universität innsbruck
Research Platform
Empirical and Experimental Economics

eeecon
[triple:e:con]

Monitoring institutions in health care markets: Experimental evidence

Silvia Angerer, Daniela Glätzle-Rützler, Christian Waibel

Working Papers in Economics and Statistics

2020-32

University of Innsbruck
https://www.uibk.ac.at/eeecon/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 71022
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
https://www.uibk.ac.at/eeecon/wopec/

For a list of recent papers see the backpages of this paper.

Monitoring institutions in health care markets: Experimental evidence*

Silvia Angerer[&], Daniela Glätzle-Rützler[§], and Christian Waibel[†]

Abstract: This paper investigates the impact of monitoring institutions on market outcomes in health care. Health care markets are characterized by asymmetric information. Physicians have an information advantage over patients with respect to the appropriate treatment for the patient and may exploit this informational advantage by over- and underprovision as well as by overcharging. We introduce two types of costly monitoring, endogenous and exogenous. When monitoring detects misbehavior, physicians have to pay a fine. Endogenous monitoring can be requested by patients, whereas exogenous monitoring is performed randomly by a third party. We present a toy model that enables us to derive hypotheses and to test them in a laboratory experiment. Our results show that introducing endogenous monitoring reduces the level of undertreatment and overcharging. Even under high monitoring costs, the threat of patient monitoring is sufficient to discipline physicians. Introducing exogenous monitoring also reduces undertreatment and overcharging when it is performed sufficiently frequently. Market efficiency increases when endogenous monitoring is introduced as well as when exogenous monitoring is implemented with sufficient frequency. Our results, therefore, suggest that monitoring may be a feasible instrument to improve outcomes in health care markets.

JEL-Codes: C91, D82, I11

Keywords: Credence goods, physician behavior, undertreatment, overtreatment, overcharging, monitoring, laboratory experiment

Date: February 26th, 2020

---

* We thank Heike Hennig-Schmidt and participants at the research seminar of the University of Augsburg, of the European University Institute in Florence, of the Copenhagen Business School, of the University of Copenhagen, participants at the 6th Workshop in Behavioral and Experimental Health Economics in Oslo, and participants at the dggö Jahrestagung in Augsburg as well as the Jahrestagung of the Verein fuer Socialpolitik for their helpful comments. Financial support from the Nachwuchsförderung at the University of Innsbruck as well as from the Austrian Central Bank (Jubilaeumsfonds Project 17805) is gratefully acknowledged.

[&] UMIT – University for Health Sciences, Medical Informatics and Technology, Institute for Management and Economics in Healthcare, Eduard-Wallnöfer-Zentrum 1, 6060 Hall in Tirol, Austria. silvia.angerer@umit.at.

[§] University of Innsbruck, Universitätsstrasse 15, 6020 Innsbruck, Austria. daniela.ruetzler@uibk.ac.at.

[†] ETH Zurich, Zuerichbergstrasse 18, 8092 Zurich, Switzerland. cwaibel@ethz.ch.

1. Introduction

A key characteristic of health care markets is information asymmetry between patients and physicians. Patients consult physicians to receive treatment without knowing what type of treatment they need to cure their illness. When diagnosing patients, physicians learn the true disease and may exploit their informational advantage over patients through three forms of misbehavior: providing more treatment than necessary (overtreatment), charging for more treatment than provided (overcharging), and providing less treatment than necessary (undertreatment). Patients must trust that their physician is providing them with appropriate treatment; as a result, health services are often referred to as credence goods (Darby and Karni, 1973; Dulleck and Kerschbamer, 2006). Empirical evidence has documented physicians' reactions to financial incentives, suggesting that over- and undertreatment as well as overcharging are a prevalent phenomenon (Brownlee, 2010; Clemens and Gottlieb, 2014; Dafny, 2005; Jürges and Köberlein, 2015; Pasero and McCaffery, 2001; Reif *et al.*, 2018; Silverman and Skinner, 2004). Recent field experiments support this empirical evidence (Currie *et al.*, 2014; Currie *et al.*, 2011; Das and Hammer, 2007; Das *et al.*, 2016; Gottschalk *et al.*, 2019).

In this paper, we investigate whether the monitoring of physicians can help reduce mistreatment in health care by using the credence goods model introduced by Dulleck and Kerschbamer (2006) to causally assess the effect of monitoring institutions on market outcomes. Interactions in health care markets frequently take place without enforceable institutions for patients. Treatments are often not verifiable by patients because they cannot directly observe the treatment – e.g., during a surgical operation – or because they do not know what kind of treatment the physician actually provides. Although physicians are liable for the treatment they apply, most patients do not consider retribution through the courts because they lack expert knowledge, and many lawsuits against experts fail (Hilger, 2016). For this reason, we use a credence goods market without institutions as our baseline condition. Introducing a monitoring institution allows patients to restore the physician's liability under such conditions. Monitoring induces costs for reviewing physicians' charging and treatment decision and incorporates a fine in case physician misbehavior is revealed.

The monitoring of physicians' treatment decisions has begun to attract more attention over the past two decades. Patients increasingly seek monitoring by patient protection agencies (see, e.g., Furrow, 2011); we also observe more monitoring by government institutions. The German government, for example, runs the "Medical Service of Health Insurers", which reviews and enforces physicians' treatment and charging behavior. The probability of monitoring by government institutions roughly amounts to 5-15% of all inpatient cases in hospitals and has increased in the last year according to surveys conducted by the German Institut of Hospitals (Deutsches Krankenhausinstitut) (Blum *et al.*, 2017; Blum *et al.*, 2018; Blum *et al.*, 2019; Medizinischer Dienst der Krankenversicherung, 2019).

We take account of these developments by implementing monitoring through a costly review of physicians' treatment decisions. This monitoring may be requested by a patient – which we will refer to as "endogenous monitoring" – or randomly performed by a third party – which we will refer to as "exogenous monitoring". If a physician deviates from the appropriate behavior, a financial penalty is imposed on the physician. We vary the conditions[1] between no monitoring, endogenous monitoring, and exogenous monitoring with different probabilities of a review that are common knowledge (65%, 35%, and 10%). We further report the results from four control conditions.

In line with our predictions, our regression analysis shows that monitoring is an effective method for reducing physician misbehavior when such monitoring is either endogenous or exogenous and sufficiently likely. Endogenous monitoring and exogenous monitoring with a monitoring probability of at least 35% both reduce the level of undertreatment and overcharging significantly compared to no monitoring. Efficiency, measured as the sum of patient and physician surplus, increases for endogenous monitoring as well as for exogenous monitoring of at least 35% compared to no monitoring. The results remain robust when we increase the costs of monitoring as well as when we relax the assumption that monitoring can perfectly verify physicians' behavior.

Our paper contributes to the existing literature by providing a systematic investigation into whether monitoring helps to overcome expert misbehavior. In contrast to field data, our laboratory experiment does not take into account all factors of a physician-patient relationship. However, the controlled environment provides the advantage to observe the patient's "true" disease. Therefore, we can unambiguously identify physicians' misbehavior. Our experiment further holds the market environment constant, enabling us to causally identify the effect of monitoring on market outcomes. Moreover, the laboratory experiment permits variations of the type of monitoring that is introduced between endogenous and exogenous monitoring without putting the health of real-world patients at risk.

1.1. Related literature

Three strands of literature relate to our paper: the literature on institutions in credence goods markets, the literature on health economic experiments, and the literature on monitoring.

Institutions in credence goods markets: Darby and Karni (1973) introduce the notion of credence goods in their seminal paper. Dulleck and Kerschbamer (2006) provide a unifying theoretical framework and investigate how different institutions shape outcomes in credence goods markets. The authors show that both experts' liability for insufficient treatment and the verifiability of treatment are sufficient to ensure efficient market outcomes. Dulleck *et al.* (2011) test these predictions in a large-scale laboratory experiment. The authors show that sellers' liability for the sufficiency of their treatments significantly improves market outcomes compared to a market without institutions, whereas the verifiability of the treatment does not. Dulleck *et al.* (2011) introduce liability in their experiment by restricting the action space of experts. Experts can only provide sufficient treatment for customers

[1] We will refer to the experimental variations as "conditions" and to physician treatment as "treatment".

with a major problem. The authors therefore implement liablity in the experiment by excluding the possibility of for experts to undertreat. However, in the real world, misbehavior often first hast to be verified in order to hold experts liable. This is, why we do not restrict the physicians' action space and therefore allow any type of misbehavior in all conditions. We test the influence of liability on the frequency of physician misbehavior by introducing endogenous and exogenous monitoring institutions. To the best of our knowledge our paper is one of the first papers applying the theoretical framework of Dulleck and Kerschbamer (2006) and Dulleck *et al.* (2011) to the health care market and therefore using a health frame. The only exception are the papers by Greiner *et al.* (2017) focusing on the separation of diagnosis and treatment and Waibel and Wiesen (2019) studying primary care physicians' referrals.

In neutral frames, the impact of competition, reputation, second opinions, price regulation, reducing insurance coverage and media have been investigated as institutions to improve market outcomes (see, e.g., Balafoutas *et al.*, 2013; Balafoutas *et al.*, 2017; Dulleck *et al.*, 2011; Huck *et al.*, 2016; Kerschbamer *et al.*, 2016; Kerschbamer *et al.*, 2019; Liu *et al.*, 2019; Mimra *et al.*, 2016a; Mimra *et al.*, 2016b; Rajgopal and White, 2019; Rasch and Waibel, 2018). In sum, this literature suggests that several institutions should theoretically be able to mitigate market inefficiencies in the credence goods market. Whether or not these instruments turn out to be effective depends on how they are implemented. Second opinions, for example, show to improve market outcomes if search costs are sufficiently low.

Laboratory experiments in health economics: Our paper is further linked to the emerging literature on experimental health economics. The seminal paper by Hennig-Schmidt *et al.* (2011) compares physician behavior under fee-for-service and capitation payment. Non-medical laboratory students act in the role of physicians. Patients are not present in the laboratory, but the monetary patient benefit is used for the treatment of genuine external patients. Hennig-Schmidt *et al.* (2011) find that physicians overtreat under fee-for-service and undertreat under capitation. In a series of laboratory experiments (Brosig-Koch *et al.*, 2017; Brosig-Koch *et al.*, 2013; Green, 2014; Kairies and Krieger, 2013; Keser *et al.*, 2014), the authors further vary the payment schemes and analyze the impact on physicians' provision behavior. Physicians appear to react to financial incentives with substantial over- and undertreatment. These results qualitatively extend to both medical students and real physicians (Brosig-Koch *et al.*, 2016). Professional norms also shape physicians' treatment behavior by shifting importance from physicians' self-interests toward patient benefits (Kesternich *et al.*, 2015). In contrast to the studies described above, in our experiment, the subjects in the role of patients are present in the laboratory. Thus we are able to investigate the interaction between patients and physicians. Patients decide whether or not to consult a physician and whether or not to monitor the physician's behavior in our endogenous monitoring condition.

Literature on monitoring: In his seminal paper Hölmstrom (1979) highlights that investments into monitoring might help to overcome informational asymmetries in a principal-agent set-up.

Complete monitoring allows observing an agent's effort. The first-best can be reached by implementing a contract that penalizes sub-optimal effort levels with a sufficiently high payment. Hölmstrom (1979) shows that if payoffs but not the effort level are observable, imperfect information only allows implementing second-best allocations. Our condition with endogeneous monitoring is similar to the contractual environment of Hölmstrom (1979) under imperfect information. Patients observe their payoffs but not physicians' actions before patients decide to request monitoring.

The effect of monitoring on human behavior has been investigated in a variety of contexts in the field. Slemrod et al. (2001), for example, employ a field experiment to investigate tax compliance in Minnesota, finding that a reminder about audits significantly increases the returns filed by low- and middle-income sole proprietors. Duflo et al. (2012) explore teachers' incentives to go to work in a field experiment in India. In the treatment group, teacher attendance was monitored daily using cameras. The authors determine that monitoring reduces teacher absenteeism by 21 percentage points. Related to the health care sector, recent results suggest that upcoding may be effectively reduced in a neonatological context by audits (Hennig-Schmidt *et al.*, 2019). Participants in the experiment take the role of neonatologists. They report birth weights either under no-audit or random auditing including fines (with a probability of 10%). Reporting lower birthweights leads to higher remuneration. The authors show that the level of upcoding almost amounts to three-quarters of all decisions under no auditing, whereas the level of upcoding is reduced to roughly 50% under auditing.[2] Another paper by Lindeboom *et al.* (2016) studies the causal effect of different audit regimes on compliance in a market for immediate (home and nursing) care by means of a field experiment conducted in the Netherlands. While different treatments with unconditional and conditional changes in audit rates were employed (in the conditional case, audit rates depend on the provider's recent performance), no significant effects on the quantity and quality of applications for immediate care could be detected. The authors conclude that the absence of direct sanctions for noncompliancce might be responsible for the null effect.

The effects of sanctions has been extensively investigated in laboratory experiments on cooperation. Fehr and Gächter (2000) were the first to show in a laboratory experiment that the potential for individuals to financially punish their trading partners reduces free-riding incentives. Masclet *et al.* (2003), Page *et al.* (2005), and Sefton *et al.* (2007) provide similar outcomes using different methods of punishment. Grosse *et al.* (2011) extend these findings to the context of teamwork.

Both Belot and Schröder (2016) and Villeval (2018) point out that monitoring can improve outcomes but may also have adverse spillover effects in unmonitored dimensions. This is also found in

[2] Note that in our study we also investigate the effect of a 10% monitoring probability and find no reduction in physician misbehavior. However, our implementation is fundamentally different. In our setting with repeated interactions, detected misbehavior in one period leads to a reduction in profits in the respective period. In addition, only 4 randomly selected periods out of 16 are payoff-relevant (see section 2.2 experimental conditions). In Hennig-Schmidt *et al.* (2019) a detection of dishonest reports leads to a loss of all earnings except the show-up fee.

Hennig-Schmidt *et al.* (2019): When introducing monitoring, dishonest reporting of birth weights in dimensions where monitoring cannot detect dishonesty, significantly increases the level of dishonest reporting compared to a situation with no monitoring. Overall, the results from previous monitoring experiments suggest that monitoring may be an effective way to change an individual's behavior. Efficiency effects are less clear, however, and depend on the costs of monitoring, the punishment applied as well as on possible spillovers.

The remainder of the paper is organized as follows: The next section introduces the experiment and presents our predictions. Section Three reports the results. Section Four discusses the results, and Section Five offers conclusions.

2. Experiment

Our experimental design is based on the credence goods framework established by Dulleck and Kerschbamer (2006) and tested in an economic laboratory experiment by Dulleck *et al.* (2011). In contrast to Dulleck *et al.* (2011), we allow for any type of misbehavior in all conditions without restricting the action space of physicians; we test the influence of reinforcing liability by introducing monitoring institutions (endogenous and exogenous) on the frequency of undertreatment, overtreatment, and overcharging. Below, we will (i) present the basic set-up and explain the adjustments made in order to more closely approximate the health care market, (ii) introduce the experimental conditions and parameters, and (iii) provide the theoretical predictions.

2.1. The basic set-up and parameterization

The set-up was framed in a health care context in which we present consumers as "patients", experts as "physicians", the quality needed as the "health problem", and the quality provided as the "treatment". This formulation is motivated by the findings of Kairies-Schwarz *et al.* (2017) and Kesternich *et al.* (2015), who show the importance of health framing on decision quality within health economic laboratory experiments. Patients suffer from a health problem that is major with probability $h = 0.5$ and minor with probability $(1 - h)$, and they must decide whether to consult a physician. The probability h is common knowledge, and the patient knows ex-ante that he has a health problem. However, the patient does not know whether he has a major or minor health problem. The physician can diagnose the health problem with certainty at no cost and then provide treatment. Treating a major health problem costs the expert $c_H = 6$, whereas treatment for the minor health problem costs $c_L = 2$. Treatment prices are $p_H = 8$ for the major treatment and $p_L = 3$ for the minor treatment. Patients with a minor health problem can be cured with either treatment. However, if the patient suffers from a major health problem, only the major treatment provides a cure. If the health problem is cured, the patient receives a value $v = 10$, whereas an insufficient treatment yields a value v of zero. We abstract from patients' health insurance.[3] When the patient consults the physician, the patient surplus amounts to v minus the price charged (p_H or p_L), whereas the physician receives the price charged (p_H or p_L) minus the cost for the provided treatment (c_H or c_L).[4] When the patient refrains from interacting, the patient receives an outside option of $o_{Pat} = -4$, whereas the physician receives $o_{Phy} = 0$. In contrast to Dulleck *et al.* (2011), we introduce exogenously fixed prices, which are common in highly regulated health care markets, and a negative surplus for patients not entering the market to illustrate the disutility of an uncured health problem.[5] In our baseline set-up, no verifiability of the type of treatment provided and no liability for the provision of insufficient treatment are assumed, allowing us to investigate all types of physician misbehavior: (i) undertreatment, (ii) overtreatment, and (iii) overcharging.[6]

The stage game of the baseline set-up is the following:

1. For each patient, nature draws the type of health problem. With probability h the patient has a major health problem, and with probability $(1-h)$ the patient has a minor health problem
2. The patient decides whether to consult the physician or not to interact.
3. If the patient decides to consult the physician, the physician is informed about the health problem and provides a treatment (q_H or q_L) and charges a price (p_H or p_L)
4. Each patient observes his payoff, and each physician observes her payoff

2.2 Experimental conditions

In total, nine conditions of interaction between physicians and patients were conducted (see Table 1). The baseline condition (***B***) acts as a control in order to measure the three types of misbehavior in the absence of liability and verifiability. In the monitoring conditions (***M***), we introduce the assumption of liability for any type of deviation from appropriate physician behavior and perfect verifiability of physician behavior by a third party (except for ***M-ExUnc***) and implement a costly monitoring institution with monitoring costs of $m = 1$ (exept for ***M-EndHc***). There are two types of monitoring institutions: (i) endogenous monitoring (***M-End***), in which the patient decides whether to

3 Note that our predictions qualitatively hold robust for any health insurance coverage with a positive co-payment. .

4 Following Dulleck *et al.* (2011), we assume large economies of scope between diagnosis and treatment. Therefore, once a patient decides to consult a physician, the patient is committed to undergoing treatment by the physician.

5 The value v when the physician provides insufficient treatment exceeds the outside option of not consulting a physician for the patient, as we assume that any treatment is better than no treatment.

6 Note, however, that without verifiability, overtreatment is a strictly dominated strategy, since overcharging does not affect the payoff to the patient (unlike overtreatment), while at the same time providing a higher payoff to the physician.

monitor the behavior of the physician[7], and (ii) exogenous monitoring (***M-Ex***), in which a third party – in our set-up the computer – implements monitoring for a random sample of physicians.[8] If monitoring reveals physician misbehavior, the monitoring costs are paid by the physician. In the ***M-End*** (***M-Ex***) conditions, the patient (third party) pays the monitoring cost in the case of appropriate treatment and pricing,. In addition to the monitoring cost, the physician must reimburse the patient for the damage caused in cases in which misbehavior is revealed. This reimbursement amounts to the difference between the patient payoff in cases of appropriate treatment and pricing and the patient payoff resulting from the treatment offered and the price charged.

In total, we have three endogenous monitoring conditions. ***M-End***, ***M-EndHc,*** where the monitoring costs are increased to $m = 3$ and ***M-EndUnc***, where the the assumption of perfect verifiability of physician behavior is relaxed and the probability of detecting physician misbehavior is 0.8.[9]

The exogenous monitoring conditions vary with respect to the probability of a physician being monitored (μ) by the third party and whether the probability is common knowledge. The three exogenous monitoring conditions with common knowledge of the monitoring probability are the following: (i) ***M-Ex65*** with $\mu = 0.65$, this reflects the likelihood of monitoring under condition ***M-End***, (ii) ***M-Ex35*** with $\mu = 0.35$ representing the highest likelihood under which the predicted behavior should not differ from the baseline (see section 2.3 and the proofs in Appendix D) and (iii) ***M-Ex10*** with $\mu = 0.10$, which can be interpreted as a feasible likelihood with which a government institution would monitor physicians and represents the average probability that the "Medical Service of Health Insurers" monitors inpatient cases (Blum *et al.*, 2017; Blum *et al.*, 2018; Blum *et al.*, 2019). In ***M-ExUnk*** with $\mu = 0.10$ we relax the assumption of common knowledge of the monitoring probability to make the exogenous monitoring conditions more realistic and to test whether the knowledge of the likelihood of monitoring has an effect on physician behavior.[10]

Finally, in order to guarantee a clean comparison between endogenous and exogenous monitoring, we also test for a further control condition (***M-End-NoInfo)*** where patients have to make a monitoring decision without any prior feedback on payoffs.

The experiment runs for 16 periods. At the beginning of the first period, the two roles are randomly assigned to the subjects; these roles are then maintained throughout the 16 periods. In each period, an anonymous patient is randomly matched with a physician from the same matching group.[11]

---

[7] In the health care market, endogenous monitoring exists in the form of patient lawsuits against physicians, with a judge acting as a third party to verify physician misbehavior. The monitoring costs m represent the court fees.

[8] An example of exogenous monitoring is the implementation of the "Medical Service of Health Insurers", which reviews and enforces physicians' treatment and charging behavior and/or the possibility of detecting physician misbehavior through the increased use of e-government in the health care sector (whereby data on diagnostic tests, the prescribed medication, the medical treatment, and the services billed can all be linked).

[9] Note that the probability of correctly verifying physician behavior in the case of appropriate treatment is always 1 if the physician is monitored.

[10] Note that we also elicited subjects' beliefs about the monitoring probability at the end of the experiment. The belief elicitation was incentivized. For a detailed description see Appendix

[11] Due to random and anonymous matching, reputation-building is excluded.

The matching groups consist of eight subjects each: four physicians and four patients. In all experimental conditions, except for condition ***M-End-NoInfo***, patients and physicians are informed about their own payoffs after the physicians' treatment choices and the prices are revealed. From this information, a patient can only infer whether the treatment was sufficient or not. Any other type of misbehavior cannot be derived from the payoff information. After the payoff feedback, the patient (third party) decides whether to exert monitoring in the ***M-End*** (***M-Ex***) conditions, whereas in condition ***B***, the period ends. In the monitoring conditions, following the monitoring decision (by the patient or by a random draw from the computer), additional payoff feedback is provided for both patients and physicians. In condition ***M-End-NoInfo*** the payoffs for patients and physicians are only revealed at the end of each period, after the monitoring decision of the patient. Ultimately, four periods are randomly selected to determine participant payments. For a description of the experimental protocol and the instructions see Appendix B.

Table 1 Experimental conditions

Conditions	**Description**			**# of subjects (# of markets)**
B	Baseline, no monitoring			48 (6)
Monitoring	**Monitoring costs**	**Detection probability**	**Monitoring frequency**	
Endogenous monitoring conditions (patients decide on monitoring)				
M-End	Low ($m = 1$)	100%	Endogenous	48 (6)
M-EndHc	High ($m = 3$)	100%	Endogenous	48 (6)
M-EndUnc	Low ($m = 1$)	80%	Endogenous	48 (6)
Exogenous monitoring conditions (3rd party randomly decides on monitoring)				
M-Ex65	Low ($m = 1$)	100%	65% (c.k.)[1]	40 (5)
M-Ex35	Low ($m = 1$)	100%	35% (c.k.)[1]	48 (6)
M-Ex10	Low ($m = 1$)	100%	10%(c.k.)[1]	48 (6)
M-ExUnk	Low ($m = 1$)	100%	10% (unknown)	48 (6)
Control condition (Patients decide on monitoring with no information on payoffs)				
M-EndNoInfo	Low ($m = 1$)	100%	Endogenous	48 (6)

Note:
[1] c.k. denotes common knowledge of the monitoring frequency

2.3 Predictions

The following section provides an equilibrium analysis and the resulting efficiency level for each of the seven conditions. The analysis is based on the assumption that physicians and patients are rational, risk-neutral, and that they maximize their own monetary payoff. We additionally assume that all information is common knowledge except for the patients' types. Table 2 summarizes the predictions.

2.3.1 Equilibria

The interaction between patients and physicians is a sequential game with incomplete information. We apply the Perfect Bayesian Equilibrium concept. Based on the parameter specifications introduced above, we provide a full description of the equilibria of our main conditions in Appendix D. Below, we summarize the results and present the expected efficiency-levels.

(i) Condition ***B***: In our baseline condition, physicians always provide minor treatment and charge the price for the major treatment. Patients anticipate the behavior of physicians and enter the market because the expected payoff is higher than the outside option.[12] The market operates at an efficiency-level of 25% of the maximum efficiency (i.e., the total surplus from the market interaction is a quarter of the maximum possible surplus).

(ii) Condition ***M-End***: Under endogenous monitoring, patients decide to monitor the behavior of physicians whenever they observe a negative payoff, they mix with probability 5/6 when observing a payoff of 2, and they do not monitor the behavior of physicians when observing a payoff of 7 (i.e., the maximum possible payoff a patient can receive). Anticipating the reaction of patients, physicians provide the appropriate treatment and charge the honest price. The market operates at an efficiency-level of 90%.

(iii) Condition ***M-EndHc***: When higher costs are introduced in our endogenous monitoring condition, patients decide to monitor the behavior of physicians whenever they observe a negative payoff. Patients mix with probability 5/8 when observing a payoff of 2, and do not monitor the behavior of physicians when observing a payoff of 7. Anticipating the reaction of patients, physicians provide the appropriate treatment and charge the honest price. The market operates at an efficiency level of 77%.

(iv) Condition ***M-EndUnc***: When monitoring detects physician misbehavior with a probability of 80%, patients decide to monitor the behavior of physicians whenever they observe a negative payoff or a payoff of 2. They do not monitor the behavior of physicians when observing a payoff of 7. Anticipating the reaction of patients, physicians provide the appropriate treatment and charge the price for the major treatment. The market operates at an efficiency-level of 81%.

(v) Condition ***M-Ex65***: When an exogenous monitoring institution with a monitoring probability of 65% is introduced, physicians provide the appropriate treatment and charge the price for the major treatment. The market operates at an efficiency level of 88%.

(vi) Condition ***M-Ex35***: When the probability of exogenous monitoring is reduced to 35%, physicians always provide minor treatment and charge the price for the major treatment. Patients always enter the market. The market operates at an efficiency level of 18%.

[12] A market breakdown as in Akerlof (1970) is excluded by our parameters due to our assumption that patients expect that receiving any treatment will be better than no treatment.

(vii) Condition ***M-Ex10***: When the probability of exogenous monitoring is further reduced to 10%, physicians always provide minor treatment and charge the price for the major treatment. Patients always enter the market. The market operates at an efficiency level of 23%.

(viii) Condition ***M-End-NoInfo***: When patients have to make the monitoring decision without any information on their payoffs, patients decide to monitor the behavior of physicians with a probability of 5/6. Anticipating the reaction of patients, physicians provide the appropriate treatment and charge the honest price. The market operates at an efficiency level of 84%.

In our condition ***M-ExUnk***, the probability of exogenous monitoring is unknown to the subjects. In the first period, both patients and physicians could have any expectation regarding the exogenous monitoring probability. Therefore, we will not provide any theoretical predictions.

Summarizing our predictions, we can conclude that in the absence of monitoring institutions, undertreatment and overcharging are prevalent. Introducing an endogenous monitoring institution leads to less physician misbehavior and an increase in efficiency. However, full efficiency cannot be achieved due to the monitoring costs that the patients bear when their health problem is major.[13] Under exogenous monitoring with a monitoring frequency of 65%, physicians will no longer undertreat but will overcharge. Efficiency decreases slightly compared to endogenous monitoring due to the more frequent use of monitoring (65% vs. 42%).[14] Once the frequency of exogenous monitoring falls below a probability of 4/11, physicians will undertreat and overcharge, as in the absence of monitoring institutions. In addition, due to the monitoring costs, efficiency is lower compared to the baseline condition.

[13] Note that in the case of a minor health problem, the patient receives a payoff of 7 in equilibrium and thus does not choose to monitor the physician.

[14] Note that in condition ***M-Ex65***, monitoring is implemented in 65% of all interactions. In the equilibrium of condition ***M-End***, however, monitoring is implemented with a probability of 5/6 in the case of a major health problem and is never implemented in the case of a minor health problem. Thus, monitoring is expected to be implemented in 42% of all interactions.

Table 2 Predictions

	B	***M-End***	***M-EndHc***	***M-EndUnc***	***M-Ex65***	***M-Ex35***	***M-Ex10***	***M-End NoInfo***
OT[1]	0%	0%	0%	0%	0%	0%	0%	0%
UT[2]	100%	0%	0%	0%	0%	100%	100%	0%
OC[3]	100%	0%	0%	100%	100%	100%	100%	0%
INT[4]	100%	100%	100%	100%	100%	100%	100%	100%
EFF[5]	25%	90%	77%	81%	88%	18%	23%	84%

Notes:
[1] OT: overtreatment; consumer needs q_l, but seller provides q_h
[2] UT: undertreatment; consumer needs q_h, but seller provides q_l
[3] OC: overcharging; seller provides q_l, but charges p_h (with $p_h > p_l$ and consumer needing q_l)
[4] INT: Interaction rate; relative frequency of market interaction
[5] EFF: Efficiency calculated as $\frac{\text{actual profit}}{\text{maximum possible profit}}$ with the actual profit being adjusted for the monitoring costs of the third party in the exogenous monitoring treatments

3. Results

In this section, we first describe the results for the baseline condition. We then describe the effects resulting from the introduction of endogenous and exogenous monitoring and the impact of changing and/or omitting information on the monitoring probability in conditions M-Ex. We also explore the influence of variation in monitoring costs and uncertainty in the verifiability of physician misbehavior. Finally, we discuss the regression analysis.

3.1. Baseline

Table 3 shows the aggregate results for each experimental condition averaged over markets and periods. When calculating the undertreatment rate, only periods with a major health problem are considered, whereas for overtreatment and overcharging, only periods with a minor health problem are considered. In line with our predictions, in condition ***B***, undertreatment and overcharging are observed. The rate of undertreatment is significantly lower than theoretically predicted: Physicians undertreat in 40% of the cases ($p < 0.05$, Wilcoxon signed ranks test (WSR)), and in 91%, they overcharge ($p > 0.1$ compared to a frequency of 100%, WSR). Overtreatment is dominated by overcharging. We only observe overtreatment once in period 1 (see Figure 1, panel A). The frequency of market interaction is 97%. Hence, as predicted, patients almost always consulted a physician. The market efficiency level is 63%, which is higher than theoretically expected ($p < 0.05$, WSR).

Table 3 Overview of results (means)

Levels of	***B***	***M-End***	***M-Ex65***	***M-Ex35***	***M-Ex10***	***M-ExUnk***	***M-EndHc***	***M-EndUnc***	***M-End NoInfo***
Overtreatment[1]	0.01[b,e*]	0.02	0.04[b,g*]	0.01[g*]	0.00[i]	0.07[e*,i]	0.06	0.04	0.04
Undertreatment[2]	0.40[a,b,c]	0.09[a]	0.12[b]	0.16[c,h]	0.41[h]	0.31	0.06	0.11	0.09
Overcharging[3]	0.91[a,b,c*,e]	0.62[a,l]	0.46[b,g]	0.73[c*,g]	0.84[i*]	0.67[e,i*]	0.54	0.53	0.41[l]
Interaction [4]	0.97[a,b,d*]	1.00[a,l]	0.99[b,m*]	0.99	0.99[d*]	0.98	0.997	0.997	0.97[l,m*]
Efficiency[5]	0.63[a,b*,c*]	0.81[a,j,l]	0.76[b*]	0.80[c*]	0.69	0.70	0.67[j]	0.77	0.75[l]
Profit physician[6]	4.33[a,b,c,e]	1.10[a,k,]	1.32[b,g]	2.53[c,g,h]	4.11[h,i*]	3.39[e,i*]	1.36	1.51[k]	1.20
Profit patient[7]	-0.11[a,b,c,e]	4.23[a,j,k,l]	3.93[b,g,m]	2.79[c,g,h]	0.75[h]	1.34[e]	2.91[j]	3.28[k]	3.46[l,m]
# subjects	48	48	40	48	48	48	48	48	48

Notes:
[1] OT: overtreatment; consumer needs q_l, but seller provides q_h
[2] UT: undertreatment; consumer needs q_h, but seller provides q_l
[3] OC: overcharging; seller provides q_l, but charges p_h (with $p_h > p_l$ and consumer needing q_l)
[4] INT: Interaction rate; relative frequency of market interaction
[5] EFF: Efficiency calculated as $\frac{\text{actual profit}}{\text{maximum possible profit}}$ with the actual profit being adjusted for the monitoring costs of the third party in the exogenous monitoring treatments
[6] PDoc: Profit physician in experimental currency units
[7] PP: Profit Patient in experimental currency units
Mann-Whitney U-tests (MWU) for pairwise differences between conditions with matching groups of 8 subjects as one independent observation
[a] ***B*** versus ***M-End*** ($p < 0.05$), [a*] ($p < 0.1$)
[b] ***B*** versus ***M-Ex65*** ($p < 0.05$), [b*] ($p < 0.1$)
[c] ***B*** versus ***M-Ex35*** ($p < 0.05$), [c*] ($p < 0.1$)
[d] ***B*** versus ***M-Ex10*** ($p < 0.05$), [d*] ($p < 0.1$)
[e] ***B*** versus ***M-ExUnk*** ($p < 0.05$), [e*] ($p < 0.1$)
[f] ***M-End*** versus ***M-Ex65*** ($p < 0.05$), [f*] ($p < 0.1$)
[g] ***M-Ex65*** versus ***M-Ex35*** ($p < 0.05$), [g*] ($p < 0.1$)
[h] ***M-Ex35*** versus ***M-Ex10*** ($p < 0.05$), [h*] ($p < 0.1$)
[i] ***M-Ex10*** versus ***M-ExUnk*** ($p < 0.05$), [i*] ($p < 0.1$)
[j] ***M-End*** versus ***M-EndHc*** ($p < 0.05$), [j*] ($p < 0.1$)
[k] ***M-End*** versus ***M-EndUnc*** ($p < 0.05$), [k*] ($p < 0.1$)
[l] ***M-End*** versus ***M-EndNoInfo*** ($p < 0.05$), [l*] ($p < 0.1$)
[m] ***M-Ex65*** versus ***M-EndNoInfo*** ($p < 0.05$), [m*] ($p < 0.1$)

With regard to the development of physician and patient choices over the 16 periods averaged over markets, Figure 1 shows no clear trend in the time path of our key variables in condition ***B***. Only for overcharging (panel C) do we find a marginal trend toward a significant increase across the 16 periods (p = 0.052, WSR for a comparison of means between the first and the second half of periods averaged over markets and periods).

Result 1. In condition ***B***, undertreatment and overcharging are observed. However, the level of undertreatment is lower than 100%.

3.2. Introduction of monitoring

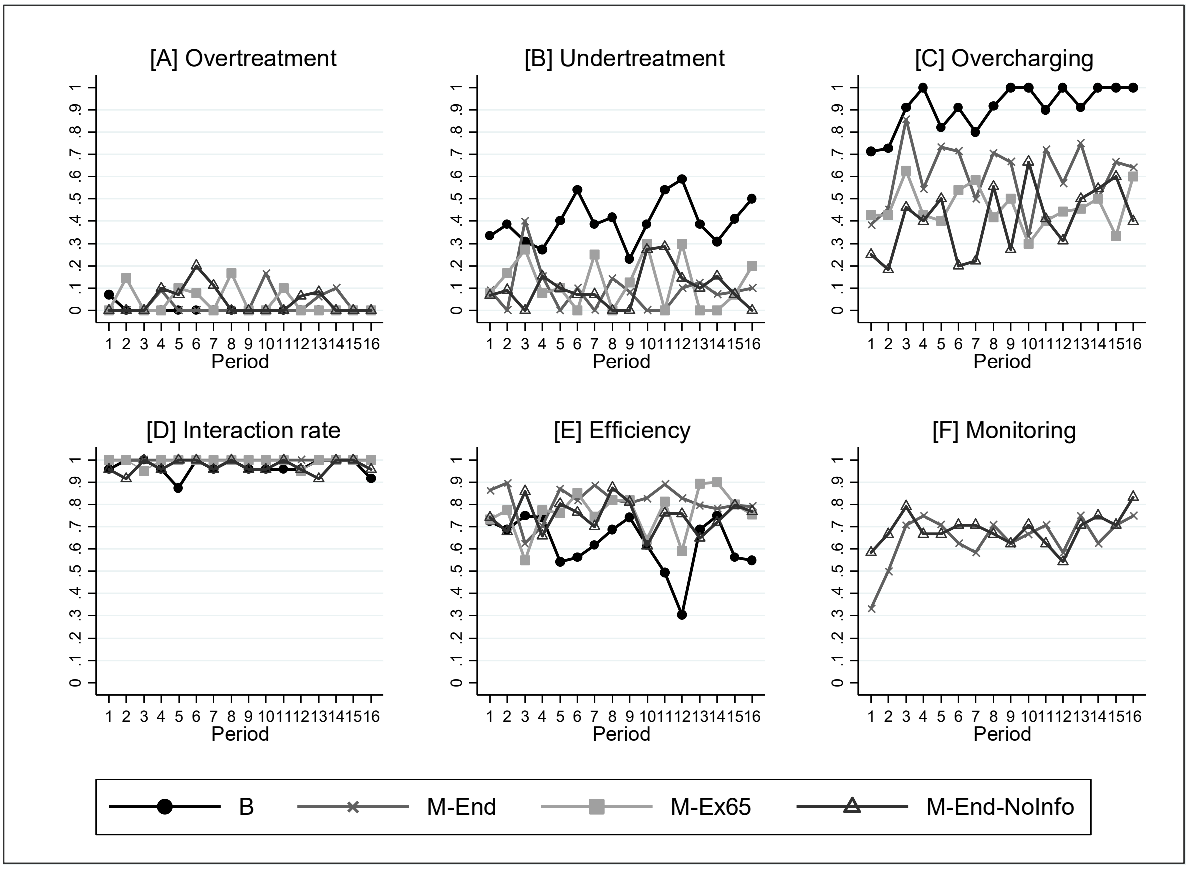
Table 3 and Figure 1 show that the introduction of endogenous monitoring or exogenous monitoring with a probability of 65% significantly reduces the rate of undertreatment, overcharging, and increases the interaction rate between patients and physicians, compared to ***B***. Due to the lower frequency of undertreatment and the resulting increase in cured health problems, efficiency significantly increases.

Regarding the decisions of patients to exert costly monitoring in ***M-End***, the institution of monitoring is used in 64.6% of periods overall. When patients receive a negative payoff feedback (i.e., undertreatment), they decide to exert monitoring in 100% of cases. When receiving a payoff feedback of 2 ECU, the monitoring frequency decreases to 78.6%, and with a payoff feedback of 7 ECU, the monitoring frequency drops to 7.5%. Thus, the behavior of patients resembles the theoretical prediction that patients will always decide to monitor with a negative payoff feedback, to mix in the case of a payoff feedback of 2, and not to monitor with a payoff feedback of 7. Analyzing the dynamics of monitoring across the 16 periods, panel F of Figure 1 shows that the overall frequency of monitoring started out at a relatively low level of 30%, increased to 70% in Period 3 (because of the sharp increase in undertreatment; see panel B), and remained at a high level thereafter (between 60% and 75%; $p < 0.05$, WSR for a comparison of means between the first and the second half of periods averaged over markets and periods). With respect to the development of the other variables across periods (panels A to E of Figure 1), there is no evidence of a time trend.

We do not find any differences between ***M-End*** and ***M-Ex65*** in any of the relevant variables (see Table 3). Also, the results from condition ***M-End-NoInfo*** show that ***M-End*** and ***M-Ex65*** are comparable with respect to the overall monitoring frequency (being 68% overall periods in ***M-End-NoInfo***) and the level of undertreatment. The level of overcharging, however, is significantly lower in ***M-End-NoInfo*** than in ***M-End***, but no difference is found to ***M-Ex65***. Due to the increased use of monitoring in case of a minor illness and appropriate physician behavior, market efficiency is significantly lower in ***M-End-NoInfo*** compared to ***M-End*** but no difference is found compared to ***M-Ex65***.

Result 2. The introduction of monitoring (endogenous or exogenous with 65% probability) decreases the levels of undertreatment and overcharging and increases the level of efficiency compared to condition ***B***.

Figure 1 Provision, charging, market interaction and efficiency by conditions ***B***, ***M-End***, ***M-Ex65*** and ***M-End-NoInfo***



3.3. Change in monitoring probabilities

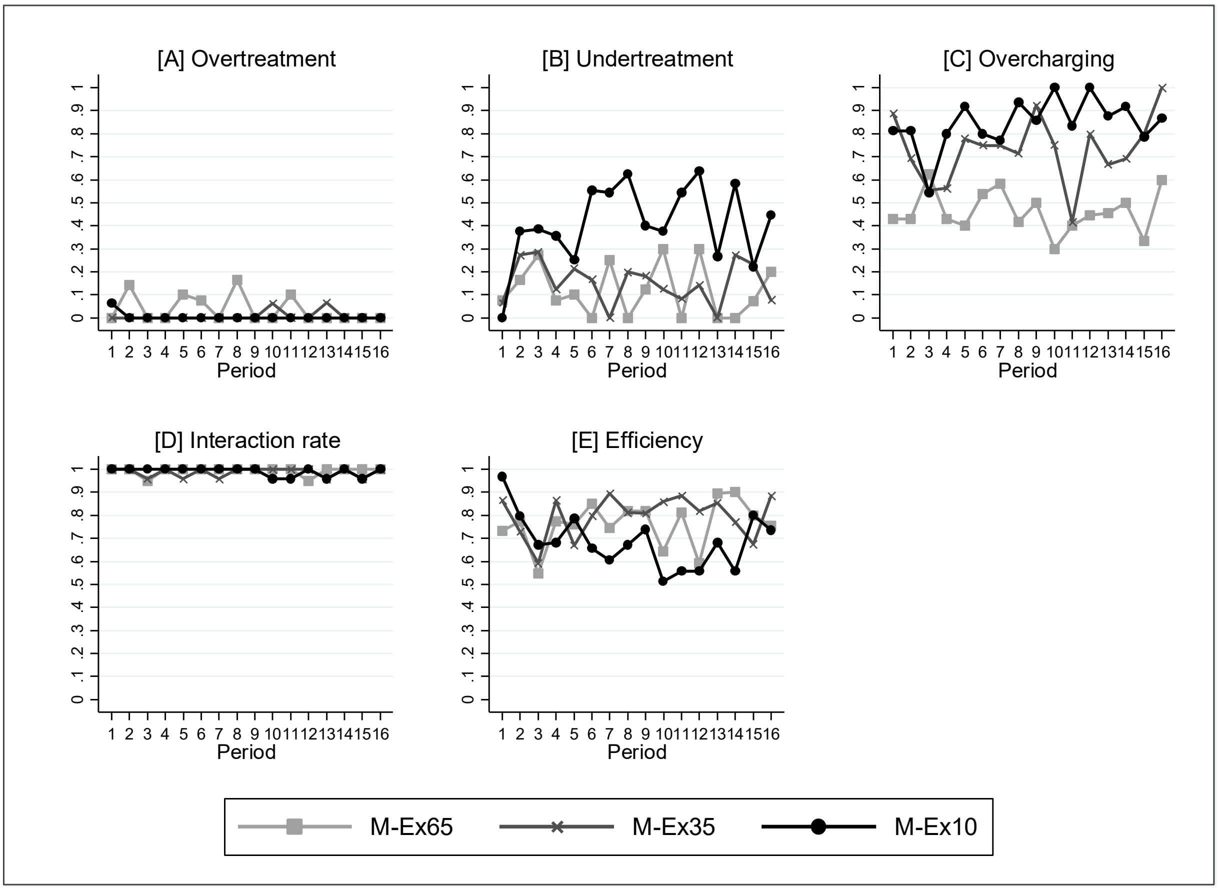
When the frequency of monitoring is lowered to 35%, the level of undertreatment, the interaction rate, and the efficiency level in ***M-Ex35*** do not change significantly compared to ***M-Ex65*** (see Table 3 and Figure 2). The rate of overcharging, however, increases significantly. Compared to condition ***B***, the level of overcharging in ***M-Ex35*** is still marginally lower. We summarize our findings as follows:

Result 3. A reduction in the monitoring frequency from 65% to 35% has no impact on the frequency of undertreatment or the efficiency level, but overcharging increases significantly.

A further reduction of the monitoring frequency to 10% significantly increases the level of undertreatment to 41%. This rate is comparable to condition ***B***. Moreover, the level of overcharging and the efficiency level do not differ significantly between ***B*** and ***M-Ex10***. Figure 2 shows the development of the most important outcomes across the 16 periods for all three exogenous monitoring conditions. There is no evidence of a time trend.

Result 4. A reduction in the monitoring frequency from 35% to 10% significantly increases the frequency of undertreatment. There is no difference in any key variable between conditions ***M-Ex10*** and ***B***.

Figure 2 Provision, charging, market interaction and efficiency by conditions ***M-Ex65, M-Ex35*** and ***M-Ex10***

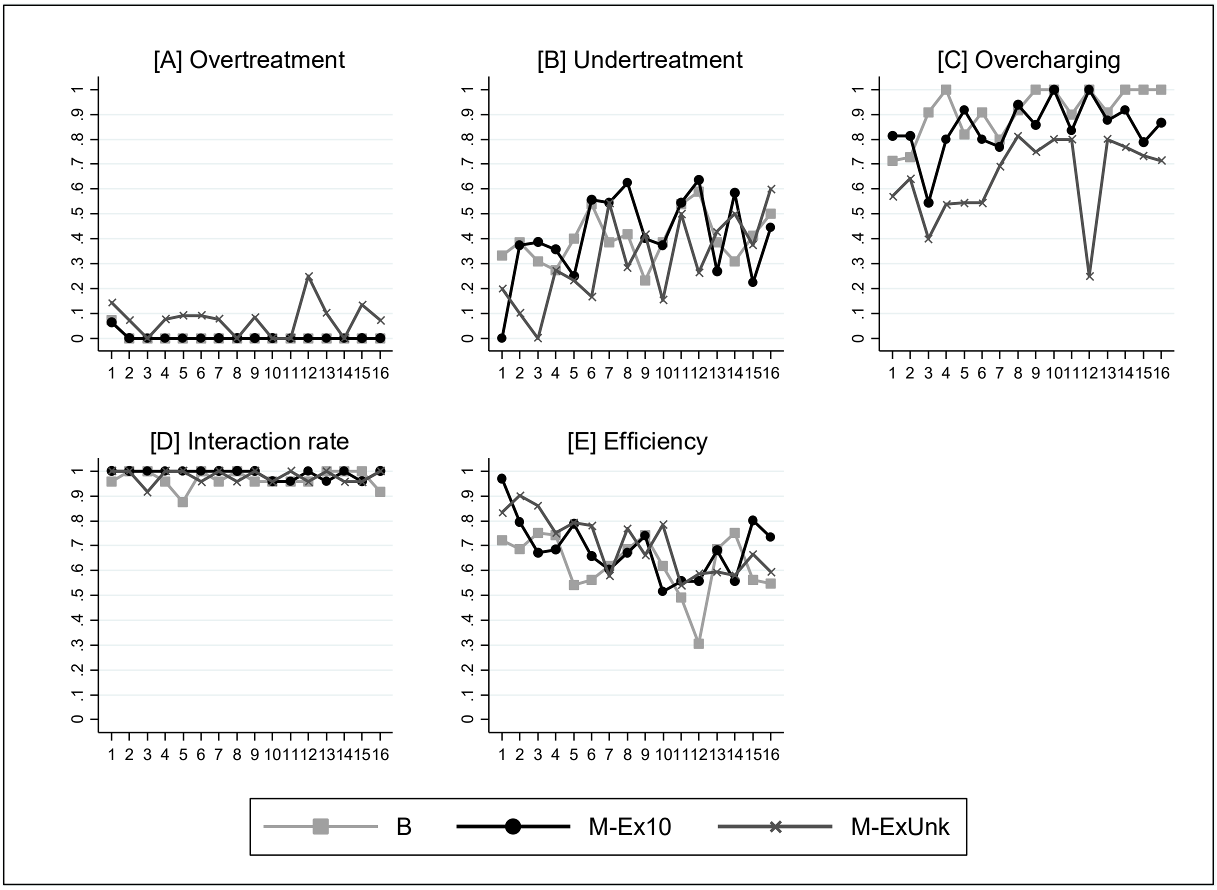


3.4. Information on monitoring probability

The effect of omitting the information concerning the probability of monitoring physician behavior over all 16 periods is shown in Table 3, column 6 and across periods in Figure 3. Over all periods, we find a significant reduction in the frequency of overcharging compared to ***B***, and a marginal trend toward a significant reduction compared to ***M-Ex10***. There is no difference in the frequency of undertreatment or in the level of efficiency between ***M-ExUnk*** and ***B***. However, when testing the behavior of physicians in periods 1-8 and periods 9-16 separately, we find that physicians undertreat in 38.3% of cases on average in the first eight periods in ***B***, whereas undertreatment occurs in 22.1% of cases in ***M-ExUnk*** ($p < 0.1$, Mann-Whitney U-test (MWU)). In the last eight periods, there is no significant statistical difference in undertreatment between ***B*** (42.2%) and ***M-ExUnk*** (39%). In line with the results on overcharging over all 16 periods, we find a significant difference in the rate of overcharging for both the first and the last eight periods between ***B*** and ***M-ExUnk*** ($p <$ 0.05, MWU).

With regard to the development of the most important outcomes across periods, Figure 3 shows that the rate of undertreatment starts out at a low level (between 0 and 20%) and increases throughout the course of the game ($p < 0.05$, WSR for a comparison of means between the first and the second half of periods averaged over markets and periods). In addition, a marginal trend toward a significant increase in the rate of overcharging from the first to the second half of periods can be observed ($p <$ 0.1, WSR). For the rest of our key variables, there is no evidence of a time trend.

Figure 3 Provision, charging, market interaction and efficiency by conditions ***B, M-Ex10*** and ***M-ExUnk***



Result 5. Omitting the information on the probability of monitoring decreases the frequency of overcharging compared to ***M-Ex10***. In comparison to ***B***, the rate of undertreatment in periods 1-8 and the rate of overcharging in all periods is significantly lower.

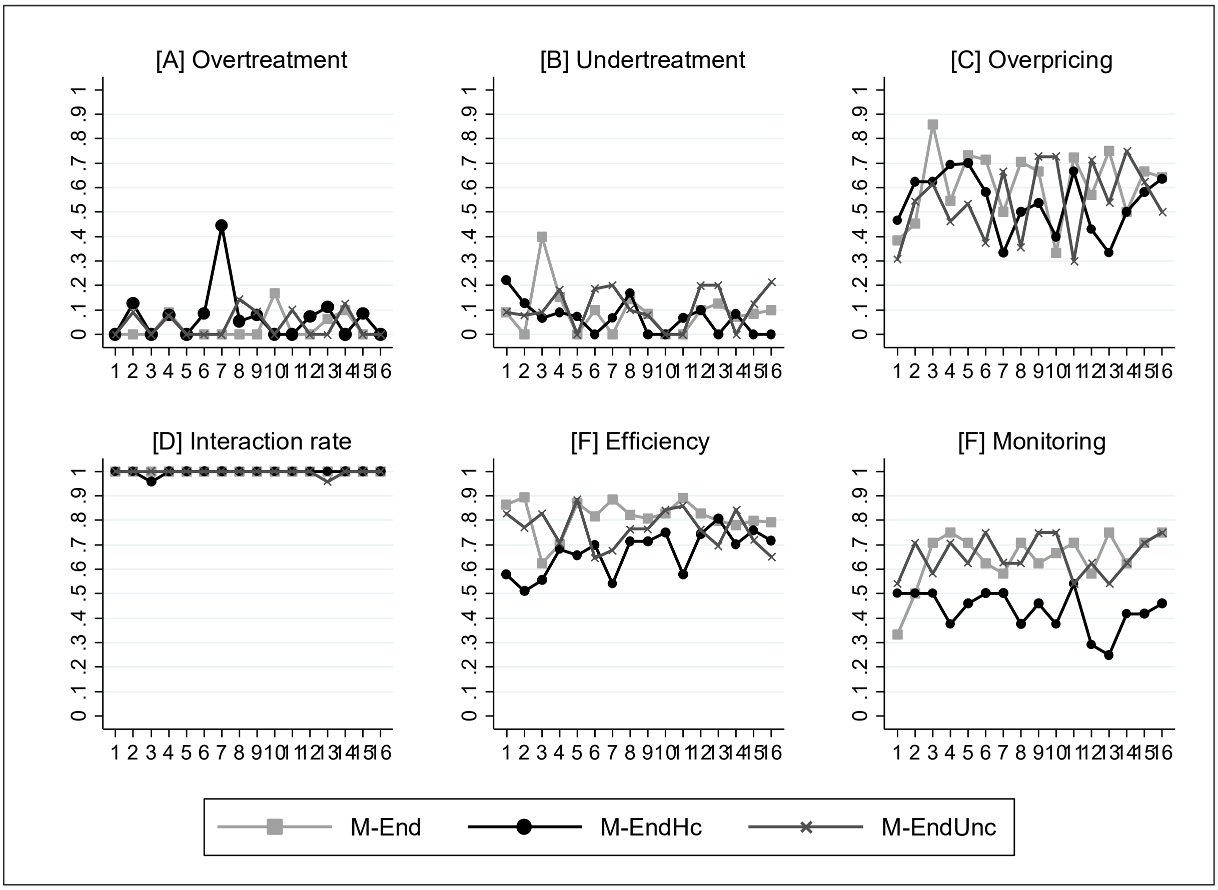
3.5. Monitoring costs and uncertainty of verifiability

The effects of an increase in the monitoring costs from 1 to 3 and the introduction of uncertainty in the verification of physician misbehavior compared to ***M-End*** are shown in columns 7 and 8 of Table 3 and in Figure 4. Overall, we find that the level of efficiency is significantly lower in ***M-EndHc*** compared to ***M-End*** due to the higher monitoring costs. No differences are found for the other key variables.

In analyzing the development of the most important key variables across periods, we do find a significant increase in the level of overcharging for ***M-EndUnc*** ($p < 0.05$, WSR for a comparison of means between the first and the second half of periods averaged over markets and periods). In addition, a marginal trend toward a significant increase in the level of efficiency from the first to the second half of the periods can be observed for ***M-EndHc*** ($p < 0.1$, WSR). For the rest of our key variables, there is no evidence of a time trend.

Result 6. Introducing uncertainty in the verifiability of physician behavior and increasing the monitoring costs have no impact on the frequency of undertreatment or overcharging compared to ***M-End***.

Figure 4 Provision, charging, market interaction and efficiency by conditions ***M-End, M-EndHc*** and ***M-EndUnc***



3.6. Regression analysis

In Tables 4 and 5, we report the results from multilevel mixed-effects probit and linear regressions in which we estimate the effects of our conditions on the relative frequency of overcharging, undertreatment, and efficiency.[15] We use condition ***B*** as the reference group. The regression results confirm our non-parametric analysis.

In line with results 2 to 6, Table 4 shows that the frequencies of overcharging and undertreatment significantly decline in our endogenous and exogenous monitoring conditions (with at least 35% monitoring frequency) relative to condition ***B*** (models (1)-(3)). However, for conditions ***M-Ex10*** and ***M-ExUnk***, we do not find a significant reduction in overcharging or undertreatment compared to ***B***, with one exception. For condition ***M-ExUnk***, we observe a significant reduction in overcharging compared to ***B*** (models (1) and (3)). In order to investigate period-dependent effects for condition ***M-ExUnk***, we introduce an interaction term in models (2) and (3). Consistent with the non-parametric analysis, we do not find an interaction effect for overcharging, but a significant interaction for undertreatment. Post-estimation Wald tests show that the level of undertreatment in the early periods is significantly lower under ***M-ExUnk*** compared to ***B*** ($p < 0.05$ periods 1-6, $p < 0.1$ periods 7-8, Wald tests) and ***M-Ex10*** ($p < 0.05$ period 1, $p < 0.1$ periods 2-4, Wald tests). There is no significant difference for the later periods ($p > 0.1$, Wald tests).

In agreement with the results from Table 4, Table 5 shows that the level of efficiency significantly increases for conditions ***M-End***, ***M-End-NoInfo***, ***M-Ex65***, ***M-Ex35***, and ***M-EndUnc*** relative to condition ***B*** (by approximately 14 to 19 percentage points). Furthermore, when introducing the interaction between periods and ***M-ExUnk*** in model (2), we find that efficiency is significantly higher in the early periods compared to ***B*** ($p < 0.05$ periods 1-6, $p < 0.1$ period 7, Wald tests) and ***M-Ex10*** ($p < 0.05$ Periods 1-2, $p < 0.1$, Period 3-4, Wald tests).

In addition to these differences between conditions, our regression results reveal certain interesting relationships between additional experimental and personality measures in model (3). We find that risk-averse subjects with a stronger preference for a guaranteed amount of money in a choice-list task (played after the credence goods game) are less likely to engage in overcharging ($p < 0.1$) and undertreatment ($p < 0.05$), which has a positive impact on efficiency. Subjects who score high on the conscientiousness scale of our 10-item BIG 5 questionnaire are less likely to undertreat patients ($p < 0.05$) and thus also have a positive effect on efficiency ($p < 0.05$). Finally, we find that subjects who are classified as liars in our lying task are significantly more likely to overcharge patients in the credence goods game ($p < 0.01$).

[15] Multilevel mixed-effects models are specifically designed to account for dependencies between observations on different hierarchical levels. In our case, we use a three-level mixed-effects model to account for the dependency of observations at the subject and market level.

Table 4 Multilevel mixed-effects probit on overcharging and undertreatment

VARIABLES	(1) OC	(1) UT	(2) OC	(2) UT	(3) OC	(3) UT
M-End (=1)	-1.492***	-1.863***	-1.525***	-1.871***	-1.371***	-1.754***
	(0.281)	(0.393)	(0.282)	(0.395)	(0.269)	(0.372)
M-EndHc (=1)	-1.757***	-2.051***	-1.790***	-2.058***	-1.724***	-1.858***
	(0.381)	(0.386)	(0.383)	(0.387)	(0.341)	(0.403)
M-EndUnc (=1)	-1.829***	-1.618***	-1.855***	-1.625***	-1.831***	-1.505***
	(0.341)	(0.306)	(0.342)	(0.306)	(0.304)	(0.385)
M-Ex65 (=1)	-1.939***	-1.401***	-1.966***	-1.405***	-1.921***	-1.218***
	(0.337)	(0.366)	(0.331)	(0.368)	(0.322)	(0.444)
M-Ex35(=1)	-0.881**	-1.257***	-0.909**	-1.263***	-0.829**	-1.305***
	(0.408)	(0.364)	(0.409)	(0.366)	(0.368)	(0.384)
M-Ex10 (=1)	-0.362	-0.093	-0.377	-0.094	-0.170	0.255
	(0.468)	(0.415)	(0.468)	(0.418)	(0.384)	(0.400)
M-ExUnk (=1)	-1.168***	-0.481	-1.469***	-1.300***	-1.367***	-1.073*
	(0.342)	(0.309)	(0.352)	(0.474)	(0.347)	(0.567)
M-End-NoInfo (=1)	-2.208***	-1.712***	-2.248***	-1.718***	-2.089***	-1.366***
	(0.364)	(0.317)	(0.363)	(0.319)	(0.320)	(0.373)
Period			0.027***	0.003	0.026***	0.003
			(0.010)	(0.013)	(0.010)	(0.014)
M-ExUnk×Period			0.036	0.091**	0.039	0.090**
			(0.028)	(0.039)	(0.029)	(0.039)
DG-Keep[1]					0.021	-0.011
					(0.036)	(0.040)
Risk aversion					-0.044*	-0.068**
					(0.026)	(0.032)
Liar					0.098***	0.030
					(0.038)	(0.039)
Conscientiousness					-0.075	-0.392**
					(0.090)	(0.165)
Add. Covariates[2]	No	No	No	No	Yes	Yes
Constant	1.856***	-0.287	1.663***	-0.315	1.965	2.237
	(0.274)	(0.226)	(0.289)	(0.265)	(1.208)	(1.790)
Observations	1,682	1,666	1,682	1,666	1,676	1,656
Number of groups	53	53	53	53	53	53

Notes:
Three-level model with two random effects equations at the market and individual level. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
UT: undertreatment
OC: overcharging
[1] The amount kept in a dictator game (DG) as a measure for altruism
[2] Additional Covariates: Trustworthiness measured in a trust game, BIG 5 personality traits (agreeableness, openness, neuroticism, extraversion and conscientiousness) measured with a 10-item BIG 5 questionnaire, gender, age, self-reported frequency of practicing religion, number of physician visits in the past 12 months, experience with incorrect physician behavior and relative school performance as a proxy for IQ.

Table 5 Multilevel mixed-effects linear regression on efficiency

VARIABLES	(1) EFF	(2) EFF	(3) EFF
M-End (=1)	0.187***	0.187***	0.158***
	(0.060)	(0.060)	(0.059)
M-EndHc (=1)	0.043	0.043	0.022
	(0.063)	(0.063)	(0.066)
M-EndUnc (=1)	0.139**	0.139**	0.127*
	(0.061)	(0.061)	(0.067)
M-Ex65 (=1)	0.138**	0.138**	0.110
	(0.060)	(0.060)	(0.071)
M-Ex35(=1)	0.172***	0.172***	0.186***
	(0.064)	(0.064)	(0.064)
M-Ex10 (=1)	0.060	0.060	0.027
	(0.074)	(0.074)	(0.074)
M-ExUnk (=1)	0.078	0.241***	0.226***
	(0.064)	(0.074)	(0.086)
M-End-NoInfo (=1)	0.121**	0.121**	0.068
	(0.058)	(0.058)	(0.068)
Period		-0.0003	-0.0003
		(0.002)	(0.002)
M-ExUnk×Period		-0.019***	-0.019***
		(0.007)	(0.007)
DG-Keep[1]			0.0004
			(0.005)
Risk aversion			0.007*
			(0.004)
Neuroticism			-0.038**
			(0.017)
Conscientiousness			0.046**
			(0.021)
Add. Covariates[2]	No	No	Yes
Constant	0.626***	0.629***	0.530**
	(0.055)	(0.056)	(0.238)
Observations	3,392	3,392	3,376
Number of id	53	53	53

Notes:
Three-level model with two random effects equations at the market and individual level. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
EFF: efficiency
[1] The amount kept in a dictator game (DG) as a measure for altruism
[2] Additional Covariates: Trustworthiness measured in a trust game, BIG 5 personality traits (agreeableness, openness, neuroticism, extraversion and conscientiousness) measured with a 10-item BIG 5 questionnaire, gender, age, self-reported frequency of practicing religion, experience with incorrect physician behavior and relative school performance as a proxy for IQ.

3.7. Dynamics

In the next step, we answer the question of whether our patient and physician subjects change their behavior based on their experiences in the previous round. To this end, we report in Table 5 the results of multilevel mixed-effects probit regressions in which we estimate the effect of previous undertreatment and overcharging choices as well as monitoring experiences on the likelihood of interaction, undertreatment, and overcharging. The results reported in model (1) answer the question of whether patients are more likely to exit the market based on the experience of undertreatment in the previous period. Indeed, we find that patient subjects who were undertreated in the previous period are more likely to opt-out of the market. The results in models (2) and (3) report the effects of monitoring experience in the previous period on undertreatment and overcharging. It appears that monitoring does not significantly alter the decision of a physician to undertreat or overcharge the patient in the following period. Moreover, the decision to undertreat (overcharge) the patient in the previous period is not significantly related to the decision to undertreat (overcharge) the patient in the current period. Finally, in models (4) and (5) we include interaction terms of undertreatment and overcharging in the previous period with monitoring in the previous period in order to investigate the effects of monitoring conditional on fraudulent behavior on overcharging and undertreatment in the following period. Once more, the results reveal that monitoring in the previous period does not have a significant effect on undertreatment (overcharging) in the current period, neither in the case of no undertreatment (no overcharging) nor in the case of undertreatment (overcharging) in the previous period (see the coefficient for MON and the interaction terms in columns 4 and 5, respectively). For an individual-level analysis of physician types across conditions see Appendix A.

Table 6 Multilevel mixed-effects probit on interaction, undertreatment and overcharging

VARIABLES	(1) INT	(2) UT	(3) OC	(4) UT	(5) OC
UT lag1	-0.387***	0.078		-0.017	
	(0.144)	(0.185)		(0.229)	
MON lag1		-0.007	-0.085	-0.074	-0.186
		(0.152)	(0.110)	(0.153)	(0.144)
UT lag1*MON lag1				0.264	
				(0.295)	
OC lag1			0.150		0.059
			(0.123)		(0.147)
OC lag1*MON lag1					0.189
					(0.173)
Constant	2.771***	-1.562***	0.372**	-1.536***	0.423***
	(0.187)	(0.190)	(0.151)	(0.192)	(0.158)
Observations	2,963	1,275	1,325	1,275	1,325
Number of groups	53	47	47	47	47

Notes:
Three-level model with two random effects equations at the market and individual level. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
UT: undertreatment
OC: overcharging
MON: monitoring
INT: interaction

4. Discussion

The present study provides evidence of the impact of monitoring on physician behavior in a health care market by using a framed laboratory experiment based on the credence goods framework of Dulleck and Kerschbamer (2006) and the experimental design of Dulleck *et al.* (2011). Similar to

Dulleck *et al.* (2011), we find that in the condition without the institutions of verifiability or liability, market outcomes are better than expected. The proportion of physicians providing the appropriate treatment (about 60%) is far greater than the theoretical prediction and also larger than in Dulleck *et al.* (2011). One difference that might explain the differences in outcomes is the health care framing in our experiment versus a neutral framing in Dulleck *et al.* (2011) (Kairies-Schwarz *et al.*, 2017; Kesternich *et al.*, 2015). An avenue for future research is to investigate the effect of a health framing in a credence goods setting. When the option of monitoring is introduced endogenously and exogenously with a 65% probability, levels of undertreatment and overcharging are significantly reduced. Unlike the theoretical predictions, however, we do not observe a difference between endogenous monitoring and exogenous monitoring with a 65% monitoring probability in the level of overcharging, or a difference between the two exogenous monitoring conditions with 65% and 35% monitoring probability in the level of undertreatment. This deviation from the theoretical prediction could be consistent with risk aversion. Finally, in line with the theoretical prediction, we find no difference in physician behavior compared to the baseline condition without monitoring when the probability of monitoring is reduced to 10%.

In the three main conditions with exogenous monitoring, the likelihood of monitoring was common knowledge for physicians and patients. The fact that both sides of the market were informed about the likelihood of monitoring ensures that our treatment effects were exclusively due to the change in our monitoring scheme and did not result from different beliefs. However, in the real world, both parties rarely know how likely it is that monitoring will occur. In order to track individuals' beliefs, we implement the ***M-ExUnk*** condition in which we elicit such beliefs. The median belief regarding the monitoring probability before the first period was played was approximately 50% (mean: 43.48, sd: 18.33). Over time, individuals can gather information from previous interactions and update their beliefs. In the final period, the median expectation concerning the monitoring probability was 11% (mean: 21.80%, sd: 24.16). Table 4 shows that physicians react to their updated beliefs by undertreating significantly more often in later periods. This observation supports our findings from the main conditions and hence highlights the robustness of our results.

The decision to enforce physician liability in the form of malpractice lawsuits can be very costly to patients, both in terms of court fees and the probability of winning the case. The results from our conditions with increased monitoring costs and uncertainty in the effectiveness of verifying physician behavior, however, show that these variations have no influence on the level of undertreatment or overcharging compared to our endogenous monitoring condition. These findings suggest that the threat of patient monitoring is sufficient to reduce undertreatment and overcharging vis-à-vis our baseline ***B***.

As a laboratory experiment, our study provides qualitative insights. In particular, the gains and losses in efficiency between the different conditions naturally depend upon the chosen parameters. Closely connected to the parameter choice is the question of external validity. We seek to investigate institutional changes in the health care market without putting anyone's health at risk. The advantages of this approach include not only the unambiguous identification of treatment effects but also the safety of patients in the real market. Nevertheless, individuals may make different decisions if there are monetary rather than health-related consequences (for a discussion, see Galizzi and Wiesen, 2018). Although the influence of this effect is still an open question, Brosig-Koch *et al.* (2016) show that non-medical participants – such as the students we recruited in our experiment – qualitatively react just as physicians do. Hence, our qualitative insights are most likely to hold also in external health care markets, although the quantitative effects must be interpreted with caution.

One important factor when designing interventions involving monetary incentives in the real world is the fact that the incentivization could lead to two effects in opposite directions: (i) a direct price effect resulting from the monetary incentives that induces the desired behavioral reaction, and (ii) a psychological effect that crowds out intrinsic motivation and therefore leads to undesired behavioral responses (Gneezy *et al.*, 2011; Kranton, 2019). Fehr and List (2004), for instance, show in a laboratory experiment that the introduction of punishment in a trust game can have a negative effect by crowding out the social norm of trust. They introduce the option of punishment by the sender in the trust game and determine that if the punishment option is used by the sender, the amount paid back by the receiver is lower compared to the situation in which the punishment option is not used. Often, the negative effect of crowding out manifests itself when monetary incentives are removed or when the price effect is lower than the psychological effect (Gneezy and Rustichini, 2000a; Gneezy and Rustichini, 2000b; Meier, 2007). In our case, the introduction of monitoring institutions led to the desired behavioral response, indicating that the price effect exceeds the psychological effect. In addition, since monitoring is not a temporary intervention but rather a permanent option to enforce physician liability, negative effects resulting from the removal of the incentives are not relevant. Nevertheless, a negative effect of monitoring on specific dimensions in the provision of health care that are unmonitored cannot be ruled out (Belot and Schröder, 2016; Villeval, 2018).

5. Conclusion

Although the monitoring of physicians' treatment decisions has become increasingly common over the past two decades, there is a lack of experimental studies that systematically investigate its effectiveness in a controlled manner. We capture these recent developments and investigate whether the monitoring of physicians can help improve market outcomes in health care employing a laboratory experiment.

In line with our predictions, we find that monitoring is an effective institution for reducing misbehavior. Compared to the baseline condition, the levels of undertreatment and overcharging are significantly reduced when monitoring is sufficiently likely. There is almost no overtreatment in any of the conditions: Overtreatment is a strictly dominated strategy, since overcharging yields a higher payoff to the physician while not affecting the payoff of the patient. Moreover, efficiency (measured as the sum of patient and physician payoffs) increases significantly for both endogenous and exogenous monitoring when monitoring is sufficiently likely compared to no monitoring. A monitoring frequency of 10% with common knowledge of the monitoring probability, however, has no impact on physician behavior compared to the baseline condition. Relaxing the assumption of perfect verifiability or increasing the costs of monitoring in the endogenous monitoring condition has no impact on the frequency of undertreatment or overcharging.

Our results thus suggest that monitoring institutions may be effective in reducing physician misbehavior. The threat of being monitored diminishes physicians' incentives to mistreat. In order to guide policymakers, however, future research should investigate the potential crowding-out effects of intrinsic motivation as well as negative spillover effects that monitoring institutions might trigger in the provision of health services.

References

Akerlof, G. A. (1970). 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics*, vol. **84**(**3**), pp. 488-500.

Balafoutas, L., A. Beck, R. Kerschbamer and M. Sutter (2013). 'What Drives Taxi Drivers? A Field Experiment on Fraud in a Market for Credence Goods', *The Review of Economic Studies*, vol. **80**(**3**), pp. 876-891.

Balafoutas, L., R. Kerschbamer and M. Sutter (2017). 'Second-Degree Moral Hazard in a Real-World Credence Goods Market', *Economic Journal*, vol. **127**(**599**), pp. 1-18.

Belot, M. and M. Schröder (2016). 'The Spillover Effects of Monitoring: A Field Experiment', *Management Science*, vol. **62**(**1**), pp. 37-45.

Blum, K., S. Löffert, M. Offermanns and P. Steffen (2017). *Krankenhausbarometer: Umfrage 2017* Düsseldorf: Deutsches Krankenhausinstitut e. V.

Blum, K., S. Löffert, M. Offermanns and P. Steffen (2018). *Krankenhausbarometer: Umfrage 2018* Düsseldorf: Deutsches Krankenhausinstitut e. V.

Blum, K., S. Löffert, M. Offermanns and P. Steffen (2019). *Krankenhausbarometer: Umfrage 2019* Düsseldorf: Deutsches Krankenhausinstitut e. V.

Bock, O., I. Baetge and A. Nicklisch (2014). 'Hroot: Hamburg Registration and Organization Online Tool', *European Economic Review*, vol. **71**, pp. 117-120.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz and D. Wiesen (2016). 'Using Artefactual Field and Lab Experiments to Investigate How Fee-for-Service and Capitation Affect Medical Service Provision', *Journal of Economic Behavior & Organization*, vol. **131**, pp. 17-23.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies-Schwarz and D. Wiesen (2017). 'The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence', *Health Economics*, vol. **26**(**2**), pp. 243-262.

Brosig-Koch, J., H. Hennig-Schmidt, N. Kairies and D. Wiesen (2013). 'How Effective Are Pay-for-Performance Incentives for Physicians?-a Laboratory Experiment'. Available at SSRN: https://ssrn.com/abstract=2278863 or http://dx.doi.org/10.2139/ssrn.2278863.

Brownlee, S. (2010). *Overtreated: Why Too Much Medicine Is Making Us Sicker and Poorer*: Bloomsbury Publishing USA.

Clemens, J. and J. D. Gottlieb (2014). 'Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?', *American Economic Review*, vol. **104**(**4**), pp. 1320-1349.

Currie, J., W. Lin and J. Meng (2014). 'Addressing Antibiotic Abuse in China: An Experimental Audit Study', *Journal of Development Economics*, vol. **110**, pp. 39-51.

Currie, J., W. Lin and W. Zhang (2011). 'Patient Knowledge and Antibiotic Abuse: Evidence from an Audit Study in China', *Journal of Health Economics*, vol. **30**(**5**), pp. 933-949.

Dafny, L. (2005). 'How Do Hospitals Respond to Price Changes?', *American Economic Review*, vol. **95**, pp. 1525-1547.

Darby, M. R. and E. Karni (1973). 'Free Competition and the Optimal Amount of Fraud', *Journal of Law & Economics*, vol. **16**(**1**), pp. 67-88.

Das, J. and J. Hammer (2007). 'Money for Nothing: The Dire Straits of Medical Practice in Delhi, India', *Journal of Development Economics*, vol. **83**(**1**), pp. 1-36.

Das, J., A. Holla, A. Mohpal and K. Muralidharan (2016). 'Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India', *American Economic Review*, vol. **106**(**12**), pp. 3765-3799.

Duflo, E., R. Hanna and S. P. Ryan (2012). 'Incentives Work: Getting Teachers to Come to School', *American Economic Review*, vol. **102**(**4**), pp. 1241-1278.

Dulleck, U. and R. Kerschbamer (2006). 'On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods', *Journal of Economic Literature*, vol. **44**(**1**), pp. 5-42.

Dulleck, U., R. Kerschbamer and M. Sutter (2011). 'The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition', *American Economic Review*, vol. **101**(**2**), pp. 526-555.

Fehr, E. and S. Gächter (2000). 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review*, vol. **90**(**4**), pp. 980-994.

Fehr, E. and J. A. List (2004). 'The Hidden Costs and Returns of Incentives-Trust and Trustworthiness among Ceos', *Journal of the European Economic Association*, vol. **2**(**5**), pp. 743-771.
Fischbacher, U. (2007). 'Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments', *Experimental Economics*, vol. **10**(**2**), pp. 171-178.
Furrow, B. R. (2011). 'Regulating Patient Safety: The Patient Protection and Affordable Care Act', *University of Pennsylvania Law Review*, vol. **159**, pp. 101-149.
Galizzi, M. M. and D. Wiesen (2018). 'Behavioral Experiments in Health Economics', in *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.
Gneezy, U., S. Meier and P. Rey-Biel (2011). 'When and Why Incentives (Don't) Work to Modify Behavior', *Journal of Economic Perspectives*, vol. **25**(**4**), pp. 191-210.
Gneezy, U. and A. Rustichini (2000a). 'A Fine Is a Prize', *Journal of Legal Studies*, vol. **29**(**1**), pp. 1-18.
Gneezy, U. and A. Rustichini (2000b). 'Pay Enough or Don't Pay at All', *Quarterly Journal of Economics*, vol. **115**(**3**), pp. 791-810.
Gottschalk, F., W. Mimra and C. Waibel (2019). 'Health Services as Credence Goods: A Field Experiment'. Available at SSRN: https://ssrn.com/abstract=3036573 or http://dx.doi.org/10.2139/ssrn.3036573
Green, E. P. (2014). 'Payment Systems in the Healthcare Industry: An Experimental Study of Physician Incentives', *Journal of Economic Behavior & Organization*, vol. **106**, pp. 367-378.
Greiner, B., L. Zhang and C. Tang (2017). 'Separation of Prescription and Treatment in Health Care Markets: A Laboratory Experiment', *Health Economics*, vol. **26 Suppl 3**, pp. 21-35.
Grosse, S., L. Putterman and B. Rockenbach (2011). 'Monitoring in Teams: Using Laboratory Experiments to Study a Theory of the Firm', *Journal of the European Economic Association*, vol. **9**(**4**), pp. 785-816.
Hennig-Schmidt, H., H. Jürges and D. Wiesen (2019). 'Dishonesty in Health Care Practice: A Behavioral Experiment on Upcoding in Neonatology', *Health Economics*, vol. **28**(**3**), pp. 319-338.
Hennig-Schmidt, H., R. Selten and D. Wiesen (2011). 'How Payment Systems Affect Physicians' Provision Behaviour--an Experimental Investigation', *Journal of Health Economics*, vol. **30**(**4**), pp. 637-646.
Hilger, N. G. (2016). 'Why Don’t People Trust Experts?', *Journal of Law and Economics*, vol. **59**(**2**), pp. 293-311.
Hölmstrom, B. (1979). 'Moral Hazard and Observability', *The Bell Journal of Economics*, vol. **10**(**1**), pp. 74-91.
Huck, S., G. Lünser, F. Spitzer and J.-R. Tyran (2016). 'Medical Insurance and Free Choice of Physician Shape Patient Overtreatment: A Laboratory Experiment', *Journal of Economic Behavior & Organization*, vol. **131**, pp. 78-105.
Jürges, H. and J. Köberlein (2015). 'What Explains Drg Upcoding in Neonatology? The Roles of Financial Incentives and Infant Health', *Journal of Health Economics*, vol. **43**, pp. 13-26.
Kairies-Schwarz, N., J. Kokot, M. Vomhof and J. Weßling (2017). 'Health Insurance Choice and Risk Preferences under Cumulative Prospect Theory – an Experiment', *Journal of Economic Behavior & Organization*, vol. **137**, pp. 374-397.
Kairies, N. and M. Krieger (2013). 'How Do Non-Monetary Performance Incentives for Physicians Affect the Quality of Medical Care? – a Laboratory Experiment', *Ruhr Economic Papers*.
Kerschbamer, R., D. Neururer and M. Sutter (2016). 'Insurance Coverage of Customers Induces Dishonesty of Sellers in Markets for Credence Goods', *Proceedings of the National Academy of Sciences*, vol. **113**(**27**), pp. 7454-7458.
Kerschbamer, R., D. Neururer and M. Sutter (2019). 'Credence Goods Markets and the Informational Value of New Media: A Natural Field Experiment', *CESifo Working Paper No. 7932*.
Keser, C., E. Peterle and C. Schnitzler (2014). 'Money Talks: Paying Physicians for Performance', *Center for European, Governance and Economic Development Research Discussion Papers*.
Kesternich, I., H. Schumacher and J. Winter (2015). 'Professional Norms and Physician Behavior: Homo Oeconomicus or Homo Hippocraticus?', *Journal of Public Economics*, vol. **131**, pp. 1-11.
Kranton, R. (2019). 'The Devil Is in the Details: Implications of Samuel Bowles's the Moral Economy for Economics and Policy Research', *Journal of Economic Literature*, vol. **57**(**1**), pp. 147-160.

Lindeboom, M., B. van der Klaauw and S. Vriend (2016). 'Audit Rates and Compliance: A Field Experiment in Care Provision', *Journal of Economic Behavior & Organization*, vol. **131**, pp. 160-173.
Liu, M., E. Brynjolfsson and J. Dowlatabadi (2019). 'Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis'.
Masclet, D., C. Noussair, S. Tucker and M.-C. Villeval (2003). 'Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism', *American Economic Review*, vol. **93**(**1**), pp. 366-380.
Medizinischer Dienst der Krankenversicherung (2019). Retrieved 12/4/2019, from https://www.mdk.de/.
Meier, S. (2007). 'Do Subsidies Increase Charitable Giving in the Long Run? Matching Donations in a Field Experiment', *Journal of the European Economic Association*, vol. **5**(**6**), pp. 1203-1222.
Mimra, W., A. Rasch and C. Waibel (2016a). 'Price Competition and Reputation in Credence Goods Markets: Experimental Evidence', *Games and Economic Behavior*, vol. **100**, pp. 337-352.
Mimra, W., A. Rasch and C. Waibel (2016b). 'Second Opinions in Markets for Expert Services: Experimental Evidence', *Journal of Economic Behavior & Organization*, vol. **131**, pp. 106-125.
Page, T., L. Putterman and B. Unel (2005). 'Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency', *Economic Journal*, vol. **115**(**506**), pp. 1032-1053.
Pasero, C. and M. McCaffery (2001). 'Pain Control: The Undertreatment of Pain', *The American Journal of Nursing*, vol. **101**(**11**), pp. 62-65.
Rajgopal, S. and R. White (2019). 'Cheating When in the Hole: The Case of New York City Taxis', *Accounting, Organizations and Society*, vol. **79**, pp. 101070.
Rasch, A. and C. Waibel (2018). 'What Drives Fraud in a Credence Goods Market? – Evidence from a Field Study', *Oxford Bulletin of Economics and Statistics*, vol. **80**(**3**), pp. 605-624.
Reif, S., S. Wichert and A. Wuppermann (2018). 'Is It Good to Be Too Light? Birth Weight Thresholds in Hospital Reimbursement Systems', *Journal of Health Economics*, vol. **59**, pp. 1-25.
Sefton, M., R. Shupp and J. M. Walker (2007). 'The Effect of Rewards and Sanctions in Provision of Public Goods', *Economic Inquiry*, vol. **45**(**4**), pp. 671-690.
Silverman, E. and J. Skinner (2004). 'Medicare Upcoding and Hospital Ownership', *Journal of Health Economics*, vol. **23**(**2**), pp. 369-389.
Slemrod, J., M. Blumenthal and C. Christian (2001). 'Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota', *Journal of Public Economics*, vol. **79**(**3**), pp. 455-483.
Villeval, M. C. (2018). 'The Spillover Effect of Monitoring Institutions on Unethical Behavior across Contexts', *Séminaire Economics and Psychology*, Université Paris 1, Paris, France.
Waibel, C. and D. Wiesen (2019). 'An Experiment on Referrals in Health Care'. Available at SSRN: https://ssrn.com/abstract=2864336 or http://dx.doi.org/10.2139/ssrn.2864336

Appendix A

Additionl Analysis: Distribution of types across conditions

Corresponding to our observation on the stable experimental behavior of our subjects in conjunction with strong treatment effects, we expect that our physician subjects will react instantaneously, from the first period onwards, to the assigned condition and thus that their behavior across the 16 periods can be classified into different types. We define seven mutually exclusive types[16] as follows: The "honest" type never undertreats or overcharges a patient across all 16 periods. The "overcharge low" type overcharges patients in fewer than 50% of all cases and never undertreats. Similarly, "overcharge high" types overcharge patients in at least 50% of all cases and never undertreat. A "mix low" type undertreats and overcharges the patient in fewer than 50% of all cases, and a "mix high" type does so at least 50% of the time.[17] Finally, whereas the two "overcharge" types only engage in misbehavior in rounds in which they have an opportunity to overcharge (i.e., when the patient suffers from the minor health problem), "dishonest" types always engage in fraudulent behavior, overcharging and undertreating patients in all rounds regardless of the severity of the health problem.[18] Figure A5 shows the distribution of types across all conditions and reveals certain market condition differences. "Honest" types (8%, averaging across all conditions), "overcharge" types (10% across all conditions) and "overcharge low" types (10% across all conditions) do not seem to react much to the condition, but we find highly significant differences for the prevalence of "overcharge high" participants – the most frequent type, with 24% across all conditions. There are significantly more "overcharge high" types in ***M-End*** conditions than in the other conditions. Thus, the participants are highly strategic. In the ***M-End*** conditions, where monitoring is endogenous, patients can directly react to their payoff information with a decision to monitor the physician. Thus, physicians can expect to be monitored with certainty if they decide to undertreat the patient, which leads to a negative payoff for the patient. Consequently, physicians opt to overcharge, which cannot be distinguished from honest behavior in terms of the final payoffs for patients. "Mix low" types (11% across all conditions) are most frequently observed in conditions ***M-Ex65, M-EndUnc*** and ***M-End-NoInfo***, whereas "mix high" types (the second most frequent type, with 12% across all conditions) are primarily found in conditions ***B***, ***M-Ex35***, ***M-Ex10***, and ***M-ExUnk***, where there is no monitoring or the detection probability is relatively low. Finally, the "dishonest" type (4% across all conditions) is only prevalent in conditions ***B*** and ***M-Ex10*** – again, where either no monitoring is possible or the detection probability is very low.

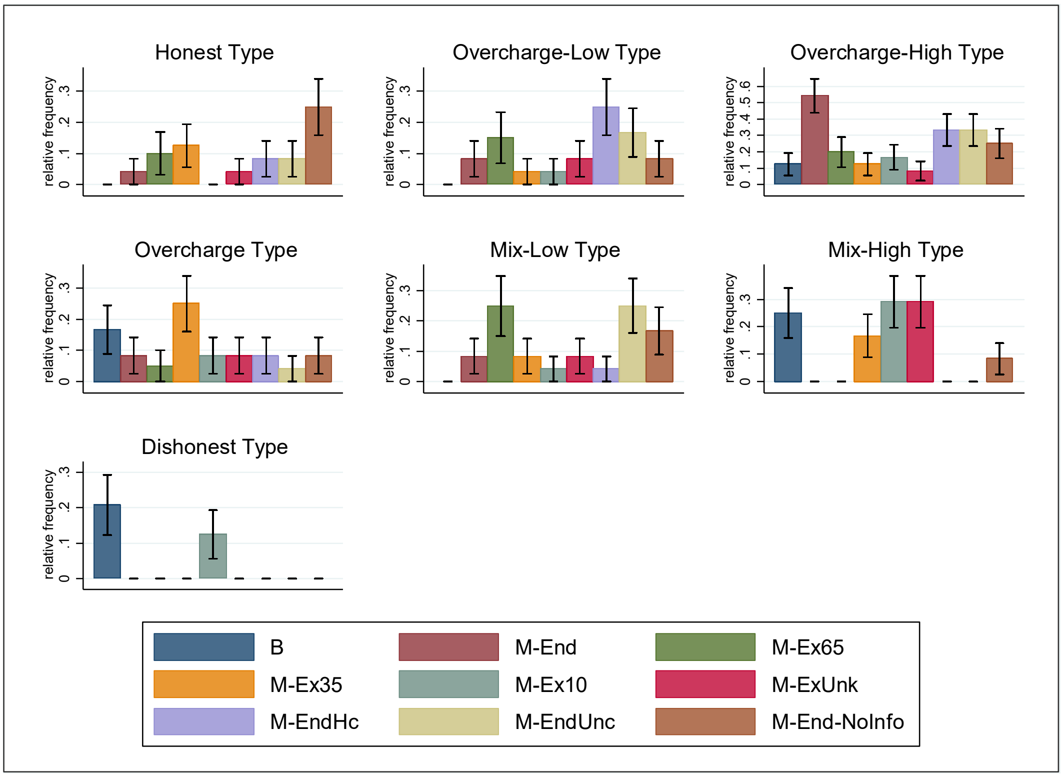
Figure A5 Distribution of types by all conditions

---

[16] 79% of all subjects can be classified into one of these seven types.

[17] The types "overcharge low" and "overcharge high" exclude the cases in which physicians never or always overcharge the patient. This rule also applies to the "mix low" and "mix high" types.

[18] We also initially defined an "undertreat" type, a type who always undertreats patients but never overcharges them, but there is no such observation in our dataset.


Honest Type
Overcharge-Low Type
Overcharge-High Type
Overcharge Type
Mix-Low Type
Mix-High Type
Dishonest Type
relative frequency
B
M-End
M-Ex65
M-Ex35
M-Ex10
M-ExUnk
M-EndHc
M-EndUnc
M-End-NoInfo

Appendix B

Experimental protocol

Overall, eighteen experimental sessions with a total of 424 undergraduate students were performed in the laboratory for experimental economic research at the University of Innsbruck. The experiment was computerized using zTree (Fischbacher, 2007), and students were recruited using hroot (Bock *et al.*, 2014).[19] For each experimental condition, two sessions were conducted, with a maximum of 24 subjects (3 matching groups) participating per session. To ensure the maximum number of participants for each session, we invited up to six subjects more than needed.[20] Once the session was full, the remaining subjects were each paid 4 Euros and dismissed. At the beginning of each session, the market set-up was explained by one of the experimenters following a standardized protocol. All subjects were given brief instructions that provided an overview of the decision problem. Afterward, they were asked to read a detailed description of the game and to answer a set of control questions that were incentivized to give each subject an initial endowment designed to offset any resulting negative period profits (see the instructions and control questions below). Once all control questions were answered correctly, the randomly assigned roles were revealed, and the subjects played the game for 16 periods. In addition to the credence goods game, subjects participated in an individual risk preferences task, a dictator game, a lying task, and a trust game; finally, they were asked to fill out a post-experimental questionnaire (see Appendix C for the additional experimental instructions and the questionnaire). At the end of the session, subjects received the payoffs from the credence goods game, one randomly selected additional task, and a lump sum payment of 2 Euros for answering the post-experimental questionnaire. The average earnings were 24.78 Euros, and the sessions lasted on average 120 minutes.

[19] Students who had previously participated in a credence goods game were excluded from the pool of subjects as well as medical students.

[20] In one session (condition ***M-Ex65***), we only had 16 subjects (2 markets) due to a shortage of participants.

Short Experimental Instructions (Credence Goods Game, endogenous monitoring condition)

Problem definition

- 16 periods
- 2 roles: **physician** and **patient**
- random role assignment (stays the same over the entire 16 periods)
- patient is suffering from an **illness** in each round
- 2 types of illness: **minor** and **major** illness
- **illness** is **randomly determined** in **each new round**
- 2 types of treatments are available for the physician: **minor** and **major treatment**
- **ATTENTION: minor and major** treatment **cures minor** illness, **BUT only major** treatment **cures major** illness.
- each period consists of a **maximum of 4 decisions** (see description below)

Information patient

- At **no time** does the patient know if he suffers from a minor or a major illness in the respective period
- The only information the patient receives is ...
  - ... his payoff at the end of decision 3
  - ... his payoff at the end of decision 4
  - and if his illness is cured

Information physician

- **Physician learns** which **illness** the patient suffers from, when the patient decides to consult the physician
- In addition, the physician receives information about ...
  - ... his payoff at the end of decision 3
  - ... the decision 4 of the patient
  - ... his payoff at the end of decision 4

Decision 1 patient

- **Consult a physician? YES or NO**
  - **YES**: Proceed to **decision 2**
  - **NO**: **Period ends** here:
    - Points patient: -4
    - Points physician: 0

Decision 2 physician

- **Type of treatment: MINOR or MAJOR?**
  - **MINOR** treatment: costs (C) for physician: $C = 2$ points
  - **MAJOR** treatment: costs (C) for physician: $C = 6$ points

Decision 3 physician

- **Price of the treatmen (P) ?**
  - **PRICE** for **MINOR** treatment: $P = 3$ points
  - **PRICE** for **MAJOR** treatment: $P = 8$ points

Decision 4 patient

- **Monitoring physician? YES od NO**
  - **YES:**
    - Monitoring costs for patient: $M = 1$ point
    - Payoffs dependent on decision 2 and 3 see below **"treatment with monitoring"**
- **NO:** Payoffs of physician and patient see below **"treatment without monitoring"**

Payoffs patient

Treatment without monitoring (respectively payoff before decision 4)

$$\boldsymbol{N - P}$$

Treatment with monitoring

Scenario 1: Payoff to the patient **is not below** the payoff received by the patient, when the physician selects a minor treatment for a minor illness and charges the price for the minor treatment respectively selects a major treatment for a major illness and charges the price for the major treatment:

$$\boldsymbol{N - P - M}$$

Scenario 2: Payoff to the patient **is below** the payoff received by the patient, when the physician selects a minor treatment for a minor illness and charges the price for the minor treatment respectively selects a major treatment for a major illness and charges the price for the major treatment:

$$\boldsymbol{N - P - M + M + compensation}$$

Illness is cured: $N = 10$ points

Illness is not cured: $N = 0$ points

Payoffs physician

Treatment without monitoring (respectively payoff before decision 4)

$$\boldsymbol{P - C}$$

Treatment with monitoring

Scenario 1: Payoff to the patient **is not below** the payoff received by the patient, when the physician selects a minor treatment for a minor illness and charges the price for the minor treatment respectively selects a major treatment for a major illness and charges the price for the major treatment:

$$\boldsymbol{P - C}$$

Scenario 2: Payoff to the patient **is below** the payoff received by the patient, when the physician selects a minor treatment for a minor illness and charges the price for the minor treatment respectively selects a major treatment for a major illness and charges the price for the major treatment:

$$\boldsymbol{P - K - M - compensation}$$

Long Experimental Instructions (Credence Goods Game, endogenous monitoring condition)

Welcome to this experiment! Thank you for your participation. Please read the following instructions carefully. It is extremely important that you now remove all items, including other reading materials and writing utensils from the table, and switch off your mobile phone, as well as any other electronic devices. If you have a question, raise your hand and one of the experimenters will come to you to answer your question privately. Please do not talk to any other participant during the experiment.
In this experiment, you can earn money. Your earnings will depend on your decisions and the decisions of others. All your decisions and answers will be treated confidentially and evaluated anonymously. To ensure your anonymity, the decisions that you make during the experiment will only be assigned to your seat number.
Your payoff will be given in points during the experiment. At the beginning of the experiment, you will receive an **endowment of 11 points**. In addition, answering the control questions gives you another **5 points**. The conversion rate of points into euros is as follows:

1 point = 60 Euro-cents

(i.e. 5 points = 3 Euro).

2 roles and 16 periods
This experiment consists of **16 periods**, and each period has the same sequence of decisions. The sequence of decisions is explained in detail below.
There are 2 roles in the experiment: **patient** and **physician**. At the beginning of the experiment you will be randomly assigned to one of these roles and you will maintain that role for the entire experiment. On the first screen you will see what role you are assigned.
A patient always interacts with a physician. However, the pairs **change** after each period. That is, in each period, the computer randomly **reassembles the pairs**.

In your group, there are 4 players who take on the role of physicians and 4 players who take on the role of patients. If you are a patient, then the 4 physicians in your group are your potential interaction partners. If you are a physician, then your potential interaction partners are the 4 patients in your group. At no point will you know which player you are interacting with.

All participants receive the same information regarding the rules of the game, including the costs and payoffs for both players.

Overview of the decision situation
Every **patient** is suffering from an **illness** in each period. There are 2 types of illnesses, a **minor** and a **major** illness. The type of illness a patient suffers from is randomly determined **each new period**. The patient suffers a **50% chance** from a **minor illness** and with a **50% chance** from a **major illness**. Imagine a coin toss in each period – if "head" shows up, then the patient suffers from a minor illness, if "number" shows up, the patient suffers from a major illness. The **patient** will **not be informed at any time** whether he suffers from a minor or a major **illness** in a particular period. The physician learns what illness a patient suffers only when the patient decides to consult the physician. The physician may then freely choose from one of two treatment types (**minor** or **major** treatment). However, a **major illness** is **only cured** by a **major treatment**. A **minor illness** is **cured** by a **minor or** a **major treatment**. Subsequently, the **patient** decides whether he wants to **monitor the behavior of the physician**.

Overview of the decisions in a period
Each single period consists of a maximum four decisions, made in succession. Decisions 1 and 4 are made by the patient, and decisions 2 and 3 are made by the physician.

Course of decisions in a given period and presentation of their consequences
Decision 1
The patient decides whether he wants to consult a physician.
If so, the assigned physician treats the patient and sets a price (see below). However, the patient will **not** be able to observe which treatment the physician chooses.
If not, this period **ends**.

Decision 2
If the patient chooses "Yes" in decision 1, **each physician will learn** the **type of illness the patient has** *before* making his decision 2. Then the physician chooses a treatment. The **patient** will **not** be informed **at any time** which treatment the physician has chosen.
The treatments incur a cost for the physician.
The **minor treatment** costs the physician **2 points** (= experimental currency unit) and only cures a minor illness.
The **major treatment** costs the physician **6 points** (= experimental currency unit) and cures both minor and the major illness.
The treatment can be chosen **independently** of the type of illness.

Decision 3
The physician **charges** a **price** for the treatment from the patient. Two prices are available:
- The price for the **minor treatment** is **3 points**.
- The price for the **major treatment** is **8 points**.

The chosen price does **not need** to be equal to the price of the treatment chosen in decision 2; it may also be the price of the other treatment.

Decision 4
The patient and the physician will be informed prior to decision 4 about their respective payoffs so far (details on payoffs will follow). The patient then decides whether to monitor the physician's behavior.
If not, this period **ends**.
If so, the **decision to monitor** costs the patient **1 point**. This results in 2 possible scenarios:
- **Scenario 1**: The monitoring of the physician's behavior shows that the physician behaved in such a way that the **payoff** to the **patient is <u>not below</u>** the **payoff** received by the patient, **when** the **physician** selects a **minor treatment** for a **minor illness** and charges the **price** for the **minor treatment** respectively, or selects a **major treatment** for a **major illness** and charges the **price** for the **major treatment**:
- **Scenario 2**: The monitoring of the physician's behavior shows that the physician behaved in such a way that the **payoff** to the **patient is <u>below</u>** the **payoff** received by the patient, **when** the **physician** selects a **minor treatment** for a **minor illness** and charges the **price** for the **minor treatment** respectively, or selects a **major treatment** for a **major illness** and charges the **price** for the **major treatment**:

Payoffs
I) No interaction (Patient decides not to consult the physician)
If **the patient** ends the period in decision 1 (*decision* "**no**" of *the patient*), then he receives **-4 points** in this period, i.e. he makes a **loss** of 4 points. The **physician** receives **0 points** for this period.

Otherwise (*decision* "**yes**" of *the patient*) the payoffs are as follows:

II) Interaction without monitoring the physician's behavior (patient decides to consult the physician without monitoring the physician's behavior)

The physician will receive the **price** (in points) chosen in decision 3, **minus** the **cost** of the treatment chosen in decision 2.

For **the patient** the payoff depends on whether the chosen treatment (by the physician in decision 2) can cure his illness.

a) The treatment has cured the illness. **The patient** receives a benefit of **10 points minus** the **price** charged in decision 3.

b) The treatment did not cure the illness. **The patient** receives a benefit of **0 points minus** the **price** charged in decision 3.

III) Interaction with monitoring the physician's behavior (patient decides to consult the physician and monitor the physician's behavior)

You have to distinguish 2 possible scenarios:

- **Scenario 1** (*see description of scenario 1 above*): This will **not affect** the payoffs (calculated in II). The **monitoring** costs (1 point) will be paid by the **patient**.
- **Scenario 2** (*see description of scenario 2 above*): In this case, the **physician** must **compensate** the patient for the lost payoffs in monetary terms. In addition, the physician pays the **monitoring** costs (1 point).

Two examples for illustration:

Example 1a:

- *The patient chooses "Yes" in decision 1.*
- *The patient suffers from a major illness.*
- *The physician chooses a major treatment and charges the price for a major treatment.*
- *The patient does not monitor the physician's behavior.*

Payoff patient: $\underbrace{10}_{\substack{benefit\\treatment}} - \underbrace{8}_{\substack{price\ major\\treatment}} = 2$

Payoff physician: $\underbrace{8}_{\substack{price\ major\\treatment}} - \underbrace{6}_{\substack{costs\ major\\treatment}} = 2$

Example 1b:

- *The patient monitors the physician's behavior in the example above (1a).*

Compensation patient:

$$\underbrace{2}_{\substack{payoff\ patient\\if\ physician\ choos\\major\ treatment\ and\\charg\quad price\ for\\major\ treatment\\(10-8=\ )}} - \underbrace{2}_{\substack{benefit\ treatment\\minus\ price\ treatment\\(10-8=\ )}} = 0$$

Scenario 1 *applies. The payoff of the patient and of the physician are calculated as follows:*

Payoff patient:: $\underbrace{10}_{\substack{benefit\\treatment}} - \underbrace{8}_{\substack{price\ major\\treatment}} - \underbrace{1}_{monitoring\ costs} = 1$

Payoff physician: $\underbrace{8}_{\substack{price\ major\\ treatment}} - \underbrace{6}_{\substack{costs\ major\\ treatment}} = 2$

Example 2a:

- *The patient chooses "Yes" in decision 1.*
- *The patient suffers from a minor illness.*
- *The physician chooses a major treatment and charges the price for a major treatment.*
- *The patient does not monitor the physician's behavior.*

Payoff patient: $\underbrace{10}_{\substack{benefit\\ treatment}} - \underbrace{8}_{\substack{price\ major\\ treatment}} = 2$

Payoff physician: $\underbrace{8}_{\substack{price\ major\\ treatment}} - \underbrace{6}_{\substack{costs\ major\\ treatment}} = 2$

Example 2b:

- *The patient monitors the physician's behavior in the example above (2a).*

Compensation patient:

$$\underbrace{7}_{\substack{payoff\ patient\\ if\ physician\ chooses\\ minor\ treatment\ and\\ charges\ price\ for\\ minor\ treatment\\ (10-3=7)}} - \underbrace{2}_{\substack{benefit\ treatment\\ minus\ price\ treatment\\ (10-8=2)}} = 5$$

Scenario 2 *applies. The payoff of the patient and of the physician are calculated as follows:*

Payoff patient: $\underbrace{10}_{\substack{benefit\\ treatment}} - \underbrace{8}_{\substack{price\ major\\ treatment}} + \underbrace{5}_{compensation} = 7$

Payoff physician: $\underbrace{8}_{\substack{price\ major\\ treatment}} - \underbrace{6}_{\substack{costs\ major\\ treatment}} - \underbrace{5}_{compensation} - \underbrace{1}_{\substack{monitoring\\ costs}} = -4$

The patient and the physician will be informed at the end of each period about their respective payoffs in this period. In addition, the patient learns whether his illness has been cured and the physician learns whether his behavior has been monitored.

At the end of the experiment, four periods will be drawn randomly for payment. For the calculation of payoffs, the initial endowment and the profits or losses over the four payoff-relevant periods are added together. From the initial endowment you can partially pay for possible losses in individual periods. Note that in this experiment there is **always** an option to avoid losses with certainty. The total number of points will be paid out in cash according to the above-mentioned exchange rate. By participating in the experiment, you agree to these terms.

Experimental receipts can be found on your table. At the end of the experiment, please insert your payoff from the experiment (which you can see on your final screen) on the receipt as well as your first and last name in block letters and sign the receipt.

Control Questions (Credence Goods Game, endogenous monitoring condition)

It is important to make sure that all participants have fully understood the experiment. Should something have remained unclear, please ask the experimenter. You will receive 5 points (= 3 Euro) for answering the questions correctly. Please answer the following questions:

Question	Correct Answer
1. How many decisions does a patient maximally make per period?	2
2. How many decisions does a physician maximally make per period?	2
Assess whether the statements below are true or false.	
3. "The patient learns what illness he suffers from in a particular period."	F
4. "If the physician cures the patient's illness, then the total payoff of the patient in that period is exactly 10 points."	F
5. "Your initial endowment of 11 points is worth 6.60 euros."	T
Please calculate the payoffs (in points) for the patient and the physician in the following examples:	
6. The patient chooses "No" in decision 1.	Patient: -4 Physician: 0
7a. The patient chooses "Yes" in decision 1. The patient suffers from a minor illness. The physician chooses a minor treatment and charges the price for a minor treatment. The patient does not monitor the physician's behavior.	Patient: 7 Physician: 1
7b. Same situation as in 7a except that the patient monitors the physician's behavior.	Patient: 6 Physician: 1
8a. The patient chooses "Yes" in decision 1. The patient suffers from a minor illness. The physician chooses a minor treatment and charges the price for a major treatment. The patient does not monitor the physician's behavior.	Patient: 2 Physician: 6
8b. Same situation as in 8a except that the patient monitors the physician's behavior.	Patient: 7 Physician: 0
9a. The patient chooses "Yes" in decision 1. The patient suffers from a major illness. The physician chooses a minor treatment and charges the price for a major treatment. The patient does not monitor the physician's behavior.	Patient: -8 Physician: 6
9b. Same situation as in 9a except that the patient monitors the physician's behavior.	Patient: 2 Physician: -5

The experiment continues as soon as all participants have answered the questions correctly.

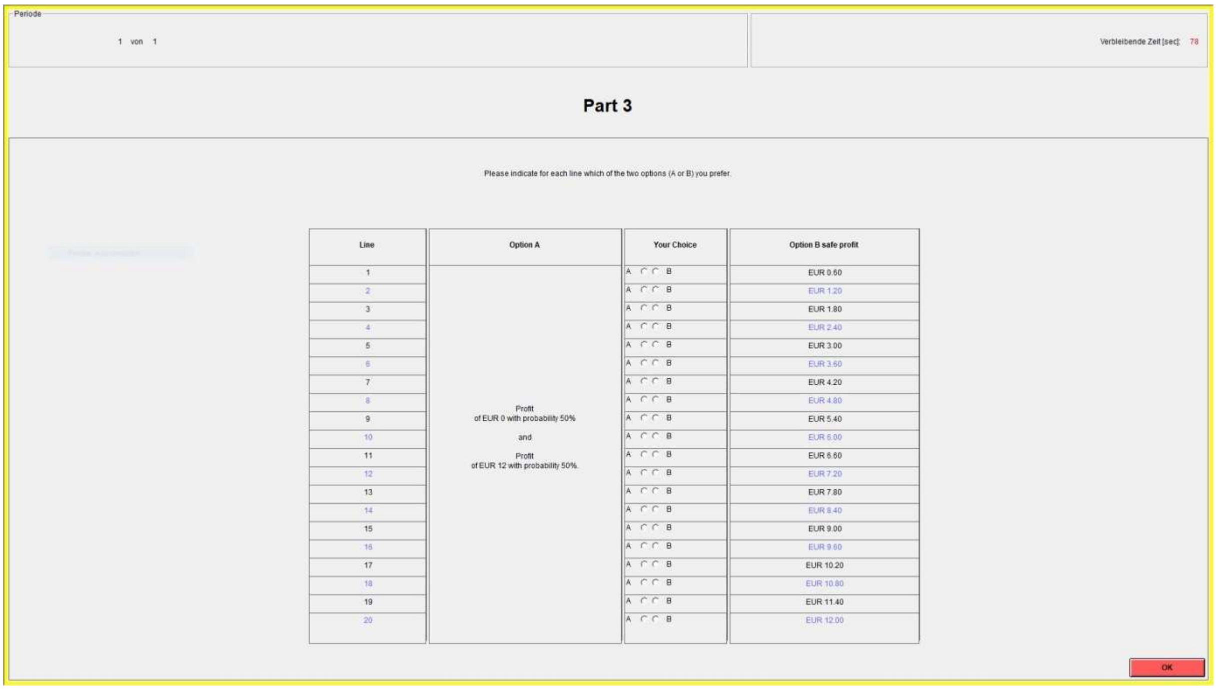
Appendix C

Experimental Instructions (part 2-5)

Part 2:
The experiment is not yet over. There are 4 more parts following. At the end of the experiment, one of these parts (part 2, part 3, part 4 or part 5) is randomly selected for payment.

In part 2, you have to make a decision regarding your payoff as well as the payoff of another person. This person is a patient who is supported by the organization "Licht für die Welt". The organization “Licht für die Welt” is known worldwide for preventing and curing preventable blindness. It enables **eye surgery** and **supplies people with eye glasses and medicines for eye diseases** in South America, Africa and Asia. You have an endowment of € 12 and you need to decide how you want to divide the money. There are two fields on your screen. One field is marked "amount for me" and the other field is marked "amount for Licht für die Welt". The amounts you enter always have to add up to € 12, in units of € 0.1 (i.e., 10 cents). The transfer will be made online at the end of the experiment. In order to be able to donate to the organization "Licht für die Welt" correctly, we kindly ask the participant with the ID 1 to confirm that the money has been transferred to the organization after the online transfer has been made. As a reminder, this part will only be paid if part 2 is randomly selected for payment at the end of the experiment. This also applies to the donation to "Licht für die Welt".

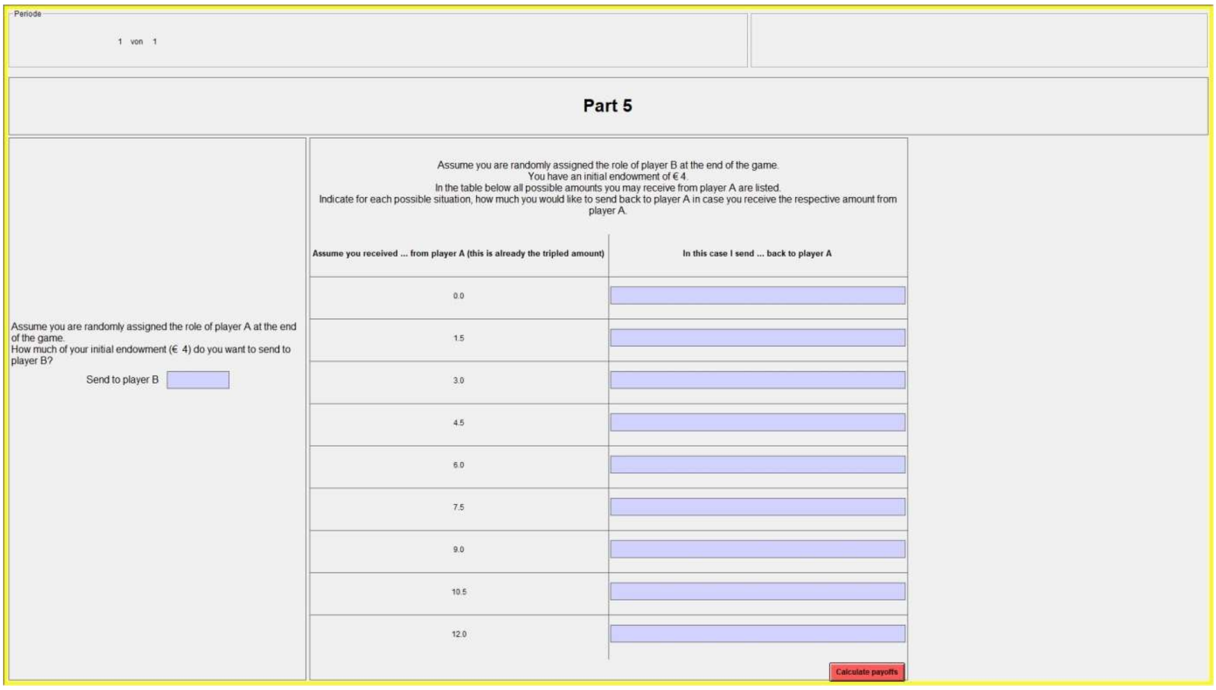
Part 3:
As a reminder, this part will only be paid if part 3 is randomly selected for payment at the end of the experiment. Part 3 consists of 20 decisions. Below, you are asked to make a decision for each situation. Each of your choices is a selection between "Option A" and "Option B".
"Option A" always offers an uncertain payoff: with 50% probability you will receive € 12, and a 50% probability you receive € 0.
"Option B" always offers a safe payoff: with 100% probability you receive an amount that varies from decision to decision (that is, you receive the guaranteed payoff of that row).
The decision situation will be presented to you on the screen as follows:



If part 3 happens to be paid out, one of the 20 decisions (lines) will be randomly selected for payment. Additionally, it will be randomly determined if you won the lottery (you receive € 12) or if you lost the lottery (you receive € 0) (if you have chosen the lottery option).
When you have made all decisions, please confirm with "OK".

Part 4:
As a reminder, this part will only be paid if part 4 is randomly selected for payment at the end of the experiment. Part 4 is about guessing the outcome of a die roll in a situation marked by randomness. You play 12 rounds of a dice guessing game. Thereby you should guess the number shown on the dice. The more outcomes you guess correctly, the more money you earn. Each round of the game works as follows:
1. First, guess what number will result from the die roll. If you have a number in your head, press the "Next" button.
2. Now you see a dice rolled randomly by the computer. Below the dice you have to enter what number you have guessed.
For each correctly guessed dice you receive 1 €. For each wrongly guessed die roll you receive 20 cents. The profits of all 12 rounds are added up at the end.

Part 5:
As a reminder, this part will only be paid if part 5 is randomly selected for payment at the end of the experiment. Part 5 works as follows: There are two roles, the role of player A and player B. Both players have an initial endowment of € 4 each. Player A has to decide how much of this endowment (between € 0 and € 4, in 50 cent increments) he wants to send to player B. The total amount sent to player B is tripled. The rest is kept by player A (without tripling). Player B may then decide how much of the tripled amount he wants to send back to player A. You have to decide in the role of player A (see left side of the decision situation on the screenshot below) as well as in the role of player B (for all possible situations, see the right side of the decision situation on the screenshot below). Only at the end of the game it will be randomly determined in which role you are in. In addition, you will be assigned to a partner playing the other role. You receive the payoff for your decisions in the role chosen for you at random, in combination with the behavior of your randomly assigned partner.



Periode

1 von 1

Part 5

Assume you are randomly assigned the role of player A at the end of the game.
How much of your initial endowment (€ 4) do you want to send to player B?

Send to player B

Assume you are randomly assigned the role of player B at the end of the game.
You have an initial endowment of € 4.
In the table below all possible amounts you may receive from player A are listed.
Indicate for each possible situation, how much you would like to send back to player A in case you receive the respective amount from player A.

Assume you received ... from player A (this is already the tripled amount)	In this case I send ... back to player A
0.0	
1.5	
3.0	
4.5	
6.0	
7.5	
9.0	
10.5	
12.0	

Calculate payoffs

Questionnaire

Belief elicitation for ***M-ExUnk***:

1) Before starting the experiment (i.e. before the first round of part 1), what was the probability that you expected the computer to monitor the physician's behavior in decision 4? Please enter a number between 0 and 100.
2) After you have completed all 16 rounds of part 1 of the experiment: What do you think, with what probability did the computer actually monitor the physician's behavior in decision 4? (You will be paid to answer this question. The closer your estimate is to the actual probability, the more money you will receive. If your estimate is +/- 5 percentage points of the actual value, you will receive an additional 1 euro if your estimate is + / - 10 percentage points of the actual value, then you will receive an additional 50 cents, otherwise you will not receive an additional payment. Please enter a number between 0 and 100.)

BIG Five Personality Test:
How well do the following statements describe your personality?

I see myself as someone who …	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
… is reserved	(1)	(2)	(3)	(4)	(5)
… is generally trusting	(1)	(2)	(3)	(4)	(5)
… tends to be lazy	(1)	(2)	(3)	(4)	(5)
… is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
… has few artistic interests	(1)	(2)	(3)	(4)	(5)
… is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
… tends to find fault with others	(1)	(2)	(3)	(4)	(5)
… does a thorough job	(1)	(2)	(3)	(4)	(5)
… gets nervous easily	(1)	(2)	(3)	(4)	(5)
… has an active imagination	(1)	(2)	(3)	(4)	(5)

Please indicate your gender: □ male □ female
How old are you?
Which field of study are you in? Which subject do you study? (If you are doing several studies, please indicate all and write the study program in parenthesis)
Semester
What was your average monthly net income over the last year, taking into account all sources of income such as scholarships, student loans, earned income, parental financial support, etcetera? Please round to the nearest ten Euro.
How often do you practice your religion? □ often □ rarely □ never
Please enter here your (average) grade from the Matura/Abitur certificate.
Which grading scale was used in your Matura/Abitur certificate? □ 1-5 (5 worst rating), □ 1-6 (6 worst rating), □ 1-10 (10 best rating), □ 0 -15 (15 best rating), □ 0 -100 (100 best rating), □ other (please specify including explanation).
How do you rate your past average school achievements compared to your former classmates? Answer on a scale from 0 -100 (0 you are the worst student in the class, 100 you are the best student in the class).
How many times have you visited a physician in the last 12 months (including all routine check-ups at the general practitioner, dentist etc.)?

Have you ever had the impression that a physician is performing more or fewer treatments than necessary or is charging for services that he has not provided? □ Yes □ No
If so, what did you do? □ no experience with monitoring, □ approaching the physician’s office, □ contacting a medical association or a patient protection organization, □ nothing, □ other:
Were the instructions clear and understandable for you? What could possibly be improved?
Do you have any other comments for us?

Appendix D

Detailed predictions and proofs:

We assume that both – patients and physicians – are rational, risk neutral, and maximize their own payoff. All information but the patients' type is common knowledge. We apply the Perfect Bayesian Equilibrium concept. We assume that physicians provide the appropriate treatment and charge the appropriate price if they are indifferent. In our proofs, we follow Dulleck *et al.* (2011) and restrict the analysis to the existence of symmetric equilibria.

Provision and charging policy of physicians.

1. ***B:*** Physicians provide the minor treatment and charge for the major treatment.
2. ***M-End:*** Physicians provide the appropriate treatment and charge the appropriate price.
3. ***M-EndHc:*** Physicians provide the appropriate treatment and charge the appropriate price.
4. ***M-EndUnc:*** Physicians provide the appropriate treatment and charge for the major treatment.
5. ***M-Ex65:*** Physicians provide the appropriate treatment and charge for the major treatment.
6. ***M-Ex35:*** Physicians provide the minor treatment and charge for the major treatment.
7. ***M-Ex10:*** Physicians provide the minor treatment and charge for the major treatment.
8. ***M-End-NoInfo:*** Physicians provide the appropriate treatment and charge the appropriate price.

Predictions on patients visiting a physician:

1. ***B:*** Patients visit the physician.
2. ***M-End:*** Patients visit the physician. Patients require monitoring if they observe a payoff-signal in {-3,-8}, patients require monitoring with probability a=5/6 if they observe a payoff-signal of 2 and patients do not require monitoring if they observe a payoff-signal of 7.
3. ***M-EndHc:*** Patients visit the physician. Patients require monitoring if they observe a payoff-signal in {-3,-8}, patients require monitoring with probability a=5/8 if they observe a payoff-signal of 2 and patients do not require monitoring if they observe a payoff-signal of 7.
4. ***M-EndUnc:*** Patients visit the physician. Patients require monitoring if and only if they observe a payoff-signal in {2,-3,-8}.
5. ***M-Ex65:*** Patients visit the physician.
6. ***M-Ex35:*** Patients visit the physician.
7. ***M-Ex10:*** Patients visit the physician.
8. ***M-End-NoInfo:***Patients visit the physician. Patients require monitoring with probability a=5/6.

B:

Patients Beliefs: For each period, **e**ach patient beliefs that each physician provides a minor treatment and charges a major treatment.

Patients Strategy: For each period, each patient visits the randomly matched physician.

Physicians' Strategy: For each period, provide the minor treatment and charge for the major treatment.

Verification: Patients' beliefs are consistent with physicians' strategy. Next turning to patients strategy: In each period it is rational for patients to interact as $0.5(10-8)+0.5(0-8)=-3>-4$.

Finally, considering physicians' strategy, providing the minor treatment and charging for the major treatment maximizes physicians' payoff given that patients always visit a physician. QED.

M-End:

Patients Beliefs: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Patients Strategy: For each period, each patient visits the randomly matched physician. Each patient requires monitoring with probability 1 if the patient receives a payoff-signal of -3 or -8. Each patient requires monitoring with probability a=5/6 if the patient receives a payoff-signal of 2. Each patient does not require monitoring if the patient receives a payoff signal of 7.

Physicians Strategy: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy:

- If patients observe a payoff-signal of -8, patients learn that they have been undertreated and been charged a major treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of -3, patients learn that they have been undertreated and been charged a minor treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of 2, patients randomize monitoring with probability a=5/6. Monitoring leads to an expected payoff of $0.5\big((10-3)a+(10-8)(1-a)\big)+0.5\big((10-8-1)a+(10-8)(1-a)\big)\ =3.667$ with $a \geq 5/6$. Maximizing his payoff, the patient monitors with probability a. Not monitoring would lead to a payoff of 2.
- If patients observe a payoff-signal of 7, patients learn that they have been treated sufficiently and have been charged the minor treatment. Hence, it is strictly better for patients not to require monitoring.

Hence, given the above patient strategy, visiting a physician is strictly better than not visiting a physician ($0.5(10-3)+0.5(10-8-a\times 1)=\ 4.083>-4.00$ with $a=5/6$).

Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the appropriate treatment and charges the appropriate price leads to the following payoff: $0.5(3-2)+0.5(8-6)=1.50$.
- Deviations:
  - Each physician provides the appropriate treatment and charges for the major treatment which leads to the following payoff: $0.5\big((8-2-5-1)a+(8-2)(1-a)\big)+0.5(8-6)=\ 1.50$ with $a=5/6$.
  - Each physician provides the minor treatment and charges the price for the major treatment which leads to the following payoff: $0.5\big((8-2-5-1)a+(8-2)(1-a)\big)+0.5(8-2-10-1)=-2$ with $a=5/6$.
  - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.

QED.

M-EndHc:

Patients Beliefs: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Patients Strategy: For each period, each patient visits the randomly matched physician. Each patient requires monitoring with probability 1 if the patient receives a payoff-signal of -3 or -8. Each patient requires monitoring with probability $a = 5/8$ if the patient receives a payoff-signal of 2. Each patient does not require monitoring if the patient receives a payoff-signal of 7.

Physicians Strategy: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy:

- If patients observe a payoff-signal of -8, patients learn that they have been undertreated and been charged a major treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of -3, patients learn that they have been undertreated and been charged a minor treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of 2, patients randomize monitoring with probability $a = 5/8$. Monitoring leads to an expected payoff of $0.5\big((10-3)a + (10-8)(1-a)\big) + 0.5\big((10-8-3)a + (10-8)(1-a)\big) = 2.625$ with $a \geq 5/8$. Maximizing his payoff, the patient monitors with probability a. Not monitoring would lead to a payoff of 2.
- If patients observe a payoff-signal of 7, patients learn that they have been treated sufficiently and have been charged the minor treatment. Hence, it is strictly better for patients not to require monitoring.

Hence, given the above patient strategy, visiting a physician is strictly better than not visiting a physician ($0.5(10-3) + 0.5(10-8-a \times 3) = 3.5625 > -4.00$ with $a = 5/8$).

Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the appropriate treatment and charges the appropriate price leads to the following payoff: $0.5(3-2) + 0.5(8-6) = 1.50$.
- Deviations:
    - Each physician provides the appropriate treatment and charges for the major treatment which leads to the following payoff: $0.5\big((8-2-5-3)a + (8-2)(1-a)\big) + 0.5(8-6) = 1.5$ with $a = 5/8$.
    - Each physician provides the minor treatment and charges the price for the major treatment which leads to the following payoff: $0.5\big((8-2-5-3)a + (8-2)(1-a)\big) + 0.5(8-2-10-3) = -3$ with $a = 5/8$.
    - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
      QED.

M-EndUnc:

Patients Beliefs: For each period, each physician provides the appropriate treatment and charges the price for the major treatment.

Patients Strategy: For each period, each patient visits the randomly matched physician. Each patient requires monitoring with probability 1 if the patient receives a payoff-signal of 2, -3 or -8. Each patient does not require monitoring if the patient receives a payoff-signal of 7.

Physicians Strategy: For each period, each physician provides the appropriate treatment and charges the price for the major treatment.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy:

- If patients observe a payoff-signal of -8, patients learn that they have been undertreated and been charged a major treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of -3, patients learn that they have been undertreated and been charged a minor treatment. Monitoring leads to a final payoff of 2. Hence, it is strictly better for patients to require monitoring.
- If patients observe a payoff-signal of 2, patients monitor. Monitoring leads to a final payoff of $0.5(10-8-1)+0.5\big(0.8(10-3)+0.2(10-8-1)\big)=3.4$ whereas not monitoring leads to a payoff of $0.5(10-8)+0.5(10-8)=2$.
- If patients observe a payoff-signal of 7, patients learn that they have been treated sufficiently and have been charged the minor treatment. Hence, it is strictly better for patients not to require monitoring.

Hence, given the above patient strategy, visiting a physician is strictly better than not visiting a physician $0.5\big(0.8(10-3)+0.2(10-8-1)\big)+0.5(10-8-1)=3.40>-4.00$.

Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the appropriate treatment and charges the price for the major treatment. This leads to the following payoff: $0.5\big(0.8(8-2-5-1)+0.2(8-2)\big)+0.5(8-6)=1.60$.
- Deviations:
    - Each physician provides the appropriate treatment and charges the appropriate price which leads to the following payoff: $0.5(3-2)+0.5(8-6)=1.50$.
    - Each physician provides the minor treatment and charges the price for the major treatment which leads to the following payoff: $0.5\big(0.8(8-2-5-1)+0.2(8-2)\big)+0.5\big(0.8(8-2-10-1)+0.2(8-2)\big)=-0.8$.
    - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
      QED.

M-Ex65:

Patients Beliefs: For each period, each patient beliefs that each physician provides the appropriate treatment and charges for the major treatment.

Patients Strategy: For each period, each patient visits the randomly matched physician.

Physicians Strategy: For each period, each physician provides the appropriate treatment and charges for the major treatment.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy, it is rational for patients to interact. Interacting leads to an expected payoff of $0.5\big(0.65(10-3)+0.35(10-8)\big)+0.5(10-8)=3.625>-4$. Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician has the following expected payoff in each period: $0.5\big(0.65(8-2-5-1)+0.35(8-2)\big)+0.5(8-6)=\ 2.05$.
- Deviations:
  - Each physician provides the appropriate treatment and charges the appropriate price would lead to the following payoff: $0.5(3-2)+0.5(8-6)=1.50$.
  - Each physician provides the minor treatment and charges the price the for major treatment leads to the following payoff: $0.5\big(0.65(8-2-5-1)+0.35(8-2)\big)+0.5\big(0.65(8-2-10-1)+0.35(8-2)\big)=\ 0.475$.
  - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
    QED.

M-Ex35:

Patients Beliefs: For each period, each physician provides the minor treatment and charges for the major treatment.

Patients Strategy: For each period, each patient visits the randomly matched physician.

Physicians Strategy: For each period, each physician provides the minor treatment and charges for the major treatment.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy, it is rational for patients to interact. Interacting leads to an expected payoff of $0.5\big(0.35(10-3)\big)+0.5\big(0.65(-8)\big)+0.5(10-8)=\ -0.375>-4$. Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the minor treatment and charges the major price leading to the following payoff: $0.5\big(0.35(8-2-5-1)+0.65(8-2)\big)+0.5\big(0.35(8-2-10-1)+0.65(8-2)\big)=\ 3.025$.
- Deviations:
  - Each physician provides the appropriate treatment and charges for the major treatment leads to the follow payoff: $0.5\big(0.35(8-2-5-1)+0.65(8-2)\big)+0.5(8-6)=$ 2.95.
  - Each physician provides the appropriate treatment and charges the appropriate price leads to the following payoff: $0.5(3-2)+0.5(8-6)=1.50$.
  - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
    QED.

M-Ex10:

Patients Beliefs: For each period, each physician provides the minor treatment and charges for the major treatment.

Patients Strategy: For each period, each patient visits the randomly matched physician.

Physicians Strategy: For each period, each physician provides the minor treatment and charges for the major treatment.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy, it is rational for patients to interact. Interacting leads to an expected payoff of $0.5\big(0.10(10-3)\big)+0.5\big(0.90(0-8)\big)+0.5(10-8)=\ -2.25>-4$. Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the minor treatment and charges the major price leading to the following payoff: $0.5\big(0.10(8-2-5-1)+0.90(8-2)\big)+0.5\big(0.1(8-2-10-1)+0.9(8-2)\big)=\ 5.15$.
- Deviations:
    - Each physician provides the appropriate treatment and charges for the major treatment leads to the follow payoff: $0.5\big(0.10(8-2-5-1)+0.90(8-2)\big)+0.5(8-6)=\ 3.70$.
    - Each physician provides the appropriate treatment and charges the appropriate price leads to the following payoff: $0.5(3-2)+0.5(8-6)=1.50$.
    - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
    QED.

M-End-NoInfo:

Patients Beliefs: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Patients Strategy: For each period, each patient visits the randomly matched physician. Each patient requires monitoring with probability a=5/6.

Physicians Strategy: For each period, each physician provides the appropriate treatment and charges the appropriate price.

Verification: Patients' beliefs are consistent with physicians' strategy. Given the above physician strategy:

- Patients randomize monitoring with probability a=5/6. Monitoring leads to an expected payoff of $0.5(10-3-a\times 1)+0.5(10-8-a\times 1)=\ 3.667$ with $a\geq 5/6$. Maximizing his payoff, the patient monitors with probability a. Not monitoring would lead to an expected payoff of $0.5(10-8)+0.5(0-8)=-3$.

Hence, given the above patient strategy, visiting a physician is strictly better than not visiting a physician ($0.5(10-3-a\times 1)+0.5(10-8-a\times 1)=\ 3.667>-4.0$ with $a=5/6$).

Finally looking at physicians' strategy for each period shows:

- In the above listed equilibrium, each physician provides the appropriate treatment and charges the appropriate price leading to the following payoff: $0.5(3-2)+0.5(8-6)=1.5$.
- Deviations:
  - Each physician provides the appropriate treatment and charges for the major treatment which leads to the following payoff: $0.5((1-a)(8-2)+a(8-2-5-1))+0.5(8-6)= 1.5$ with $a=5/6$.
  - Each physician provides the minor treatment and charges the price for the major treatment which leads to the following payoff: $0.5\big((1-a)(8-2)+a(8-2-5-1)\big)+0.5\times\big((1-a)(8-2)+a(8-2-10-1)\big)=-1.083$ with $a=5/6$.
  - Note that providing always the major treatment is dominated for any pricing strategy by providing the appropriate treatment.
    QED.

Efficiency levels:

In equilibrium, the efficiency calculated as $h\sum_i \frac{(total\ market\ surplus)_i}{(maximum\ possible\ surplus)_i}$ with $i\in\{H,L\}$ for our conditions is shown below. The total market surplus for both illnesses is the sum of the payoff from the patient and the physician minus the monitoring costs in case of exogenous monitoring and the maximum possible surplus is the sum of the payoff from the patient and the physician if the physician provides the appropriate treatment.

- ***B***: $0.5(-2/4+8/8)=0.25$
- ***M-End***: $0.5\left(\frac{19/6}{4}+8/8\right)=0.90$
- ***M-EndHc***: $0.5(2.125/4+8/8)=0.77$
- ***M-EndUnc***: $0.5(3/4+7/8)=0.81$
- ***M-Ex65***: $0.5((4-0.65)/4+7.35/8)=0.88$

  where 0.65 are the monitoring costs that the exogenous institution has to cover in cases of honest provision and charging behavior of the physician. Since in ***M-Ex65*** physicians provide the appropriate treatment and charge honestly when Nature draws the major health problem, the monitoring costs that the exogenous institution absorb amount to $0.65\times 1=0.65$.

- ***M-Ex35***: $0.5(-2.35/4+7.65/8)=0.18$
- ***M-Ex10***: $0.5(-2.1/4+7.9/8)=0.23$
- ***M-End-NoInfo***: $0.5\left(\frac{19/6}{4}+\frac{43/}{8}\right)=0.84$

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

https://www.uibk.ac.at/eeecon/wopec/

2020-32 **Silvia Angerer, Daniela Glätzle-Rützler, Christian Waibel:** Monitoring institutions in health care markets: Experimental evidence

2020-31 **Jana Friedrichsen, Katharina Momsen, Stefano Piasenti:** Ignorance, Intention and Stochastic Outcomes

2020-30 **Esther Blanco, Alexandra Baier, Felix Holzmeister, Tarek Jaber-Lopez, Natalie Struwe:** Substitution of social concerns under the Covid-19 pandemic

2020-29 **Andreas Hackethal, Michael Kirchler, Christine Laudenbach, Michael Razen, Annika Weber:** On the (ir)relevance of monetary incentives in risk preference elicitation experiments

2020-28 **Andrej Gill, Matthias Heinz, Heiner Schumacher, Matthias Sutter:** Trustworthiness in the Financial Industry

2020-27 **Matthias Sutter, Michael Weyland, Anna Untertrifaller, Manuel Froitzheim:** Financial literacy, risk and time preferences - Results from a randomized educational intervention

2020-26 **Rene Schwaiger, Jürgen Huber, Michael Kirchler, Daniel Kleinlercher, Utz Weitzel:** Unequal Opportunities, Social Groups, and Redistribution

2020-25 **Roman Inderst, Martin Obradovits:** Competitive Strategies when Consumers are Relative Thinkers: Implications for Pricing, Promotions, and Product Choice

2020-24 **Martin Obradovits, Philipp Plaickner:** Price-Directed Search and Collusion

2020-23 **Helena Fornwagner, Oliver P. Hauser:** Climate action for (my) children

2020-22 **Esther Blanco, Natalie Struwe, James M. Walker:** Incentivizing public good provision through outsider transfers: experimental evidence on sharing rules and additionality requirements

2020-21 **Loukas Balafoutas, Helena Fornwagner, Rudolf Kerschbamer, Matthias Sutter, Maryna Tverdostup:** Diagnostic Uncertainty and Insurance Coverage in Credence Goods Markets

2020-20 **Anna Ulrichshofer, Markus Walzl:** Customer Disputes, Misconduct, and Reputation Building in the Market for Financial Advice

2020-19 **Anna Ulrichshofer, Markus Walzl:** Social Comparison and Optimal Contracts in the Competition for Managerial Talent

2020-18 **Martin Obradovits, Philipp Plaickner:** Searching for Treatment

2020-17 **Jun Honda:** The Gender-Punishment Gap revisited

2020-16 **Jun Honda:** The Relation between Rankings and Risk-Taking in the Labor Market for Financial Advice

2020-15 **Christina Bannier, Eberhard Feess,Natalie Packham, Markus Walzl:** Differentiation and Risk-Aversion in Imperfectly Competitive Labor Markets

2020-14 **Felix Holzmeister, Rudolf Kerschbamer:** oTree: The Equality Equivalence Test

2020-13 **Parampreet Christopher Bindra, Graeme Pearce:** The effect of priming on fraud: Evidence from a natural field experiment

2020-12 **Alessandro De Chiara, Marco A. Schwarz:** A Dynamic Theory of Regulatory Capture

2020-11 **Christoph Huber, Jürgen Huber, Michael Kirchler:** Market shocks and professionals' investment behavior - Evidence from the COVID-19 crash

2020-10 **Elisabeth Gsottbauer, Daniel Müller, Samuel Müller, Stefan T. Trautmann, Galina Zudenkova:** Social class and (un)ethical behavior: Causal versus correlational evidence

2020-09 **Parampreet Christopher Bindra, Rudolf Kerschbamer, Daniel Neururer, Matthias Sutter:** Reveal it or conceal it: On the value of second opinions in a low-entry-barriers credence goods market

2020-08 **Robert Steiger, Eva Posch, Gottfried Tappeiner, Janette Walde:** Effects of climate change on tourism demand considering individual seasonal preferences

2020-07 **Fang Liu, Alexander Rasch, Marco A. Schwarz, Christian Waibel:** The role of diagnostic ability in markets for expert services

2020-06 **Matthias Stefan, Jürgen Huber, Michael Kirchler, Matthias Sutter, Markus Walzl:** Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect

2020-05 **Michael Razen, Jürgen Huber, Laura Hueber, Michael Kirchler, Matthias Stefan:** Financial Literacy, Economic Preferences, and Adolescents' Field Behavior

2020-04 **Christian König-Kersting, Johannes Lohse, Anna Louisa Merkel:** Active and Passive Risk-Taking

2020-03 **Christoph Huber, Jürgen Huber:** Bad bankers no more? Truth-telling and (dis)honesty in the finance industry

2020-02 **Dietmar Fehr, Daniel Müller, Marcel Preuss:** Social Mobility Perceptions and Inequality Acceptance

2020-01 **Loukas Balafoutas, Rudolf Kerschbamer:** Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions

2019-21 **Felix Holzmeister, Martin Holmen, Michael Kirchler, Matthias Stefan, Erik Wengström:** Delegated Decision-Making in Finance

2019-20 **Julia Rose, Michael Kirchler, Stefan Palan:** Status and Reputation Nudging

2019-19 **Felix Holzmeister, Matthias Stefan:** The risk elicitation puzzle revisited: Across-methods (in)consistency?

2019-18 **Katharina Momsen, Markus Ohndorf:** Information Avoidance, Selective Exposure, and Fake(?) News-A Green Market Experiment

2019-17 **Stjepan Srhoj, Bruno Skrinjaric, Sonja Radas, Janette Walde:** Closing the Finance Gap by Nudging: Impact Assessment of Public Grants for Women Entrepreneurs

2019-16 **Adam Farago, Martin Holmen, Felix Holzmeister, Michael Kirchler, Michael Razen:** Cognitive Skills and Economic Preferences in the Fund Industry

2019-15 **Christopher Kah, Daniel Neururer:** Generiert der stationäre Buchhandel positive Nachfrageeffekte und verhilft dadurch dem Kulturgut Buch bei seiner Verbreitung? - Ein natürliches Experiment

2019-14 **Stjepan Srhoj, Michael Lapinski, Janette Walde:** Size matters? Impact evaluation of business development grants on SME performance

2019-13 **Andrea M. Leiter, Engelbert Theurl:** Determinants of prepaid systems of healthcare financing - A worldwide country-level perspective

2019-12 **Michael Razen, Michael Kirchler, Utz Weitzel:** Domain-Specific Risk-Taking Among Finance Professionals

2019-11 **Jonathan Hall, Rudolf Kerschbamer, Daniel Neururer, Eric Skoog:** Uncovering sophisticated discrimination with the help of credence goods markups - evidence from a natural field experiment

2019-10 **Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Collective intertemporal decisions and heterogeneity in groups

2019-09 **Morten Hedegaard, Rudolf Kerschbamer, Daniel Müller, Jean-Robert Tyran:** Distributional Preferences Explain Individual Behavior Across Games and Time

2019-08 **Daniel Müller, Sander Renes:** Fairness Views and Political Preferences - Evidence from a representative sample

2019-07 **Florian Lindner, Michael Kirchler, Stephanie Rosenkranz, Utze Weitzel:** Social Status and Risk-Taking in Investment Decisions

2019-06 **Christoph Huber, Julia Rose:** Individual attitudes and market dynamics towards imprecision

2019-05 **Felix Holzmeister, Jürgen Huber, Michael Kirchler, Florian Lindner, Utz Weitzel, Stefan Zeisberger:** What Drives Risk Perception? A Global Survey with Financial Professionals and Lay People

2019-04 **David M. McEvoy, Tobias Haller, Esther Blanco:** The Role of Non-Binding Pledges in Social Dilemmas with Mitigation and Adaptation

2019-03 **Katharina Momsen, Markus Ohndorf:** When do people exploit moral wiggle room? An experimental analysis in a market setup

2019-02 **Rudolf Kerschbamer, Daniel Neururer, Matthias Sutter:** Credence goods markets and the informational value of new media: A natural field experiment

2019-01 **Martin Geiger, Eric Mayer, Johann Scharler:** Inequality and the Business Cycle: Evidence from U.S. survey data

University of Innsbruck

Working Papers in Economics and Statistics

2020-32

Silvia Angerer, Daniela Glätzle-Rützler, Christian Waibel

Monitoring institutions in health care markets: Experimental evidence

Abstract
This paper investigates the impact of monitoring institutions on market outcomes in health care. Health care markets are characterized by asymmetric information. Physicians have an information advantage over patients with respect to the appropriate treatment for the patient and may exploit this informational advantage by over- and underprovision as well as by overcharging. We introduce two types of costly monitoring, endogenous and exogenous. When monitoring detects misbehavior, physicians have to pay a fine. Endogenous monitoring can be requested by patients, whereas exogenous monitoring is performed randomly by a third party. We present a toy model that enables us to derive hypotheses and to test them in a laboratory experiment. Our results show that introducing endogenous monitoring reduces the level of undertreatment and overcharging. Even under high monitoring costs, the threat of patient monitoring is sufficient to discipline physicians. Introducing exogenous monitoring also reduces undertreatment and overcharging when it is performed sufficiently frequently. Market efficiency increases when endogenous monitoring is introduced as well as when exogenous monitoring is implemented with sufficient frequency. Our results, therefore, suggest that monitoring may be a feasible instrument to improve outcomes in health care markets.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)