# Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision

*Jeannette Brosig-Koch*
University of Duisburg-Essen and
Health Economics Research Center

*Heike Hennig-Schmidt*
University of Bonn and
University of Oslo

*Nadja Kairies-Schwarz,*
University of Duisburg-Essen and
Health Economics Research Center

*Daniel Wiesen*
University of Cologne and
University of Oslo

HERO

# Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision

by

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

**Health Economics Research Network at the University of Oslo
HERO September 2014**

# Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision[*]

Jeannette Brosig-Koch[†], Heike Hennig-Schmidt[‡], Nadja Kairies-Schwarz[§], Daniel Wiesen[¶]

September 30, 2014

## Abstract

We analyze how physicians, medical students, and non-medical students respond to financial incentives from fee-for-service and capitation. We employ a series of artefactual field and conventional lab experiments framed in a physician decision-making context. Physicians, participating in the field, and medical and non-medical students, participating in lab experiments, respond to the incentives in a consistent way: Significantly more medical services are provided under fee-for-service compared to capitation. Our findings are robust regarding subjects' gender, age, and personality traits.

**Keywords:** artefactual field experiment, laboratory experiment, fee-for-service, capitation, physician behavior

**JEL-Classification:** C91, I11

[†]Faculty of Economics and Business Administration, University of Duisburg-Essen and CINCH—Health Economics Research Center, Germany

[‡]bonneconlab—Laboratory for Experimental Economics, Department of Economics, University of Bonn, Germany, and Department of Health Management and Health Economics, University of Oslo, Norway

[§]Chair for Experimental Economics, Faculty of Economics and Business Administration, University of Duisburg-Essen and CINCH—Health Economics Research Center, Germany

[¶]Seminar of Personnel Economics and Human Resource Management, Faculty of Management, Economics and Social Sciences, University of Cologne, Germany, and Department of Health Management and Health Economics, University of Oslo, Norway; e-mail: daniel.wiesen@uni-koeln.de

# 1 Introduction

Laboratory experiments have, only recently, entered the field of health economics.[1] A stream of experimental research in health economics focuses on payment systems for physicians[2], which is important in light of increasing health care expenditures (e.g., Baicker and Goldman, 2011). For example, Hennig-Schmidt et al. (2011) find that fee-for-service (FFS) incentivizes too many and capitation (CAP) too few medical services. Brosig-Koch et al. (2014) show that mixed payment systems mitigate the incentive effects from FFS and CAP as predicted by Ellis and McGuire's (1986) seminal model. Lagarde and Blaauw (2014a), in addition to CAP and FFS, introduce a salary payment. Pay-for-performance systems are analyzed by Brosig-Koch et al. (2013), Keser and Schnitzler (2013), Lagarde and Blaauw (2014b), and Green (forthcoming).

All of these experiments are conventional lab experiments according to the taxonomy of Harrison and List (2004), in that they either use a standard student subject pool or medical students, or both subject pools.[3] Harrison and List (2004) and Levitt and List (2009) convincingly make the point that lab experiments in isolation have limited relevance for predicting field behavior but if combined with field data they permit more compelling inferences. We find this point to be particularly important for a young but rapidly developing research field like experimental health economics.

In this paper, we investigate whether findings from lab experiments with student subject pools also hold for artefactual field experiments with physicians and thereby addressing the concern outlined above. We compare how different subject pools (i.e., real physicians, medical students, and non-medical students) respond to two common physician payment systems, FFS or CAP (see, e.g., McGuire, 2000). We gradually approach the field by starting with a conventional lab experiment using a standard subject pool of non-medical students. We then proceed by having prospective physicians, i.e., medical students, participate in our experiment. Finally, we bring the lab to the field by introducing an artefactual field experiment to analyze how real physicians respond to incentives inherent in FFS and CAP. We completely mimic the lab setting in the field.

Even though artefactual field experiments are commonplace in areas like public economics, environmental economics, finance, industrial organization, and game theory (see,

---

[1] The only early exception that we are aware of is the experiment by Fan et al. (1998), who analyze physician payment systems under a global budget.

[2] Other experiments investigate, for example, health insurance choice (Schram and Sonnemans, 2011), health care financing (e.g., Buckley et al., 2012), or the salience of the Hippocratic Oath (Kesternich et al., 2014).

[3] Except for Green (forthcoming), the experiments are framed in a medical context insofar as participants decide in the role of physicians. The rationale is to avoid the experimenter's lack of "control for the context that subjects might themselves impose on the abstract experimental task" (Harrison and List, 2004, p. 1028).

e.g., Levitt and List, 2009; List, 2011; Voors et al., 2011), in health economics we are not aware of any such experiment with real physicians. Our paper therefore marks a first step in providing external validity of findings from the lab.

In both the lab and the field, we use a framed setting in which participants decide in the role of physicians on the provision of medical services. A subject's quantity choice determines his or her own profit and a patient's health benefit. Decisions are incentivized by monetary rewards determined by the respective payment method. Real patients' health outside the lab is affected by these decisions. We randomly assign participants to the two payment schemes, thereby excluding selection biases.

Behavioral data show that all subject pools respond to financial incentives in a similar and consistent way. In particular, significantly more medical services are provided under FFS compared to CAP. This result is robust with regard to subjects' characteristics such as age or gender and personality traits, as measured by a ten-item personality inventory (Gosling et al., 2003).

The remainder of the paper is organized as follows. In Section 2, we describe the experimental design and procedure. Section 3 presents the results, and Section 4 concludes.

# 2 Experimental design and procedure

## 2.1 Basic setup and decision situation

Overall, 137 non-medical students (FFS: 68, CAP: 69) and 76 medical students (FFS: 38, CAP: 38) took part in our lab experiments. In the artefactual field experiments, 29 physicians participated (FFS: 13, CAP: 16). Each subject was randomly assigned to only one of the two payment conditions. Our $3 \times 2$ design allows us to compare behavior of the three subject groups and between payment conditions.[4]

In FFS, subjects receive a fee of $p = 2$ Taler—our experimental currency—for each service provided to a patient. In CAP, subjects are paid a lump-sum of $L = 10$ Taler irrespective of the quantity of medical services provided. The conversion rate is 1 Taler = EUR 0.8 in the lab experiment and 1 Taler = EUR 3.4 in the artefactual field experiment. Compared to the lab, the payment in the field experiment was increased by a factor of 4.25 to provide adequate incentives for the physicians to take the task seriously.[5]

The decision situation is the same in both payment conditions. Participants decide

---

[4]Notice that the general experimental design follows Brosig-Koch et al. (2014).

[5]The amount physicians could earn in the experiment was set such that it reflects the average net hourly wage of a physician in Germany, bearing in mind potential differences, for example, across physicians' specialization and seniority. We set this factor after a consultation with Dr. Harald Herholz of the Association of Statutory Health Insurance Physicians in Hesse (Germany), who has been involved in budget negotiations for physicians' remuneration.

on the quantity of medical services $q \in [0, 10]$ for nine different patients ($j = 1, \ldots, 9$). Patients differ in illnesses and in severities of illness. They are assumed to be passive and fully insured, accepting each level of medical service. In FFS and in CAP, physicians are confronted with the same patients. Thus, the patient population for which a physician provides services remains constant under both payment systems.

When deciding on $q$, subject $i$ simultaneously determines her own profit $\pi(q)$ and the health benefit $B(q)$ of patient $j$. A subject's profit is $\pi(q) = \mu L + (1 - \mu)pq - c(q)$ with $c(q) = q^2/10$, $\mu = 0$ in FFS and $\mu = 1$ in CAP. The payment systems are designed such that the maximum and marginal profits are the same in both payment conditions; see Figure 1. Common to the health benefits of all nine patients is a global optimum at $q^*$ on $q \in (0, 10)$. The health benefit is varied systematically for illness $k$ and severity of illness $l$.[6] Figure 2 illustrates the patient health benefits for the nine patients. The full set of parameters is shown in Table 1 in Appendix A.1.

All experimental parameters are common knowledge. In particular, when choosing the quantity of medical services, subjects are informed about their payment, costs, profit, and the patient's health benefit for each quantity that can be chosen. Details on the decision situation are given in the instructions in Appendix A.2.

While all participants in the experiment decide in the role of physicians on service provision for abstract patients, real patients' health outside the lab is affected by their choices. Subjects are informed that the monetary equivalent of the patient health benefit resulting from their decisions is transferred to a charity caring for ophthalmic patients; for procedural details see Subsection 2.2.

## 2.2 Experimental protocol

Both the artefactual field as well as the lab experiment are computerized, programmed with z-Tree (Fischbacher, 2007). Physicians and students were confronted with identical computer screens, instructions, and control questions. The only difference was the higher conversion factor from Taler to Euro for physicians compared to students.

The artefactual field experiment was conducted in 2012 and 2013 using the mobile lab of *elfe*, Essen Laboratory for Experimental Economics, at the Academy for Training and Education of Physicians (Akademie für Ärztliche Fort- und Weiterbildung) in Bad Nauheim, Germany. Here, German physicians take mandatory, annual advanced education and training courses. The physicians were recruited by announcements in their courses. They

---

[6]In particular, for illnesses $A$ and $B$ the absolute marginal health benefit is 1 and for illness $C$ it is 2. For illnesses $A, B, C$ the maximum health benefit is $B_{Al}(q^*) = 7$, $B_{Bl}(q^*) = 10$, and $B_{Cl}(q^*) = 14$, respectively. The patient-optimal quantity $q^*$ varies with severities of illness $l$. For low ($x$), intermediate ($y$), and severe ($z$) severities, the patient-optimal quantities are $q^* = 3$, $q^* = 5$, and $q^* = 7$, respectively.

voluntarily participated before or after their courses. The lab experiment was conducted at *elfe*, where participants were recruited via the online recruiting system ORSEE (Greiner 2004). Our mobile *elfe* mimics the lab in Essen, which is a typical computer laboratory; see Figure 1.

The experimental procedure was the same in the lab and in the field, except for the

Figure 1: The mobile lab and the computer laboratory in Essen



*Notes:* The left panel of this figure shows the mobile lab mimicking the lab setup in the Essen Laboratory for Experimental Economics at the University of Duisburg-Essen, which can be seen in the right panel.

minor deviations described in footnote 7. Upon arrival[7], participants were randomly assigned to workstations separated by panels to ensure that decisions could be made in full anonymity. Participants were then given ample time to read the instructions and ask clarifying questions, which were answered in private. To check for participants' understanding of the decision task, they had to answer a set of control questions. The experiment did not start unless all subjects had answered the control questions correctly. Participants decided on the quantity of medical services for each of the nine patients. The order of patients was randomly determined and kept constant for all participants and both payment conditions. The experiment actually consisted of two parts. Subjects only knew that there would be a second part but were left ignorant about of the content of this part. In this paper, however, we focus on behavior in the first part of the experiment where subjects are paid according to either FFS or CAP.

Before making their decision for a specific patient, participants were informed about their remuneration, their cost and profit, as well as about the patient's benefit for each quantity of care, from 0 to 10. At the end of the experiment—after all subjects had made their decisions—we randomly determined one decision in each part of the experiment to be relevant for a participant's actual payoff and the related patient benefit. This procedure was chosen to rule out income effects.

---

[7]Before the experiments, physicians were briefly introduced to the experimental economics method, the universities involved in running the experiment, and the funding institution (DFG, German Research Foundation) of our research project. After the experiment, physicians were debriefed and informed about results of previous related health economic experiments.

Before paying participants in private and according to the randomly determined decisions, we asked them to complete a questionnaire on social demographics such as age and gender. In addition, we elicited personality traits by a ten-item personality inventory (TIPI; Gosling et al., 2003), which includes a subscale for each of the five personality dimensions extraversion, agreeableness, conscientiousness, neuroticism, and openness. We administered the well-validated German version of the TIPI following Rammstedt and John (2007).[8]

To verify that the monetary amount corresponding to the sum of the patient benefits in a session was actually transferred, we applied a procedure similar to Hennig-Schmidt et al. (2011) and Eckel and Grossman (1996). To this end, one of the participants was randomly chosen to be a monitor. After the experiment, the monitor verified that an order for the amount of the aggregate benefit was written to the finance department of the University of Duisburg-Essen to transfer the money to the Christoffel Blindenmission. The money supports surgical treatments for cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists from the charity. Notice that we did not inform the participants that the money was assigned to a developing country (see instructions in Appendix A.2). The order was sealed in an envelope, and the monitor and experimenter then walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional EUR 5.

The first part of the experiment, which we analyze in this paper, lasted for about 30 minutes. Students earned, on average, EUR 6.95, and physicians EUR 26.90. The average benefit per patient was EUR 6.18 and EUR 32.36 for students and physicians, respectively. In total, EUR 2198.08 were transferred to the Christoffel Blindenmission. Average costs for a cataract operation amount to about EUR 30. Thus, our experiment provided money for the treatment of 75 patients.

# 3    Results

On the aggregate, 119 participants in FFS provide, on average, 6.91 (s.d. 2.07, $N = 1071$) medical services. In CAP, 123 participants choose an average quantity of 3.37 (s.d. 2.04, $N = 1107$) medical services.[9] Participants, thus, do respond to incentives inherent in FFS and CAP, which is in line with the theoretical health economics literature (e.g., Ellis and McGuire, 1986) and findings from earlier empirical studies (e.g., Gaynor and Pauly, 1990;

---

[8]Notice that previous studies showed that personality dimensions are related to, for example, preferences regarding incentives (e.g., Bartling et al., 2009), behavior in dictator games (e.g., Ben-Ner et al., 2004), and in public good games (e.g., Volk et al., 2012).

[9]Notice that the data for medical and non-medicial students is taken from the first part of the experiments of Brosig-Koch et al. (2013) and Brosig-Koch et al. (2014).

Hennig-Schmidt et al., 2011).

Figure 2: Average quantities of medical services by payment systems and subject pools
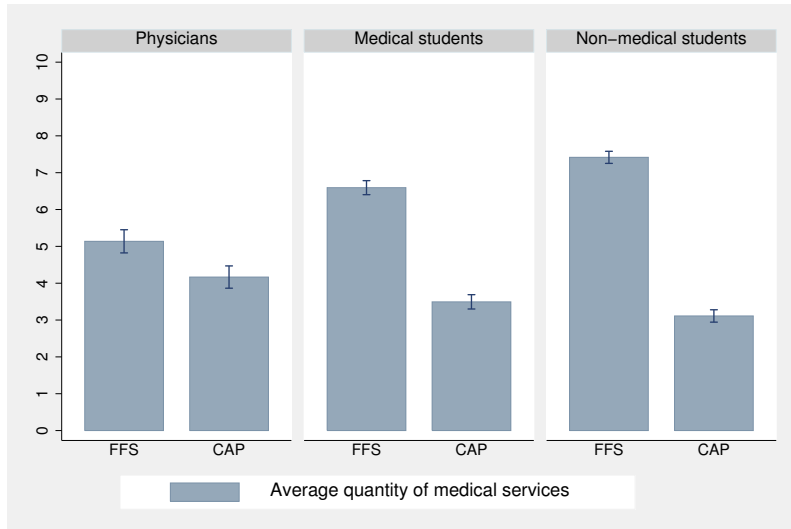


*Notes:* This figure shows average quantities of medical services by payment systems and subject pools. From left to right, averages are shown for choices of physicians (FFS: $N = 117$, CAP: $N = 144$), medical students (FFS: $N = 342$, CAP: $N = 342$), and non-medical students (FFS: $N = 612$, CAP: $N = 621$). Standard deviations are also indicated.

Figure 2 shows that the behavioral difference between FFS and CAP exists for physicians, medical and non-medical students. Physicians provide on average, about 19% more medical services in FFS compared to CAP. For medical and non-medical students, the difference is even more pronounced—medical students and non-medical students choose to provide 47% and 58% more services in FFS, respectively. Differences between payment systems are highly significant for all subject pools ($p \leq 0.0004$, two-sided Mann-Whitney U-test). Accordingly, the effect size (Cohen's $d$) is substantial for all subject groups, being $d = 0.54$ for physicians, $d = 1.71$ for medical students, and $d = 2.05$ for non-medical students. The effect sizes for medical and non-medical students are comparable to findings of Hennig-Schmidt and Wiesen (2014). In their experiment, they find $d = 1.26$ for medical students and $d = 2.22$ for non-medical students.

We further investigate in a series of regression analyses whether the effect of the payment system is driven by patients' attributes (illness and severity of illness), participants' demographics (gender, age), and/or participants' personality traits. We estimate four models in Table 1. Non-medical students are the reference category throughout. In model (1), subjects' medical service provision is explained by dummies for physician, medical student, and the payment system (i.e., FFS = 1). Consistent with our descriptive analysis, the estimation results show that significantly more medical services are provided

Table 1: OLS regression analyses

| Dependent variable | (1) Quantity of med. services | (2) Quantity of med. services | (3) Quantity of med. services | (4) Quantity of med. services |
|---|---|---|---|---|
| Medical student | -0.217 | -0.217 | -0.267 | -0.285 |
| | (0.205) | (0.205) | (0.202) | (0.207) |
| Physician | -0.477 | -0.477 | -0.772 | -0.854 |
| | (0.319) | (0.319) | (0.575) | (0.600) |
| Payment system (if FFS = 1) | 3.530*** | 3.530*** | 3.555*** | 3.538*** |
| | (0.202) | (0.202) | (0.205) | (0.204) |
| Severity of illness | | 0.531*** | 0.531*** | 0.531*** |
| | | (0.026) | (0.026) | (0.026) |
| Illness | | 0.070*** | 0.070*** | 0.070*** |
| | | (0.025) | (0.026) | (0.026) |
| Age | | | 0.013 | 0.015 |
| | | | (0.020) | (0.020) |
| Gender (if female =1) | | | 0.277 | 0.312 |
| | | | (0.217) | (0.220) |
| Extraversion | | | | 0.035 |
| | | | | (0.127) |
| Neuroticism | | | | -0.078 |
| | | | | (0.109) |
| Openness | | | | -0.019 |
| | | | | (0.117) |
| Conscientiousness | | | | 0.033 |
| | | | | (0.141) |
| Agreeableness | | | | -0.088 |
| | | | | (0.146) |
| Constant | -0.034 | -2.830*** | -3.321*** | -3.051** |
| | (0.335) | (0.266) | (0.608) | (1.221) |
| | | | | |
| Observations | 2,178 | 2,178 | 2,178 | 2,178 |
| $R^2$ | 0.430 | 0.523 | 0.535 | 0.537 |

*Notes:* The table displays coefficients from ordinary least square-regression models, clustered for 242 individual subjects. The reference category is non-medical students. Robust standard errors are given in parentheses below the coefficients. Wald tests show that coefficients for medical student and physician dummies do not differ significantly ($p \geq 0.3417$). Using alternative estimation techniques, like ordered probit and ordered logit-models, we arrive at very similar results.

***Significant at the 1-percent level.

**Significant at the 5-percent level.

*Significant at the 10-percent level.

in FFS compared to CAP. Also, medical students and physicians tend to choose fewer services, which might be driven by the fact that non-medical students are affected by incentives of FFS to a larger extent. This effect is not significant, however. In model (2), we account for patients' characteristics (illness and severity). We find that with increasing severity of illness, i.e., the higher the patient optimum, the quantity increases by half a unit. Also the illness, i.e., the level of the patient's health benefit, affects quantity choices. The significant effect of the payment system remains unchanged, however.

We further account for subjects' age and gender in model (3). Neither of them significantly affect subjects' quantity choices. The size of the effect of the payment system remains almost the same after controlling for age and gender. Further, in model (4) we add subjects' personality traits to the regression model. None of the traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) influence subjects' quantity choices significantly.

In sum, we find that the significant effect of the payment system—higher medical service provision in FFS compared to CAP—holds for all subject pools whether physicians, medical or non-medical students. We can therefore conclude that real physicians' behavioral responses to the payment systems are similar to those of medical and non-medical students. The effect is robust when accounting for participants' characteristics and personality traits.

# 4 Concluding remarks

We explore behavioral differences between physicians, medical students, and non-medical students in controlled experiments framed in a physician decision-making context. To this end, we gradually approach the field by starting with a conventional lab experiment using non-medical students. We then proceed by having medical students participate in our experiment. Finally, we bring the lab to the field by introducing an artefactual field experiment to analyze how real physicians respond to incentives from FFS and CAP payment schemes. We actually completely mimic the lab setting in the field. The design of all experiments follows Brosig-Koch et al. (2014). To the best of our knowledge, we are the first to conduct an artefactual field experiment with physicians to analyze behavioral responses to fee-for-service and capitation, which are common forms for paying physicians (see, e.g., McGuire, 2000).

Physicians, medical students, and non-medical students respond to incentives inherent in the payment systems in a consistent way: More medical services are provided in fee-for-service compared to capitation. This finding is in line with the theoretical health economics literature (e.g., Ellis and McGuire, 1986) and corresponds to results from earlier

empirical and experimental studies (e.g., Gaynor and Pauly, 1990; Hennig-Schmidt et al., 2011). Moreover, our findings are robust regarding gender, age, and personality traits.

By comparing behaviors from the artefactual field and the lab experiments, we make a methodological contribution that informs research in the burgeoning field of experimental health economics. In particular, with our artefactual field experiment we address a common criticism of the relevance of inferences about responses to physician payment systems drawn from laboratory experiments which implies that an experiment needs to be conducted with "real" physicians, not with medical or non-medical students; for a general discussion see Harrison and List, (2004, p.1014). Our main finding that subjects in artefactual field and lab experiments in a physician decision-making context respond to incentives in a very similar way makes a case for the validity of inferences drawn from the latter.

We also observe, however, that the intensity of subjects' responses to incentives in fee-for-service and capitation slightly differs for our three subject pools. Thus, inferences based on effect sizes' face value have to be viewed with caution. Moreover, further artefactual field experiments could account for the possible heterogeneity in physicians' characteristics, such as professional background, specialization and work environment (e.g., hospital, group practice or private practice).

# References

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

BARTLING, B., E. FEHR, M. A. MARÉCHAL, AND D. SCHUNK (2009): "Egalitarianism and competitiveness," *American Economic Review*, 99, 93–98.

BEN-NER, A., L. PUTTERMAN, F. KONG, AND D. MAGAN (2004): "Reciprocity in a two-part dictator game," *Journal of Economic Behavior & Organization*, 53, 333–352.

BROSIG-KOCH, J., H. HENNIG-SCHMIDT, N. KAIRIES, AND D. WIESEN (2013): "How Effective are Pay-for-performance Incentives for Physicians? A Laboratory Experiment," Ruhr Economic Papers, No. 413.

——— (2014): "The effects of introducing mixed payment systems for physicians: Experimental evidence," Working paper.

BUCKLEY, N. J., K. CUFF, J. HURLEY, L. MCLEOD, S. MESTELMAN, AND D. CAMERON (2012): "An experimental investigation of mixed systems of public and

private health care finance," *Journal of Economic Behavior & Organization*, 84, 713–729.

ECKEL, C. AND P. GROSSMAN (1996): "Altruism in Anonymous Dictator Games," *Games and Economic Behavior*, 16, 181–191.

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

FAN, C., K. CHEN, AND K. KAN (1998): "The Design of Payment Systems for Physicians under Global Budget - An Experimental Study," *Journal of Economic Behavior and Organization*, 34, 295–311.

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

GAYNOR, M. AND M. V. PAULY (1990): "Compensation and Productive Efficiency of Partnerships: Evidence from Medical Group Practice," *Journal of Political Economy*, 98, 544–73.

GOSLING, S. D., P. J. RENTFROW, AND W. B. SWANN JR (2003): "A very brief measure of the Big-Five personality domains," *Journal of Research in personality*, 37, 504–528.

GREEN, E. (forthcoming): "Payment systems in the healthcare industry: An experimental study of physician incentives," *Journal of Economic Behavior & Organization*.

GREINER, B. (2004): "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen : Beiträge zum Heinz-Billing-Preis 2003*, ed. by K. Kremer and V. Macho, Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, 79–93.

HARRISON, G. W. AND J. A. LIST (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009–1055.

HENNIG-SCHMIDT, H., R. SELTEN, AND D. WIESEN (2011): "How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation," *Journal of Health Economics*, 30, 637–646.

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

KESER, C. AND C. SCHNITZLER (2013): "Money talks: Paying physicians for performance," Discussion Papers, Center for European Governance and Economic Development Research, No. 173.

Kesternich, I., H. Schumacher, and J. Winter (2014): "Professional norms and physician behavior: homo oeconomicus or homo hippocraticus?" Discussion Paper Series of SFB/TR 15 Governance and the Efficiency of Economic Systems, No. 456.

Lagarde, M. and D. Blaauw (2014a): "Testing the effects of doctors' remuneration schemes in a multitasking environment: A real effort laboratory experiment," Working paper, London School of Hygiene & Tropical Medicine.

——— (2014b): "Carrot, stick or competition? The relative effect of Pay for performance schemes in health care," Working paper, London School of Hygiene & Tropical Medicine.

Levitt, S. D. and J. A. List (2009): "Field experiments in economics: The past, the present, and the future," *European Economic Review*, 53, 1–18.

List, J. A. (2011): "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off," *Journal of Economic Perspectives*, 25, 3–16.

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

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

Schram, A. and J. Sonnemans (2011): "How Individuals Choose Health Insurance: An Experimental Analysis," *European Economic Review*, 55, 799–819.

Volk, S., C. Thöni, and W. Ruigrok (2012): "Temporal stability and psychological foundations of cooperation preferences," *Journal of Economic Behavior & Organization*, 81, 664 – 676.

Voors, M., E. Bulte, A. Kontoleon, J. A. List, and T. Turley (2011): "Using artefactual field experiments to learn about the incentives for sustainable forest use in developing economies," *American Economic Review*, 101, 329–333.

# A    Appendix

## A.1    Additional tables and figures

Figure 1: Profit parameters in FFS and CAP
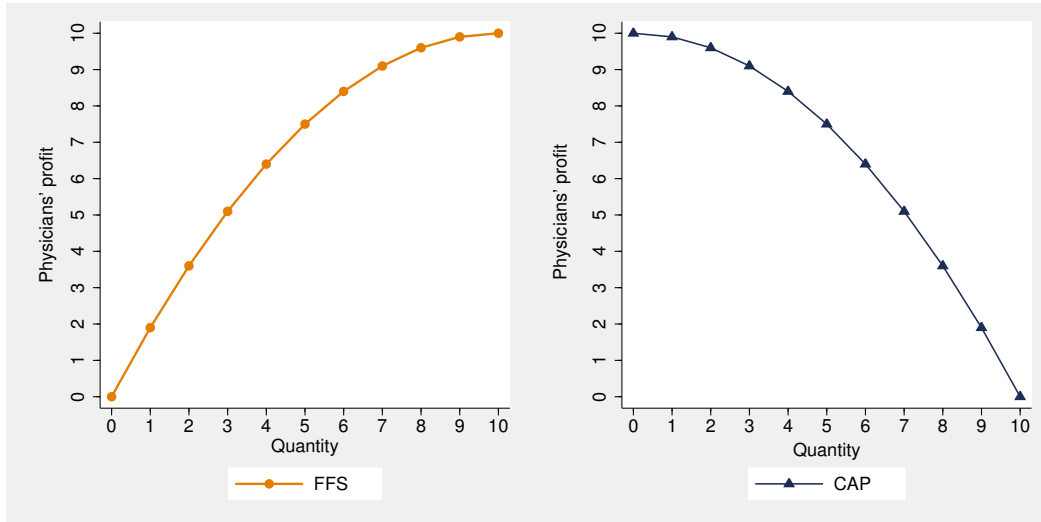


*Notes:* This figure illustrates profit parameters in different payment conditions of the experiment. Profits in FFS and CAP are shown in the left panel and the right panel, respectively.

Figure 2: Patient health benefits
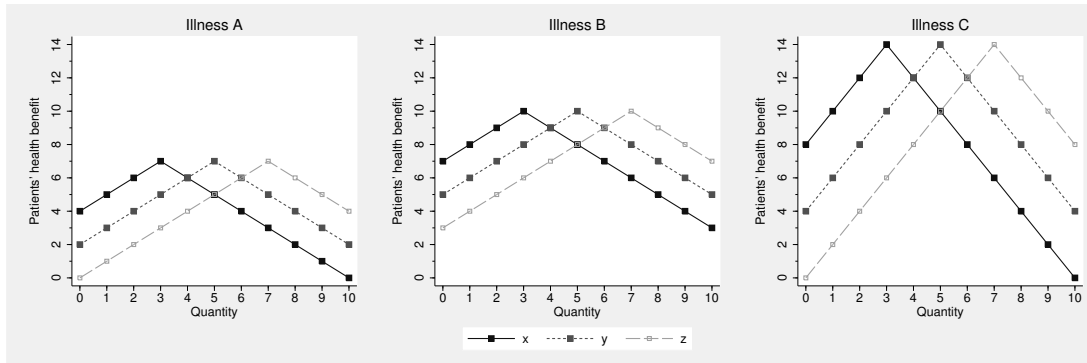


*Notes:* This figure illustrates patient benefit parameters for illnesses $k = A, B, C$ and severites of illness $l = x, y, z$, which are kept constant for all payment conditions.

Table 1: Experimental parameters

| | | Quantity ($q$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Patient | $B_{Ax}$ | 4 | 5 | 6 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| benefit | $B_{Ay}$ | 2 | 3 | 4 | 5 | 6 | 7 | 6 | 5 | 4 | 3 | 2 |
| | $B_{Az}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 6 | 5 | 4 |
| | $B_{Bx}$ | 7 | 8 | 9 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 |
| | $B_{By}$ | 5 | 6 | 7 | 8 | 9 | 10 | 9 | 8 | 7 | 6 | 5 |
| | $B_{Bz}$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 9 | 8 | 7 |
| | $B_{Cx}$ | 8 | 10 | 12 | 14 | 12 | 10 | 8 | 6 | 4 | 2 | 0 |
| | $B_{Cy}$ | 4 | 6 | 8 | 10 | 12 | 14 | 12 | 10 | 8 | 6 | 4 |
| | $B_{Cz}$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 12 | 10 | 8 |
| Costs | $c$ | 0.0 | 0.1 | 0.4 | 0.9 | 1.6 | 2.5 | 3.6 | 4.9 | 6.4 | 8.1 | 10.0 |
| | | | | | | | | | | | | |
| FFS | $p$ | 0.0 | 2.0 | 4.0 | 6.0 | 8.0 | 10.0 | 12.0 | 14.0 | 16.0 | 18.0 | 20.0 |
| | $\pi$ | 0.0 | 1.9 | 3.6 | 5.1 | 6.4 | 7.5 | 8.4 | 9.1 | 9.6 | 9.9 | 10.0 |
| CAP | $L$ | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| | $\pi$ | 10.0 | 9.9 | 9.6 | 9.1 | 8.4 | 7.5 | 6.4 | 5.1 | 3.6 | 1.9 | 0.0 |

*Notes:* This table shows the parameters used in our experiment for all payment conditions.

## A.2 Instructions

**Welcome to the Experiment!**

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

**Part $I$ of the experiment**

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

*Decision situations*

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

*Payment*

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

In each round you will receive detailed information on your screen (see below) for the

respective patient, the illness, your amount of fee-for-service (capitation) remuneration—for each possible amount of medical treatment—your costs, profit, as well as the benefit for the patient with the corresponding illness and severity.

Screen in FFS

**Patient 1 with illness B**

| Quantity of medical treatment | Your fee-for-service payment (in Taler) | Your costs (in Taler) | Your profit (in Taler) | Benefit of the Patient with illness B and severity x (in Taler) |
|---|---|---|---|---|
| 0 | | | | |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Which quantity of medical treatment do you want to provide?

Your decision: [ ]

OK

Screen in CAP

**Patient 1 with illness B**

| Quantity of medical treatment | Your capitation payment (in Taler) | Your costs (in Taler) | Your profit (in Taler) | Benefit of the patient with illness B and severity x (in Taler) |
|---|---|---|---|---|
| 0 | | | | |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |

Which quantity of medical treatment do you want to provide?

Your decision: [ ]

OK

*Payoff*

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

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

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

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

*Comprehension Questions*

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

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