Mylène Lagarde and Duane Blaauw

Physicians' responses to financial and social incentives: a medically framed real effort experiment

Article (Published version)
Refereed

Original citation:
Lagarde, Mylène and Blaauw, Duane (2017) *Physicians' responses to financial and social incentives: a medically framed real effort experiment.* Social Science & Medicine, 179 . pp. 147-159. ISSN 0277-9536

DOI: 10.1016/j.socscimed.2017.03.002

Social Science & Medicine 179 (2017) 147–159

Contents lists available at ScienceDirect

Social Science & Medicine

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

Physicians' responses to financial and social incentives: A medically framed real effort experiment

Mylène Lagarde [a, *], Duane Blaauw [b]

[a] Department of Social Policy, London School of Economics, UK
[b] Centre for Health Policy, University of the Witwatersrand, South Africa

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form
19 February 2017
Accepted 1 March 2017
Available online 2 March 2017

Keywords:
South Africa
Real effort experiment
Fee-for-service
Capitation
Salary
Altruism

ABSTRACT

Because compensation policies have critical implications for the provision of health care, and evidence of their effects is limited and difficult to study in the real world, laboratory experiments may be a valuable methodology to study the behavioural responses of health care providers. With this experiment undertaken in 2013, we add to this new literature by designing a new medically framed real effort task to test the effects of different remuneration schemes in a multi-tasking context. We assess the impact of different incentives on the quantity (productivity) and quality of outputs of 132 participants. We also test whether the existence of benefits to patients influences effort. The results show that salary yields the lowest quantity of output, and fee-for-service the highest. By contrast, we find that the highest quality is achieved when participants are paid by salary, followed by capitation. We also find a lot of heterogeneity in behaviour, with intrinsically motivated individuals hardly sensitive to financial incentives. Finally, we find that when work quality benefits patients directly, subjects improve the quality of their output, while maintaining the same levels of productivity. This paper adds to a nascent literature by providing a new approach to studying remuneration schemes and modelling the medical decision making environment in the lab.

1. Introduction

When planning radical health care reforms, governments often focus on the compensation structure of providers (typically FFS, capitation, or salary), as it can have an impact on the efficiency of health care expenditure, as well as the quantity and quality of care delivered. While the effects of these compensation policies have been well described in theory, it has been challenging to study them empirically (Gosden et al., 1999; Scott et al., 2011). Obstacles have included the difficulty of obtaining data not biased by self-selection problems or the confounding effect of other contextual factors, and the challenge of obtaining accurate measures of provider performance. Because of these issues, several studies have recently used laboratory experiments to explore the behavioural responses of doctors under alternative remuneration mechanisms.

Most of these health experimental studies adopt the approach pioneered by Hennig-Schmidt et al. (2011) where participants face a number of decision situations, with outcomes depending on specific cost and benefit functions. Taking the role of physicians, participants choose to deliver a hypothetical quantity of services q to patients, which determines simultaneously their profit and patients' benefit. The experiment is incentivised in two ways. First, participants receive monetary gains. Second, real patients outside the lab derive benefits, since monetary proceedings from the experiment are used to fund care for patients.

While this experimental approach presents the advantage that effort is not distorted by personal variables such as ability and experience, such 'chosen effort' experiments poorly reproduce some aspects of real work where effort is not hypothetical and always negative (van Dijk, Sonnemans et al., 2001), but instead can sometimes yield utility. As such intrinsic motives can influence responses to financial incentives, some experimental economists prefer to use 'real effort' experiments where participants are paid for performing simple tasks, such as simple mathematical calculations (Dohmen and Falk, 2011; Niederle and Vesterlund, 2007), moving sliders on a screen (Gill and Prowse, 2012), or entering data (Greiner et al., 2011; Tonin and Vlassopoulos, 2015).

In the health experimental literature, only one study has used a

* Corresponding author. London School of Economics, Houghton Street, London, WC2A 2AE, UK.
E-mail address: M.Lagarde@lse.ac.uk (M. Lagarde).

http://dx.doi.org/10.1016/j.socscimed.2017.03.002

real effort experiment to test the impact of physicians' remuneration mechanisms (Green, 2014). While this study explores the effect of payment remunerations used in the health care industry, it uses a helping frame that has no relationship to health.

Our study contributes to this nascent literature in several ways. We designed a new medically framed real effort experiment, where two dimensions of performance are observed (quantity and quality). We assess the relative effects of the three traditional payment mechanisms for doctors (salary, capitation and FFS), and test how the presence of patients' benefits affects performance in the task. We find that productivity is highest under FFS, but that quality is maximised under salary. We also show that some individuals are intrinsically motivated to work well, and do not react to financial incentives, while social incentives improve subject performance. While quantity of output is lower under capitation than FFS in the absence of patients' benefits, this difference disappears with social incentives.

2. Related literature

2.1. Doctors' remuneration in the health economics literature

The potential effects of doctors' remuneration have been well described in the economic literature (McGuire, 2000). Under FFS, a throughput-based remuneration, if the FFS rate exceeds the marginal cost of delivering additional services, doctors will over-serve patients. Capitation systems provide an incentive to increase the numbers of patients served, but conditional on this, doctors have an incentive to reduce their effort and minimise the cost per patient. Because capitation systems with a uniform rate introduce the incentive for providers to select healthier (less costly) patients, risk-adjusted rates are often used to reflect the effort required by different types of patients (Newhouse, 1998). Finally, salary, a time-based remuneration scheme (the provider receives a set amount to work for a specified period of time), creates an incentive to exert little effort.

These conclusions derive largely from models of physician behaviour which consider performance as a one-dimensional output, ignoring the fact that doctors' output, like that of hospitals (Chalkley and Malcomson, 1998), is multi-dimensional. Even when restricting the focus to clinical care, doctors decide not only how many patients to see (quantity of effort), but also how much time to dedicate to each patient, whether to examine them thoroughly, etc (quality). Agency theory suggests that doctors are likely to neglect quality at the expense of quantity in such a multi-tasking context, because quality of care is much more difficult to observe (Holstrom and Milgrom, 1991), especially when their remuneration is linked to the quantity of services provided (e.g. FFS). By contrast, salary schemes, which provide low-powered incentives for quantity of output should have less negative consequences for quality of care.

2.2. Doctors' remuneration in the health experimental literature

The small but growing literature studying physician behaviour in the lab has mostly followed the design of Hennig-Schmidt et al. (2011)described above. Most studies have focused on two payment mechanisms, FFS and capitation, and on one unique outcome q interpreted as the services delivered to patients. Hennig-Schmidt et al. (2011) found that participants provided a quantity of services higher than optimal in FFS, but lower in capitation. They also found that with a uniform capitation rate, more costly patients end up with fewer services than healthier ones. Extensions of this experiment have sought to study the effects of blended payment mechanism (Brosig-Koch et al., 2017), and pay-for-performance schemes (Brosig-Koch et al., 2013; Keser et al., 2013). Two aspects have not been tested with this experimental set-up: the effect of capitation on the number of patients treated, and the impact of the multi-tasking environment faced by providers.

These two gaps were partly addressed in the experiment designed by Green (2014) which is more closely related to our approach. The study uses a real effort task where participants are asked to correct spelling mistakes on behalf of others, under five payments: salary, capitation, FFS, report cards with FFS, and capitation. The results suggest that the highest quantity of services is produced under FFS, and when subjects are paid by salary or capitation they reach the same productivity. The results also indicate no difference in quality (number of correct edits) under the three payment mechanisms, which the author interprets as evidence of intrinsic motivation.

There are several distinctions between our design and the one used by Green (2014). First, the task we employ (data entry) is less likely to depend on individuals' prior knowledge or abilities, allowing a sharper evaluation of the causal effect of incentives. Second, our experiment is closer to the health setting as it adopts a medical framing (subjects are asked to enter the blood test results of patients into a computer), is played with medical students, and social incentives are implemented benefits to real patients outside the lab. Third, we look at a broader range of outcomes (including number of patients treated and average services per patient). Fourth, we investigate the impact of risk-adjusted capitation rates.

2.3. Doctors' financial incentives and pro-social motivation

In economic models physicians are traditionally assumed to maximise income. This assumption leads to the conclusion that, due to the asymmetry of information between themselves and patients, doctors are likely to recommend unnecessary treatment under FFS (Evans, 1974). Yet, professional norms and health care providers' altruism are recognised sources of pro-social motivation, which have led more recent models to incorporate patients' benefits into doctors' utility function (Chone and Ma, 2011; Liu and Ma, 2013; Makris and Siciliani, 2013).

There is empirical evidence that doctors take into account patients' benefits, for example when they forego profits for higher quality of care (Kolstad, 2013) or when they accept posts in hardship areas where they can serve more patients (Lagarde and Blaauw, 2014). The experimental literature has studied further physicians' prosocial motives, showing that medical and nursing students' altruism is more powerful than other students' (Hennig-Schmidt and Wiesen, 2014; Jacobsen et al., 2011), even though some evidence emphasises substantial heterogeneity in their altruistic concerns (Godager and Wiesen, 2013). Kesternich et al. (2015) show that when professional values are made more salient (with the Hippocratic Oath) or when social incentives benefit actual patients (rather than students), medical students behave more altruistically. While these studies show the existence of physicians' prosocial motivation, they do not compare the performance of subjects with and without social incentives. In our study, we follow Tonin and Vlassopoulos (2015) who compare the performance of workers facing social incentives to those who do not, to isolate the impact of prosocial motivation on productivity.

3. Methods

3.1. A medically-framed real effort task

The experiment involved a real effort task framed in a medical context and constructed to reproduce the main characteristics of the medical decision-making environment. The experiment consisted of several periods of 8 min during which subjects had to enter data. We chose data entry as it was easy to relate it to a medical context and it permitted measurement of performance in terms of both quantity and quality of output.

In each period participants were handed a pile of 15 hardcopy laboratory test reports containing a series of blood test results: 5 short and 10 long reports (more details of the task including report content and computer screen snapshots can be found in the online appendix). While the 15 reports were different across periods, in a given period, all subjects were given the same 15 reports, in the same order.

Subjects had to type blood results into an input mask developed in the software z-Tree (Fischbacher, 2007). During the task, they would type the reference number of the report they wanted to enter, leading them to an input mask formatted on the specific laboratory report (i.e. short mask for a short report). At any point, subjects could decide to move on to another report even if entries were missing, but they could not go back to edit previous entries. Reports could be entered in any order but each report could be entered only once.

We intended this task to parallel the main characteristics of the medical decision-setting, with a report representing an interaction with a patient during which a doctor treats the patient's case by performing a number of tasks (ask questions, take the patient's blood pressure, examine the patient, etc.). Having two types of reports imitates the fact that some patients require less effort than others. Allowing the subjects to enter laboratory reports in any order allows them to select shorter ones first, as some doctors might be inclined to select healthier patients.

3.2. Design of the experiment

The experiment design includes a within-subject and a between-subject component.

For the within-subject dimension, subjects took part in three consecutive data entry periods of 8 min each, after doing a 3-min training session. Participants were paid differently for each period, the sequence of payments being randomised to control for possible ordering effects.

We first set the rate of the salary at R125 (ZAR 10 ≈ USD 0.965) to reflect the cost of employing a junior doctor for half an hour. We then conducted some piloting to estimate what would be an average quantity achieved (number of entries made over the 8-min period). First, we asked research assistants and colleagues to carry out the task, and on average they entered 139.4 numbers. Then we organised a pilot with 11 medical students (results and demographic characteristics can be found in Table A1 of the online appendix). Using pilot results as the expected performance, we set the capitation and FFS rates so that all three payments would be roughly equivalent:

- For the FFS scheme, each individual test result entered was paid R1, whether accurate or not;
- For capitation, a long report was paid R20 and a short report R15, independent of the performance (quantity or quality of data). This is similar to a risk-adjusted capitation system, with higher payments for patients requiring more effort.

In addition to this within-subject variation in payment mechanism, subjects were randomly assigned to one of two treatments to test whether their performance was affected by social incentives. In the first treatment (‘SELF’), the decisions made by participants create benefits only to themselves (the monetary rewards described above). In the second treatment (‘PATIENT’), subjects' performance could also translate into benefits for real patients. Subjects started the experiment by choosing their preferred charity from a list of five (Witwatersrand Hospice, SOS Children's villages, South African National Tuberculosis Association, Cancer Association of South Africa, Thusanani Children's Foundation). Then they were informed that for each individual test results they would enter correctly, a donation of R0.50 would be made to the charity of their choice and earmarked for the provision of health care services to patients. The rate of R0.50 was chosen based on the R1 rate for the FFS rate. For each correct entry made, this rate is equivalent to a 1/3–2/3 split, which is not dissimilar to the voluntary contributions made in Dictator Games to ‘deserving’ recipients (Engel, 2011; Kesternich et al., 2015). Credibility in economic experiments is key, and overall, 83% of subjects said they believed we would pay the charities (note that all results presented in the paper are robust to the exclusion of the 17% who had doubts).

3.3. Hypotheses

Because there is a higher incentive to enter individual results under FFS than under a flat payment rate (salary), we hypothesize that quantities will be larger under FFS [***Hypothesis 1, piece-rate vs fixed rate***], as observed in field (Lazear, 2000) and lab experiments (Greiner et al., 2011). Following health economics theory (McGuire, 2000), we also hypothesize that the quantity of services provided under FFS will be higher than under capitation [***Hypothesis 2, FFS vs capitation***].

In the capitation scheme, there is an incentive to process as many forms (patients) as possible, since payment is received for each one, but do little per form. Hence, we hypothesize that under capitation participants will aim to maximise the number of reports [***Hypothesis 3, patient enrolment***], while making minimal effort per form [***Hypothesis 4, low effort per patient***].

In our experiment we introduced risk-adjusted capitation rates, paying a higher rate for the more costly patient reports (seeking to equalise the rates for the two types of cases). As a result, we hypothesize that participants will not be incentivised to select easier (shorter) reports first [***Hypothesis 5, no cream-skimming***].

Because FFS links remuneration to quantity of output, we hypothesize that quality will be the lowest [***Hypothesis 6, multitasking with high-powered incentive***], followed by CAP where the link between remuneration and quantity of output is weaker than in FFS [***Hypothesis 7, multitasking with intermediate incentive***].

If we follow the assumption that prosocial motivation plays a role in the agency relationship between doctors and patients (Arrow, 1963), the marginal cost of making correct entries will be lower for subjects in the PATIENT treatment who derive utility from sending benefits to patients. Therefore, our final hypothesis [***Hypothesis 8, altruistic physician***] is that if physicians are altruistic, quality will be higher in the PATIENT treatment than in the SELF treatment.

3.4. Outcome measures

To test these hypotheses, five outcome measures are used.

Focusing on the quantity of output, we observe for each individual in a given period (1) the total number of entries where an entry is a number recorded for a particular test result (i.e. number of acts performed, following the medical analogy); (2) the number of laboratory forms done (total number of patients seen in a given period); and (3) the average number of entries made per form (average number of services provided to each patient). To evaluate the quality of output, we measure (4) the total number of correct entries, excluding unnecessary ones (number of services correctly performed), which is also directly related to patients' benefits under the PATIENT treatment. Finally, to study cream-skimming, we construct a categorical variable (5) distinguishing whether subjects entered the cases as given, chose to re-order reports starting with easier ones, or chose to treat the longer cases first.

3.5. Implementation and sample

Ethics approval was granted by the ethics committees at the London School of Hygiene and Tropical Medicine and the University of the Witwatersrand. The experiment was conducted in 2013 in a computer laboratory with medical students from the University of the Witwatersrand (South Africa). Subjects were recruited amongst 3rd and 4th year students, through adverts posted on their intranet and leaflets distributed in class. To limit selection bias, the adverts did not mention the monetary incentives, and talked about “short tasks” instead of “experiments”. Although a self-selection effect cannot be completely ruled out, the high response rate (65.6%) and the heterogeneity in performance in the task provide reassuring evidence suggesting that self-selection of more prosocial students was not an issue, as shown in other settings (Falk et al., 2013).

On the day of the experiment, participants were given a show-up fee of R50 and were randomly allocated to a workstation. Most instructions were presented on the computer screen (see online appendix). At the end of the session, one participant was invited to throw a die to determine which period would be chosen for payment. Payments to subjects were then calculated and made to each participant anonymously in a sealed envelope.

In total, we recruited 132 students, 66 in each of the two treatments. A session lasted approximately 60 min and on average participants earned R167 per session (in addition to the show-up fee).

Looking at the demographic characteristics of participants in the two treatment arms (see Table 1), no difference could be detected between the two groups. Overall, participants were young (21.5 year old), about 40% of them were male, 53% described themselves as black and 22% as white. 44% of participants were fourth year students and their average grade the year before was 66%. Finally, subjects were perfectly randomised to one of the six sequences of remuneration schemes over the three periods, allowing the average remuneration effect to be unconfounded by the ordering effect in the within-treatment dimension of the experiment.

4. Results

4.1. Effects of remuneration schemes on performance

4.1.1. Descriptive analysis

Fig. 1 provides an overview of the subjects’ performance in the SELF treatment (see Table A2 in online appendix).

Looking at the overall productivity, subjects entered nearly 200 test results under FFS, and 188 under capitation, and 148 under salary. Although individuals under FFS appear slightly more productive than under capitation, a Wilcoxon signed-rank (WSR) test fails to reject equality of the average number of entries made ($p = 0.447$). Using the same test, we find strong evidence that participants were the least productive under salary (WSR test $p < 0.001$ for salary vs. FFS and salary vs. capitation). Turning to the number of patients treated, the results look similar. We find no difference in terms of number of patients treated under FFS and capitation (WSR test p-value: 0.584), but the number of patients treated under salary is lower than under capitation and FFS (WSR test p-values <0.0001 in both tests). For the last measure of quantity of output (average output per patient), the same pattern emerges. On average participants entered 17.40 results per patient under salary, but 18.45 under FFS and 18.10 under capitation (WSR test p-values <0.0001 for salary vs. FFS and salary vs. capitation, but WSR test $p = 0.494$ between FFS and capitation).

Looking at the quality of participants’ output, the results appear quite different. Subjects achieved the highest performance under salary with 101 correct entries, followed by capitation with 78 entries (WSR test p-value: 0.024) and 65 entries under FFS (WSR test p-value: 0.005). We find also evidence that the quality of output seems higher under capitation than FFS (WSR test p-value: 0.005).

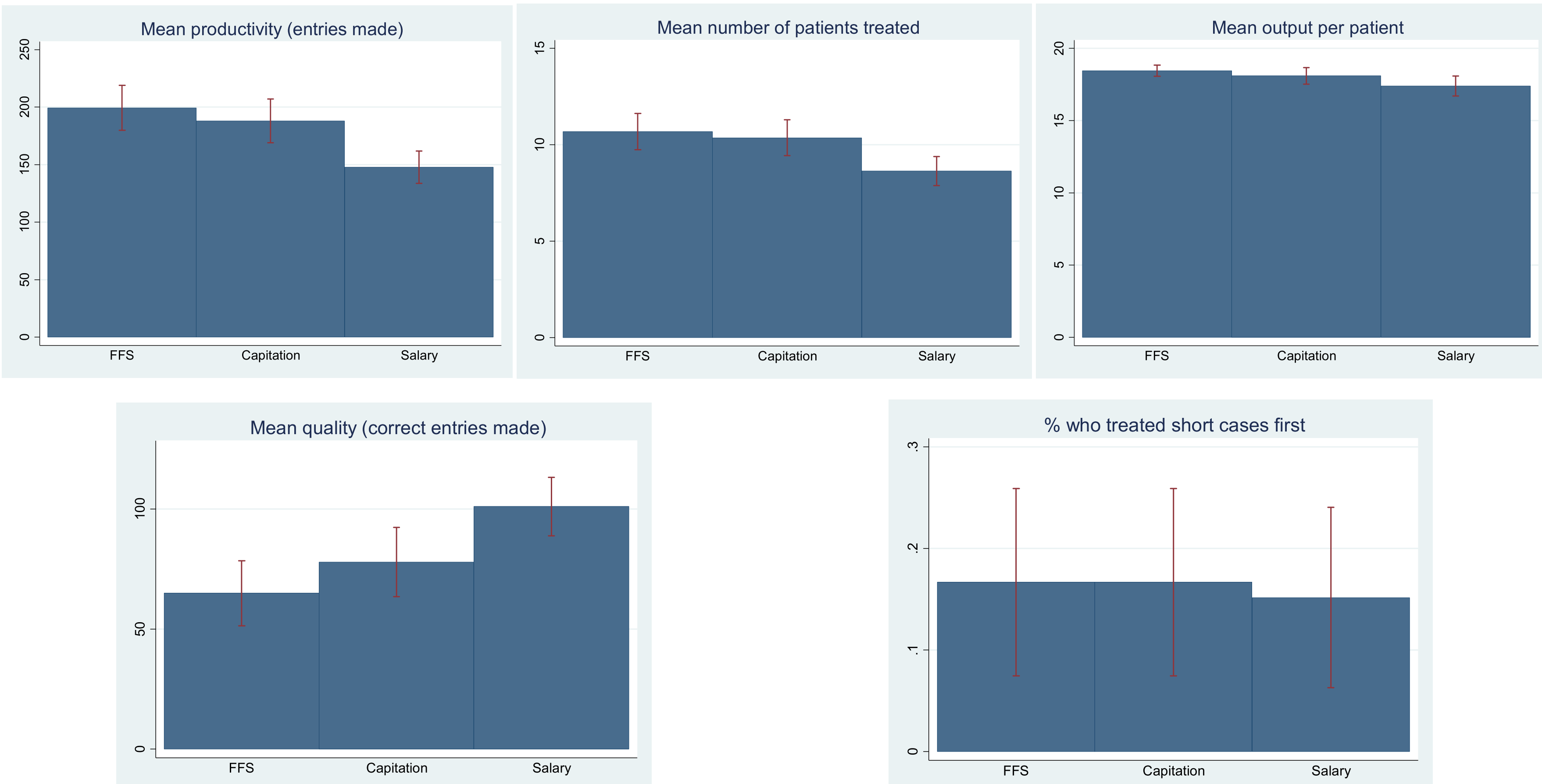
Finally, the last graph in Fig. 1 shows the proportion of respondents who chose to treat the shorter cases first, and there appears to be no difference across the three treatments (the proportions are 10.6%, 13.6% and 13.6% respectively for FFS, salary and capitation - all p-values of the MW tests are higher than the 10% threshold).

These unconditional means give an incomplete picture of the results since they do not control for individual characteristics or

Table 1
Characteristics of respondents.

	All (n = 132)	SELF (n = 66)	PATIENT (n = 66)	MW test
		Mean	Mean	
Participant demographic characteristics				
Age	21.705	21.636	21.773	0.759
Male	0.409	0.394	0.424	0.540
Black	0.530	0.515	0.545	0.546
White	0.219	0.197	0.242	0.275
Is in fourth year	0.439	0.455	0.424	0.544
Grade (%) obtained the previous year	66.83	67.015	66.652	0.428
Choice of charities (%)				
Witwatersrand Hospice	–	–	24.24	–
SOS Children's villages	–	–	15.15	–
National Tuberculosis Association	–	–	1.52	–
Cancer Association of South Africa	–	–	40.91	–
Thusanani Children's Foundation	–	–	18.18	–
Sequences of remuneration payments				
FFS – CAP – SAL	0.167	0.167	0.167	–
FFS – SAL – CAP	0.167	0.167	0.167	–
CAP – SAL – FFS	0.167	0.167	0.167	–
CAP – FFS – SAL	0.167	0.167	0.167	–
SAL – CAP – FFS	0.167	0.167	0.167	–
SAL – FFS – CAP	0.167	0.167	0.167	–

Note: This table contains the mean of participant characteristics. MW test column reports the p-value of a Mann-Whitney *U* test to compare the means of the two treatment arms where relevant.



Note: Each plot reports the mean performance for one outcome of interest (indicated in the plot title) across the three payment conditions for subjects in the SELF treatment.

Fig. 1. Average performance (standard error bars) under the three remuneration schemes, SELF treatment.

correlations over periods. To address these issues, we turn to the results of the regression analysis.

4.1.2. Regression analysis

Our regression analysis consists of estimating panel data random-effects model of the form:

$$y_{ip} = \alpha + \beta_1 SAL + \beta_2 CAP + \theta Z_i + \sum_{p=2}^{3} \eta_p PERIOD_p + \phi SEQU_i + u_i + \varepsilon_{ip} \quad (1)$$

where y_{ip} is the variable of interest for individual i in period p; SAL and CAP are dummies for the salary and capitation payment respectively, Z_i is a vector of individual characteristics, $PERIOD_p$ are period fixed effects that capture trends in productivity over the three periods, and $SEQU_i$ is a set of dummies included to control for the six different ordering sequences used; u_i is an individual-specific random element that captures within-subject correlation over the three rounds, and ε_{ip} an error term. Standard errors are clustered at the individual level in all specifications (alternative specifications clustering at the level of the session provided similar results).

Table 2 presents the results of the regressions run with the four outcome measures defined earlier. The results in column 1 allow us to detect the impact of remuneration schemes on the overall productivity of participants. The estimates show provide support to Hypothesis 1, as we find that, compared to FFS, subjects made nearly 52 fewer entries when they received a salary, corresponding to a reduction in productivity of nearly 25%. The results also support Hypothesis 2, since subjects are found to be less productive under capitation than FFS (about 11 fewer entries).

Turning to column 2, we consider the impact of remuneration mechanisms on patients treated. We find that our experiment provides no statistical support for Hypothesis 3, since there is no difference in the number of patients treated (reports done) under capitation and FFS ($p = 0.253$). In line with their lower overall productivity, we find that subjects paid by salary treated two fewer patients.

Column 3 presents the impact of remuneration schemes on average effort per patient. Our experiment also fails to support Hypothesis 4 on the effect of capitation, as we find that subjects provided the same effort per patient under FFS and capitation. However, we find that participants paid by salary made on average one less entry per patient than under FFS.

Column 4 analyses the impact of remuneration scheme on whether participants chose to deal with easier cases first. The results support Hypothesis 5 which proposed that under an effort-adjusted capitation scheme, cream-skimming would not occur in that we fail to detect a difference between capitation and the other two payment mechanisms.

In line with Hypotheses 6 and 7, the last column of Table 2 provides evidence that quality is higher under salary and capitation than FFS (respectively 55% and 17% more). The difference in the effects of salary and capitation is statistically significant (Wald test, $p < 0.001$), meaning that the highest quality is reached when subjects are paid by salary.

Interestingly, there seems to be a learning effect over the three periods with regard to the number of entries made. Looking at this effect for the different remuneration schemes (see Table A4 in online appendix), it seems marked for the FFS and capitation schemes, but not for salary. This could suggest that subjects learn how to earn more money under the different schemes.

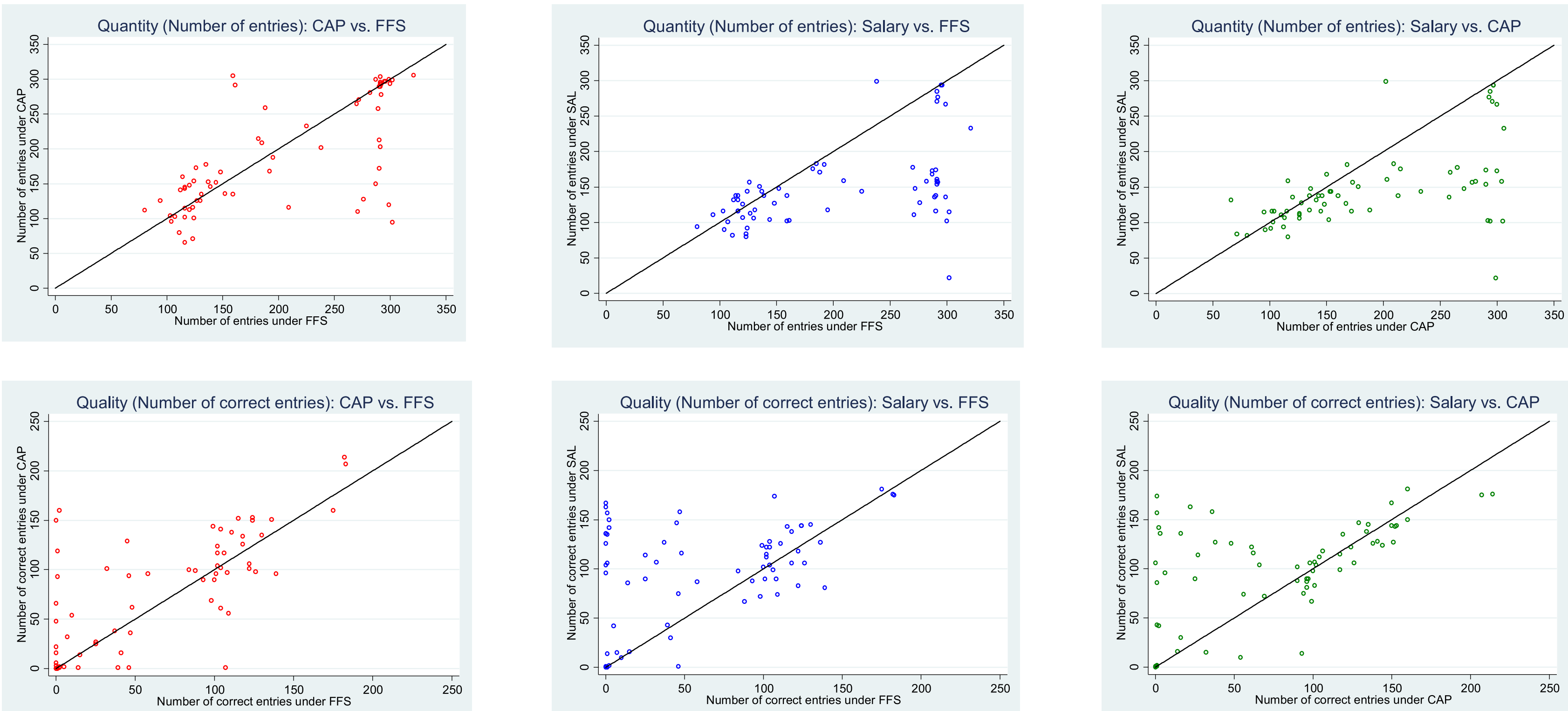
To investigate the heterogeneity in provider behaviour under the different remuneration schemes, we plot the performance of respondents in one scheme against their own performance under one of the other two payments. Fig. 2 presents the resulting graphs for the two main outcomes (total quantity and total quality). Several behavioural patterns emerge. Looking at quantity of output (upper row), there seems to be three groups of performers. First, the high performers (upper right hand corner), achieve high levels of output regardless of the payment. Second, one can identify 'average' performers whose output is clustered around the 45-degree line, indicating consistent behaviour under different schemes, but at lower levels of output than the high performers. Finally, a third group is made of individuals who respond to the payment used, achieving high levels of output under FFS, but much lower under salary and capitation.

We can regard participants' quality of output as a sign of intrinsic motivation since compensation is not linked to the number of correct entries. Two behaviours emerge (bottom graphs of Fig. 2): individuals can be categorised as highly-motivated subjects who reach high levels of quality under the different incentives (points on the right-hand side, clustered around the 45-degree line), or poorly-motivated individuals with lower levels of quality,

Table 2
Impact of remuneration schemes in the absence of patient benefit.

	Total output	Number of patients	Output per patient	Treated short cases first	Number of correct entries
	(1)	(2)	(3)	(4)	(5)
SAL	−51.636***	−2.045***	−1.059**	−0.218	36.152***
	(9.105)	(0.398)	(0.424)	(0.614)	(7.374)
CAP	−11.273*	−0.318	−0.358	−0.039	13.000**
	(6.800)	(0.285)	(0.313)	(0.569)	(5.488)
Period 2	22.455***	1.061***	0.206	−0.314	−5.667
	(7.557)	(0.330)	(0.344)	(0.627)	(6.067)
Period 3	22.182***	1.212***	0.107	−0.369	1.273
	(8.539)	(0.399)	(0.363)	(0.565)	(7.341)
Constant	227.475**	13.261**	17.150***	−7.235	−1.486
	(98.970)	(5.401)	(3.475)	(6.502)	(88.655)
Individual controls	Yes	Yes	Yes	Yes	Yes
Sequence controls	Yes	Yes	Yes	Yes	Yes
No of subjects	66	66	66	66	66
No of observations	198	198	198	198	198

Notes: This table reports the random-effects GLS regression of five different outcomes (indicated in the first row) on indicator variables for two of the payment experimental conditions (omitted category: FFS), and period dummies. The sample is made of all participants in the SELF treatment. Individual controls include age, gender, ethnicity, grade obtained the previous year and year of study. Sequence controls are a set of five dummy variables included to control for the effect of the six different ordering sequences that we used. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.



Note: Each figures show of all subjects in the SELF treatment under two payment conditions (CAP vs FFS for the first column, Salary vs. FFS for the 2nd column and Salary vs, CAP for the third column). Graphs in the top row use the main performance indicator for quantity (number of entries) while graphs on the bottom row use the main indicator of quality of performance (number of correct entries).

Fig. 2. Within-subject comparisons of performance under different remuneration schemes, SELF treatment.

particularly under FFS.

To investigate this further, we split the sample between high motivation (above median) and low motivation (below median) subjects, based on their performance (number of correct entries) under FFS, the payment that provides the highest incentive to divert effort *away* from quality. An interesting pattern emerges from this analysis: low motivation subjects appear responsive to financial incentives (Table 3), while high motivation subjects are not, and behave in the same way under the different payment schemes (Table 4). The results for the low-motivation group are qualitatively the same as the results for the whole sample, although the size of the effects is larger, and there is evidence that individuals make less effort per patient under capitation compared to FFS (Hypothesis 4).

4.2. Impact of social and financial incentives on performance

4.2.1. Descriptive analysis

In this section, we use data from the PATIENT treatment to test the impact of the three remuneration schemes while appealing to participants' prosocial motivation.

Fig. 3 presents the average performance of participants randomised to the PATIENT treatment (see numerical results in Table A3 in the online appendix), in which there was a social incentive, in the form of a donation to a charity looking after patients for each correct item entered.

Looking at differences across the three remuneration schemes in this treatment, the results appear similar to those obtained under the SELF treatment. The overall productivity, subjects entered nearly 180, 172 and 145 test results under FFS, capitation and salary respectively. The non-parametric test results provide no evidence of a difference between FFS and capitation (WSR test: $p = 0.352$), while we reject equality of performance between salary and FFS ($p = 0.001$) and salary and capitation ($p = 0.005$). The same pattern of results emerges for the number of patients treated: we find no evidence of difference between FFS and capitation (WSR test: $p = 0.784$), while we reject equality of performance between salary and FFS ($p = 0.001$) and salary and capitation ($p = 0.005$).

By contrast, we find that subjects made nearly 18 entries per patient under both salary and capitation (WSR test p-value: 0.786), while under FFS they worked a bit harder reaching nearly 18.5 entries (we reject equality of performance of salary vs. FFS: $p = 0.049$, and FFS vs. capitation: $p = 0.048$). Looking at the quality of participants' output, subjects achieved the highest performance under salary with 118 correct entries, followed by capitation with 105 entries (WSR test p-value: 0.004) and 94 entries under FFS (WSR test p-value: 0.001). We also find evidence that quality of output seems higher under capitation than FFS (WSR test p-value: 0.005). Finally, looking at the proportion of subjects who chose to process the shorter reports first, we find that this proportion was similar under the three payments.

Using Mann-Whitney tests to compare the performance across the two treatments, we find that the social incentives had an effect on the quality of output, with an increase in the number of correct entries made for the three remuneration schemes (the p-values are respectively 0.004 for FFS, 0.087 for salary and 0.007 for capitation). However, the tests fail to provide evidence of any difference in the other outcomes.

4.2.2. Regression analysis

To analyse the data more formally, we estimate two types of specifications. First, we run the same model as before using data from the PATIENT treatment. This allows us to test the hypotheses derived earlier in the presence of patients' benefits. Then, to estimate the effect of the social incentive, we pool data from the two treatments and estimate data random-effects models of the form:

$$y_{ip} = \alpha + \beta_1 SAL + \beta_2 CAP + \beta_3 PATIENT_i + \theta Z_i + \sum_{p=2}^{3} \eta_p PERIOD_p + \phi SEQU_i + u_i + \varepsilon_{ip} \quad (2)$$

where the only difference with specification (1) is the addition of a dummy variable $PATIENT_i$ indicating whether the subject was randomised to the PATIENT treatment. The coefficient β_3 can be interpreted as the effect of the social incentive, or the effect of the subject' prosocial motivation on their performance.

Table 5 presents the results for the five outcomes of interest. We first consider the first half of the results (columns 1–5) where we can test the Hypotheses laid out in section 3.3, in the presence of social incentives.

As in the SELF treatment, the results (column 1) support Hypothesis 1, as we find that under salary, subjects made nearly 34 fewer entries than under FFS. Unlike in the SELF treatment, we do not find evidence supporting Hypothesis 2, as subjects are found to

Table 3
Impact of remuneration schemes for individuals with low intrinsic motivation.

	Total output	Number of patients	Output per patient	Treated short cases first	Number of correct entries
	(1)	(2)	(3)	(4)	(5)
SAL	−96.176***	−3.693***	−2.032***	1.100	70.238***
	(14.272)	(0.657)	(0.752)	(1.023)	(11.616)
CAP	−23.491*	−0.585	−0.962*	0.104	23.312***
	(12.851)	(0.496)	(0.544)	(0.910)	(7.729)
Period 2	33.991**	1.604***	0.399	−0.935	−19.615*
	(13.351)	(0.553)	(0.571)	(0.852)	(10.165)
Period 3	28.815*	1.474**	0.466	−1.182	−18.471*
	(14.964)	(0.736)	(0.622)	(0.853)	(11.117)
Constant	278.168**	13.873*	22.998***	−48.832**	80.238
	(121.449)	(7.098)	(5.403)	(19.690)	(95.771)
Individual controls	Yes	Yes	Yes	Yes	Yes
Sequence controls	Yes	Yes	Yes	Yes	Yes
No of subjects	99	99	99	84	99
No of observations	33	33	33	28	33

Notes: This table reports the random-effects GLS regression of five different outcomes (indicated in the first row) on indicator variables for two of the payment experimental conditions (omitted category: FFS). The sample includes participants who entered a number of correct entries below the median under FFS. Individual controls include age, gender, ethnicity, grade obtained the previous year and year of study. Sequence controls are a set of five dummy variables included to control for the effect of the six different ordering sequences that we used. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 4
Impact of remuneration schemes for individuals with high intrinsic motivation.

	Total output	Number of patients	Output per patient	Treated short cases first	Number of correct entries
	(1)	(2)	(3)	(4)	(5)
SAL	−6.717	−0.378	−0.094	−1.677	2.354
	(4.149)	(0.266)	(0.375)	(1.049)	(4.667)
CAP	2.633	0.024	0.294	−0.359	−0.144
	(3.506)	(0.241)	(0.341)	(0.833)	(5.236)
Period 2	11.254***	0.562**	−0.042	−0.028	8.357
	(3.966)	(0.274)	(0.413)	(0.841)	(5.113)
Period 3	18.596***	1.096***	−0.250	−0.209	19.035***
	(4.564)	(0.273)	(0.312)	(0.951)	(6.951)
Constant	138.307	8.129	17.421***	3.164	54.973
	(123.574)	(7.122)	(4.712)	(10.187)	(84.505)
Individual controls	Yes	Yes	Yes	Yes	Yes
Sequence controls	Yes	Yes	Yes	Yes	Yes
No of subjects	99	99	99	84	99
No of observations	33	33	33	28	33

Notes: This table reports the random-effects GLS regression of five different outcomes (indicated in the first row) on indicator variables for two of the payment experimental conditions (omitted category: FFS). The sample includes participants who entered a number of correct entries above the median under FFS. Individual controls include age, gender, ethnicity, grade obtained the previous year and year of study. Sequence controls are a set of five dummy variables included to control for the effect of the six different ordering sequences that we used. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

be equally productive under capitation and FFS. Mirroring the findings of the SELF treatment, there is no statistical support for Hypothesis 3, as the number of patients (reports done) is similar in both FFS and capitation ($p = 0.776$), albeit higher than under salary where subjects treated 1.6 fewer patients. Data from the PATIENT treatment partly support Hypothesis 4 on the effect of capitation on average effort per patient: we find that subjects provided less effort per patient under capitation compared to FFS, but their effort was similar under capitation and salary. As before, we find strong evidence that an effort-adjusted capitation scheme prevents cream-skimming: we fail to detect a difference in the propensity to treat easier cases first between capitation and the other two payment mechanisms.

Finally, in line with Hypotheses 6 and 7, there is evidence that quality is lowest under FFS, although the difference with salary appears smaller in the PATIENT treatment compared to results presented in Table 2: subjects paid by salary (capitation) entered nearly 25 (11) more numbers than when paid by FFS, while this difference was 36 (13) in the SELF treatment. Quality is also lower under capitation than salary (Wald test $p = 0.0014$).

The second half of the table (columns 6–10), investigates the effect of introducing patient benefits on performance. The regression suggest that prosocial motivation increased the quality of output, by nearly 26 more correct entries (column 10) at the expense of a decrease in productivity of about 15 entries (column 6) and 0.9 patients treated (column 7). These results can be explained by the existence of a trade-off between quantity and quality of output: making a correct entry takes more time and effort than making an inaccurate one, so because individuals work harder to enter data accurately, they have less time to make inaccurate entries. The results also suggest that introducing the patient benefit had no effect on the average effort per patient (column 8) or the propensity to treat easier cases first (column 9).

Looking at the social incentive effect separately for the three payments (see Table A5 in online Appendix), we find that patients' benefits only reduce the quantity of output for capitation and FFS, the two payment schemes that link remuneration to quantity and produce the highest performance in the absence of a social incentive. We also find that the effect of the social incentive on quality of work was the same under FFS and capitation, but smaller under salary (probably because quality was already high under salary creating less room for improvement).

4.3. Impact of the choice of charity

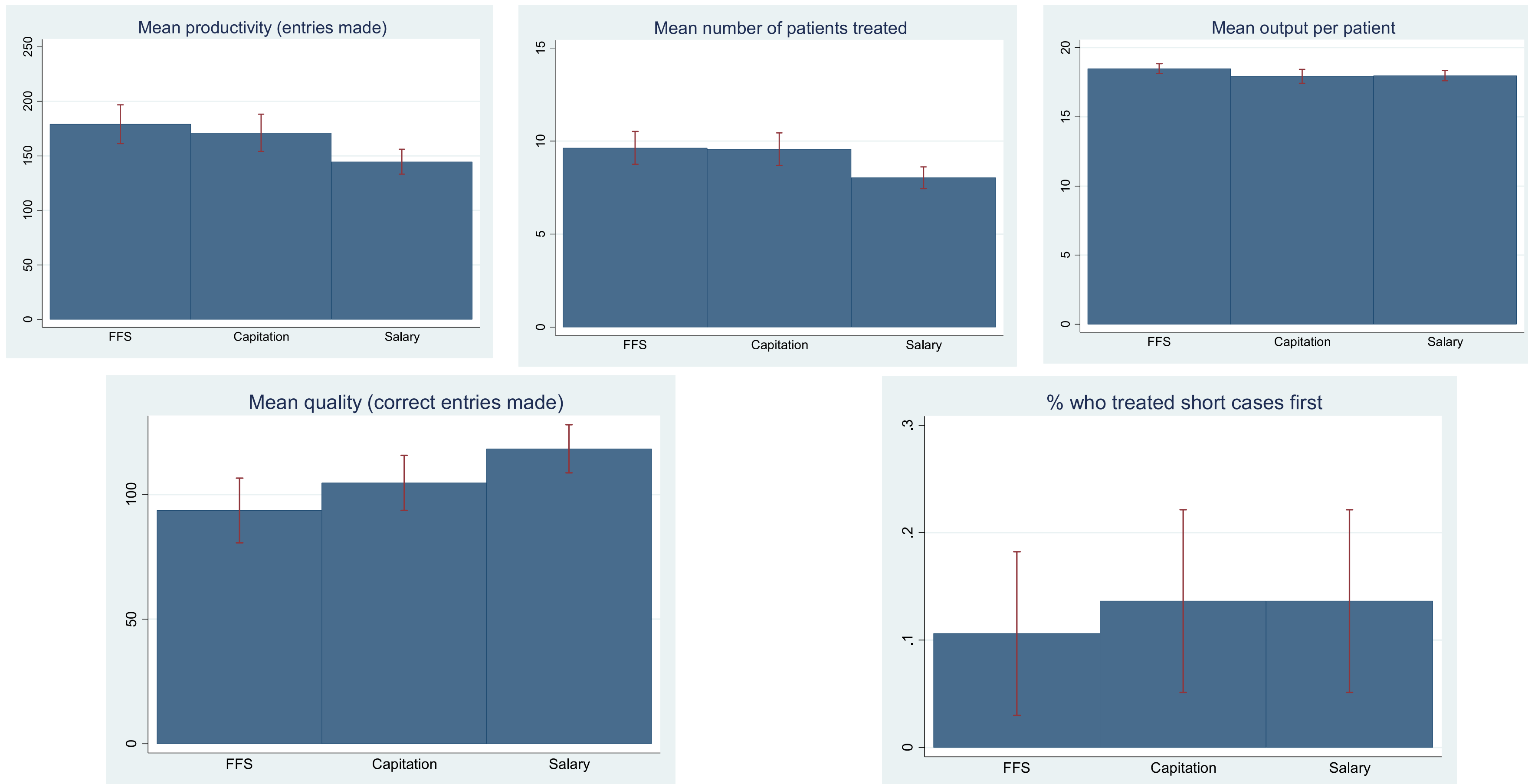
Since participants in the PATIENT treatment were able to choose which charity would receive patients' benefits, we explored the association between this choice and performance. Fig. 4 shows that the quality of output is roughly similar across all choices, while one charity stands out for its high productivity levels. These results are confirmed by regression analysis (see Table A6 in online appendix).

5. Discussion and conclusion

In this paper, we presented a novel medically-framed real effort experiment designed to reproduce the main aspects of clinical decision making, while allowing us to exogenously change the different incentives influencing subjects' decisions.

We found strong supporting evidence for five of the eight hypotheses derived from theory (see Section 3.3): with or without the presence of patients' benefits, productivity is higher under FFS than salary (Hyp 1); risk-adjusted capitation prevents cream-skimming (Hyp 5); quality of output is lowest under FFS due to its link to quantity of output (Hyp 6); quality is also higher with salary compared to capitation (Hyp 7). We also showed that social incentives, in the form of benefits received by patients linked to quality of work, improved the performance of subjects, confirming the altruistic physician's hypothesis (Hyp 8).

However, evidence on the relative productivity under FFS and capitation (Hyp 2) was mixed, with higher productivity under FFS in the SELF treatment but no difference in the PATIENT treatment. This result is at odds with other studies that have showed a lower productivity under capitation than FFS (Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz and Wiesen, 2015; Brosig-Koch et al., 2016; Hennig-Schmidt et al., 2011). Two reasons might explain our result. First, this could be linked to the combined effect of two different incentives. On the one hand, respondents sought to maximise their own income by treating many patients; on the other hand, for each patient treated, the subjects could be sensitive to the social incentive and not shirk their effort (this last point is supported by the fact that we did not find evidence that subjects minimised effort per case (Hyp 4)). These two effects combined would have resulted in high productivity. Another explanation relates to the complexity and lack of familiarity of participants with the capitation payment. Although initial instructions clearly indicated that subjects could always skip entries and move to the next



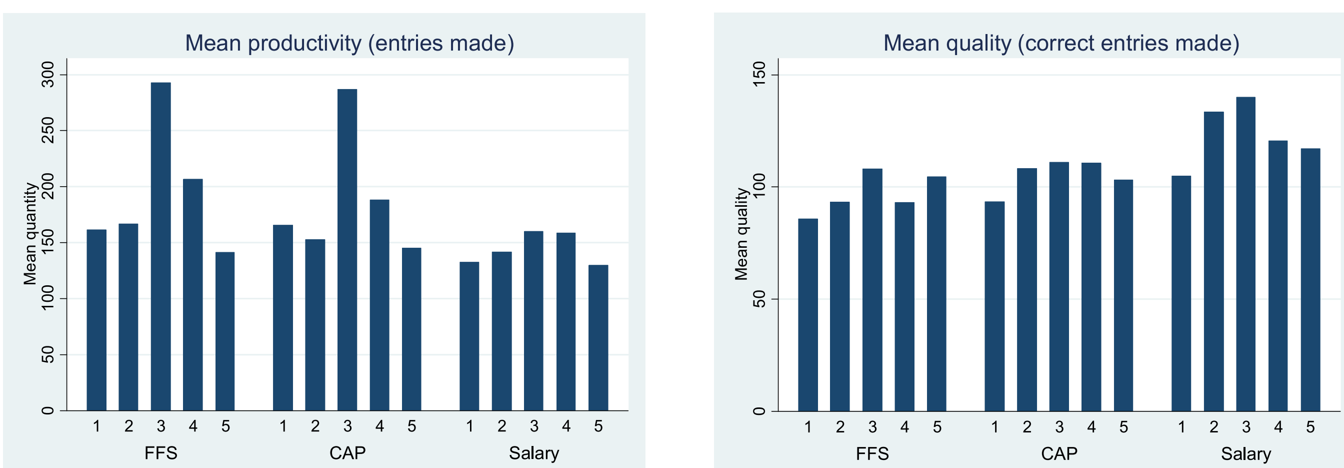
Note: Each plot reports the mean performance for one outcome of interest (indicated in the plot title) across the three payment conditions for subjects in the PATIENT treatment.

Fig. 3. Average performance (standard error bars) under the three remuneration schemes, PATIENT treatment.

Table 5
Impact of patient benefit and remuneration schemes.

	PATIENT treatment					Whole sample				
	Total output	Number of patients	Output per patient	Treated short cases first	Number of correct entries	Total output	Number of patients	Output per patient	Treated short cases first	Number of correct entries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SAL	−34.485***	−1.606***	−0.498**	0.547	24.712***	−43.061***	−1.826***	−0.779***	0.141	30.432***
	(7.591)	(0.379)	(0.224)	(0.737)	(5.210)	(5.865)	(0.270)	(0.235)	(0.452)	(4.454)
CAP	−8.030	−0.076	−0.551**	0.589	11.061***	−9.652**	−0.197	−0.455**	0.228	12.030***
	(5.961)	(0.266)	(0.267)	(0.716)	(3.637)	(4.389)	(0.189)	(0.202)	(0.452)	(3.244)
PATIENT						−14.898*	−0.855**	0.081	−0.374	25.702***
						(8.116)	(0.411)	(0.232)	(0.558)	(7.128)
Period 2	29.348***	1.530***	0.155	1.057	−2.697	25.902***	1.295***	0.181	0.195	−4.182
	(7.557)	(0.381)	(0.191)	(0.881)	(4.812)	(5.183)	(0.246)	(0.192)	(0.488)	(3.767)
Period 3	27.364***	1.561***	−0.058	1.551*	7.379*	24.773***	1.386***	0.024	0.407	4.326
	(6.496)	(0.269)	(0.295)	(0.861)	(4.359)	(5.268)	(0.236)	(0.230)	(0.469)	(4.209)
Constant	167.658**	9.106***	17.760***	−0.380	63.673	210.412***	11.632***	17.610***	−1.369	37.899
	(67.632)	(3.366)	(1.571)	(5.423)	(55.951)	(59.217)	(2.921)	(1.364)	(3.570)	(44.181)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sequence controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of subjects	66	66	66	66	66	132	132	132	132	132
Number of observations	198	198	198	198	198	396	396	396	396	396

Notes: This table reports the random-effects GLS regression of five different outcomes (indicated in the first row) on indicator variables for two of the payment experimental conditions (omitted category: FFS). Columns (1) to (5) report results for the PATIENT treatment only. Columns (6) to (10) report results for both SELF and PATIENT treatments and add a dummy variable capturing the effect of the patient benefit. Individual controls include age, gender, ethnicity, grade obtained the previous year and year of study. Period controls are a set of two dummies for period 2 and period 3. Sequence controls are a set of five dummy variables included to control for the effect of the six different ordering sequences that we used. Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.



Note: charity 1 is Witwatersrand Hospice ; charity 2 is SOS Children's villages ; charity 3 is the South African National Tuberculosis Association; charity 4 is Cancer Association of South Africa; Charity 5 is Thusanani Children's Foundation.

Fig. 4. Average productivity and quality across the five charities, PATIENT treatment.

case, this element was not repeated again and could have been overlooked by respondents. Finally, we did not find that under capitation subjects sought to maximise the number of cases (Hyp 3). This result is probably related to a characteristic of our design, whereby respondents could only ‘treat’ a maximum of 15 patients. This relatively low number of cases available limited the opportunity to differentiate the number of patients treated under FFS and capitation, since income-maximisers could treat all 15 cases under 8 min by quickly entering random data (we had not anticipated this aspect during the design phase, as pilot participants had not shown similar behaviours under FFS).

Our results also showed that financial incentives are strong for those who need to be motivated, but do not affect subjects who are intrinsically motivated to work well (achieving high quality output in the absence of any form of incentive). This finding echoes some of the results of Tonin and Vlassopoulos (2015), who found that subjects who were highly productive in the absence of incentives did not respond to financial incentives, in contrast to low-productivity subjects whose performance was improved by financial incentives. Although we cannot compare the effects of the intrinsic and social motivation directly, our results suggest that intrinsic motivation is more powerful than prosocial motivation, since in the PATIENT treatment the effect of social incentives did not completely muted the impact of financial incentives.

Generalising findings from the lab to the real world has long been a contentious debate (Levitt and List, 2007). Compared to the recent experimental literature, this paper takes lab experiments one step further towards the real world, by adding more realism and reproducing some key features of a real job (intrinsic motivation, boredom etc.). Yet, this contribution remains limited by the constraints and the abstraction of the lab. In the experiment, the existence of the patient benefits is not only made obvious, they are clearly quantified. In real life, the link between their actions and patients' benefits may be obvious, but it is hardly quantified or even indeed quantifiable. The longer time horizon of reality allows individuals to learn how incentives work over time, as they make the link between their effort and remuneration. The choice of the subject pool has also been shown to have important implications on behaviour, especially on prosocial motivation. Non-students have been found to be more motivated by social incentives than students (Falk et al., 2013), so it is possible that the impact of the social incentives might be higher for physicians. Brosig-Koch et al. (2016) suggest that medical students and physicians react in a similar way to financial incentives.

We presented a new experimental design that allowed us to compare the relative effects of the three traditional physician payment mechanism, and isolate the impact of pro-social motivation on performance. We were also able to replicate several important aspects of the health care setting which had not been studied before, such as the existence of risk-adjusted capitation rates, or the multi-tasking environment where both quantity and quality of output are important, but cannot necessarily be incentivised simultaneously in the same way. Our real effort task design is quite flexible, and can easily be adapted to accommodate new features (Lagarde and Blaauw, 2015a, 2015b). Although we purposefully adopted a medical framing partly to increase the external validity of the task, we did not study its impact. Such framing effect may exist if one believes that participants' motivation relates to certain professional norms that will be triggered by the framing (e.g. “serving patients”). Evidence on the medical framing effect of experiments is mixed. In an experiment on resource allocation with nursing students in Ethiopia, Barr et al. found that framing did not change average performance although it increased the variance in subjects' behaviours, suggesting that different people adhere to different norms (Barr et al., 2009). Kesternich et al. (2015) found that the combination of the medical framing and mention of the Hippocratic Oath significantly increased prosocial behaviour, and the authors suggest that the professional norms are mostly conveyed through the Oath. We believe that our experiment was less likely to have created a significant framing effect because the instructions created less personal identification or involvement from participants, compared to the public servant's game. There was also no priming of professional medical norms similar to that created by the Hippocratic Oath. However, this aspect could be investigated in future research.

Health economists remain more reluctant than other economists to the abstraction of laboratory experiments. While the external validity of economic experiments outside the laboratory is an issue, the results obtained in a controlled environment can be useful to think through the complex challenges posed by the complex interaction of actors (providers, patients, payers) and the incentives emerging from their relationships. It is hoped that this paper will add to this nascent literature by providing a new approach to studying remuneration schemes and modelling the medical decision making environment.

Acknowledgments

This manuscript is an output from a project funded by the UK Aid from the UK Department for International Development (DFID) for the benefit of developing countries. However, the views expressed and information contained in it are not necessarily those of or endorsed by DFID, which can accept no responsibility for such views or information or for any reliance placed on them.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.socscimed.2017.03.002.

References

Arrow, K., 1963. Uncertainty and the welfare economics of medical care. Am. Econ. Rev. 53 (5), 941–943.
Barr, A., Lindelow, M., Serneels, P., 2009. Corruption in public service delivery: an experimental analysis. J. Econ. Behav. Organ. 72 (1), 225–239.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2015. The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence. Health Economics. http://dx.doi.org/10.1002/hec.3292 n/a-n/a.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Wiesen, D., 2016. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. J. Econ. Behav. Organ. 131 (Part B), 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. Ruhr Economic Papers #413. http://repec.rwi-essen.de/files/REP_13_413.pdf. Ruhr-Universität Bochum (RUB), Department of Economics. Essen, Germany.
Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2017. The effects of introducing mixed payment systems for physicians: experimental evidence. Health Econ. 26 (2), 243–262.
Chalkley, M., Malcomson, J.M., 1998. Contracting for health services with unmonitored quality. Econ. J. 108 (449), 1093–1110. http://dx.doi.org/10.1111/1468-0297.00331.
Chone, P., Ma, C.T.A., 2011. Optimal health care contract under physician agency. Ann. Econ. Statistics (101/102), 229–256. http://dx.doi.org/10.2307/41615481.
van Dijk, F., Sonnemans, J., van Winden, F., 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45, 187–214.
Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: productivity, preferences, and gender. Am. Econ. Rev. 101 (2), 556–590.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14, 583–610.
Evans, R., 1974. Supplier Induced Demand: some empirical evidence and implications. In: Perlman, E. (Ed.), The Economics of Health and Medical Care. John Wiley and Sons, New York, pp. 162–173.
Falk, A., Meier, S., Zhender, C., 2013. Do lab experimentsmisrepresent social preferences? The case of self-selected student samples. J. Eur. Econ. Assoc. 11 (4), 839–852.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. Am. Econ. Rev. 102, 469–503.
Godager, G., Wiesen, D., 2013. Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. J. Health Econ. 32 (6), 1105–1116. http://dx.doi.org/10.1016/j.jhealeco.2013.08.008.
Gosden, T., Pedersen, L., Torgerso, D., 1999. How should we pay doctors? A systematic review of salary payments and their effect on doctor behaviour. Q. J. Med. (92), 47–55.
Green, E.P., 2014. Payment systems in the healthcare industry: an experimental study of physician incentives. J. Econ. Behav. Organ. 106 (0), 367–378. http://dx.doi.org/10.1016/j.jebo.2014.05.009.
Greiner, B., Ockenfels, A., Werner, P., 2011. Wage transparency and performance: a real-effort experiment. Econ. Lett. 111 (3), 236–238. http://dx.doi.org/10.1016/j.econlet.2011.02.015.
Hennig-Schmidt, H., Selten, R., Wiesen, D., 2011. How payment systems affect physicians' provision behaviour—an experimental investigation. J. Health Econ. 30 (4), 637–646. http://dx.doi.org/10.1016/j.jhealeco.2011.05.001.
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 (0), 156–165. http://dx.doi.org/10.1016/j.socscimed.2014.03.001.
Holstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: incentives contracts, asset ownership and job design. J. Law, Econ. Organ. 7, 24–52.
Jacobsen, K., Eika, K., Helland, L., Lind, J., Nyborg, K., 2011. Are nurses more altruistic than real estate brokers? J. Econ. Psychol. 32, 818–831.
Keser, C., Peterle, E., Schnitzler, C., 2013. Money Talks: Paying Physicians for Performance: Discussion Papers. Center for European Governance and Economic Development Research. No. 173.
Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? J. Public Econ. 131, 1–11. http://dx.doi.org/10.1016/j.jpubeco.2015.08.009.
Kolstad, J.T., 2013. Information and quality when motivation is intrinsic: evidence from surgeon report cards. Am. Econ. Rev. 103 (7), 2875–2910. http://dx.doi.org/10.1257/aer.103.7.2875.

Lagarde, M., Blaauw, D., 2014. Pro-social preferences and self-selection into jobs: evidence from South African nurses. J. Econ. Behav. Organ. 107 (Part A), 136–152. http://dx.doi.org/10.1016/j.jebo.2014.09.004.

Lagarde, M., Blaauw, D., 2015a. Carrot, Stick or Competition? Experimental Evidence of Performance Contracts in a Multitasking Setting *RESYST Working Paper*. London School of Hygiene and Tropical Medicine, mimeo, London.

Lagarde, M., Blaauw, D., 2015b. Financial incentives, social pressure and pro-social motives in health: evidence from the lab *RESYST Working paper*. London School of Hygiene and Tropical Medicine, mimeo, London.

Lazear, E.P., 2000. Performance pay and productivity. Am. Econ. Rev. 90 (5), 1346–1361. http://dx.doi.org/10.1257/aer.90.5.1346.

Levitt, S., List, J., 2007. Viewpoint: on the generalizability of lab behaviour to the field. Can. J. Econ. 40, 347–370.

Liu, T., Ma, C.-t. A., 2013. Health insurance, treatment plan, and delegation to altruistic physician. J. Econ. Behav. Organ. 85, 79–96. http://dx.doi.org/10.1016/j.jebo.2012.11.002.

Makris, M., Siciliani, L., 2013. Optimal Incentive Schemes for Altruistic providers. J. Public Econ. Theory 15 (5), 675–699. http://dx.doi.org/10.1111/jpet.12047.

McGuire, T.G., 2000. Physician agency. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. North-Holland Elsevier, Amsterdam.

Newhouse, J.P., 1998. Risk adjustment: where are we now? Inquiry 35 (2), 122–131.

Niederle, M., Vesterlund, L., 2007. Do women shy away from Competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101. http://dx.doi.org/10.1162/qjec.122.3.1067.

Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J., Young, D., 2011. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database Syst. Rev. (9), CD008451 doi: 008410.001002/14651858.CD14008451.pub14651852.

Tonin, M., Vlassopoulos, M., 2015. Corporate philanthropy and productivity: evidence from an online real effort experiment. Manag. Sci. 61 (8), 1795–1811. http://dx.doi.org/10.1287/mnsc.2014.1985.