Separation of Prescription and Treatment in Health Care Markets - A Laboratory Experiment*

Ben Greiner[†], Le Zhang[‡], and Chengxiang Tang[§]

Abstract

Health care is a credence good, and its market is plagued by asymmetric information. In this paper, we use a laboratory experiment to test the performance of a potential remedy discussed in the applied literature, the separation of prescription and treatment activities. We observe a significant amount of overtreatment (and a smaller non-predicted amount of undertreatment) in our baseline environment. Requiring a different than the treating physician to provide diagnosis and prescription for free is an effective way to reduce overtreatment in our laboratory setting. This effect, however, is partially offset by an increased frequency of undertreatment. Allowing prescription and treatment physicians to independently set prices for their services reduces efficiency due to coordination failures: in sum, prices are often higher than expected benefit of patients, who in turn do not attend to the physician. Also contrary to theory, bargaining power does not play a significant role for the distribution of profits between physicians.

Keywords: credence goods; overtreatment; health care; separation of prescribing and dispensing

JEL Classification: C72, C91, D82, I11

*Financial support through an Australian Research Council Discovery Grant is gratefully acknowledged. This research has been approved by the Human Research Ethics Committee of UNSW Australia under approval number HREC 10239. We thank two anonymous reviewers and audiences at Curtin U, U Adelaide, U Alicante, U Amsterdam, U Cologne, UNSW, U Sydney, WU Vienna, and at the 2014 ESA conference in Florida for helpful comments and suggestions. Greiner enjoyed the hospitality of UNSW Australia while completing this paper.

†Wirtschaftsuniversität Wien, Institute for Markets and Strategy, Welthandelsplatz 1, 1020 Vienna, Austria, e-mail: bgreiner AT wu.ac.at, and University of New South Wales, School of Economics.

‡Macquarie University, Macquarie Graduate School of Management, NSW 2113, Australia, e-mail: Lyla.Zhang AT mgsm.edu.au.

§Guangzhou University, School of Public Administration, Guangzhou, Guangdong, China, e-mail: tang.chengxiang AT gmail.com.

I Introduction

> *The confectionnarii* [apothecaries] *will prepare medicines at their own expense, under the control of Physicians, ... The stationarius* [local apothecary] *will receive money for his drugs ... That he* [the physician] *must not have any partnership with Confectionnarii* [apothecaries]. *... That he* [the physician] *could not agree to cure a patient, including the preparation of medicines for a price already determined, and that he could not have a store of his own.* – Emperor Frederic II in 'Liber Augustalis' (or 'The Constitutions of Melfi'), promulgated in 1231, Titulus 46 (added as 'Novae Constitutiones Regni Siciliae' in 1240/41).

In 2015, the Michigan physician Farid Fata was sentenced to 45 years in federal prison. He had confessed to falsely diagnosing patients with cancer and prescribing expensive chemotherapy cancer treatment (with US$34 million fraudulent charges altogether), motivated by the kickbacks involved for him. Fata's example is an extreme case, but is also symptomatic for the incentive problems involved in the health care industry, which drive up costs and lower welfare.[1] Health care is considered a credence good. Patients have limited knowledge about the reasons for their ill-being and the proper treatment. Physicians who possess this knowledge may – depending on the exact circumstances – have incentives to exploit the information asymmetry through providing more or more expensive treatment than necessary, which is difficult to verify ex-post. The late Arrow (1963) was among the first to apply economic methods to the study of health care provision, and to note the specialities of this market, while Darby and Karni (1973) pioneered the analysis of credence goods in general. In the health economics literature, the problem has been studied under the keyword 'supplier-induced demand' (e.g. Dranove, 1988; Eggleston, 2012; Fuchs, 1978; Lim, Emery, Lewis and Sunderland, 2009; McGuire, 2000). More recent empirical evidence for overtreatment is provided, among others, by Baker (2010), Gruber, Kim and Mayzlin (1999), Hughes and Yule (1992), Iizuka (2012), and Lundin (2000).

In this paper, we use a laboratory experiment to test a potential remedy discussed in the applied literature that may help to align health care provider incentives with patient preferences. Namely, we provide evidence on the effectiveness of separating agency in prescription and treatment. This issue has been studied by a number of empirical studies, but we are not aware of existing evidence from the experimental laboratory.

[1] There are more examples of health care fraud in terms of overtreatment, many of which were discovered by tax authorities and publicly reported on in newspapers. In 2002, the physicians Kent Brusett, Chae Moon, Ricardo Javier Moreno-Cabral, and Fidel Realyvasquez from the Redding Medical center were convicted for unnecessary heart surgeries. From 2007 to 2013, a cardiologist named Mark Midei implanted medically unnecessary cardiac stents into patients. The same is true for cardiologist John R. McLean, convicted 2011. Abubakar Atiq Durrani convinced patients in 2013 to undergo medically unnecessary spinal surgeries. Shadrach Gonqueh (in 2015) and Amanda Hoover (in 2016) were charged for the prescription and implementation of unnecessary radical dental treatments, crowns and teeth restorations, partly on underage patients. As a particularly disgusting example, the doctor Oscar Huachillo, convicted in 2014, administered medically unnecessary expensive treatments at a highly diluted dose to HIV/AIDS patients.

The departure point of our study is a simple theoretical framework for credence goods.[2] A patient meets a physician.[3] The patient has either a severe or a non-severe illness with some probability. The patient cannot observe his own illness, but the physician can. There are two types of treatment available, a non-severe or a severe treatment. The former is less costly to perform and is only effective for the non-severe illness. The latter imposes a higher cost on the physician and cures both a non-severe and a severe illness. The physician first posts prices for the possible treatments. The patient decides whether to see the physician. If the patient arrives, the physician performs a diagnosis and proposes a treatment. The patient decides whether he takes up the treatment, and is healed conditional on receiving a treatment that is effective for his illness. Under our parameter configuration, economic theory predicts overtreatment in this baseline setup (B).

Our experimental study implements this framework in the laboratory, with students as subjects, repeated random anonymous matching, and monetary payoffs, but instructions framed as a physician-patient setup. Within this framework, we experimentally test two theoretical assertions. First, we follow a prominent policy suggestion in the literature and separate prescribing and treating agents (S). In particular, we have each patient seeing two physicians, one of whom provides diagnosis and prescription (for free) while the other then provides the prescribed treatment. This separation breaks the link between a physician's prescription and profits, thus should align incentives and result in full treatment efficiency. Second, we allow the two physicians to independently set their prescription and treatment fees, respectively (ST-x). We vary the bargaining power between physicians, such that in theory either the prescribing physician can fully extract the patient's willingness to pay (ST-P), or the treating physician reaps all the profits (ST-T), or (in a Nash Bargaining type of game) the profits are expected to be shared equally between the two physicians (ST-B).

The effect of separating the agency over prescription and treatment has also been studied in the empirical health economics literature. Evidence for overtreatment from integrated physicians stems from a number of countries such as the U.S., Japan, and Switzerland (see Afendulis and Kessler, 2007; Iizuka, 2007; Kaiser and Schmid, 2016, respectively). While a physician's integration of both prescribing and dispensing drugs has been banned in Western Europe as early as the 13th century (see our introductory quote from legislation imposed by Emperor Frederic II in 1240), it is still a persistent institution in many Eastern Asian countries, as well as some remote areas in the U.S. and European countries (Rodwin and Okamoto, 2000). Interestingly, even in Western countries the prevention of incentive conflicts in health care has not carried over (yet) to separating modern surgeon's diagnosis

[2] We use the framework of Dulleck and Kerschbamer (2006). We follow Fong, Liu and Wright (2014) in allowing for different patient utility outcomes for different untreated illnesses and by dropping the assumption that a patient who takes up diagnosis is committed to any treatment the physician may propose. For details please see our theoretical discussion in Section II. Earlier contributions to the analysis of credence goods were made by Emons (1997, 2001), and Wolinsky (1993, 1995), among others.

[3] Throughout this paper we use a health care frame, referring to physician-patient interactions, since this is the most important application we have in mind. However, in principle our analysis can be analogously applied to other credence good problems. We refer to 'prescription' as the act of diagnosis and treatment suggestion, and we refer to 'treatment' as the actual act of remedy (against a fee), which also includes surgery or the dispension of drugs, for example.

and treatment. Some Asian countries introduced policies to separate the prescription and dispensation of drugs (e.g. Japan (Hayase, 2003), Korea (Kwon, 2003; Lee and Malone, 2003), and Taiwan (Chen, Gertler and Yang, 2016; Chou, Yip, Lee, Huang, Sun and Chang, 2003)), with overwhelmingly positive results due to less prescriptions and lower health spending. Most of these empirical studies, however, are limited to before-after comparisons of health care reforms, and suffer from differences in other institutional details.

Laboratory experiments have a number of methodological advantages in this setting. First, they allow us to create environments which are very close to the economic models which are being tested. If a model's predictions and suggestions do not work in such a controlled environment, then their usefulness for even more complex real-world environments and policy applications is questionable. Second, laboratory experiments allow us to empirically test different market setups without endangering actual people's lives and health, which would be at stake if we experimented with real-world health care markets. And third, laboratory experiments allow us to observe variables (such as the true condition of a patient) which cannot be perfectly observed or ex-post verified in a real-world setting; they even allow us to externally and randomly induce such conditions. This feature of laboratory experiments allows for causal inferences with respect to research questions for which any empirical field-data analysis can only be suggestive in nature. The caveat of laboratory experiments is their external validity. However, if we are careful in generalizing laboratory results to real-world settings, in particular with respect to point predictions, then they can give us important insights, and help us direct the focus of theoretical and empirical research.

Our main findings are as follows. First, we observe a significant amount of overtreatment (and a smaller non-predicted amount of undertreatment) in our baseline environment (B). Second, the most effective way to reduce overtreatment in our laboratory setting is to require a different than the treating physician to provide diagnosis and prescription for free (S). This effect, however, is partially offset by an increased frequency of undertreatment (patients with severe illness receiving insufficient but costly treatment). Third, allowing prescription and treatment physicians to independently set prices for their services (ST-x) reduces efficiency due to coordination failures: the sum of prices is often too high such that many patients do not attend to a physician due to negative expected payoffs. Fourth, also contrary to theory, bargaining power does not play a significant role for the distribution of profits between physicians.

Not all integrated physicians give in to the incentive to overtreat, and behave honestly. On the other hand, in none of our models undertreatment was predicted, but we observe a significant albeit small share of 7% with integrated physicians (B), and an increased share of 11-24% when separating prescription and treatment (S, ST-x). The share of undertreatment is highest when we require the prescription physician to provide her services for free (S).

Our study complements a recent laboratory-experimental literature on incentive problems of providers in a health care context. One set of studies examines physician incentives from different payment schemes (such as fee-for-services and capitation payments) in a non-interactive/non-strategic context (e.g., Brosig-Koch, Hennig-Schmidt, Kairies-Schwarz and Wiesen, 2015, 2016; Godager, Henning-Schmidt and Iversen, 2016; Green, 2014; Hennig-Schmidt, Selten and Wiesen, 2011; Hennig-Schmidt and Wiesen, 2014). A number of recent field experiments test the susceptivity of credence good providers to overtreating their clients under different information conditions and incentives, e.g. Balafoutas, Beck, Kerschbamer and Sutter (2013) for cab drivers, Kerschbamer, Neururer and Sutter (2016) for computer experts, Lu (2014) for endocrinology and cardiology specialists, and Gottschalk, Mimra and Waibel (2017) for dentists. Beck, Kerschbamer, Qiu and Sutter (2014) invite car mechanics to the laboratory and find them to be more overtreating than students.

More closely related to our paper are a set of laboratory experiments that study the effect of institutions on credence good market outcomes. Dulleck, Kerschbamer and Sutter (2011) investigate a general credence good framework and observe that liability plays a more crucial role than verifiability in increasing trade and efficiency, and that reputation only matters when both liability and verifiability are absent. Kerschbamer, Sutter and Dulleck (2017) provide evidence that heterogenous social preferences may explain why the (non-)existence of verifiability has so little impact on market outcomes. Huck, Lünser, Spitzer and Tyran (2016) study the effects of (medical) insurance and competition on credence good provision. They find that insurance increases both physician attendance and overtreatment, while competition reduces the occurrence of overtreatment. Mimra, Rasch and Waibel (2016*b*) allow clients to obtain costly second opinions, which leads to less overtreatment, although customers make use of it less often as predicted. Mimra, Rasch and Waibel (2016*a*) study the interaction of competition and reputation in a credence good context. They observe that the level of fraud (undertreatment and overcharging) is significantly higher under competitive compared to fixed prices, and that reputation does have little impact in this environment. Waibel and Wiesen (2016) make a case for kickbacks for GPs who refer patients to specialists, since they reduce undertreatment and overtreatment by compensating GPs for forgone opportunistic profits. Greiner and Zhang (2017) test and find no evidence for a theoretical conjecture by Fong et al. (2014) that a two-part tariff can address the efficiency issues introduced by allowing patients to reject a prescribed treatment.

II Theoretical framework and experimental design

We base our framework on Dulleck and Kerschbamer (2006) and Fong et al. (2014).[4] A patient meets a physician. Patients are ex-ante homogenous and suffer from either a severe illness s (with probability θ) or from a non-severe illness n (with probability $1-\theta$). The patient does not observe his own illness, but the physician does. The physician can prescribe one of two treatments, a severe treatment s

[4]Dulleck and Kerschbamer (2006) assume that patients are homogenous and committed to treatment (that is, if they decide to visit the physician, they also commit to take up the treatment which the physician prescribes) and explore the effects of verifiability (the patient observes which treatment he gets and can condition payment upon it) and liability (the patient observes the treatment outcome and can condition payment upon it) on the existence of an efficient market solution and incentives to overtreat, undertreat, or overcharge the patient. Fong et al. (2014) drop the assumption of patient commitment and explore the set of Perfect Bayesian Nash Equilibria in a large range of parameter constellations. In our setup, we assume verifiablity of the received treatment, but no liability for the treatment outcome. As in Fong et al. (2014), patients are not committed to take up the prescribed treatment, which under our restricted parameter space (see assumptions discussed below) gives rise to overtreatment in our baseline model.

which cures both kinds of illnesses or a non-severe treatment n which cures only the non-severe illness. Treating the patient is costly (C_s and C_n, respectively, with $C_s > C_n$) and the physician charges the patient a fee depending on provided treatment (P_s and P_n, respectively). A cured and healthy patient enjoys utilities U_s and U_n, respectively, while an untreated patient endures U_s^0 and U_n^0, with $U_s > U_s^0$ and $U_n > U_n^0$.

In our baseline case B, the interaction takes place as follows. First, the physician posts treatment prices P_s and P_n for the severe and non-severe treatment, respectively. Then, the patient decides whether to obtain a diagnosis/prescription. If the patient decides to see the physician, then the physician will privately observe the patient's illness and prescribe a treatment. Only after the patient decides to see the physician, the physician privately observes the patient's illness. Then, the physician prescribes the treatment. If the patient takes up the prescribed treatment, he will pay the respective treatment price P_s or P_n to the physician, who in turn provides the treatment and incurs treatment costs of C_s or C_n, respectively. If the treatment cures the patient's disease, he will enjoy utility U_s or U_n, respectively. If the patient does not obtain a prescription or does not take up the prescribed treatment or receives insufficient treatment, then his utilities are U_s^0 or U_n^0, respectively.

TABLE 1: SEQUENCE OF DECISIONS IN OUR FIVE EXPERIMENTAL CONDITIONS

	B	S	ST-B	ST-P	ST-T
Physician sets diagnosis fee d			P1/	P1	P1
Physician sets treatment prices P_s and P_n	○	P2	P2	P2	
Patient decides to see physician, or not	○	○	○	○	○
(Physician sets treatment prices P_s and P_n)					P2
Physician observes the patient's illness and then prescribes treatment s or n	○	P1	P1	P1	P1
Patient decides to take up treatment, or not	○	○	○	○	○

Column B in Table 1 shows the sequence of decisions in our baseline condition B. In addition, we make the following assumptions.

1. Appropriate treatments are efficient. That is, patient utility gains from being treated for a severe (non-severe) illness are larger than the costs of a severe (non-severe) treatment, $U_s - U_s^0 > C_s$ and $U_n - U_n^0 > C_n$, respectively.

2. The patient's utility gain from receiving treatment for a severe illness may be larger than or equal to the utility gain from being treated for a non-severe illness, $U_s - U_s^0 \geq U_n - U_n^0$. Dulleck and Kerschbamer (2006) assume equality in this relation. For example, there may be different reasons why a car does not work, but the utility when the car is broken and the utility when the car is fixed are independent of the reason. Fong et al. (2014) relax this assumption as we do. In particular, we allow that, keeping the utility of a healthy person fixed ($U_s = U_n$), enduring an untreated severe illness takes a higher toll than suffering an untreated non-severe illness ($U_s^0 < U_n^0$).

3. The welfare gains from successfully treating a patient with a severe illness using a severe treatment are larger than the welfare gains from successfully treating a patient with a non-severe illness using a non-severe treatment, $(U_s - U_s^0) - C_s > (U_n - U_n^0) - C_n$. This simply reflects the parameterization of our experiment, and results in the equilibrium prediction of overtreatment (rather than undertreatment) in our baseline model B, since these welfare gains are the profits a physician can extract in an interaction.

4. We assume that when monetarily indifferent, a physician acts honestly rather than dishonestly. This is a standard assumption in the literature (e.g. Dulleck and Kerschbamer, 2006), and required to be able to solve the model versions S and ST-x, where the prescription physician's payoff function in equilibrium does not depend on the honesty of prescription.

In our experimental study, we consider five versions of the game.

Baseline (B). Our theoretical prediction for the baseline condition follows the analysis of Fong et al. (2014). Dulleck and Kerschbamer (2006) show that when the patient is committed to take up whatever treatment the physician prescribes (and there are ex-ante homogenous patients and treatments are verifiable), then there will be efficient health care in equilibrium, with the physician providing honest prescription and treatment. The reason is that honest treatment maximizes the patient's willingness to pay. The physician will set treatment prices such that she can be honest in the prescription stage (which implies setting prices such that both treatments yield the same profit) and at the same time extract the full information rent from the patient.

Part of Dulleck and Kerschbamer (2006)'s model, and a point stressed by Fong et al. (2014), is that when the patient is *not* committed to taking up the prescribed treatment, as in our baseline condition (B), then the market outcome is not efficient anymore. To see this intuitively, consider the case of honest treatment as described above. In that equilibrium, the price for the non-severe treatment P_n exceeds the utility gain of the patient from being treated for a non-severe illness. As a result, with that set of prices, a patient (who assumes the physician to prescribe honestly) would only accept the severe treatment but reject the non-severe treatment. This lowers the physician's expected profits.

Fong et al. (2014) provide a complete analysis of our baseline game, for the full range of parameter values. Each pair of treatment prices opens a proper subgame. Since the physician is informed about the illness at the time of treatment prescription, when deciding about treatment take-up the patient will have to form beliefs about the treatment strategy of the physician. In other words, these subgames are Bayesian games. Fong et al. (2014)'s approach is to characterize the Perfect Bayesian Equilibria of the subgames that follow the price-setting stage, and then, when considering the prices to be set in the first stage, to look for *optimal equilibria*, i.e. the equilibria that maximize the price-setting physician's profits.

Proposition 2 in Fong et al. (2014) shows that for our parameter configuration, namely $(U_s - U_s^0) - C_s > (U_n - U_n^0)C_n$ and $\theta \in \big((C_s - C_n)/((U_s - U_s^0) - (U_n - U_n^0)), 1\big]$, the optimal equilibrium is the one with full overtreatment and a price of $P_s = \theta(U_s - U_s^0) + (1-\theta)(U_n - U_n^0)$. This corresponds to a price for the severe illness of 70, while the price of minor/non-severe illness (P_n) is not uniquely defined and determines the existence of other, payoff-dominated equilibria.[5] Thus, our prediction for condition B is that the physician will charge a severe treatment price equal to the patient's expected utility gain from being treated with the severe treatment for *any* illness, the patient will take up a diagnosis, the physician will always provide severe treatment (i.e. overtreat if the patient has a non-severe illness), and the patient will accept that treatment. The market is inefficient, but the patient will always be successfully treated.[6]

Separation of agency in prescription and treatment (S). A potential fix to the arising incentive problems in credence goods markets such as health care is to separate the agency over prescription and treatment. That is, in our model version S we assume that a patient meets two different physicians. The prescription physician P1 only determines the treatment. The treatment physician P2 provides the treatment based on the prescription of P1. We require the prescription physician to provide the diagnosis for free.[7]

It is immediately clear that in theory, this change in the market structure solves the problem of potential overtreatment (as long as collusion between physicians can be controlled for). The prescription physician P1 is free of moral hazard problems and provides honest treatment. The treatment physician P2 can only set the prices for her treatment. Given the external honest prescription, P2 charges the patient at his willingness to pay, namely $P_s = U_s - U_s^0$ and $P_n = U_n - U_n^0$.

Endogenous diagnosis and treatment prices (ST-P, ST-T, ST-B). When different physicians are responsible for prescription and treatment, a natural market structure also allows them to set their own prices independently, such that the prescription physician can charge a diagnosis fee (unlike in our model S where she is forced to provide diagnosis for free) and the treatment physician charges treatment prices (as before). Our model versions ST-P, ST-T, ST-B vary the bargaining power assigned to the two physicians, respectively.

[5]With $P_n \leq 40$, the perfectness of the profit-maximizing full overtreatment equilibrium needs to be sustained with an out-of-equilibrium belief that the true illness is severe whenever a non-severe treatment would be prescribed (and thus such a non-severe treatment prescription would be rejected, see Fong et al., 2014, discussion of Lemma 3). When $P_s = 70$ and $P_n \leq 40$, there is also an equilibrium in the subgame with partial overtreatment (Lemma 5), and when $P_s = 70$ and $P_n \leq 20$, then there is an additional equilibrium in the subgame with full undertreatment (Lemma 2). These equilibria yield lower payoffs for the price-setting physician than the full overtreatment equilibrium.

[6]The expected loss in welfare in this equilibrium, compared to efficient treatment, equals $((1-\theta)(C_s - C_n)) / (\theta(U_s - C_s) + (1-\theta)(U_n - C_n))$, i.e. the relative costs of treating with the more costly severe treatment when the patient has only the non-severe illness. For our experimental parameters, this equals 9%.

[7]In terms of strategic analysis, this is equivalent to setting a fixed but positive fee for diagnosis. We think that our parameter choice of a zero diagnosis fee and thus no income for the prescription physician puts a lower bound on the efficiency gains to be expected from agency separation. A prescription physician who receives a positive fee may behave more honestly.

In model version ST-P, the prescription physician has all the bargaining power. In the interaction sequence, the prescription physician P1 first announces a diagnosis fee d. Then, after observing that fee, the treatment physician P2 announces her prices P_s and P_n. Then the patient decides to obtain a diagnosis or not, P1 prescribes the treatment, the patient decides to take up treatment or not, and P2 correspondingly provides the treatment or not. P2 will want the patient to take up treatment whenever the price she charges is larger or equal to the cost of providing the respective treatment. The patient will take up treatment when the respective price is smaller than the utility gain from being treated. P1 is unaffected by later choices and thus will provide honest prescriptions. Given this, the patient will obtain a diagnosis according to the participation constraint $d + \theta P_s + (1-\theta)P_n \leq \theta(U_s - U_s^0) + (1-\theta)(U_n - U_n^0)$, i.e. when the patient's expected payment is not larger than his expected utility from being treated. This implies that P2 will set treatment prices such that this condition is fulfilled, as long as $\theta P_s + (1-\theta)P_n \geq \theta C_s + (1-\theta)C_n$. In turn, P1 can now maximize her profits by setting a diagnosis fee of $d = \theta(U_s - U_s^0 - C_s) + (1-\theta)(U_n - U_n^0 - C_n)$. This diagnosis fee extracts all information rents, and forces P2 to set (expected) treatment prices at (expected) costs with zero profits.[8]

The interaction sequence of model version ST-T differs from ST-P only in that the treatment physician P2 now announces her treatment prices P_s and P_n *after* the patient decided to take up the diagnosis (not before, as in ST-P). This change moves all the bargaining power to P2. As discussed above, the patient will accept an honestly prescribed treatment when the respective price is smaller than the utility gain from being treated, i.e. when $P_s \leq U_s - U_s^0$ and $P_n \leq U_n - U_n^0$. Thus, in this model version, P2 will set her prices at the maximum level allowed by these inequalities. However, the patient's participation constraint for obtaining a diagnosis still holds, in turn allowing the prescription physician P1 only to charge a diagnosis fee of $d = 0$.

Finally, in model version ST-B, physicians P1 and P2 announce their prices d, P_s, and P_n simultaneously. These choices are still subject to P2's participation constraint of $\theta P_s + (1-\theta)P_n \geq \theta C_s + (1-\theta)C_n$ and the patient's participation constraint of $d + \theta P_s + (1-\theta)P_n \leq \theta(U_s - U_s^0) + (1-\theta)(U_n - U_n^0)$. However, the simultaneous procedure of price setting (assuming subgame behavior as discussed above) opens up a variant of a Nash bargaining game. The Nash bargaining solution prescribes equal utility gains over outside options. This implies a diagnosis fee of $d = 0.5(\theta(U_s - U_s^0 - C_s) + (1-\theta)(U_n - U_n^0 - C_n))$ and prices P_s and P_n such that $\theta P_s + (1-\theta)P_n = 0.5(\theta(U_s - U_s^0 - C_s) + (1-\theta)(U_n - U_n^0 - C_n))$.[9,10]

[8] Technically, the treatment physician will set treatment prices such that $\theta P_s + (1-\theta)P_n = \theta C_s + (1-\theta)C_n$, which allows for multiple solutions. However, an epsilon risk aversion would be sufficient to make the unique solution of $P_s = C_s$ and $P_n = C_n$ optimal, which is thus our prediction for this experiment condition.

[9] Once again, this allows for multiple solutions, but an epsilon risk aversion yields a unique optimal solution where the two medical treatments lead to the same profits for the treatment physician.

[10] While our experimental design of different bargaining power conditions is mainly motivated as an experimental robustness check, there are also some corresponding real-world scenarios. When introducing of a separation policy, an authority may want to keep diagnosis fees as low as previously mandated (as in our condition ST-T), and then over time shift bargaining power to a more balanced state (as our condition ST-B). Large profits of the pharmaceutical industry could motivate even further shifts of bargaining power away from pharmaceutical companies to health care professionals (as in our condition ST-P).

III Experimental Procedures

Our experiment implements all five versions of our model of physician-patient interactions in a between-subjects design. Experiment condition B serves as our baseline.[11] Experiment condition S separates agency over prescription and treatment, such that one physician provides the diagnosis (for free) and the other physician provides the prescribed treatment. In experiment conditions ST-P, ST-B, and ST-T, agency is still separated, but both physicians can freely set prices for their services. The three experiment conditions vary in the bargaining power which they assign to the two physicians. In experiment condition ST-P, in equilibrium the prescription physician can extract all the information rents from the patient, while the treatment physician prices at costs. In ST-T, the treatment physician is able to extract all information rents in equilibrium, while the prescription physician provides diagnosis for free. Finally, in experiment condition ST-B, the Nash bargaining solution applies, in that both physicians have equal bargaining power and should be able to extract half of the information rents.

In the experimental implementation of our theoretical framework, the likelihood of the patient having the severe illness θ was set to 50%, such that both illnesses were equally likely to happen. The utility of being healthy was set to $U_s = U_n = 100$ points, with 30 points being worth 1 AUD. The utility of an uncured severe illness equals $U_s^0 = 0$, and the utility of an uncured non-severe illness was $U_n^0 = 60$. The costs for curing the illnesses were set to $C_s = 20$ and $C_n = 5$, respectively. These parameters ensure that our model assumptions are fulfilled. The upper panel of Table 2 lists the numerical equilibrium predictions of our model for prices, patient and physician choices, and (expected) efficiency in the different experiment conditions.

The experiment sessions took place at the BizLab of the University of New South Wales. Using ORSEE (Greiner 2015), we recruited 300 participants (60 participants for each of our five experiment conditions, 30 participants in each session, and two sessions for each experiment condition). In the experimental instructions, we used a "Doctor-Patient" frame.[12] In each session, participants were randomly assigned to be either a 'Doctor' or a 'Patient', and they kept their role throughout the 20 rounds of the experiment. In experiment condition B, participants were randomly re-matched to groups of two at the beginning of each round: a doctor and a patient. In the two-physician experiment conditions S, ST-P, ST-T, and ST-B, each physician participant played two roles, prescription physician and treatment physician, in different groups. At the beginning of each round, groups of three were randomly matched together: one patient, one prescription physician, and one treatment physician. The matching procedure ensured not only that the two physicians were different participants, but also that the physician they interacted with in one group was not the same as in the other group.

[11] Greiner and Zhang (2017) reuse the data collected for baseline condition B in order to test the effect of treatment commitment and two-part tariffs.

[12] Our initial draft of the instructions used a neutral frame. However, some pre-testing resulted in the insight that the setup of our experiment was very difficult to understand from these neutrally framed, abstract instructions. We then tested instructions where the game was framed as a doctor-patient interaction, with overwhelming success in terms of comprehension. Our testers now had no problem understanding the game structure, the parameters, and the choices to be made. We decided to use these framed instructions in the experiment.

Fixing behavior, our procedure kept average expected profits constant across single-physician and two-physician experiment conditions (and in particular avoided overall zero profits to prescription physicians in experiment condition S).

At the beginning of each session, participants read and signed their consent forms. They then received the experimental instructions and could ask questions. After all questions had been answered (in most sessions, there weren't any), the experimental program started. The experiment was implemented in zTree (Fischbacher 2007). At the end of the experiment, participants filled in a short questionnaire asking for demographics, and were paid out in cash. Sessions lasted 74 minutes on average, and the average earnings were about $24.53 per person (including a $5 fee to show up, StdDev $6.04).

TABLE 2: AGGREGATE RESULTS

	B	S	ST-B	ST-P	ST-T
Theoretical predictions					
Diagnosis fee	-	-	28.75	57.5	0
Non-severe treatment price	<55	40	33.75	5	40
Severe treatment price	70	100	48.75	20	100
Diagnosis take-up	100%	100%	100%	100%	100%
Non-severe illness: Overtreatment	100%	0%	0%	0%	0%
Severe illness: Undertreatment	0%	0%	0%	0%	0%
Treatment take-up	100%	100%	100%	100%	100%
Profit patient	30	30	30	30	30
Profit integrated physician	50				
Profit prescription physician		-	28.75	57.5	0
Profit treatment physician		57.5	28.75	0	57.5
Efficiency loss	9%	0%	0%	0%	0%
Laboratory evidence					
Diagnosis fee	-	-	31.3 (10.9)	28.8 (8.0)	28.7 (9.5)
Non-severe treatment price	39.0 (13.0)	33.7 (10.9)	28.8 (11.9)	23.5 (9.9)	27.3 (11.1)
Severe treatment price	70.8 (15.7)	65.6 (16.0)	51.2 (15.4)	50.6 (11.7)	50.1 (15.4)
Diagnosis take-up	83%	95%	54%	58%	59%
Non-severe illness: Overtreatment	51%	20%	25%	18%	14%
Severe illness: Undertreatment	7%	24%	11%	20%	18%
Non-severe prescription: Treatment take-up	70%	60%	94%	91%	89%
Severe prescription: Treatment take-up	70%	85%	96%	95%	88%
Overall treatment take-up	70%	72%	95%	93%	89%
Profit patient	36.6 (26.0)	40.2 (29.9)	30.1 (29.0)	26.9 (30.5)	27.0 (31.2)
Profit integrated physician	24.5 (24.2)				
Profit prescription physician		-	15.3 (15.2)	15.8 (14.6)	15.4 (14.2)
Profit treatment physician	-	24.3 (20.8)	12.0 (15.1)	12.1 (14.0)	12.8 (15.2)
Efficiency loss	32%	27%	36%	38%	38%

Note: The lower panel displays averages, with standard deviations in brackets.

IV Results

The lower panel of Table 2 shows the aggregate outcomes in our five experiment conditions for the most important variables. We will first discuss whether separating the agency over prescription and treatment with regulated diagnosis fees has an effect on prices, treatment choices, and efficiency. Then we will examine the role that the price determination mechanism (as a bargaining game between physicians) plays in this environment. We find that separating prescription and treatment with a fixed (zero) diagnosis fee increases efficiency. However, allowing physicians to negotiate prices leads to coordination failures and consequently to lower efficiency. Interestingly, the exact bargaining protocol plays much less a role for prices and the distribution of profits than theoretically predicted.

TABLE 3: Effects on diagnosis fees, treatment prices, and profits

Model	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Dependent variable	Diagnosis fee	Non-severe treatment price	Severe treatment price	Non-severe treatment price incl. diag fee	Severe treatment price incl. diag fee	Patient's round income
Constant	31.31***	38.99***	70.81***	38.99***	70.81***	36.62***
	(1.40)	(1.99)	(2.79)	(1.99)	(2.79)	(1.57)
S		-5.32**	-5.22	-5.32**	-5.22	3.54
		(2.08)	(3.55)	(2.08)	(3.55)	(2.44)
ST		-10.23***	-19.64***	21.07***	11.67***	-6.52***
		(2.84)	(3.58)	(3.00)	(3.14)	(2.31)
ST × STP	-2.56	-5.27**	-0.61	-7.83**	-3.17	-3.25
	(2.10)	(2.53)	(2.94)	(3.19)	(2.86)	(2.56)
ST × STT	-2.63	-1.47	-1.11	-4.10	-3.74*	-3.11
	(1.99)	(2.48)	(3.36)	(2.45)	(1.96)	(2.61)
N	1800	3000	3000	3000	3000	3000
Adj. R^2	0.015	0.181	0.257	0.331	0.118	0.031
Post-estimation tests, p-values						
S=ST	-	0.028	0.000	0.000	0.000	0.052
ST×STP=ST×STT	0.975	0.078	0.874	0.143	0.841	0.966

Notes: Standard errors are clustered at the matching group level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

Separation of agency in prescription and treatment. Tables 3 and 4 display results from estimations of experiment condition effects on treatment prices, diagnosis and treatment take-up, physician's diagnosis choices, and overall efficiency. In all regressions, we include experiment condition dummies in the following way. Experiment condition B serves as a baseline for all effects. The dummy S equals 1 in experiment condition S, and thus captures differences between experiment conditions S and B. The dummy ST equals 1 in all three ST-experiment conditions, while the dummies STP and STT (included as interactions on ST) equal 1 for experiment conditions ST-P and ST-T, respectively. Thus, the coefficient on ST represents the difference between B and ST-B (that is, overall effect of separating prescription and treatment *and* at the same time allowing physicians to charge their own prices with equal bargaining power). The effects ST×STP and ST×STT break out the effects of moving the bargaining power to the prescribing and treating physician, respectively, relative to the equal bargaining power condition ST-B. For each regression, we additionally conduct two post-estimation Wald tests, the results of which are displayed in the last two rows of Tables 3 and 4, respectively. The first test detects differences between experiment conditions S and ST, the second test inspects differences between experiment conditions ST-P and ST-T.

With respect to effects of experiment condition S vs. experiment condition B, we make the following main observations:

1. Both the prices of severe and non-severe treatment are (about 5-6 points) lower in experiment condition S than in experiment condition B (Table 3, Models 2 and 3). Statistically, this is only significant for the price of non-severe treatment. Possibly as a result of lower prices, the likelihood of the patient obtaining a diagnosis is statistically significantly higher (about 22%) in condition S than in condition B (Table 4, Model 7). This result is interesting, since in both conditions S and B there is no diagnosis fee, and a prescribed treatment can be rejected later on.[13]

2. In condition S compared to condition B, the likelihood of overtreatment (when the true illness is non-severe) is significantly reduced by 26%, while the likelihood of undertreatment (when the true illness is severe) is significantly increased by 19% (Table 4, Models 8 and 9). Simultaneously (and very likely as a consequence), the acceptance of a severe treatment prescription is higher and the acceptance of a non-severe treatment prescription is lower in condition S compared to condition B (regressions analogous to Table 4 Model 10, but separately for cases of severe and non-severe treatment prescriptions). However, the overall likelihood of treatment acceptance is not significantly affected (Table 4, Model 10).

3. Overall health care market efficiency is 5% higher in experiment condition S compared to experiment condition B. This difference, however, is statistically not significant (Table 4, Model 11). Patients seem to be the main benefactors of these efficiency gains, resulting in higher patient profits (although once again, the difference is not statistically significant, see Table 3 Model 6).

[13] For both prices, physicians could choose any integer number between 0 and 100. Our results on treatment prices are robust when using Tobit models with censoring at 0 and 100.

TABLE 4: Effects on diagnosis and treatment take-up, over- and undertreatment, and efficiency

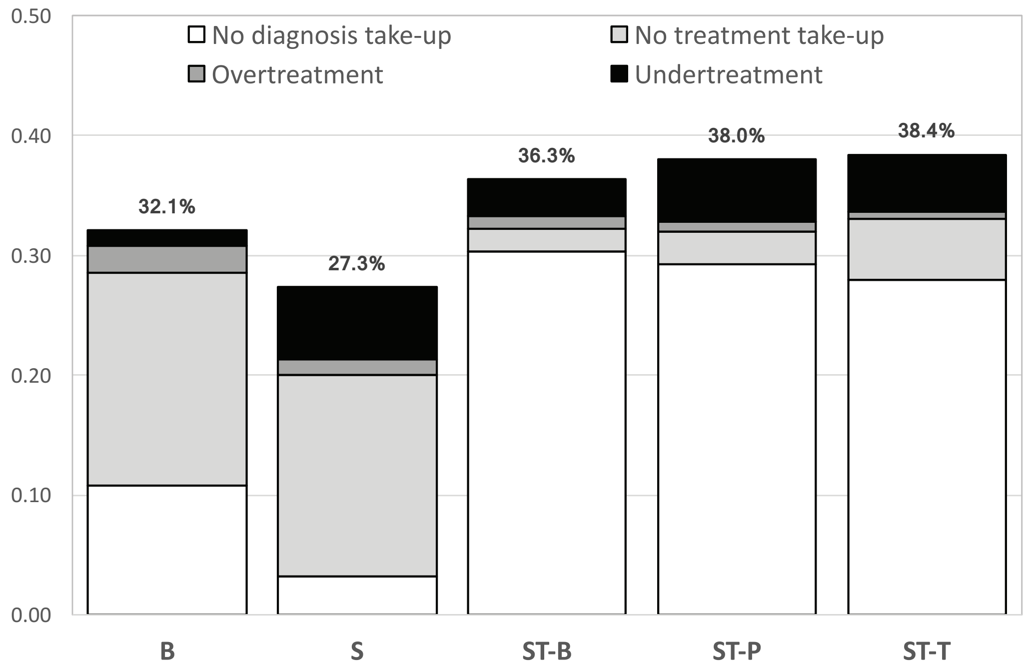
Model	(7)	(8)	(9)	(10)	(11)
	Probit	Probit	Probit	Probit	OLS
Dependent variable	Diagnosis take-up	Over-treatment	Under-treatment	Treatment take-up	Efficiency
Constant					0.686
					(0.153)
S	0.223***	-0.262***	0.189***	0.016	0.041
	(0.086)	(0.047)	(0.071)	(0.036)	(0.041)
ST	-0.256***	-0.204***	0.064	0.274***	-0.044*
	(0.056)	(0.070)	(0.067)	(0.028)	(0.023)
ST × STP	0.027	-0.076	0.090	-0.042	-0.030
	(0.030)	(0.082)	(0.056)	(0.063)	(0.040)
ST × STT	0.038	-0.129	0.079	-0.108***	-0.027
	(0.037)	(0.082)	(0.072)	(0.028)	(0.046)
N	3000	1031	1057	2088	3000
Adj. R^2	0.119	0.08	0.041	0.085	0.011
Post-estimation tests, p-values					
S=ST	0.000	0.458	0.071	0.000	0.052
ST×STP=ST×STT	0.791	0.426	0.862	0.281	0.966

Notes: For Probit regressions, the table reports average marginal effects. Standard errors are clustered at the matching group level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

The overall picture that emerges is that separating agency over prescription and treatment with a regulated fixed fee for prescription increases efficiency through reducing overtreatment, having more patients seeing a physician, and inducing a higher acceptance rate for severe treatments.

However, when physicians receive a zero fee for providing a diagnosis, as in our experiment condition S, these positive effects may be offset by an efficiency loss from more cases of undertreatment and less acceptances of (appropriate) non-severe treatment. Undertreatment has a particularly negative effect on efficiency, since it implies the provision of a (non-severe) treatment at a positive cost with no effect on the health of the patient. Figure 1 decomposes the sources of efficiency losses (compared to the benchmark of received honest treatment) for both experimental conditions S and B. We define efficiency as the sum of payoffs (over both physician and patient) realized in an interaction, relative to the sum of payoffs that would be realized with appropriate treatment of the patient.[14] The figure demonstrates that the positive efficiency effect of less overtreatment in condition S is overshadowed by efficiency losses due to undertreatment. The effect of undertreatment may even be understated by these numbers, as undertreatment may have a 'chilling' effect, such that for fear of undertreatment patients do not take up the treatment in the first place. The efficiency gains of the separation of agency come mainly through *indirect* effects. People who have previously denied to obtain a diagnosis or treatment (such that overtreatment could not take place in the first place) seem to be more willing to engage in the market when physicians' agency is separated.

FIGURE 1: Sources of total efficiency loss relative to fully efficient treatment



Endogenous diagnosis and treatment prices. Based on the theoretical model, our main prediction is that there should be no difference in patient and physician *behavior* between experiment conditions S and the three ST-conditions as well as within the ST-conditions, but that we would expect significant differences across the three ST-conditions in diagnosis and treatment *prices* and consequently in the distribution of profits (see also the upper part of Table 2).

Our results do not line up well with these predictions. Table 2 shows that there are surprisingly little differences *within* the three ST-conditions in terms of patient and physician behavior, prices, and profits. The average diagnosis fee varies between 28.7 and 31.3, the average prices for severe and non-

[14] With our parameters, the efficient total payoffs are $U_n - C_n = 95$ and $U_s - C_s = 80$ for the non-severe and severe illness, respectively. Rejection of diagnosis or treatment result in payoffs of $U_n^0 = 60$ and $U_s^0 = 0$, respectively, which correspond to efficiency losses of 36.8% and 100%, respectively. Overtreatment results in an efficiency loss of 15.8% (1-(80/95)) and undertreatment yields an efficiency loss of 106.25% (1-(-5/80)). To calculate the percentages displayed in Figure 1, we weigh these efficiency losses with the number of the occurrences of the respective outcomes.

severe treatments fluctuate from 50.1 to 51.2 and 23.5 to 28.8, respectively. Correspondingly, average earnings of patients and physicians vary little between the three experiment conditions. This stands in stark contrast to the theoretical analysis, which predicts that the treatment physician extracts all information rents through prices in experiment condition ST-T, that analogously the prescription physician extracts all gains through the diagnosis fee in experiment condition ST-P, while both share equal bargaining power and profits in experiment condition ST-B.[15] We also do not observe large differences in the likelihood of obtaining a diagnosis (54-58%), overtreatment (14-25%), undertreatment (11-18%), and treatment take-up (89-95%).[16]

There are, however, major differences *between* the ST-conditions with endogenous prices for diagnosis and treatment on the one hand and the S condition where the diagnosis fee is regulated to be zero on the other hand. As Table 2 and Figure 1 show and the regression analysis in Table 4 Model 11 statistically confirms, we observe significantly lower efficiency (higher efficiency losses compared to proper treatment) in the ST-conditions than in treatment S (and baseline B). These negative effects are mainly driven by much lower levels of diagnosis take-up rates, which drop from 95% in condition S to less than 60% in experiment conditions ST-B, ST-P, and ST-T (see also Model 7 in Table 4).

The reason for these efficiency losses lies in the price setting behavior of the disintegrated physicians. While the treatment physicians in the ST-conditions charge lower treatment prices than the integrated physician in experiment condition S (Models 2 and 3 in Table 3), the *total* prices which include the diagnosis fee are significantly higher when both physicians set their prices than if the diagnosis fee is regulated (to be zero).

A rational, risk-neutral patient should only take up a diagnosis if his expected profit from doing so is at least as high as his outside option of remaining untreated. Even when assuming to be treated honestly, the patient will reject the diagnosis if the expected price to be paid in total for diagnosis and treatment is larger than 70, which equals the expected loss from not being treated from the two equally likely illnesses. Figure 2 displays the empirically estimated likelihood to obtain diagnosis, conditional on the expected total price under honest treatment, in our five experiment conditions.[17] We observe that patients are very likely to obtain a diagnosis when the expected price is below 70, and much less likely to do so when the price is 70 or above. However, we do not observe large differences in this behavior across our different experiment conditions.

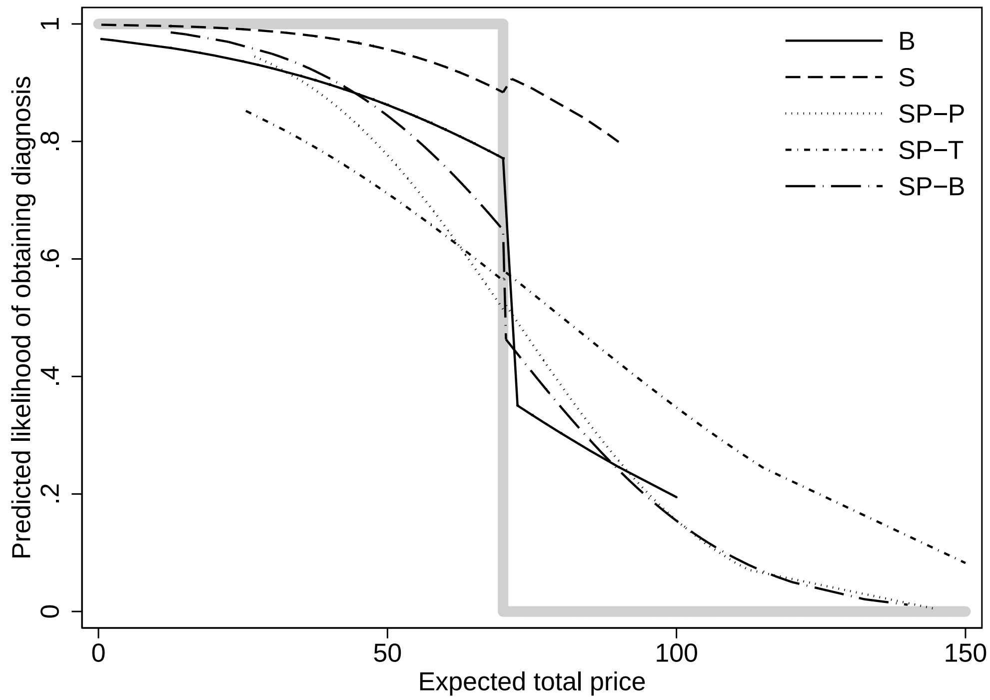
Figure 3 shows scatterplots, one for each ST condition. On the y-axis of each plot we denote the diagnosis fee set by the prescription physician. On the x-axis, we denote the expected profit of the treatment physician (equalling expected income from chosen treatment prices minus expected

[15]The regressions reported in Tables 3 and 4 support this result, with a few minor exceptions: The price for the non-severe treatment is about 5 points lower in ST-P than in ST-B and ST-T, and the total price for the severe treatment (including diagnosis fee) is about 4 points lower in ST-T than in ST-B (but both are not different to ST-P).

[16]Statistically, the regressions reported in Table 4 indicate that the treatment take-up rate in experiment condition ST-T is lower than in experiment condition ST-B (but both are not different to ST-P).

[17]Over all experiment conditions, we estimate a Probit model that includes the expected price as well as a dummy for the expected price being above 70 as independents, and allows these independents to vary by treatment. Thus, the estimation allows for a structural break at a price of 70 in terms of an absolute shift in the likelihood to obtain diagnosis.

FIGURE 2: Likelihood of diagnosis take-up conditional on expected profit (assuming honest treatment)

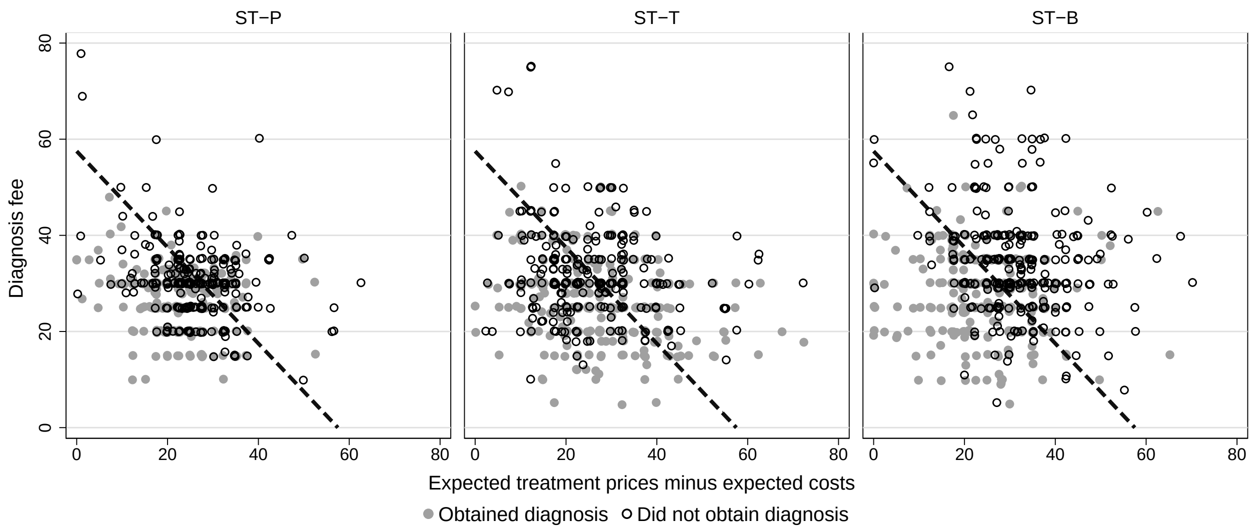


Note: The grey solid line represents the behavior of a risk-neutral rational patient who expects to be treated honestly.

treatment costs), assuming that the prescription physician provides an honest prescription. Each dot represents the price offers of the physicians in a physician-physician-patient triad in one round. The line connecting the points (0, 57.5) and (57.5, 0) represents the break-even points for the patient: if he expects honest treatment, he should accept price combinations below that line and reject price combinations above that line. Grey dots are the offers that were indeed accepted by a patient, black circles are price combinations that were indeed rejected. The Figure shows that many offers from the physicians are located above the break-even line. That is, the physicians fail to coordinate on a set of prices which make it worthwhile for the patient to obtain a diagnosis. In fact, 55% (34%, 41%) of price sets lie above the break-even line in experiment conditions ST-B (ST-P, ST-T). Correspondingly, many of the offers were rejected by the patient, by not obtaining a diagnosis.

Effects of demographics, learning, and experience. All our results are robust to including controls for demographic characteristics of our subjects, such as age, gender, or field of study. While Hennig-Schmidt and Wiesen (2014) find that medical students are more patient-regarding than non-medical students, we do not find such an effect. The 11 medical students in our sample seem not to behave differently than their peers in other majors.

FIGURE 3: Acceptances and rejections for different sets of prices in ST experiment conditions



Note: The dashed line represents the break-even line, below which a patient who assumes honest treatment should obtain diagnosis, and above which he should reject it.

Despite some coordination and convergence in the first 5 rounds, we do not detect strong effects over time that would indicate some kind of learning or unraveling in the market. The Supplementary Appendix A contains a number of figures which show the development of key outcome variables (treatment prices, diagnosis fees, frequencies of diagnosis and treatment take-up as well as over- and undertreatment, and efficiency) over the course of the 20 rounds in our experiment. A related question is how patients react to being over- or undertreated. Table 5 shows the results of Probit estimations of the likelihood to obtain a diagnosis, not only depending on the diagnosis fee and treatment prices being asked for by the physician(s), but also conditional on the frequency of having been under- or overtreated in previous rounds (relative to having received a prescription of a non-severe and severe treatment, respectively). For clarity, we run these estimations separately by experiment condition. As expected and shown before, the higher the prices for diagnosis and treatment, the lower the likelihood of the patient taking up a diagnosis (all price effects are estimated to be negative, albeit not all of them turn out to be statistically significant). Interestingly, except for the baseline treatment B, the effect of having been overtreated in previous rounds is not negative. The effect of having been undertreated (receiving a costly but ineffective treatment when having a severe illness) is negative in all conditions (but statistically significant only in two of them).[18]

[18]We ran similar regressions to explain treatment take-up based on previous experience of under- and overtreatment, and report results in Supplementary Appendix A, Table 6. We find that previously experienced undertreatment negatively affects the take-up of a non-severe treatment in all experiment conditions (statistically significant in all treatments except baseline), while previously experienced overtreatment negatively affects the take-up of a severe illness only in conditions B and S. These regression results, however, have to be interpreted cautiously, since they ignore the selection due to diagnosis take-up examined above.

TABLE 5: Diagnosis take-up conditional on prices and previous experience

Model	(1)	(2)	(3)	(4)	(5)
	Probit	Probit	Probit	Probit	Probit
Experiment condition	B	S	ST-B	ST-P	ST-T
Diagnosis fee			-0.164***	-0.173***	-0.183***
			(0.034)	(0.041)	(0.064)
Non-severe treatment price	-0.058***	-0.014	-0.049	-0.067*	
	(0.012)	(0.011)	(0.032)	(0.038)	
Severe treatment price	-0.075***	-0.028*	-0.146***	-0.097***	
	(0.020)	(0.016)	(0.023)	(0.010)	
Frequency of overtreatment in rounds 1 to $t-1$	-0.321***	-0.008	0.030	0.040	0.188
	(0.094)	(0.035)	(0.096)	(0.211)	(0.222)
Frequency of undertreatment in rounds 1 to $t-1$	-0.070	-0.124**	-0.112	-0.302*	-0.154
	(0.135)	(0.057)	(0.095)	(0.162)	(0.171)
N	427	493	420	435	457
Pseudo R^2	0.307	0.263	0.181	0.084	0.091

The table reports average marginal effects. Standard errors are clustered at the matching group level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

V Conclusions

We experimentally tested the effect of separating agency over prescription and treatment in a credence good market. In general, and in line with previous findings, we observe that physicians are more honest than predicted by theory, in particular in cases where they could make more profits from prescribing a less efficient treatment.[19] Nevertheless, they respond to incentives. Compared to the integrated physicians in our baseline condition B, we observe less overtreatment in the condition where prescription and treatment are separated and the prescription physician provides her services for free. This results in (statistically not significant) higher efficiency, also due to higher diagnosis and treatment take-up. These positive effects are limited, however, by the occurrence of more undertreatment. When we allow both physicians to set prices independently (experiment conditions ST-x), we observe no changes in over- or undertreatment, but significantly lower diagnosis take-up and thus lower efficiency. The latter roots in coordination failures in the price setting process. The prices set by the physicians are so high such that in total they exceed the benefit patients can expect from attending to the

[19] This result is consistent with empirical evidence that physicians take patients' benefits into consideration and forgo profits (e.g. Kolstad, 2013; Lagarde and Blaauw, 2017). However, different from findings in Hennig-Schmidt and Wiesen (2014), the few medical students in our experiment did not behave differently from other students.

physician and receiving proper treatment. While experiment conditions ST-P, ST-T, and ST-B assign most bargaining power to the prescription physician, to the treatment physician, and to both equally, respectively, we do not observe significant effects of these different bargaining power conditions on prices and profits.

Our main results fit consistently into the existing experimental literature on credence goods. The observation that physicians are heterogeneous in their response to overtreatment incentives has also been made in Beck, Kerschbamer, Qiu and Sutter (2013); Dulleck et al. (2011); Gottschalk et al. (2017); Kerschbamer et al. (2016), with a particular discussion and estimation of heterogeneity in Kerschbamer et al. (2017). Nevertheless, agents respond to incentives. Our finding that a separation of prescription and treatment reduces overtreatment is closely related to the observation in Mimra et al. (2016*b*) that the option of considering second opinions helps making physicians more honest, and the finding of Huck et al. (2016) that competition from other physicians increases market efficiency. Unpredicted by theory but consistent with other experimental studies (e.g. Dulleck et al., 2011), we observe a substantial amount of undertreatment in our experiment. We attribute the increased amount of undertreatment in condition ST to the fact that physicians earn a mandated fee of zero, which may have made them less inclined towards patients.

Our findings complement evidence from the empirical literature that a separation of prescription and treatment may be effective in reducing overtreatment. These benefits will have to be weighed against the costs of implementing and enforcing such separation. Our experiment also highlights the importance of carefully considering the market institutions that frame the separated credence good market. On the one hand, regulating diagnosis fees to a fixed amount or even zero may reduce the motivation of prescription physicians, leading to less careful or even malicious treatment suggestions. On the other hand, allowing physicians to freely set their prices may lead to market failures in that patients, who feel exploited through too high prices, may refuse to attend to health care in the first place.

Our theoretical and experimental framework assumes away many details and complexities of real-world health care markets, such as health insurance and adverse selection, competition and reputation, etc. For example, we do not model any uncertainty on the physician's side in terms of imperfect diagnosis. Combined with prosocial preferences towards patients, such uncertainties may have large effects in the health care market. Future studies may extent our framework to such richer environments. We believe that this and further laboratory experiments will prove useful in studying health care markets.

References

Afendulis, C. C. and Kessler, D. P. (2007), 'Tradeoffs from integrating diagnosis and treatment in markets for health care', *The American Economic Review* **97**(3), 1013–1020.

Arrow, K. J. (1963), 'Uncertainty and the welfare economics of medical care', *The American Economic Review* **53**(5), 941–973.

Baker, L. C. (2010), 'Acquisition of MRI equipment by doctors drives up imaging use and spending', *Health Affairs* **29**(12), 2252–2259.

Balafoutas, L., Beck, A., Kerschbamer, R. and Sutter, M. (2013), 'What drives taxi drivers? A field experiment on fraud in a market for credence goods', *The Review of Economic Studies* **80**(3), 876–891.

Beck, A., Kerschbamer, R., Qiu, J. and Sutter, M. (2013), 'Shaping beliefs in experimental markets for expert services: guilt aversion and the impact of promises and money-burning options', *Games and Economic Behavior* **81**, 145–164.

Beck, A., Kerschbamer, R., Qiu, J. and Sutter, M. (2014), 'Car mechanics in the lab–investigating the behavior of real experts on experimental markets for credence goods', *Journal of Economic Behavior and Organization* **108**, 166–173.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2015), 'The effects of introducing mixed payment systems for physicians: experimental evidence', *Health Economics* **26**(2), 243–262.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2016), 'Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision', *Journal of Economic Behavior and Organization* **131**, 17–23.

Chen, B. K., Gertler, P. J. and Yang, C.-Y. (2016), 'Physician ownership of complementary medical services', *Journal of Public Economics* **144**, 27–39.

Chou, Y., Yip, W. C., Lee, C.-H., Huang, N., Sun, Y.-P. and Chang, H.-J. (2003), 'Impact of separating drug prescribing and dispensing on provider behaviour: Taiwan's experience', *Health Policy and Planning* **18**(3), 316–329.

Darby, M. R. and Karni, E. (1973), 'Free competition and the optimal amount of fraud', *The Journal of Law and Economics* **16**(1), 67–88.

Dranove, D. (1988), 'Demand inducement and the physician/patient relationship', *Economic Inquiry* **26**(2), 281–298.

Dulleck, U. and Kerschbamer, R. (2006), 'On doctors, mechanics, and computer specialists: the economics of credence goods', *Journal of Economic Literature* **44**(1), 5–42.

Dulleck, U., Kerschbamer, R. and Sutter, M. (2011), 'The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition', *The American Economic Review* **101**(2), 526–555.

Eggleston, K. (2012), 'Prescribing institutions: explaining the evolution of physician dispensing', *Journal of Institutional Economics* **8**(2), 247–270.

Emons, W. (1997), 'Credence goods and fraudulent experts', *The RAND Journal of Economics* **28**(1), 107–119.

Emons, W. (2001), 'Credence goods monopolists', *International Journal of Industrial Organization* **19**(3), 375–389.

Fong, Y.-f., Liu, T. and Wright, D. J. (2014), 'On the role of verifiability and commitment in credence goods markets', *International Journal of Industrial Organization* **37**, 118–129.

Fuchs, V. R. (1978), 'The supply of surgeons and the demand for operations', *Journal of Human Resources* **13**, 35–56.

Godager, G., Henning-Schmidt, H. and Iversen, T. (2016), 'Does performance disclosure influence physicians' medical decisions? an experimental analysis', *Journal of Economic Behavior and Organization* **131**(Part B), 36–46.

Gottschalk, F., Mimra, W. and Waibel, C. (2017), 'Health services as credence goods: a field experiment', Mimeo.

Green, E. P. (2014), 'Payment systems in the healthcare industry: an experimental study of physician incentives', *Journal of Economic Behavior and Organization* **106**, 367–378.

Greiner, B. and Zhang, L. (2017), 'On the role of commitment and two-part tariffs in credence good markets: An experimental study in behavioral mechanism design', Mimeo.

Gruber, J., Kim, J. and Mayzlin, D. (1999), 'Physician fees and procedure intensity: the case of cesarean delivery', *Journal of Health Economics* **18**(4), 473–490.

Hayase, Y. (2003), 'Problems of the separation of prescription and dispensing', *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan* **123**(3), 121–132.

Hennig-Schmidt, H., Selten, R. and Wiesen, D. (2011), 'How payment systems affect physicians' provision behaviour? an experimental investigation', *Journal of Health Economics* **30**(4), 637–646.

Hennig-Schmidt, H. and Wiesen, D. (2014), 'Other-regarding behavior and motivation in health care provision: an experiment with medical and non-medical students', *Social Science & Medicine* **108**, 156–165.

Huck, S., Lünser, G., Spitzer, F. and Tyran, J.-R. (2016), 'Medical insurance and free choice of physician shape patient overtreatment: a laboratory experiment', *Journal of Economic Behavior and Organization* **131**, 78–105.

Hughes, D. and Yule, B. (1992), 'The effect of per-item fees on the behaviour of general practitioners', *Journal of Health Economics* **11**(4), 413–437.

Iizuka, T. (2007), 'Experts' agency problems: evidence from the prescription drug market in Japan', *The RAND Journal of Economics* **38**(3), 844–862.

Iizuka, T. (2012), 'Physician agency and adoption of generic pharmaceuticals', *The American Economic Review* **102**(6), 2826–2858.

Kaiser, B. and Schmid, C. (2016), 'Does physician dispensing increase drug expenditures? empirical evidence from Switzerland', *Health Economics* **25**(1), 71–90.

Kerschbamer, R., Neururer, D. and Sutter, M. (2016), 'Insurance coverage of customers induces dishonesty of sellers in markets for credence goods', *Proceedings of the National Academy of Sciences* **113**(27), 7454–7458.

Kerschbamer, R., Sutter, M. and Dulleck, U. (2017), 'How social preferences shape incentives in (experimental) markets for credence goods', *Economic Journal* **127**, 393–416.

Kolstad, J. T. (2013), 'Information and quality when motivation is intrinsic: Evidence from surgeon report cards', *American Economic Review* **103**(7), 2875–2910.

Kwon, S. (2003), 'Pharmaceutical reform and physician strikes in Korea: separation of drug prescribing and dispensing', *Social Science & Medicine* **57**(3), 529–538.

Lagarde, M. and Blaauw, D. (2017), 'Physicians' responses to financial and social incentives: A medically framed real effort experiment', *Social Science & Medicine* **179**, 147–159.

Lee, E. K. and Malone, D. C. (2003), 'Comparison of peptic-ulcer drug use and expenditures before and after the implementation of a government policy to separate prescribing and dispensing practices in South Korea', *Clinical Therapeutics* **25**(2), 578–592.

Lim, D., Emery, J., Lewis, J. and Sunderland, V. B. (2009), 'A systematic review of the literature comparing the practices of dispensing and non-dispensing doctors', *Health Policy* **92**(1), 1–9.

Lu, F. (2014), 'Insurance coverage and agency problems in doctor prescriptions: evidence from a field experiment in China', *Journal of Development Economics* **106**, 156–167.

Lundin, D. (2000), 'Moral hazard in physician prescription behavior', *Journal of Health Economics* **19**(5), 639–662.

McGuire, T. G. (2000), 'Physician agency', *Handbook of Health Economics* **1**, 461–536.

Mimra, W., Rasch, A. and Waibel, C. (2016*a*), 'Price competition and reputation in credence goods markets: experimental evidence', *Games and Economic Behavior* **100**, 337–352.

Mimra, W., Rasch, A. and Waibel, C. (2016*b*), 'Second opinions in markets for expert services: experimental evidence', *Journal of Economic Behavior and Organization* **131**, 106–125.

Rodwin, M. A. and Okamoto, A. (2000), 'Physicians' conflicts of interest in Japan and the United States: lessons for the United States', *Journal of Health Politics, Policy and Law* **25**(2), 343–375.

Waibel, C. and Wiesen, D. (2016), 'Kickbacks, referrals and efficiency in health care markets: experimental evidence', Working Paper.
URL: *https://ssrn.com/abstract=2864336*

Wolinsky, A. (1993), 'Competition in a market for informed experts' services', *The RAND Journal of Economics* **24**(3), 380–398.

Wolinsky, A. (1995), 'Competition in markets for credence goods', *Journal of Institutional and Theoretical Economics* **151**(1), 117–131.

A Additional figures and tables

FIGURE 4: Prices and profits over time

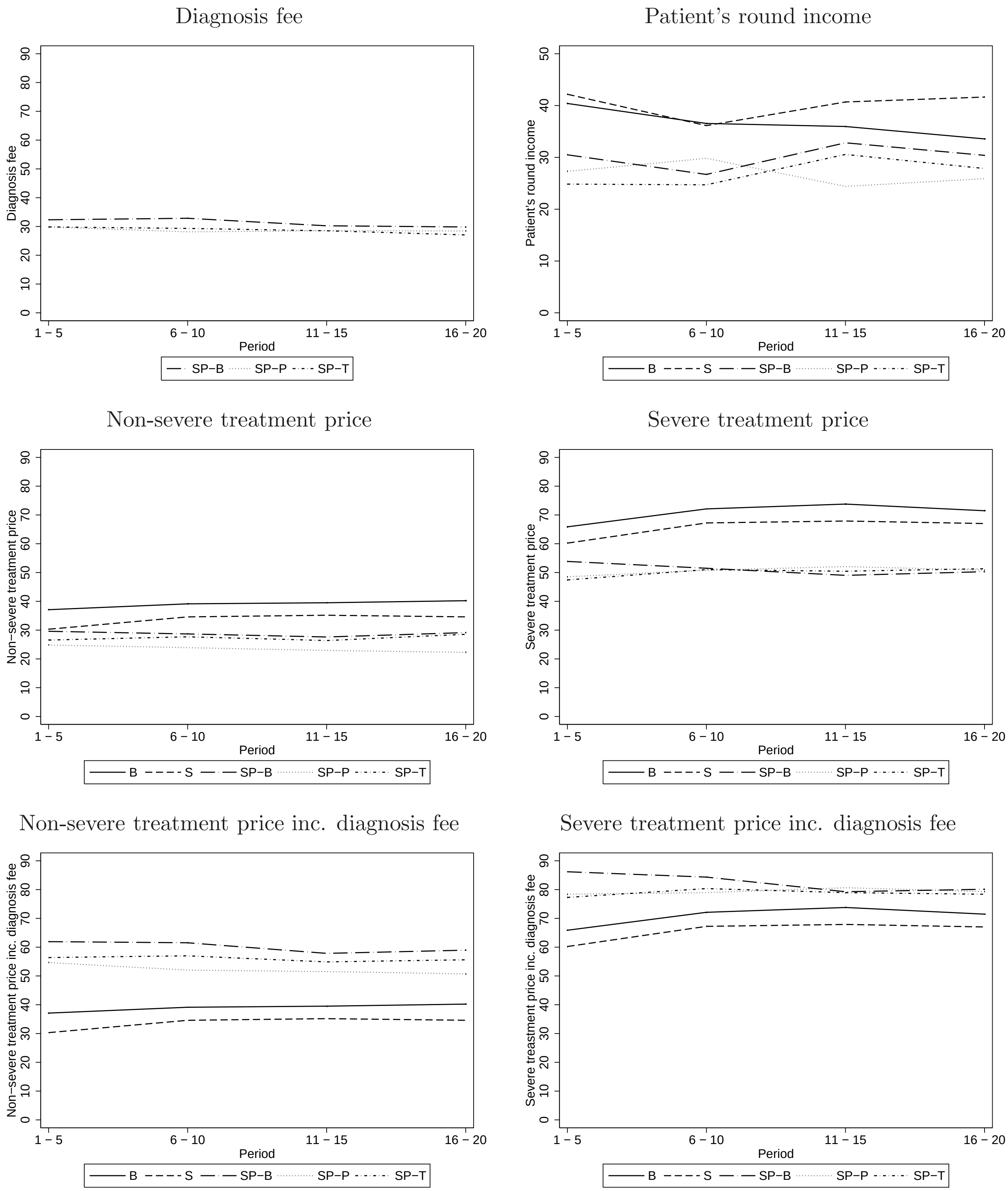


FIGURE 5: Diagnosis and treatment take-up, treatment choices, and efficency over time

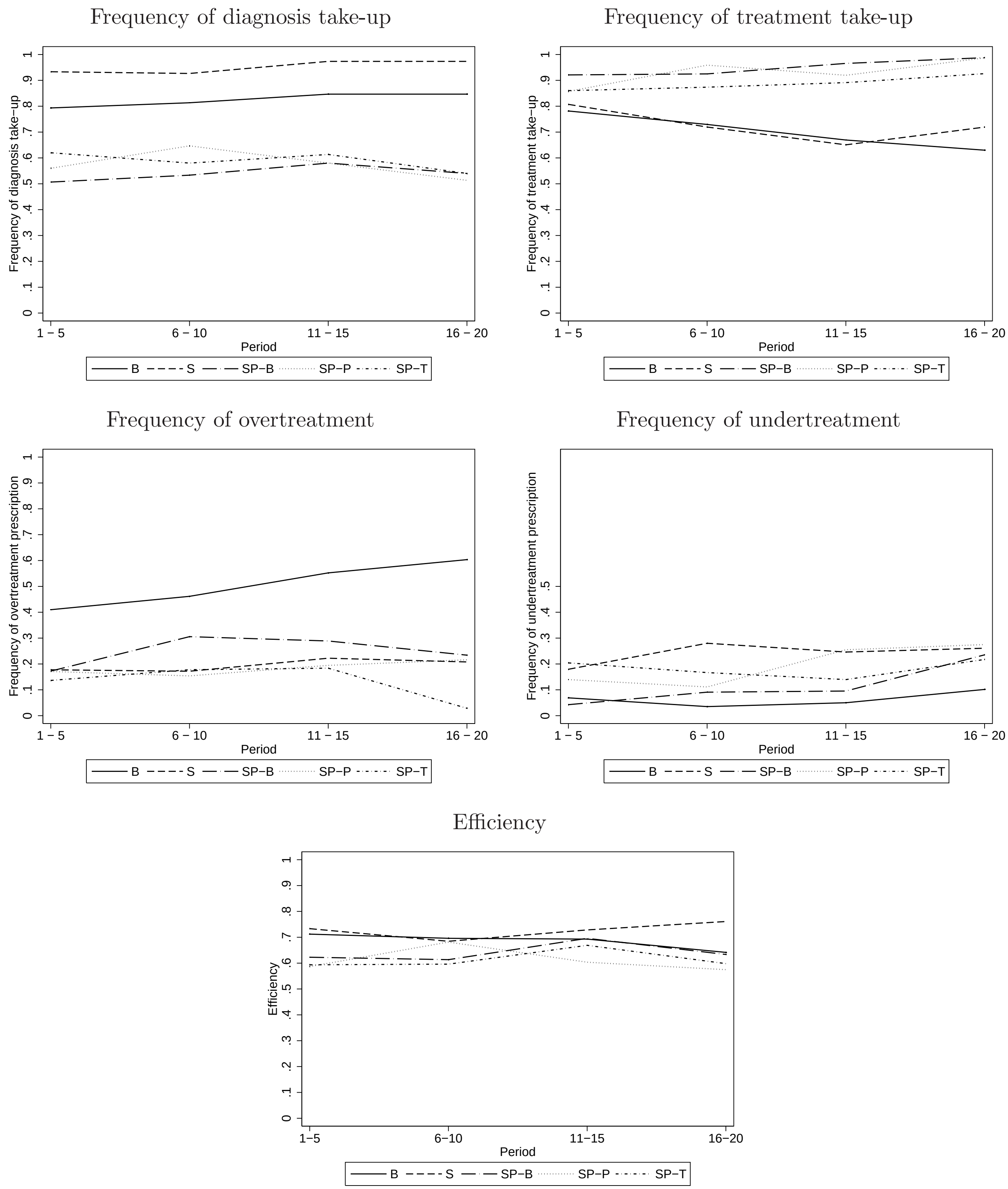


TABLE 6: Treatment take-up conditional on prices and previous experience

Non-severe treatment	Probit B	Probit S	Probit ST-B	Probit ST-P	Probit ST-T
Diagnosis fee			-0.025	-0.001	-0.043
			(0.016)	(0.044)	(0.040)
Non-severe treatment price	-0.136*	-0.166***	-0.030	-0.042	-0.076**
	(0.074)	(0.059)	(0.018)	(0.047)	(0.035)
Severe treatment price	0.025	0.031	-0.002	0.012	-0.012
	(0.043)	(0.041)	(0.003)	(0.021)	(0.028)
Frequency of undertreatment	-0.417	-0.604***	-0.106***	-0.164***	-0.231***
in rounds 1 to $t-1$	(0.510)	(0.114)	(0.025)	(0.056)	(0.086)
N	111	267	105	141	152
Pseudo R^2	0.133	0.14	0.377	0.106	0.218

Severe treatment	Probit B	Probit S	Probit ST-B	Probit ST-P	Probit ST-T
Diagnosis fee			-0.029*	-0.034	-0.078**
			(0.015)	(0.027)	(0.037)
Non-severe treatment price	-0.008	0.007	0.004	-0.024*	-0.002
	(0.021)	(0.023)	(0.012)	(0.014)	(0.031)
Severe treatment price	-0.188***	-0.112***	-0.042**	-0.044*	-0.078***
	(0.007)	(0.028)	(0.019)	(0.023)	(0.021)
Frequency of overtreatment	-0.252**	-0.309***	-0.026	0.035	-0.028
in rounds 1 to $t-1$	(0.129)	(0.057)	(0.047)	(0.101)	(0.073)
N	325	244	159	145	142
Pseudo R^2	0.181	0.262	0.142	0.244	0.208

The table reports average marginal effects. Standard errors are clustered at the matching group level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

B EXPERIMENTAL INSTRUCTIONS

Thank you for participating in this experiment. Please do not to talk to any other participant until the experiment is over. Please switch off your mobile phone and put it away. If you do not adhere to these rules, we will have to exclude you from any payments.

Your identity in this experiment will not be revealed to any other participant, and the identity of other participants will not be revealed to you. In this sense, your decisions are anonymous.

Your earnings in this experiment are determined by your decisions and the decisions of other participants. At the end of the session, the points you earned will be converted into cash using an exchange rate of 30 points = 1 AUD. Your final payoff is a $5 show up fee plus the initial endowments minus the loss from all rounds.

The experiment will run for 20 rounds. In each round, you will be randomly matched to a group of **[B:** two**; S/ST-P/ST-T/ST-B:** three**]** participants. One of the participants will be a patient, and the **[B:** other participant will be a doctor**; S/ST-P/ST-T/ST-B:** other two participants will be doctors, Doctor 1 and Doctor 2**]**. If you are a patient, you will be a patient in all 20 rounds; if you are a doctor, you will be a doctor in all 20 rounds. Each round will proceed according to the following rules.

Overview

A patient starts a round with an **endowment of 100 points**. However, the patient will contract an illness, which is either a **minor illness** or a **major illness**. The type of the illness is randomly determined, with a **50%** chance of a minor illness and a **50%** chance of a major illness. Imagine a coin-flip: if the result is "head", the Patient has a minor illness; if the result is "tail", he has a major illness. If the illness **remains untreated**, the Patient will **lose 40 points if the illness is minor** and **lose 100 points if the illness is major**.

The Patient doesn't know whether he has a minor illness or a major illness until the end of the round. **[B:** The Doctor**; S/ST-P/ST-T/ST-B:** Doctor 1**]** will find out whether it is minor or major when she diagnoses the Patient.

[B: The Doctor**; S/ST-P/ST-T/ST-B:** Doctor 2**]** can treat the Patient's illness by **[B:** prescribing**; S/ST-P/ST-T/ST-B:** providing**]** either a **minor treatment** or a or **major treatment[S/ST-P/ST-T/ST-B:** according to Doctor 1's prescription**]**. The minor treatment remedies **only the minor illness** and imposes a cost of 5 points on the doctor, while the major treatment that is capable to remedy **any illness** (minor or major) and imposes a cost of 20 points on the doctor.

[B: The Doctor will charge for the treatment, and will post the prices for the two treatments before the patient decides to get a diagnosis. After he receives the diagnosis,

the Patient can decide whether or not to receive the treatment the doctor prescribes.]
[S: Doctor 1 will provide the diagnosis for free. Doctor 2 will charge for the treatment, and will post the prices for the two treatments before the patient decides to get a diagnosis. After he receives the diagnosis, the patient can decide whether or not to receive the treatment the doctor prescribes.]
[ST-P: Doctor 1 will charge for the diagnosis, and Doctor 2 will charge for the treatment. Doctor 1 will post the diagnosis fee first, then Doctor 2 will post the prices for the two treatments. Both will post their prices before the patient decides to get a diagnosis. After he receives the diagnosis, the patient can decide whether or not to receive the treatment the doctor prescribes.]
[ST-T: Doctor 1 will charge for the diagnosis, and Doctor 2 will charge for the treatment. Doctor 1 will post the diagnosis fee, before the patient decides to get a diagnosis. After he receives the diagnosis, Doctor 2 will post the prices for the two treatments, and the patient can decide whether or not to receive the treatment the doctor prescribes.]
[ST-B: Doctor 1 will charge for the diagnosis, and Doctor 2 will charge for the treatment. Doctor 1 will post the diagnosis fee and at the same time Doctor 2 will post the prices for the two treatments. Both will post their prices before the patient decides to get a diagnosis. After he receives the diagnosis, the patient can decide whether or not to receive the treatment the doctor prescribes.]

Detailed decisions procedures

Each round consists of multiple stages.

At the beginning of each round, the computer determines randomly whether the Patient has a minor illness or a major illness. Each illness has a probability of 50% to be selected. At this point of time, both Doctor and Patient are not informed yet about the type of illness of the Patient.

1. [ST-B: **Diagnosis fee and treatment prices**
   Simultaneously, Doctor 1 posts the *diagnosis fee* and Doctor 2 posts two treatment prices for the Patient: *PriceMinor* for Minor treatment and *PriceMajor* for Major treatment.]
2. [ST-P/ST-T: **Diagnosis fee**
   Doctor 1 posts a *diagnosis fee* for the Patient.
   ]
3. [B/S/ST-P: **Treatment prices**
   [ST-P: Doctor 2 observes Doctor 1's *diagnosis fee* for the Patient.]
   [B: The Doctor; S/ST-P: Doctor 2] posts two treatment prices for the Patient: *PriceMinor* for Minor treatment and *PriceMajor* for Major treatment.
   ]
4. **Patient decides to get diagnosis or not**
   The Patient observes the [ST-T: diagnosis fee; ST-P/ST-B: diagnosis fee and the] [B/S/ST-P/ST-B: two treatment prices: PriceMinor and PriceMajor]. The Patient

decides whether or not to get the diagnosis **[ST-P/ST-T/ST-B:** (and pay the diagnosis fee)**]** in this round.

If the Patient chooses NOT to see the Doctor, then this round ends in this group. **[B:** The Doctor**; S/ST-P/ST-T/ST-B:** Both Doctor 1 and Doctor 2**]** will receive 0 points in this round, and the Patient will receive his endowment of 100 points minus the costs of the untreated illness: either 20 points (if he has the minor illness) or 80 points (if he has the major illness).

If the Patient chooses to see the Doctor, then the round continues**[ST-P,ST-T,ST-B:** and the Patient pays the *diagnosis fee* to Doctor 1**]**.

5. **Diagnosis and treatment prescription**
   [B: The Doctor**; S/ST-P/ST-T/ST-B:** Doctor 1**]** diagnoses the Patient and is now informed about the type of illness that Patient has. The Patient remains uninformed. **[B:** The Doctor**; S/ST-P/ST-T/ST-B:** Doctor 1**]** prescribes either a *minor treatment* or a *major treatment*.

6. **[ST-T: Treatment prices**
   Doctor 2 observes Doctor 1's *diagnosis fee* for the Patient. Before knowing the actual treatment prescription, Doctor 2 posts two treatment prices for the Patient: *PriceMinor* for Minor treatment and *PriceMajor* for Major treatment.
   ]
7. **Patient decides to receive treatment or not**
   The Patient is informed about the treatment prescription **[S/ST-P/ST-T/ST-B:** of Doctor 1**]** and **[B/S/ST-P/ST-B:** is reminded of**]** the treatment prices **[S/ST-P/ST-T/ST-B:** of Doctor 2**]**. The Patient chooses whether or not to receive the prescribed treatment from **[B:** the Doctor**; S/ST-P/ST-T/ST-B:** Doctor 2**]**.

   If the Patient chooses NOT to receive the treatment, then this round ends in this group. **[B:** The Doctor**; S:** Both Doctor 1 and Doctor 2**; ST-P/ST-T/ST-B:** Doctor 1 will receive the diagnosis fee and Doctor 2**]** will receive 0 points in this round. The Patient will receive his endowment of 100 points **[ST-P,ST-T,ST-B:** minus the *diagnosis fee***]** minus the costs of the untreated illness: either 40 points (if he has the minor illness) or 100 points (if he has the major illness).

   If the Patient chooses to receive the treatment, then the round continues and, depending on the prescription, the Patient pays *PriceMinor* or *PriceMajor* to **[B:** the Doctor**; S/ST-P/ST-T/ST-B:** Doctor 2**]]**.
8. **Treatment provision**
   [B: The Doctor**; S/ST-P/ST-T/ST-B:** Doctor 2**]** provides the prescribed treatment to the Patient and receives the price she posted for that treatment (*PriceMinor* or *PriceMajor*, depending on the prescription).

After this last stage, the round payoffs will be as follows:

[B: The Doctor will receive

Provided minor treatment		Provided major treatment	
PriceMinor		***PriceMajor***	
minus			
costs of *5 points*		**costs of *20 points***	

]

[S/ST-P/ST-T/ST-B: Doctor 1 will receive [S: ***0 points***; ST-P/ST-T/ST-B: **the *diagnosis fee***].

Doctor 2 will receive

Provided minor treatment		Provided major treatment	
PriceMinor		***PriceMajor***	
minus			
costs of *5 points*		**costs of *20 points***	

]

The Patient will receive

his endowment of **100 points**

[ST-P,ST-T,ST-B: **minus the *diagnosis fee***]

minus

Received minor treatment	Received major treatment
PriceMinor	***PriceMajor***

minus

Received minor treatment		Received major treatment	
Illness was minor	Illness was major	Illness was minor	Illness was major
0 points	***100 points***	***0 points***	***0 points***

This ends the round. At the end of the round, the Patient and [B: the Doctor; S/ST-P/ST-T/ST-B: the two Doctors] are informed about the illness of the patient, the decisions of each participant in the group, and their payoffs.

Your final payoff will be your payoff from all 20 rounds.

[S/ST-P/ST-T/ST-B:

Note

As described above, each randomly matched group will consist of three participants, one Patient and two doctors, Doctor 1 and Doctor 2. However, in each round, each doctor will take part in two different groups: in one group she will take over the role of Doctor 1 and diagnose the Patient in that group, and in the other group she will be Doctor 2 and provide treatment to the Patient in that group. A doctor's round earnings will consist of her earnings from both groups. The computer will make sure that a doctor interacts with a different other doctor (and a different patient) in the two groups.
Patients will only be matched to one group in each round, and meet one Doctor 1 and one Doctor 2.]