UK).
Cromwell, J., Smith, K.W., 2011. Evaluating pay for performance interventions. Pay for Performance in Health Care: Methods and Approaches 267.
Culyer, A.J., Wagstaff, A., 1993. Equity and equality in health and health care. J. Health Econ. 12, 431–457.
Di Guida, S., Gyrd-Hansen, D., Oxholm, A.S., 2019. Testing the myth of fee-for-service and overprovision in health care. Health Econ. 28, 717–722.
Eggleston, K., 2005. Multitasking and mixed systems for provider payment. J. Health Econ. 24, 211–223.

Eijkenaar, F., 2013. Key issues in the design of pay for performance programs. Eur. J. Health Econ. 14, 117–131.
Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O., 2013. Effects of pay for performance in health care: a systematic review of systematic reviews. Health Pol. 110, 115–130.
Ellis, R.P., McGuire, T.G., 1986. Provider behavior under prospective reimbursement: cost sharing and supply. J. Health Econ. 5, 129–151.
Emanuel, E.J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., et al., 2020. Fair allocation of scarce medical resources in the time of covid-19. N. Engl. J. Med. 382, 2049–2055.
Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Galizzi, M.M., Wiesen, D., 2017. Behavioural experiments in health: an introduction. Health Econ. 26, 3–5.
George, S., Gerada, C., 2019. Stressed GPs: a call for action. Br. J. Gen. Pract. 69, 116–117.
Giuffrida, A., Gosden, T., Forland, F., Kristiansen, I., Sergison, M., Leese, B., et al., 1999. Target payments in primary care: effects on professional practice and health care outcomes. Cochrane Database Syst. Rev. 1–10.
Godager, G., Hennig-Schmidt, H., Iversen, T., 2016. Does performance disclosure influence physicians' medical decisions? An experimental study. J. Econ. Behav. Organ. 131, 36–46.
Godager, G., Hennig-Schmidt, H., Li, J.J., Wang, J., Yang, F., 2021. Does Gender Affect Medical Decisions? Results from a Behavioral Experiment with Physicians and Medical Students. University of Oslo, Health Economics Research Programme.
Godager, G., Wiesen, D., 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. J. Health Econ. 32, 1105–1116.
Hennig-Schmidt, H., Selten, R., Wiesen, D., 2011. How payment systems affect physicians' provision behaviour—an experimental investigation. J. Health Econ. 30, 637–646.
Hennig-Schmidt, H., Wiesen, D., 2014. Other-regarding behavior and motivation in health care provision: an experiment with medical and non-medical students. Soc. Sci. Med. 108, 156–165.
Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: homo oeconomicus or homo hippocraticus? J. Publ. Econ. 131, 1–11.
Krasnik, A., Groenewegen, P.P., Pedersen, P.A., von Scholten, P., Mooney, G., Gottschau, A., et al., 1990. Changing remuneration systems: effects on activity in general practice. Br. Med. J. 300, 1698–1701.
Lagarde, M., Blaauw, D., 2017. Physicians' responses to financial and social incentives: a medically framed real effort experiment. Soc. Sci. Med. 179, 147–159.
Martinsson, P., Persson, E., 2019. Physician behavior and conditional altruism: the effects of payment system and uncertain health benefit. Theor. Decis. 87, 365–387.
McGuire, T.G., 2000. Physician agency. Handb. Health Econ. 1, 461–536.
McGuire, T.G., Pauly, M.V., 1991. Physician response to fee changes with multiple payers. J. Health Econ. 10, 385–410.
Mehrotra, A., Sorbero, M.E., Damberg, C.L., 2010. Using the lessons of behavioral economics to design more effective pay-for-performance programs. Am. J. Manag. Care 16, 497.
Milstein, R., Schreyoegg, J., 2016. Pay for performance in the inpatient sector: a review of 34 P4P programs in 14 OECD countries. Health Pol. 120, 1125–1140.
Ministry of Higher Education and Science, 2021. Ansøgere Og Optagne Fordelt På Køn, Alder Og Adgangsgrundlag Ansøgere Og Optagne Fordelt På Køn, Alder Og Adgangsgrundlag Søgning Og Optag På Videregående Uddannelser.
National Institute for Health and Care Excellence, 2020. NICE Quality and Outcomes Framework Indicator.
Oxholm, A.S., 2016. Physician response to target-based performance payment. In: COHERE Discussion Papers, vol. 9. University of Southern Denmark, pp. 1–21.
Oxholm, A.S., Di Guida, S., Gyrd-Hansen, D., Olsen, K.R., 2019. Taking care of high-need patients in capitation-based payment schemes–an experimental investigation into the importance of market conditions. Appl. Econ. 51, 5174–5184.
Oxholm, A.S., Kristensen, S.R., Sutton, M., 2018. Uncertainty about the effort–performance relationship in threshold-based payment schemes. J. Health Econ. 62, 69–83.
Scott, A., 2000. Economics of general practice. Handb. Health Econ. 1, 1175–1200.
Scott, A., Liu, M., Yong, J., 2018. Financial incentives to encourage value-based health care. Med. Care Res. Rev. 75, 3–32.
Tsuchiya, A., Dolan, P., 2009. Equality of what in health? Distinguishing between outcome egalitarianism and gain egalitarianism. Health Econ. 18, 147–159.
Wang, J., Iversen, T., Hennig-Schmidt, H., Godager, G., 2020. Are patient-regarding preferences stable? Evidence from a laboratory experiment with physicians and medical students from different countries. Eur. Econ. Rev. 125, 103411.