



**Kickbacks, referrals and
efficiency in health care
markets: Experimental
evidence**

Christian Waibel

Department of Management,
Technology and Economics,
ETH Zurich
Switzerland

Daniel Wiesen

Department of Health Management
and Health Economics,
Institute of Health and Society,
University of Oslo
&
Department of Business
Administration and Healthcare
Management,
University of Cologne
Germany

**UNIVERSITY
OF OSLO**

HEALTH ECONOMICS
RESEARCH NETWORK

Working paper 2016: 8

ISSN 1501-9071 (print version)
ISSN 1890-1735 (online)
ISBN 978-82-7756-257-5

HERO

Kickbacks, referrals and efficiency in health care markets: Experimental evidence*

Christian Waibel Daniel Wiesen

November 1, 2016

Abstract

We analyse the causal effect of kickbacks (referral payments) on general practitioners' behaviours and efficiency. In a stylized model, we derive behavioural predictions for general practitioners' diagnostic efforts and referrals to secondary care (specialized physicians), which we test in a series of controlled laboratory experiments. We exogenously vary the level of regulated referral payments in our experimental treatments. We find that introducing referral payments significantly improves efficiency. An increase in payments leads to less undertreatment of severely ill patients, but also to more unnecessary referrals of mildly ill patients. The net effect is positive, as the former outweighs the latter. Interestingly, the increase in efficiency is mainly driven by behavioural changes of barely altruistic general practitioners.

JEL classification: I11, D47, C91

Keywords: Kickback, referral payment, diagnostic effort, referrals, altruism, efficiency, laboratory experiment

*Waibel: Department of Management, Technology, and Economics, ETH Zurich, Zürichbergstrasse 18, 8092 Zurich, Switzerland (e-mail: cwaibel@ethz.ch); Wiesen: Department of Business Administration and Healthcare Management, University of Cologne, Albertus-Magnus-Platz, 50223 Cologne, Germany, and Institute of Health and Society, University of Oslo, Norway (e-mail: wiesen@wiso.uni-koeln.de). For helpful comments and suggestions, we thank Jeannette Brosig-Koch, Geir Godager, Jerry Hurley, Olena Izhak, Ludwig Kuntz, Albert Ma, Wanda Mimra, Lise Rochaix, Luigi Siciliani and Achim Wambach, as well as seminar and conference participants at Basel, Cologne, ETH Zurich, ASHEcon Philadelphia, dggö Berlin, and EuHEA Hamburg. We thank Emanuel Castillo, Anne Classen, Lena Kuhne, and Laurin Stenz for their excellent programming, research assistance and help in conducting the experiments. Financial support by the Research Council of Norway (IRECOHEX, Project-No. 231776), the European Centre for Liberal Professions, and the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

1 Introduction

Inefficiencies are pervasive in healthcare markets (Garber and Skinner, 2008). A major source of inefficiency is the general practitioners' (GP) inappropriate referrals of patients to specialists (e.g., Scott, 2000; Chandra and Skinner, 2012). To enhance efficiency, referral payments (or kickbacks) from the specialist to the GP are often described as an effective means in the theoretical economics literature (see, e.g., Pauly, 1979; Garicano and Santos, 2004; Inderst and Ottaviani, 2012). Surprisingly, referral payments from specialists to GPs are prohibited in most healthcare markets, as they are considered to be unethical.¹ In the UK, for example, incentives for GPs might even be such that referrals to specialist care are discouraged, which could imply considerable adverse effects for patients' health.²

A prohibition of referral payments in healthcare markets seems too broad. According to Pauly's (1979) seminal paper, inappropriate referrals from the GP to the specialist have two sides: On the one hand, referral payments encour-

¹The World Medical Association's Statement on Conflict of Interest states: "*Kickbacks (or fee-splitting) occur when a physician receives financial consideration for referring a patient to a specific practitioner or for a specific service for which a fee is charged. This practice is not acceptable.*" For more, see: <http://www.wma.net/en/30publications/10policies/i3>. In the US, for example, an anti-kickback statute in the Social Security Act (§1128B) prohibits payments to reward referrals of patients participating in federal healthcare programs. The so-called 'Stark Law' prevents physicians from referring Medicare and Medicaid patients for health services from entities in which physicians have a financial relationship (see, e.g., Morreim, 1989; Wilensky et al., 2007).

²In light of the current pressure to make substantial savings (£22bn by 2020), a recent target of leaders of the National Health Service is to cut patient referrals that are deemed inappropriate, particularly when patients could be cared for locally. Helen Briggs of BBC News writes on October 2, 2015: "*At least nine clinical commissioning groups (CCGs) were offering GPs payments to reduce referrals to specialists. In one case, Birmingham South Central CCG was offering practices more than £11,000 to reduce new outpatient attendances, follow-ups, A&E attendances and emergency admissions by 1%, compared with 2014/15. It said the schemes, which exclude cancer referrals, were designed to incentivize best quality practice and drive improvements in the quality of primary medical care.*" For more, see <http://www.bbc.com/news/health-34421115>.

age GPs to refer patients to specialists even if medical conditions suggest they should not. On the other hand, in the absence of referral payments, GPs may *not* refer patients when medically needed. In order to incentivize more efficient referrals, setting the optimal referral payment is therefore crucial (see, e.g., Garcia-Mariñoso and Jelovac, 2003; González, 2010).

Moreover, it is important better to understand GPs' referral behaviour within (vertically integrated) health care organizations. For example, in US Accountable Care Organizations (ACOs),³ which offer medical services through general practitioners and specialized physicians, (internal) referral payments may be an appropriate means to improve efficient referral behaviour (see, e.g., DeCamp and Lehmann, 2015; Grassi and Ma, forthcoming; Song et al., 2014).

Empirically, how referral payments (from specialists to GPs) affect the GPs' behaviour and market efficiency is not well understood. Field data are very limited (due to the prohibition of referral payments) and if available might suffer from multiple confounding effects, due to lack of control and self-selection. For example, GPs' referral payments are adopted endogenously in the field, rendering causal inferences about their effects difficult (for a general discussion, see Falk and Heckman, 2009). Even in field experiments, randomly assigning different referral payments to GPs seems to be prohibitively challenging and might even cause unintended consequences for patients.

How a GP responds to referral payments might also depend on her altruism towards the patient. That a physician is (at least partially) altruistic on behalf of her patient has become a common assumption in the health economics literature since Arrow (1963) coined the importance of a physician's patient-regarding motivation (see, e.g., Ellis and McGuire, 1986; Farley, 1986; Chalkley and Malcolmson, 1998; Allard et al., 2011). Measuring individual physicians' altruism seems difficult, however, in the field due to limited availability of appropriate data.

³Accountable Care Organizations are groups of hospitals, general practitioners and specialized physicians, and other health care providers, who form an organization to provide coordinated medical care to Medicare patients; see, for example, McClellan et al. (2010).

In this paper, we report a comprehensive study comprising theory and laboratory experimentation, in which we establish the causal link between regulated referral payments and the GPs' referral behaviour. Employing a laboratory experiment gives us control over the decision environment and allows us to *exogenously* vary the level of a GP's referral payment. In particular, we address the following research questions: How does the introduction of a referral payment affect a GP's diagnostic effort, a GP's referral decision and the efficiency? How do different levels of referral payments affect a GP's behaviour and the efficiency? How does individual altruism towards patients affect GPs' behaviour?

In a stylized theoretical model, we formalize a GP's behaviour and derive behavioural hypotheses for the GPs who vary in their altruistic motivation towards their patients. In a series of controlled laboratory experiments ($N = 252$), we then analyse the causal effect of referral payments on GPs' diagnostic efforts, referrals, and the efficiency. In particular, we systematically vary the size of the referral payments. This allows us to compare the GPs' behaviours to the baseline (no referral payment), which mirrors the current practice, and for different payment levels. Employing a lab experiment allows us to tightly control the decision environment, to randomly assign subjects to experimental treatments, and to implement clean measures of efficiency. We also elicit individual physicians' altruistic motivations towards their patients using a simple experimental task proposed by Godager et al. (2016). This allows us to distinguish between behaviours of GPs according to their altruistic (patient-regarding) motivation.

The decision situation in our framework comprises four stages. In the first stage, nature randomly determines the patient's severity of illness, which could either be a mild or severe illness. The patient then consults a GP. Both do not observe the severity of the patient's illness. In the second stage, the GP can exert a costly diagnostic effort to learn (with certainty) about the patient's severity. In the third stage, the GP decides whether to treat or to refer the patient. Here, we assume that a GP can only heal a mild illness, whereas the specialist heals both severities, but at a higher cost. When referring the patient, the GP receives a referral payment. For simplicity and to avoid confounding motives for GPs'

behaviours (e.g., due to uncertainty about a specialist’s treatment decision), we assume that specialists always provide the utility-maximizing treatment for a patient. For treating a patient, both the GP and the specialist receive a lump-sum payment.

In our medically-framed experiment, we vary the referral payment at a between-subject level—being either zero, low or high. The decision situation closely mimics our theoretical model. According to our stylized model, we hypothesize that: First, GPs are more likely to exert diagnostic effort under a low referral payment; second, GPs refer more patients the higher the referral payment is; and, third, efficiency increases when introducing a referral payment and is highest under the low referral payment.

Behavioural data from our experiment show that introducing referral payments significantly improves the efficiency if payments are sufficiently high. Without referral payments, many patients with a severe illness are treated by GPs, leading to an inefficient outcome. GPs make more efficient decisions under low referral payments. In particular, GPs exert diagnostic effort and treat mildly ill patients themselves, while referring severely ill patients to specialists in about 40% of the cases under low referral payments. Further, our results show that introducing low referral payments especially motivates barely altruistic GPs to refer severely ill patients, whereas intermediately or highly altruistic GPs hardly change their referral behaviour. Increasing the referral payment to a high level also increases the number of referrals for both patient types. In line with theory, we find that almost all patients are referred under high referral payments. As total efficiency increases more substantially when GPs treat a severely ill patient than in an unnecessary referral of a mildly ill patient, it reaches its maximum under high referral payments. Taken together, our behavioural results suggest that regulated referral payments affect the GPs’ referral behaviour and enhance efficiency. In particular, the referral behaviour of barely altruistic GPs can be incentivized towards more efficiency when introducing referral payments.

The paper proceeds as follows: In Section 2, we relate our paper to the literature. Section 3 introduces our stylized theoretical model. In Section 4, we

describe the experiment and derive behavioural predictions. In Section 5, we present the results from the experiment. Section 6 concludes.

2 Related literature

Our paper relates to several streams of the literature. First, it relates to the abundant literature on kickbacks. For example, Owen (1977) describes the role of kickbacks in the context of conveyancing services paying brokers to steer homebuyers and critically discusses their regulation. that providers of conveyancing services pay to brokers to steer homebuyers in real estate markets. In health care, Pauly (1979) models the role of kickbacks paid by one physician to another in return for patient referrals. Patients follow the referral advice up to a given maximum level. Whenever this maximum level is above the social optimal level, GPs might over-refer patients to specialists in order to receive the kickbacks. On the other hand, however, Pauly shows that referral payments can enhance efficiency because they incentivize GPs to refer patients to more cost-efficient specialists. We contribute to this early literature by explicitly analyzing the causal effect of referral payments (and different levels therein) on GPs' referral behaviour. To the best of our knowledge, we are the first to do so.

Second, we add to the theoretical economics literature on referrals. Garicano and Santos (2004) study referrals between two experts with different productivities and costs in generating revenue from a project by exerting efforts. An expert can choose between implementing a project himself, or referring it to the other expert. The referral of a project is subject to asymmetric information because a project's potential can be either high or low, which is private information of experts. In our parsimonious framework, we model a GP and a specialist with varying costs of treatment. We also consider asymmetric information about patients' types (either mild or severe illness) between patients and GPs. However, patients are passive in our set-up, which is a common assumption in the health economics literature (see, e.g., McGuire, 2000). GPs can exert a costly diagnostic effort to learn the severity of a patient's illness and decide whether to refer

or treat a patient. Inderst and Ottaviani (2012) analyse competition through kickbacks. In their model, two firms compete through kickbacks paid to an advisor. The advisor issues a recommendation to a customer regarding which of two products to purchase, on the basis of private information about the match between the customer's needs and the product's characteristics. The advisor is paid by the firms. In addition, the advisor cares about the customer purchasing the most suitable product, because of liability, ethical or reputational concerns. Firms set product prices, taking into account the advice customers receive. In our model, we do not consider competition among specialists and endogenous setting of referral payments by specialists, as we analyse the mere effect of *exogenously* given (regulated) referral payments at different levels on GPs' behaviour. Moreover, as we clearly focus on GPs, we assume that the specialist is the patient's perfect agent in that he always chooses the optimal treatment.

Third, we closely relate to the theoretical literature on referrals in health care markets. Barros and Olivella (2005) analyse cream skimming by physicians in public services who self-refer patients to their own private practices. Biglaiser and Ma (2007) examine the welfare effects of allowing dual practice and self-referrals. In our framework, GPs are only allowed to refer patients to a specialist; self-referrals of patients are not considered. Moreover, several gatekeeping models consider incentive contracts for gatekeeping GPs in a setting where the diagnostic precision is endogenous (e.g., Garcia-Mariñoso and Jelovac 2003; González, 2010; Malcomson, 2004). In our framework, we do not explicitly consider endogenous diagnostic precision, as we simply assume that a physician learns for sure the patient's type when exerting a costly diagnostic effort. Garcia-Mariñoso and Jelovac (2003) analyse optimal payment schemes under gatekeeping and direct-access systems. Allard et al. (2011) consider how referral to secondary care is affected by incentive contracts for primary care physicians. Brekke et al. (2007) analyse the effect of GP gatekeeping on equilibrium quality in an imperfectly competitive secondary care market. Godager et al. (2015) consider the effect of competition among GPs on referral behaviour. We neither consider competition in the market for primary care services (among GPs) nor

in the market for secondary care (among specialists). Our model framework is parsimonious and clearly focuses on the effect of referral payments on the GPs' diagnostic effort and referrals. Yet, it derives a set of hypotheses for the GPs' behaviour in our experiment.⁴

Finally, our experiment contributes to the recent experimental literature analyzing physicians' behaviour. The main focus of this literature is on the effect of financial incentives on an individual physician's decision to provide medical services; analysed incentives are, for example, fee-for-service and capitation (e.g., Hennig-Schmidt et al., 2011; Keser et al. 2014a), mixed payment systems (Brosig-Koch et al., forthcominga), and pay for performance (e.g., Brosig-Koch et al., 2013; Keser et al., 2014a; Lagarde and Blaauw, 2015b). The effect of professional norms on the provision of medical services is considered in Kesternich et al. (2015). A physician's decision to exert a diagnostic effort and to refer a patient has not been considered in the experimental health economics literature. We not only consider a physician's medical service provision, but we also explicitly introduce a physician's decision on diagnostic effort and whether to treat or to refer a patient to a more specialized physician. We therefore augment a physician's decision situation and choice alternatives in our experiment.

3 Theoretical model

We now introduce a stylized model in which GPs act as gatekeepers to secondary care (i.e. to specialists). In our model, we assume that patients may suffer from a mild or severe illness (similar to Allard et al., 2011). A GP (he) does not observe the patient's type. If exerting a costly diagnostic effort, the GP learns the severity of a patient's illness. The GP then decides whether to treat or

⁴Notice that we also relate to the operations management literature analyzing referral behaviour. Shumsky and Pinker (2003) derive the optimal referral rate given deterministic customer inter-arrival and service times and propose incentive structures that induce system optimal gatekeeping behaviour in a principal-agent setting. Lee et al. (2012) use the same framework to explore the problem from an outsourcing perspective. Zhang et al. (2011) present a two-tier system for security-check queues.

refer the patient to a specialist. We assume that the GP treats a patient with a standard treatment, which only heals the mild illness. The specialist (she) always treats the mild-severity patient with a standard treatment and the high-severity patient with a comprehensive treatment, respectively. Whenever the GP fails to refer a severely ill patient to the specialist (instead undertreats the high-severity patient with the standard treatment), then the patient's health deteriorates. If the GP refers a patient with a mild severity of illness to the specialist unnecessary expenses will be incurred without improving the patient's health or utility. The GP's decision to refer a patient to the specialist depends on the profit margin from treatment, the referral payment and the GP's altruism towards the patient.⁵

In the following, we present the timing of the game while specifying the objective functions of the patient, the GP, and the specialist. Figure 1 depicts the stages of the game; the timing is as follows:

Stage 1. Nature independently draws the severity of a patient's illness. The patient either suffers from a severe illness (H -type patient) with probability h or from a mild illness (L -type patient) with probability $1 - h$.

Stage 2. The GP decides whether to exert a diagnostic effort or not. If the GP exerts a diagnostic effort, he learns with certainty the patient's type of illness (mild or severe). If he does not exert a diagnostic effort, the GP does not learn about the severity of a patient's illness.

Stage 3. The GP decides whether to treat the patient with the standard treatment T_L or to refer the patient to a specialist. If the GP did not exert a diagnostic effort in stage 2, he remains uninformed about patient's severity when deciding to treat or to refer the patient.

Stage 4. If the GP referred the patient to the specialist (SP) or undertreated the H -type patient, the SP provides the standard treatment T_L

⁵In the theoretical health economics literature, assuming the physician to be altruistic has become quite common in models of physician behaviour (see, e.g., Chalkley and Malcomson, 1998; Biglaiser and Ma, 2007; Allard et al., 2011; Choné and Ma, 2011; Brekke et al., 2012).

and the comprehensive treatment T_H to the L -type patient and H -type patient, respectively.

Let U_i^P denote the utility of a patient with severity of illness $i \in \{L, H\}$. U_i^P comprises an initial health H_0 , a loss M_i due to the illness and a treatment utility of T_i . An L -type patient's utility is:

$$U_L^P = H_0 - M_L + T_L.$$

The H -type's utility depends on the GP's referral decision and is given by:

$$U_H^P = \begin{cases} H_0 - M_H + T_H & \text{if GP chooses to refer,} \\ H_0 - M_H - \Lambda + T_H & \text{if GP chooses } T_L, \end{cases}$$

with Λ denoting the loss a severely ill patient suffers when being treated with T_L by the GP. The loss reflects the deteriorating health of severely ill patients due to the delay compared to an immediate referral by the GP. After being (under-) treated by the GP, the H -type patient still has to visit a specialist.

GP j maximizes his utility which consists of his profit and an altruistic component accounting for the patient's utility from treatment. The former is given by the lump-sum payment for the GP's treatment p^{GP} , the costs for providing the treatment T_L , c^{GP} , and the costs of exerting diagnostic effort, c_e^{GP} . If a GP decides to exert effort, $c_e^{GP} > 0$; otherwise $c_e^{GP} = 0$. The latter represents the non-monetary utility a GP receives from treating a patient. GPs are heterogeneous with respect to their degree of altruism towards the patient, α_j . When referring the patient, the GP receives a referral payment R from the SP. Hence, GP j 's utility function is as follows:

$$U_j^{GP} = \begin{cases} p^{GP} - c^{GP} - c_e^{GP} + \alpha_j(H_0 - M_i + T_i(-\Lambda)) & \text{if GP chooses } T_L, \\ R - c_e^{GP} + \alpha_j(H_0 - M_i + T_i) & \text{if GP chooses to refer.} \end{cases}$$

For the specialist, we assume that the SP is a perfect agent of the patient and always provides the utility-maximizing treatment for both the L -type patient and

the H -type patient. We make this simplifying assumption to avoid confounding the GP's diagnosis effort and referral decision with beliefs about the specialist's treatment choices in the experiment. The SP's utility is as follows:

$$U^{SP} = \begin{cases} p^{SP} - c^{SP} - R & \text{if GP chooses to refer,} \\ p^{SP} - c^{SP} & \text{if GP chooses } T_L \text{ for } H\text{-Type patient,} \\ 0 & \text{if GP chooses } T_L \text{ for } L\text{-Type patient.} \end{cases}$$

For treating a patient, the SP receives the lump-sum payment p^{SP} and bears cost c^{SP} , with $p^{SP} > p^{GP}$ and $c^{SP} > c^{GP}$.⁶ If the GP refers a patient, the SP pays the referral payment R to the GP. We define the expected efficiency as the patient's health benefit from treating less the treatment costs and possible costs from a delay in H -type patients' treatment. Assuming a uniform distribution of patient types, $E = \frac{1}{2}(T_L - c_L^{GP} - c_e^{GP} + T_H - c_H^{SP} - c_e^{GP} - (\Lambda))$. The following lemma describes a GP's best response to different levels of referral payments.

Lemma 1. *GP j 's optimal behaviour denoted as $s_j(R) := \{\text{diagnostic effort, treatment or referral of } L\text{-Type, treatment or referral of } H\text{-Type}\}$ for varying referral payments is:*

$$s_j(R) = \begin{cases} \{\text{no effort, treat, treat}\} & \text{if } \begin{aligned} R < p^{GP} - c_L^{GP} + 2c_e^{GP} - \alpha_j\Lambda & \text{ and} \\ R < p^{GP} - c_L^{GP} - \frac{1}{2}\alpha_j\Lambda, \end{aligned} \\ \{\text{no effort, refer, refer}\} & \text{if } \begin{aligned} R > p^{GP} - c_L^{GP} - \frac{1}{2}\alpha_j\Lambda & \text{ and} \\ R > p^{GP} - c_L^{GP} - 2c_e^{GP}, \end{aligned} \\ \{\text{effort, treat, refer}\} & \text{if } \begin{aligned} R < p^{GP} - c_L^{GP} - 2c_e^{GP} & \text{ and} \\ R > p^{GP} - c_L^{GP} + 2c_e^{GP} - \alpha_j\Lambda. \end{aligned} \end{cases}$$

Proof. See Appendix A. □

⁶We assume that the specialist bears a higher cost than the GP for treating a patient, irrespective of the patient's severity of illness. In particular, an intuition for this assumption is that specialists might be capacity-constrained and have, for example, higher fixed costs for treating a L -type patient.

The intuition of Lemma 1 is as follows: If there is no referral payment, a barely and intermediately altruistic GP treats his patients without exerting diagnostic effort. A highly altruistic GP is sufficiently altruistic even to exert effort without a referral payment. This allows a highly altruistic GP to not undertreat H -type patients while not foregoing the positive payoff on treating L -type patients. Introducing referral payments motivates intermediately altruistic GPs also to perform a diagnosis and to treat L -type patients and refer H -type patients if diagnosis costs are sufficiently small. If diagnosis costs are high, intermediately altruistic GPs refer both types of patients. For a high referral payment, there is no incentive for any GP type to treat the patient himself, as specialists always provide the appropriate treatment.

Figure 2 illustrates the GPs' best responses to different levels of referral payments for different degrees of the GPs' altruism. For different combinations of referral payments and altruism, we expect different behaviours in the experiment; for a detailed description of the behavioural predictions, see Subsection 4.2.

4 Experimental Design

The experiment closely relates to our stylized theoretical model. We now describe the decision situation and the treatments of the experiment, derive behavioural hypotheses, and present the procedure of the experiment.

4.1 General design

In our medically framed experiment, subjects are randomly allocated either to the role of the GP or the specialist. The role remains constant throughout the 20 rounds of the experiment. Before each round, pairs of GPs and specialists

are randomly and anonymously (re-) matched.⁷ In each round, a random draw determines whether the patient suffers from a mild or a severe illness. Patients seek medical services, first from the GP gatekeeper. A patient’s severity of illness is unknown to the GP. There are no subjects in the role of patients in the lab. Real patients’ health outside the lab, however, is affected by subjects’ decisions in the lab. Benefits accrued in the experiment translate into monetary transfers to a charity providing surgeries for ophthalmic patients. This mechanism ensures that subjects in the lab also take a patient’s health into account.⁸

In each round, GPs make two decisions. First, they decide whether to exert effort to diagnose a patient and, therefore, to learn about the severity of a patient’s illness. The cost for the diagnostic effort is $c_e^{\text{GP}} = 10$ Taler (the experimental currency with 1 Taler = 0.05 Euro). Second, a GP decides whether to treat a patient with a standard treatment or to refer a patient to the specialist.

⁷Due to the anonymous re-matching mechanism, reputation building is absent in our experiment. We also argue that learning should not be an issue in our experiment as all parameters, subjects’ possible actions, resulting payoffs and patient benefits are common knowledge to subjects. Strategic interactions between GPs and specialists are also absent as specialists are restricted to always treating the patient optimally. Showing subjects an summary of their individual payoffs after each round (for more, see Subsection 4.3), is therefore very unlikely to affect a GP’s decision. After each round, GPs are informed about the resulting payoff and the patient benefit from their decisions. We also employ a random-choice payment technique, which prevents incentives for averaging or end behaviour. Taken together, the stranger matching therefore allows us to derive very similar behavioural predictions, as in a “one-shot” design, at the same time giving us more observations per subject; for more, see Camerer (2003).

⁸Patients’ health benefits are measured in monetary terms. The accumulated benefits are then transferred to a charity caring for real patients. Notice that the mechanism is particularly attributed to the treatment of patients, which makes it different from mere donations in the charitable giving literature; see, for example, Andreoni, (1989) or DellaVigna et al. (2012). This ‘mechanism’ of patient benefit transfer introduced by Hennig-Schmidt et al. (2011) has been applied in several experiments in health economics, as it embeds an incentive for subjects in the lab to account for real patients’ health outside the lab. An equivalent setup is used, for example, in Brosig-Koch et al. (2013), Godager and Wiesen (2013), Hennig-Schmidt and Wiesen (2014) and Brosig-Koch et al. (forthcominga, forthcomingb). In Kesternich et al. (2015), Keser et al. (2014a, 2014b) and Lagarde and Blaauw (2015a 2015b), subjects could choose from several (medical) charities to which a donation was transferred.

If the GP refers the patient he receives a referral fee R from the specialist. We exogenously vary the referral fee in our experimental treatments; see Subsection 4.2.

We assume that the GP’s medical treatment only heals a patient with a mild illness. The GP’s costs for treating a patient with a low treatment intensity is $c_L^{\text{GP}} = 100$ Taler. The specialist, however, heals both types of patients, but at a higher costs, which are $c_L^{\text{SP}} = c_H^{\text{SP}} = 150$ Taler, and she does not incur diagnosis costs. In the experiment, a specialist’s choice is restricted to provide the patient’s benefit maximizing treatment.

Both, the GP and the specialist, receive a lump-sum payment for treating a patient. The payment is $p^{\text{GP}} = 250$ Taler and $p^{\text{SP}} = 420$ Taler for the GP and the specialist, respectively.

The severely ill patient gains a higher benefit from medical treatment than the mildly ill patient. The health loss from a severe illness is $M_H = 470$ Taler and a mild illness is $M_L = 250$ Taler. We assume that the patient’s benefit from treatment is such that the initial health status ($H_0 = 210$ Taler) is restored—in particular, $T_H = M_H = 470 > 250 = M_L = T_L$. Whenever a patient with a high severity of illness is treated by a GP instead of being referred, this patient suffers a disutility of 210 Taler—for example, due to the delay in receiving the appropriate medical treatment. For more details of the decision situation of the experiment, see the instructions in Appendix D.

4.2 Referral payments and behavioural predictions

To analyse the causal effect of a referral payment on the GPs’ referral decision, diagnostic effort and efficiency, we exogenously vary the level of referral payments from the specialist to the GP. In the baseline treatment, we set $R_0 = 0$, which reflects current policies. In treatment LOW, we introduce a low referral payment of $R_1 = 100$. We also introduce two high referral payments of $R_2 = 160$ and $R_3 = 200$ in treatments HIGH and HIGH-2, respectively.

In the following, we classify GPs based on their degree of altruism towards the patient to derive behavioural predictions: barely altruistic GPs with $\alpha_j \in$

$[0, 0.33)$, intermediately altruistic GPs with $\alpha_j \in [0.33, 0.81]$, and highly altruistic GPs with $\alpha_j \in (0.81, 1]$.⁹ We carefully chose parameters of the experiment such that the treatments “Baseline”, LOW and HIGH lead to different predictions about the behaviour of GPs and the efficiency for different levels of GP altruism. In particular, introducing low referral payments should change intermediately altruistic GP’s diagnostic effort and referral behaviour while GPs with low and high altruism towards their patients should not change behaviours.

Table 1 shows the experimental treatments, the predicted behaviour of GPs according to Lemma 1 and the predicted efficiency. Without a referral payment (baseline treatment), GPs with low and intermediate altruism neither have an incentive to diagnose patients and, thus, to learn about a patient’s type nor to refer patients to specialists. Only highly altruistic GPs will exert a diagnostic effort, treat L -type patients and refer H -type patients if no referral payments are in place. The inefficiency arises as patients suffering from a severe illness are undertreated by barely and intermediately altruistic GPs. On the other hand, under sufficiently high referral payments (treatments HIGH and HIGH-2), GPs, irrespective of their level of altruism, will always choose to refer the patient. This implies an inefficiency as a specialist’s treatment cost to heal mild illnesses is higher than a GP’s treatment cost. A low level of referral payment (treatment LOW) should improve efficiency, as it shifts the intermediately altruistic GPs towards diagnosing their patients and treating patients with a mild illness, while referring patients with a severe illness. In sum, we state the following hypotheses regarding the GPs’ diagnostic efforts, referral decisions, and efficiency, which we test in our laboratory experiment:

Hypothesis 1. *On the aggregate, the GP exerts more frequently a diagnostic effort with a low referral payment than without a payment. Further, GPs exert diagnostic effort more often without a referral payment than with high referral payments.*

When differentiating between barely, intermediately and highly altruistic GPs,

⁹Note that the classification of GPs’ degree of altruism in the experiment is based on the 33 and 66 percentile.

Table 1: Experimental treatments and predictions for GPs' behaviour

Treatment	Referral payment	Type of GP's altruism	Predicted GP's behaviour (per patient and period of the experiment)	Predicted efficiency (per patient and period of the experiment)	Subjects
Baseline	$R_0 = 0$	barely and intermediate	no diagnostic effort; treat both patient types with T_L	$E_0 = \frac{1}{2}(T_L - c_L^{GP} + T_H - \Lambda - c_L^{GP}) = 105$	64
		high	diagnostic effort; treat L -type patients and refer H -type patients iff c_e^{GP} is sufficiently small	$E_1 = \frac{1}{2}(T_L - c_L^{GP} - c_e^{GP} + T_H - c_H^{SP} - c_e^{GP}) = 225$	
LOW	$R_1 = 100$	barely	no diagnostic effort; treat both patient types with T_L	$E_0 = \frac{1}{2}(T_L - c_L^{GP} + T_H - \Lambda - c_L^{GP}) = 105$	64
		intermediate and high	diagnostic effort; treat L -type patients and refer H -type patients iff c_e^{GP} is sufficiently small compared to R	$E_1 = \frac{1}{2}(T_L - c_L^{GP} - c_e^{GP} + T_H - c_H^{SP} - c_e^{GP}) = 225$	
HIGH	$R_2 = 160$	barely, intermediate, and high	no diagnostic effort; refer both patient types	$E_2 = \frac{1}{2}(T_L - c_L^{SP} + T_H - c_H^{SP}) = 210$	60
HIGH-2	$R_3 = 200$	low, intermediate, and high	no diagnostic effort; refer both patient types	$E_3 = \frac{1}{2}(T_L - c_L^{SP} + T_H - c_H^{SP}) = 210$	64
<i>Subjects</i>					

Subjects

252

Notes. This table indicates the experimental variations in the referral payments and provides predictions about GPs' behaviour and the efficiency. We differentiate between predictions for GPs with low, intermediate, and high altruism towards their patients. Monetary values are indicated in Taler, our experimental currency.

we are able to formulate the following hypothesis regarding a GP's diagnostic effort:

Hypothesis 2. *Highly altruistic GPs exert diagnostic effort without referral payments and with low referral payments. Intermediately altruistic GPs exert diagnostic effort with low referral payments. Barely altruistic GPs never exert diagnostic effort.*

With respect to a GPs' referral behaviour, Lemma 1 allows us to derive the following hypothesis:

Hypothesis 3. *GPs, on the aggregate, refer more severely ill (H-type) patients to the specialist with low referral payments than without referral payment. Under high referral payment ($R_2 = 160$ and $R_3 = 200$), all GPs, regardless of their level of altruism, refer both patient types.*

Differentiating behaviour for the three different altruistic types of GPs, we state the following hypothesis regarding a GP's referral decision:

Hypothesis 4. *Highly altruistic GPs refer high severity (H-type) patients regardless of the referral payment and refer low severity (L-type) patients only under high referral payments. GPs with an intermediate level of altruism always refer H-type patients, but not without referral payment and L-type patients only under high referral payments. Barely altruistic GPs refer both H-type and L-type patients only under high referral payments.*

We now turn to the efficiency which we defined as the patients' treatment benefit less the costs for the treatment and the possible health loss due to undertreatment. A GP makes efficient decisions if he exerts diagnostic effort to learn the patient type and then treats L-type patients and refers H-type patients to the specialist. Individual incentives are, however, only aligned for the GPs with a high level of altruism under treatments baseline and LOW and for the intermediately altruistic GP under treatment LOW. For all other treatments and types, we do expect a lower number of efficient decisions. The aggregate level of efficiency, however, depends on the distribution of GP types. If there

are sufficiently many intermediate and high altruistic GPs ($> 87.5\%$), efficiency is highest in the LOW treatment. Otherwise, efficiency reaches its maximum under HIGH and HIGH-2. We therefore hypothesize:

Hypothesis 5. *On the aggregate, the number of efficient decisions is highest under low referral payments. Efficiency is highest with low referral payments if there are sufficiently many GPs with an intermediate and high level of altruism. Otherwise, efficiency is highest under high referral payments.*

Again, differentiating between the different GPs' altruistic types, we can state the following hypothesis:

Hypothesis 6. *Barely altruistic GPs make the most efficient decisions under high referral payments. GPs with an intermediate level of altruism make the most efficient decisions without referral payment and with low referral payment. For intermediately altruistic GPs, efficiency is highest under low referral payment. For highly altruistic GPs, efficiency is not affected by the referral payments.*

4.3 Procedure

The computerized experiments programmed with z-Tree (Fischbacher, 2007) were conducted at the Cologne Laboratory for Experimental Economics between October 2015 and January 2016. In total, 252 students from the University of Cologne participated in our experimental sessions. Among those were 50 medical students. All subjects were recruited through the online recruiting system ORSEE (Greiner, 2004).

Besides our main experiment (as described above), we also employed an experimental task to elicit subjects' altruism according to Godager et al. (2016). In this task, subjects decided in the role of physicians on the quality of medical service provision for eight patients. Each decision implied a trade-off between the own profits and the patients' health benefits. This allows us to elicit individuals' altruism in a medical context. We deliberately kept the medical context constant across the main experiment and the altruism elicitation task. For an illustration of the decision situations in this part of the experiment, see also Appendix D.

For each GP, we calculate the average quality of medical services provided across the eight decisions in the altruism task. Based on this average quality choice, we classify GPs as barely altruistic if their average quality choice is in the 33rd percentile, as intermediately altruistic if their choice is between the 34th and 66th percentile and as highly altruistic if their choice is larger than the 66th percentile. We systematically vary the order of the parts of the experiments within an experimental session. The experiments are concluded by comprehensive post-experimental questionnaires on social demographics and personality traits.

More specifically, the procedure was as follows: Upon arrival, subjects were randomly allocated to cubicles. They were then given ample time to read the instructions for the first part and to ask clarifying questions, which were answered in private. Subjects were informed that the experiment consisted of two parts, but received detailed instructions for the second part only after having finished the first part of the experiment. To check for the subjects' understanding of the decision task, they had to answer a set of control questions. The experiment did not start unless all subjects had answered the control questions correctly.

We employed a random-choice payment technique. At the end of the experiment—when all subjects had made their decisions—we randomly determined one decision in each part of the experiment to be relevant for a subject's actual payoff and the patient benefit. This was done to rule out income effects. Subjects were paid in private according to the two randomly determined decisions.

To verify that the money corresponding to the sum of patient benefits in a session was actually transferred, we applied a procedure similar to Hennig-Schmidt et al. (2011). One of the subjects was randomly chosen to be a monitor. After the experiment, the monitor verified that a payment order in the amount of the aggregate benefit was written to the financial department of the University of Cologne, to transfer the money to the Christoffel Blindenmission, which used the monetary transfers exclusively to support surgical treatments of cataract patients in a hospital in Masvingo (Zimbabwe) staffed by ophthalmologists from the charity. Notice that we did not inform the subjects that the money was assigned to a developing country. The order was sealed in an envelope, and the

monitor and experimenter then walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional EUR 5.

Sessions lasted for about 90 minutes. Subjects earned, on average, about EUR 19.9. In total, EUR 2,807 were transferred to the Christoffel Blindenmission. The average cost for a cataract operation amounts, according to the Christoffel Blindenmission, to about EUR 30. Thus, our experiment allowed 94 patients to be treated.

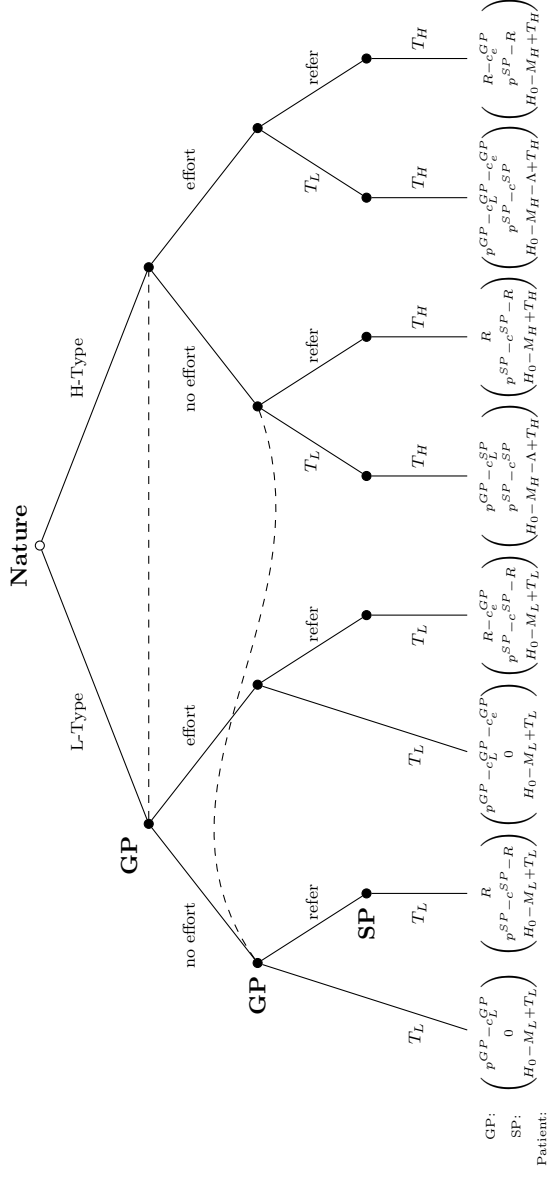


Figure 1: Illustration of the decision situation

Notes. This figure illustrates the four stages of our three-player game. In the first stage, nature randomly determines the severity of a patient's illness, which could either be mild L or severe H and is ex-ante unknown to the GP. In the second stage, the GP decides whether to treat a patient with T_L for remuneration p^{GP} and at cost c_L^{GP} or to refer the patient, for which the GP receives payment R from the SP. When the GP treats the mildly ill patient, the game ends. In the fourth stage, the SP heals both severities for which she receives p^{SP} and bears cost c^{SP} . When referring the patient, the GP receives a referral payment. Outcomes for the GP, the SP and the patient are shown at the end nodes.

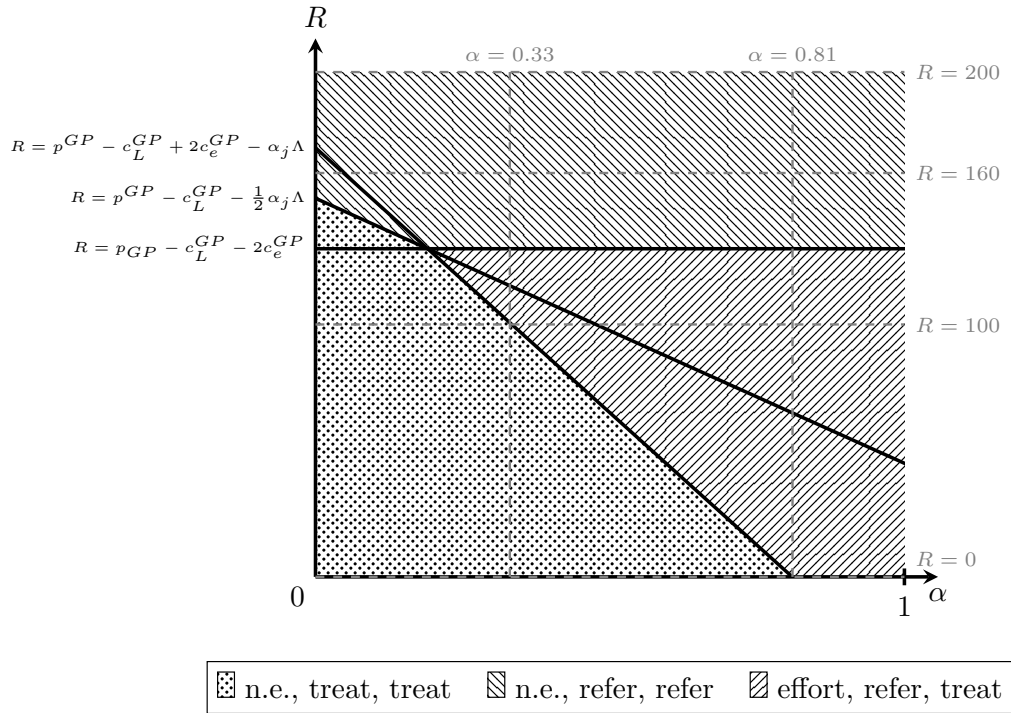


Figure 2: GPs' best responses by the degree of altruism for different referral payments

Notes. This figure shows the GPs' best responses dependent on their degree of altruism towards their patients and the level of the referral payment. The grey shaded area illustrates the GP's strategy {no effort, treat, treat}; that means, the GP exerts no diagnostic effort and treats both patient types. The blue shaded area depicts the strategy {no effort, refer, refer}; that means, the GP exerts no diagnostic effort and refers both patient types. The red colored area indicates the strategy {effort, refer, treat}; that means the GP exerts diagnostic effort, refers L -type patients and treats H -type patients. Notice that dominated strategies are discarded.

5 Results

5.1 GPs' characteristics and their degree of altruism

First, we provide a short overview on the characteristics of subjects in the role of GP. As subjects were mostly students, the average age amounted to about 24 years. About 57% of the subjects were female. A share of about 20% of the subjects were medical students. Regarding GPs' altruism towards their patients, 42 subjects were classified to be barely, 43 intermediately and 41 highly altruistic.

Table 2 shows the distribution of the GPs' altruistic types across experimental treatments. We observe that the distribution of GPs' altruism is fairly balanced for the Baseline, the LOW and the HIGH-2 treatment. In treatment HIGH, there are more intermediately altruistic than barely and highly altruistic GPs. Given that the GPs' optimal behaviour should not vary across different levels of GP's altruism under HIGH, we do not expect this cumulation to confound our results.

Table 2: Distribution of GPs' altruistic types across treatments

Experimental treatments	Barely altruistic GPs	Intermediately altruistic GPs	Highly altruistic GPs	Total number of GPs
Baseline ($R_0 = 0$)	11	8	13	32
LOW ($R_1 = 100$)	13	9	10	32
HIGH ($R_2 = 160$)	7	17	6	30
HIGH-2 ($R_3 = 200$)	11	9	12	32
Overall	42	43	41	126

5.2 GPs' diagnostic efforts

We now consider the GPs' aggregate diagnostic effort. Figure 3 shows that GPs provide a diagnostic effort for a substantial number of patients irrespective of the referral payment. The average share of patients diagnosed is around 43% in treatments Baseline ($R_0 = 0$), LOW ($R_1 = 100$), and HIGH ($R_2 = 160$). Only under HIGH-2 ($R_3 = 200$) is the share of diagnostic decisions slightly lower (i.e., 31%). The difference of the latter to the other three treatments is not significant,

however (Baseline and LOW and HIGH vs. HIGH-2: $p = 0.2443$, Mann Whitney U-test, two-sided).¹⁰ To account for effects due to repetition and learning over the 20 periods of the experiment, we ran a set of Logit regressions. Estimations indicate a reduction in the share of diagnostic effort over time; see Table C.1 in Appendix C. In sum, our behavioural data do not support Hypothesis 1, which suggested ‘no diagnostic effort’ in treatments Baseline, HIGH, and HIGH-2, and ‘diagnostic effort’ in treatment LOW. We state the following result:

Result 1. *On the aggregate, referral payments do not significantly affect the GPs’ diagnostic effort.*

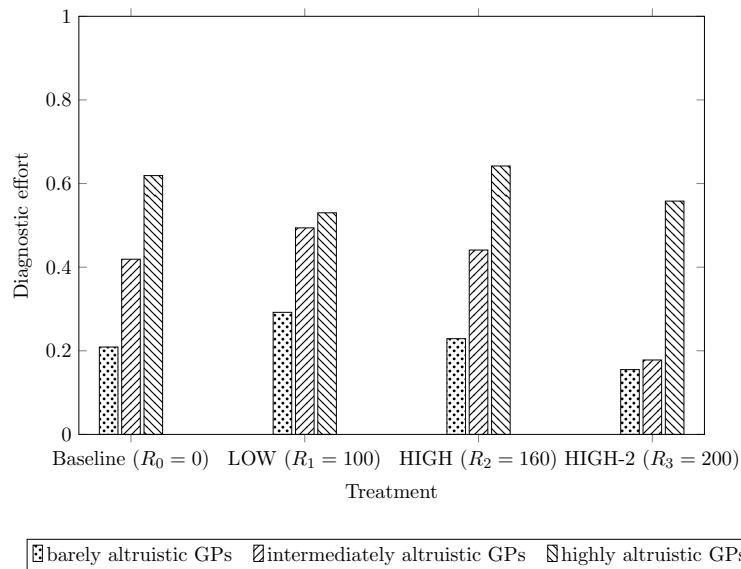


Figure 3: GPs’ aggregated diagnostic efforts by level of referral payments

Notes. This figure shows GPs’ aggregate share of diagnosed patients; each GP decides to exert diagnostic effort or not for 20 patients. Blue bars show diagnosed patients by barely, gray bars by intermediately and red bars by highly altruistic GPs, respectively. In the baseline treatment, there are 312 L -type and 328 H -type patients. In treatment LOW, there are 321 L -type and 319 high type patients. There are in HIGH and in HIGH-2 306 L -type as well as 294 H -type patients and 336 L -type as well as 304 H -type patients, respectively.

We next explore the heterogeneity in the GPs’ altruism and investigate, for the three different GPs’ altruistic types, how the diagnostic effort decision is affected by the levels of the referral payment. Comparing the effort levels between the

¹⁰Note that all p -values are reported for two-sided Mann Whitney U tests if not indicated otherwise.

different GP types suggests that our classification is consistent with theory, as barely altruistic GPs exert on average across the four treatments significantly less diagnostic effort than intermediately altruistic GPs ($p = 0.0408$), and intermediately altruistic GPs exert significantly less diagnostic effort than highly altruistic GPs ($p = 0.0348$).

Figure 3 shows that there are only small differences induced by different referral payments on GPs with a high level of altruism. Highly altruistic GPs exert a diagnostic effort in about roughly 60% of the patient visits across all treatments. GPs with a low and intermediate level of altruism, however, can be motivated to increase their diagnostic effort weakly significantly if referral payments are LOW and HIGH compared to Baseline and HIGH-2 (Baseline and HIGH-2 vs. LOW and HIGH: $p = 0.0763$). In sum, we state the following result:

Result 2. *GPs with a high level of altruism exert a diagnostic effort in about 60% of the patient visits across all treatments. Barely and intermediately altruistic GPs can be best motivated to provide diagnostic effort by implementing low ($R_1 = 100$) and high referral payments ($R_2 = 200$).*

Result 2 is somewhat surprising. Following theory, even altruistic GPs should not exert effort under high referral payments. One possible explanation, partially supported by our data, is that GPs are not only altruistic, but also have efficiency concerns. In order to make an efficient treatment decision, GPs have to perform the diagnosis to learn the patients' type of disease. Our results in fact show that GPs who exerted effort in the HIGH-2 treatment make significantly more efficient referral decisions than GPs who did not exert effort (see Result 2). Our model could easily capture GPs' tastes for efficiency (see, for example, Charness and Rabin (2002) and Andreoni and Miller (2002)). The GP's utility function then not only depends on patient benefit, but also on treatment costs. Adding treatment costs to the altruism term would allow us to interpret α_j as a taste of efficiency of GP j instead of altruism. In Appendix B, we derive the extended model. The intuition is as follows: On the one hand, referring without diagnosing becomes more unattractive for a GP because he also internalizes the possibly

unnecessarily high costs of treatment if referring an L -Type patient. On the other hand, treating without diagnosing becomes more unattractive because the GP internalizes the additional costs that are caused due to treating an H -type patient by a GP and a specialist. Hence it becomes, overall, more attractive to diagnose a patient in order to prevent unnecessarily high costs.

5.3 GPs' referral behaviour

We now analyse the GPs' referral behaviour according to Hypothesis 3. In particular, we focus on the GPs' binary decisions between referring or treating a patient after deciding whether or not to diagnose a patient.

Table 3 shows average referral rates. On the aggregate, about 19% of patients (not distinguishing between patients' severity) are referred in the absence of a referral payment. Under a low referral payment ($R_1 = 100$), the aggregate referral rate increases slightly to 24%. However, when introducing high referral payments (i.e., $R_2 = 160$ and $R_3 = 200$), the referral rate increases significantly to about 83% and 90%, respectively; see Table 3.

When differentiating between patients' severities of illness, we observe that

Table 3: Average referral rates by treatment and patient types

Referral payment	Overall	L -type patients	H -type patients
$R_0 = 0$	0.188 (0.391)	0.006 (0.080)	0.360 (0.481)
$R_1 = 100$	0.242 (0.429)	0.025 (0.156)	0.461 (0.499)
$R_2 = 160$	0.832 (0.374)	0.739 (0.440)	0.929 (0.258)
$R_3 = 200$	0.895 (0.306)	0.824 (0.381)	0.974 (0.160)

Notes. This table shows the average referral rates per referral payment. Standard deviations are reported in parentheses.

if there is no referral payment ($R_0 = 0$), almost no L -type patients, but about 36% of the H -patients are referred. Under a low referral payment ($R_1 = 100$), we find that almost no L -type patients but almost half of the H -type patients are referred. For the high referral payments $R_2 = 160$ and $R_3 = 200$, we observe

that about 93% and 97% of H -type patients are referred, respectively (Baseline vs. HIGH: $p < 0.001$; Baseline vs. HIGH-2: $p < 0.001$). Also for the L -type patients, the referral rate significantly increased to about 74% under $R_2 = 160$ and 82% under $R_3 = 200$ (Baseline vs. HIGH: $p < 0.001$; Baseline vs. HIGH-2: $p < 0.001$); see also Table 3. Referrals of H -type patients under low referral payment ($R_1 = 100$) increase significantly compared to $R_0 = 0$ (Baseline vs. LOW: $p = 0.010$). For L -type patients, we find only a weak significantly increase in referral rates (Baseline vs. LOW: $p = 0.097$). These behaviours are in line with Hypothesis 3; we state:

Result 3. *GPs tend to only refer high-severity patients when receiving a low referral payment. When receiving a high referral payment, GPs refer both, mild and high-severity patients, to the specialist.*

We now analyse whether referral behaviour differs with the degree with GPs' degrees of altruism. Figure 4 shows referral rates differentiated by the patients' severity and the GPs' altruism. We first investigate differences between different levels of altruism for a given level of referral payment: For L -type patients, we do not find significant differences in referral rates between low, medium and high altruistic GPs except for the HIGH-2 treatment. Highly altruistic GPs refer L -type patients significantly less often than GPs with an intermediate ($p = 0.0352$) and a low level of altruism ($p = 0.0398$). Remember that highly altruistic GPs provide significantly more effort in HIGH-2 than barely and intermediately altruistic GPs. These findings support the hypothesis that GPs are not only driven by altruism, but may also account for efficiency. For H -type patients, there are substantial behavioural differences. Highly altruistic GPs refer more than 60% of H -type patients if there is no referral payment. In contrast, barely altruistic GPs refer only two of their 115 H -type patients if there is no referral payment ($p < 0.001$). GPs with an intermediate level of altruism refer 37.5% of their H -type patients. A similar pattern holds true in the LOW treatment. When referral payments are high (i.e., $R_2 = 160$ and $R_3 = 200$), the difference in referral behaviour diminishes between GPs with low and high altruism.

A comparison of the GPs' referral behaviours between treatments shows that barely altruistic GPs refer significantly more H -type patients in the LOW than the baseline treatment (Baseline vs. LOW: $p = 0.0207$). Barely altruistic GPs also refer H -type patients significantly more often in the HIGH and HIGH-2 than in treatment LOW (LOW vs. HIGH: $p = 0.0032$, LOW vs. HIGH-2: $p = 0.0002$). GPs with an intermediate or high level of altruism do not significantly differ in their referral behaviour of H -type patients between the treatments baseline and LOW (Baseline vs. LOW: $p = 0.7633$ for intermediately and $p = 1.000$ for highly altruistic GPs) but between the LOW and the HIGH treatment (LOW vs. HIGH: $p = 0.0038$ for intermediately and $p = 0.0120$ for highly altruistic GPs). We therefore conclude:

Result 4. *Highly altruistic GPs refer H -type patients significantly more often than barely altruistic GPs. The introduction of referral payments incentivizes barely altruistic GPs to refer H -type patients.*

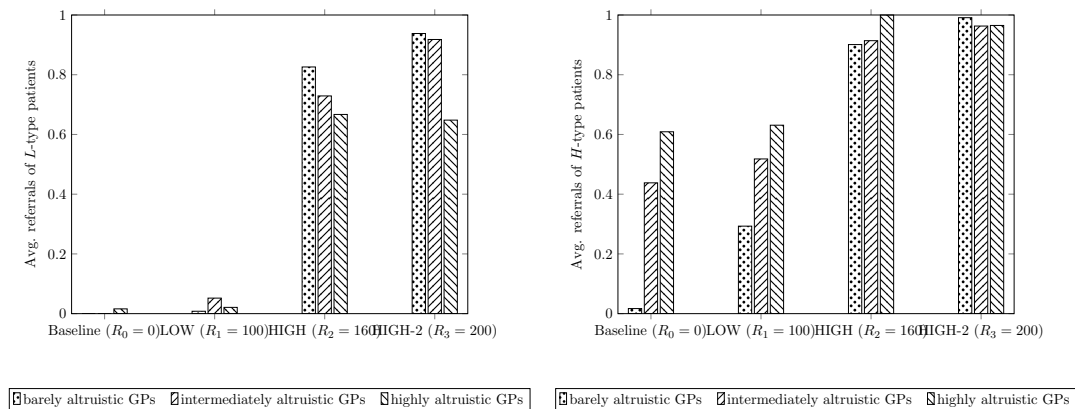


Figure 4: Referral rates by patients' severity of illness and GPs' altruism.

Notes. Average referral rates for the low-severity patients are shown in the left panel and for high-severity patients in the right panel. Blue bars show patients referred by barely, gray bars by intermediately and red bars by highly altruistic GPs, respectively. Bars are grouped by treatment.

Further, Table 4 indicates that our main results are robust towards the period of the experiment, subjects' characteristics such as gender, age and medical major. Whether subjects have exerted a diagnostic effort before choosing to refer or treat a patient does not change our estimation results for L - and H -type

patients; see models (5) and (6).

Taken together, Result 3 implies that GPs with a low altruism level can be motivated to refer patients efficiently by implementing a referral payment between the specialist and the GP. The GPs with a high altruism level hardly change their behaviour, but provide fewer unnecessary referrals of L -type patients.

5.4 Efficiency

After analyzing the GPs' diagnostic effort and referral decisions separately, we now combine the different parts in our efficiency analysis. As mentioned above, we define efficiency as the patients' treatment benefit, minus the costs for the treatment and the possible health loss due to undertreatment. A GP makes efficient decisions if he exerts a diagnostic effort to learn the patient type and then treats L -type patients and refers H -type patients to the specialist. Individual incentives, however, are such that the GPs' incentives to make efficient decisions are aligned for most GPs under a low referral payment ($R_1 = 100$). For the three remaining treatments, we expect fewer efficient decisions.

Table 5 presents the share of efficient decisions, the average absolute observed and the normalized observed efficiency levels per treatment. The maximum number of efficient decisions is reached in the LOW treatment, which is in line with Hypothesis 5. However, differences in the number of efficient decisions are not significant between treatments. The observed efficiency level is lowest in the status quo, the baseline treatment, with 131 (with 225 being the maximum level of efficiency). The main driving factor is that barely altruistic GPs do only refer two of the 115 treated H -type patients and thus cause a considerable decrease in efficiency. Introducing low referral payments ($R_1 = 100$) starting from the status quo does not increase the level of efficiency on the aggregate to a significant extent (Baseline vs. LOW: $p = 0.2886$). This finding is in contrast to Hypothesis 5. We observe that barely altruistic GPs decide weak significantly more efficiently (Baseline vs. LOW: $p = 0.0555$) when introducing referral payments, but the ef-

Table 4: Random effects panel logit regressions on GP's referral behavior

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Patients:	<i>L</i> -type patients	<i>L</i> -type patients	<i>L</i> -patients	<i>H</i> -type patients	<i>H</i> -type patients	<i>H</i> -type patients
Dependent variable:	GPs' referrals	GPs' referrals	GPs' referrals	GPs' referrals	GPs' referrals	GPs' referrals
LOW ($R_1 = 100$)	0.100 (0.065)	0.100 (0.066)	0.100* (0.059)	0.069 (0.059)	0.073 (0.067)	0.108*** (0.038)
HIGH ($R_2 = 160$)	0.521*** (0.055)	0.517*** (0.056)	0.466*** (0.049)	0.511*** (0.055)	0.520*** (0.047)	0.483*** (0.043)
HIGH-2 ($R_3 = 200$)	0.566*** (0.058)	0.562*** (0.058)	0.504*** (0.047)	0.533*** (0.060)	0.533*** (0.050)	0.534*** (0.044)
Diagnosis (= 1 if GP exerted effort)			-0.185*** (0.024)			0.254*** (0.041)
Gender (= 1 if female)			-0.079*** (0.027)			-0.026 (0.034)
Age			-0.001 (0.003)			-0.000 (0.003)
Med (= 1 if medical student)			-0.016 (0.038)			0.076* (0.042)
Intermediate altruism			0.035 (0.033)			0.087** (0.041)
High altruism			0.010 (0.037)			0.150*** (0.043)
Period dummies	No	Yes	Yes	No	Yes	Yes
Observations	1275	1275	1275	1245	1245	1245

Notes. Coefficients represent average marginal effects. Standard errors are clustered on the individual level and are reported in parentheses. The reference treatment is 'Baseline' ($R_0 = 0$). The reference altruism level is 'low'. Note that results from panel Probit regression with robust standard errors clustered on subject level do not differ qualitatively. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. p -values are based on two-tailed tests.

Table 5: Absolute observed and normalized observed efficiency per treatment

Treatment	Number of efficient decisions	Absolute observed efficiency	Normalized observed efficiency
Baseline ($R_0 = 0$)	0.379	131	0.475
LOW ($R_1 = 100$)	0.411	147	0.565
HIGH ($R_2 = 160$)	0.310	199	0.865
HIGH-2 ($R_3 = 200$)	0.230	202	0.893

Notes. Efficient decision and absolute observed efficiency is averaged over all subjects' decisions. Normalized observed efficiency implies a normalization to the interval $[0, 1]$, in which 0 refers to the minimum absolute efficiency and 1 refers to the maximum absolute efficiency given the distribution of patient types.

iciency of decisions of intermediately and highly altruistic GPs does not improve significantly (Baseline vs. LOW for intermediately respectively highly altruistic GPs: $p = 0.8852$ resp. $p = 0.8281$). Increasing referral payments to a high level ($R_2 = 160$) improves the level of the aggregate efficiency significantly (LOW vs. HIGH: $p = 0.004$) and leads to an average efficiency level that is close to maximum efficiency (210 out of 225). Once the level of referral payments is high, increasing referral payments further to $R_3 = 200$ does not significantly change the efficiency level (HIGH vs. HIGH-2: $p = 0.9158$). Our above results are robust towards normalizing the absolute observed efficiency levels on the interval between minimum and maximum efficiency based on the actual distribution of L -type and H -type patients.

With the provided parametrization of the experiment, treating H -type patients as a GP decreases the total efficiency considerably more than referring an L -type patient. As almost all patients are referred under high referral payments, total efficiency is highest under high referral payments (see Table 5). In sum, we therefore state the following result:

Result 5. *Efficiency increases in referral payments. The highest efficiency is observed for high referral payments.*

6 Conclusion

Inefficiencies due to GPs' inappropriate referrals of patients to specialists are pervasive in many health care markets (e.g., Scott, 2000; Chandra and Skinner, 2012). Although suggested by economic theory (e.g., Pauly 1979 Garicano and Santos, 2004) as an effective means to enhance efficiency, referral payments are prohibited in most health care markets. In the absence of referral payments, however, there is an incentive for GPs to treat patients with a severe illness which leads to inefficient outcomes. Empirical evidence on the effect of referral payments on GPs' behavior and efficiency is (obviously) lacking.

This paper analyses the causal effect of referral payments from specialists to GPs on GPs' diagnostic effort, GPs' referrals and efficiency. We introduce a controlled incentivized experiment in which we exogenously vary the amount of the referral payment. Understanding how referral payments affect behavior of GPs interacting with specialists is not only important when taking a market perspective, but even more so within (vertically integrated) health care organizations.

We find evidence that GPs choose efficiently most often under low referral payments. They exert a diagnostic effort and treat mildly ill patients themselves, whereas they refer severely ill patients to specialists. Further, our results show that introducing low referral payments especially motivates the physicians who are barely altruistic whereas highly altruistic physicians do not change their referral behaviour. Increasing the referral payment to a high level increases the number of referrals of both patient types significantly. In line with theory, we find that almost all patients are referred under high referral payments.

From our experiments, we also learn that referrals from specialists to GPs could be an effective means to enhance efficiency. As total efficiency decreases more substantially in a GP's treatment of a severely ill patient than in an unnecessary referral of a mildly ill patient, efficiency reaches its maximum under high referral payments. Taken at face value, referral payments help to increase efficiency by about 40% under high referral payments.

An appealing feature of our parsimonious experimental design is that it lends itself to further research. For example, an interesting follow-up question, from a market and organizational design point of view, would be on the effect of an endogenous choice of the referral payment by specialists. Further, understanding how competition among GPs and specialists affect referrals and efficiency is an important question for future research.

Appendix A Proof of Lemma 1

First note that it is never optimal for a GP to diagnose and not to differentiate between L -type and H -type patients in his referral decision. If it was optimal for a GP not to differentiate between the two patient types in the second stage, diagnosing the patient in the first stage would decrease expected utility without consequences for the referral decision. Hence, three action sets remain to be considered:

- (1) not exerting effort and treating L - and H -type patients,
- (2) not exerting effort and referring L - and H -type patients, and
- (3) exerting effort and treating L -type and referring H -type patients.

These three action sets lead to the expected utilities

$$(1) \rightsquigarrow U_{\{\text{no effort, treat, treat}\}}^{GP} = p^{GP} - c_L^{GP} + \alpha_j \left[H_0 - \frac{1}{2}(M_L - T_L) - \frac{1}{2}(M_H + \Lambda - T_H) \right], \quad (\text{A.1})$$

$$(2) \rightsquigarrow U_{\{\text{no effort, refer, refer}\}}^{GP} = R + \alpha_j \left[H_0 - \frac{1}{2}(M_L - T_L) - \frac{1}{2}(M_H - T_H) \right], \quad (\text{A.2})$$

$$(3) \rightsquigarrow U_{\{\text{effort, treat, refer}\}}^{GP} = \frac{1}{2} \left[p^{GP} - c_L^{GP} + \alpha_j(H_0 - M_L + T_L) - c_e^{GP} \right] + \frac{1}{2} \left[R + \alpha_j(H_0 - M_H + T_H) - c_e^{GP} \right]. \quad (\text{A.3})$$

Hence, (1) leads to a higher expected utility than (2) if

$$R < p^{GP} - c_L^{GP} - \frac{1}{2}\alpha_j\Lambda, \quad (\text{A.4})$$

(2) leads to a higher expected utility than (3) if

$$R > p^{GP} - c_L^{GP} - 2c_e^{GP}, \quad (\text{A.5})$$

(3) leads to a higher expected utility than (1) if

$$R > p^{GP} - c_L^{GP} + 2c_e^{GP} - \alpha_j \Lambda. \quad (\text{A.6})$$

□

Appendix B Extension of the Model: GPs' taste for Efficiency

Again, three action sets are to be considered:

- (1) not exerting effort and treating L - and H -type patients,
- (2) not exerting effort and referring L - and H -type patients, and
- (3) exerting effort and treating L -type and referring H -type patients.

These three action sets lead to the expected utilities

$$(1) \rightsquigarrow U_{\{\text{no effort, treat, treat}\}}^{GP} = p^{GP} - c_L^{GP} + \alpha_j \left[H_0 - \frac{1}{2} (M_L + c_L^{GP} - T_L) - \frac{1}{2} (M_H + \Lambda + c_L^{GP} + c_H^{SP} - T_H) \right], \quad (\text{B.1})$$

$$(2) \rightsquigarrow U_{\{\text{no effort, refer, refer}\}}^{GP} = R + \alpha_j \left[H_0 - \frac{1}{2} (M_L + c_L^{GP} - T_L) - \frac{1}{2} (M_H + c_H^{SP} - T_H) \right], \quad (\text{B.2})$$

$$(3) \rightsquigarrow U_{\{\text{effort, treat, refer}\}}^{GP} = \frac{1}{2} \left[p^{GP} - c_L^{GP} + \alpha_j (H_0 - M_L - c_L^{GP} + T_L) - c_e^{GP} \right] + \frac{1}{2} \left[R + \alpha_j (H_0 - M_H - c_H^{SP} + T_H) - c_e^{GP} \right]. \quad (\text{B.3})$$

Hence, (1) leads to a higher expected utility than (2) if

$$R < p^{GP} - c_L^{GP} - \frac{1}{2} \alpha_j (c_L^{GP} + \Lambda), \quad (\text{B.4})$$

(2) leads to a higher expected utility than (3) if

$$R > p^{GP} - c_L^{GP} - 2c_e^{GP}, \quad (\text{B.5})$$

(3) leads to a higher expected utility than (1) if

$$R > p^{GP} - c_L^{GP} + 2c_e^{GP} - \alpha_j (c_L^{GP} + \Lambda). \quad (\text{B.6})$$

Appendix C Additional Tables

Table C.1: Random effects panel logit regressions on GP's effort decision with robust standard errors clustered on subject level.

Model:	(1)	(2)	(3)
Dependent Variable:	Diagnostic Effort	Diagnostic Effort	Diagnostic Effort
LOW ($R_0 = 100$)	-0.013 (0.088)	-0.014 (0.102)	0.021 (0.089)
HIGH ($R_1 = 160$)	0.036 (0.089)	0.039 (0.102)	0.094 (0.085)
HIGH-2 ($R_2 = 200$)	-0.098 (0.078)	-0.101 (0.089)	-0.053 (0.080)
Gender (= 1 if female)			0.061 (0.054)
Age			0.012*** (0.004)
Med (= 1 if medical student)			0.095 (0.077)
Intermediate altruism			0.151** (0.065)
High altruism			0.366*** (0.065)
Observations	2520	2520	2520

Coefficients represent average marginal effects. Standard errors are clustered on the individual level and are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. p-values are based on two-tailed tests. The reference treatment is *Baseline*. Note that results from panel probit regression with robust standard errors clustered on subject level do not differ qualitatively.

Appendix D Instructions of the experiment

[Instructions are translated from German. Insertions for different treatments are indicated in squared brackets.]

You are now taking part in an economic decision-making experiment. Please carefully read the instructions for the experiment. It is very important that you do not talk with other participants for the duration of the experiment. If you disregard the rules, you can be excluded from the experiment and you won't receive payment.

If there is something you don't understand, please have a look again at these instructions. If you still have questions, please raise your hand. We will come to your desk and answer your question personally.

For showing up you receive an expense allowance in the amount of 2.50 EUR. Over the course of the experiment you can earn additional money. The amount of your earnings is dependent on your own decisions and the decisions of the other participants. You will not find out the names of the other participants at any time. The other decision-makers will not learn your identity either.

In this experiment, all amounts of money are shown in Taler, at a rate of 20 Taler = 1 EUR

At the end of this experiment, the amount of money you've earned will be paid cash. You are going to make your decisions on the computer monitor in a cubicle. All data and responses will be evaluated anonymously. You drew a personal cubicle number to guarantee your anonymity.

The experiment lasts about 90 minutes and consists of two parts. Before each of the parts, you will receive detailed instructions. Note that your decisions in the first part of the experiment and your decisions in the second part of the experiment don't have an influence on each other.

In both parts of the experiment, one decision will be selected randomly to determine your payment for this part of the experiment. At the end of the experiment,

we ask you to answer a few questions.

First part of the experiment

Procedure and decision situation

The first part of the experiment comprises a total of 20 rounds. In this experiment, there are two types of decision-makers: general practitioner or specialist physician. At the beginning of the experiment, it will be randomly defined in which role—either general practitioner or specialized physician—you will make your decisions over the course of the entire experiment. The role in which you make decisions remains the same for the whole first part of the experiment.

In each round, a general practitioner and a specialized physician will be randomly paired and will then interact in one round. You don't know the identity of the other decision-makers at any time, and you make your decisions anonymously.

In each round, one patient demands for medical treatment. This patient has either a simple disease or a serious disease. In half of the cases it's a serious disease, and in the other half it's a simple disease. The patient himself can't observe whether he has a simple or a serious disease. A patient with a simple disease receives the highest benefit from a simple treatment, while a patient with a serious disease receives the highest benefit from a specialized treatment. Each round consists of two stages.

Stage 1: Decision of the general practitioner

In the first stage, the general practitioner decides whether or not he performs a diagnosis for the patient who demands medical services. A diagnosis gives the general practitioner certain information about whether the patient has a simple or a serious disease. If the general practitioner performs a diagnosis, costs in the amount of 10 Taler are incurred. If the general practitioner doesn't perform a diagnosis, he won't receive information about the patient's type of the disease. If the general practitioner doesn't perform a diagnosis, there are no costs for him.

Afterwards the general practitioner can decide whether he treats the patient himself or whether he refers the patient to the specialist physician. The decision is independent of the general practitioner's decision about the diagnosis.

The general practitioner can only perform a simple treatment. If the general practitioner treats a patient with a serious disease, the patient's state of health will deteriorate even further. The seriously ill patient then has to be treated by the specialist physician.

If the general practitioner treats the patient himself, he gets a lump sum payment in the amount of 250 Taler. The treatment incurs costs in the amount of 100 Taler.

- For a referral of the patient to the specialist physician, the general practitioner receives an amount of 200 Taler from the specialist physician. There are no costs incurred by the general practitioner for the referral.

The earnings of the general practitioner for one patient comprise as follows:

If the general practitioner conducts the treatment himself:

- Earnings of the general practitioner = lump sum payment for the treatment costs of the treatment (if applicable) costs for the diagnosis

If the general practitioner refers:

- Earnings of the general practitioner = Payment to the specialist physician for the referral - (if applicable) costs for the diagnosis

With his decisions, the general practitioner determines not only his own earnings but also the benefit experienced by the patient.

Stage 2: Decision of the specialist physician

In the second stage, the specialist physician chooses the medical treatment for the patient, in the event that the patient was referred by the general practitioner or the patient is seriously ill and was treated by the general practitioner in stage 1. The specialist physician can observe in the course of his treatment whether the patient has a simple or serious disease.

For patients with the simple disease, the specialist physician always conducts a simple treatment and, for patients with a serious disease, a specialized treatment; in other words, the specialist physician must always choose the treatment that offers maximum benefit to the patient. The specialist physician receives compensation for a treatment in the amount of 420 Taler. For each of the two treatments, the specialist physician incurs costs in the amount of 150 Taler. The specialist physician pays the general practitioner 200 Taler for a patient referral.

The earnings of the specialist physician for one patient comprises as follows:

- Earnings of the specialist physician = lump sum payment for the treatment costs for the treatment - payment to the general practitioner for the referral

With his decision, the specialist physician determines not only his own earnings, but also the benefit experienced by the patient.

At the end of each round, the general practitioner as well as the specialized physician receive information about the amount of their earnings and about the amount by which the patient benefited from the treatment(s) in this round.

Patient benefit

In this experiment there are no participants in the laboratory who assume the role of the patient. The benefit that the patient receives from your treatment from one randomly selected round will benefit a real patient: The amount that arises from patient benefit based on decisions will be provided to the Christoffel Blindenmission (Christoffel Blindenmission Deutschland e.V., 64625 Bensheim). It is engaged in enabling the treatment of patients with cataracts, an eye disease. The patient benefit is dependent on the disease and the treatment from the general practitioner and/or the specialist physician and can therefore be different in each round. The state of health before the disease of both patients is identical and is expressed by 210 Taler.

Patient's loss of benefit due to the disease:

- A patient with a simple disease has a loss of benefit due the disease in the amount of 250 Taler, and a patient with a serious disease has a loss of benefit in the amount of 470 Taler.

Patient benefit from the treatment:

- A patient with a simple disease has a benefit of 250 Taler from the treatment— independent of whether the patient is treated by the general practitioner or if he is referred to and treated by the specialist physician.
- A patient with a serious disease has a benefit of 470 Taler from the specialized treatment, if the general practitioner referred him to the specialized physician. The specialist physician provides the specialized treatment to the patient.
- A patient with a serious disease receives a benefit of 260 Taler from treatment if the general practitioner first treated the patient himself (so that the patient benefit by the specialized treatment is reduced) and the patient then gets a specialized treatment from the specialist physician.

The overall benefit of the patient comprises as follows:

- Overall patient benefit = state of health - loss of benefit of the patient due to the disease + patient benefit from the treatment (by the general practitioner and/or specialist physician)

A concluding overview to the composition of the overall benefit of the patient is shown in the following table:

State of health (in Taler)	Severity of the patient's disease	Loss of patient benefit due to the disease (in Taler)	Decision of the general practitioner	Patient benefit from the treatment by the general practitioner (in Taler)	Patient benefit from the treatment by the specialized physician (in Taler)	Overall patient benefit (in Taler)
210	Simple	-250	Treat patient	+ 250	-	210
210	Simple	-250	Referral to the specialist physician	-	+250	210
210	Serious	-470	Treat patient	-210	+470	0
210	Serious	-470	Referral to the specialist physician	-	+470	210

Payment procedure

The proper disbursement to the participants and the transfer of the amount to the Christoffel Blindenmission e.V. will be confirmed by an overseer. After the disbursement of earnings to all participants, the overseer will sign to verify for the finance department of the University of Cologne that all participants were paid correctly. For the transfer to the Christoffel Blindenmission, the overseer will enter the amount in Euro that arises from the resulting patient benefit in the randomly selected round into a payment order to the Christoffel Blindenmission. The payment of the amount from this experiment to the Christoffel Blindenmission will be arranged by the finance department of the University of Cologne.

The order will be addressed to the finance department of the University of Cologne and put into a stamped envelope. Together, the overseer and the experimenter will throw this envelope in the closest letterbox.

After the second part of the experiment, one participant will be randomly selected as the overseer. The overseer receives beyond his payment from the experiment an additional expense allowance in the amount of 5 Euro. With her signature on a form, the overseer confirms that she has correctly performed the tasks as described.

On request, a copy of this form as well as a copy of the confirmation of the receipt of payment from the Christoffel Blindenmission will be sent from the Department of Business Administration and Health Care Management via e-mail to all participants.

Comprehension questions

Before the decision-making in the experiment begins, we would like to ask you to respond to some practice questions. These practice questions should help to familiarize you with the decision situation. If you have questions, please raise your hand. After all participants have correctly answered the practice questions, the first part of the experiment will start.

Second part of the experiment (following Godager et al., 2016)

In this part of the experiment, you will take on the role of a physician and make decisions about the treatment of various patients. In total, you will determine the quality of care that you would like to provide for eight different types of patients. For each of these patients you can choose quality of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10.

The demand for medical care by the various patient types is determined only after you have made your decisions about the quality of care for all eight types. In total, 100 patients of each type demand medical care. It will only be determined after you have made your decisions about the quality of care for all eight

types how many of the 100 patients of each type wish to seek treatment from you.

Earnings

For each patient who seeks medical care from you, you receive a lump sum that is independent of the quality of care you have selected. You incur costs with your selection of the quality of care. These costs depend on the quality level you choose and can vary between the different patient types. Your earnings for each patient type are as follows:

$$\text{Earnings} = (\text{Lump sum} - \text{Costs}) \times \text{Number of patients who seek medical care from you}$$

(when read: your earnings are equal to the difference between the lump sum and the costs that arise from the quality of care you have chosen, multiplied by the number of patients who seek treatment from you.)

With the quality of care you choose, you determine not only your own earnings, but also the utility enjoyed by the patient. The amount of the lump sum, your costs, your earnings, and the patient's utility will be displayed on your screen (as illustrated below) for each patient type.

Before you choose the quality of care for each patient type, you have the opportunity to click on the "calculator" button and thereby calculate patients' potential demand for treatment (as illustrated below). You can enter the quality you would like to provide as many times as you want. Clicking on the "calculate" button provides you with information about the number of patients who would seek care given the quality level you entered. In addition, you receive information about the resulting earnings and patient utility. You define the quality of care that you wish to provide by entering that quality in the field "your decision" and confirming this entry with OK."

Payment

After the conclusion of the experiment, one of the 8 decisions will be randomly chosen to function as the relevant round for determining your payment for this

Verbleibende Zeit [sec]:

Patiententyp 1
Pauschale Vergütung: 10

Qualität	Kosten	Profit	Patientennutzen
0	0.00	10.00	0
1	0.10	9.90	1
2	0.40	9.60	2
3	0.90	9.10	3
4	1.60	8.40	4
5	2.50	7.50	5
6	3.60	6.40	6
7	4.90	5.10	7
8	6.40	3.60	8
9	8.10	1.90	9
10	10.00	0.00	10

Ihre Entscheidung:

part of the experiment. The earnings from this randomly chosen round will be converted into Euro at the end of the experiment and paid out to you in cash. There are no participants present in the lab who take on the role of patients. An actual patient will benefit from the patient utility resulting from the quality of care you selected in the randomly chosen round: A monetary value equaling the patient utility derived from your decision, multiplied by the number of patients who seek treatment from you, will be transferred to Christoffel Blindenmission Deutschland e.V., 64625 Bensheim. This organization will use the funds to enable the treatment of patients suffering from cataracts, a serious eye condition.

Comprehension questions

Before proceeding to the decisions in the experiment, we would like to ask you to answer several control questions. These control questions should make it easier for you become acquainted with the decision-making situation. If you have questions about this, please raise your hand. The second part of the experiment will begin after all participants have correctly answered the control questions.

Payment procedure

The proper disbursement to the participants and the transfer of the amount to

Verbleibende Zeit [sec]:

Patiententyp 1
Pauschale Vergütung: 10

Qualität	Kosten	Profit	Patientennutzen
0	0.00	10.00	0
1	0.10	9.90	1
2	0.40	9.60	2
3	0.90	9.10	3
4	1.60	8.40	4
5	2.50	7.50	5
6	3.60	6.40	6
7	4.90	5.10	7
8	6.40	3.60	8
9	8.10	1.90	9
10	10.00	0.00	10

	Meine Qualität	
Qualität	<input type="text"/>	<input type="button" value="Berechnen"/>
Anzahl der Patienten		
Profit		
Patientennutzen		

Ihre Entscheidung:

the Christoffel Blindenmission e.V. will be confirmed by an overseer. After the disbursement of earnings to all participants, the overseer will sign to verify for the finance department of the University of Cologne that all participants were paid correctly. For the transfer to the Christoffel Blindenmission, the overseer will enter the amount in Euro that arises from the resulting patient benefit in the randomly selected round into a payment order to the Christoffel Blindenmission. The payment of the amount from this experiment to the Christoffel Blindenmission will be arranged by the finance department of the University of Cologne.

The order will be addressed to the finance department of the University of Cologne and put into a stamped envelope. Together, the overseer and the experimenter will throw this envelope in the closest letterbox.

After the second part of the experiment, one participant will be randomly selected as the overseer. The overseer receives beyond his payment from the experiment an additional expense allowance in the amount of 5 Euro. With her signature on a form, the overseer confirms that she has correctly performed the tasks as described.

On request, a copy of this form as well as a copy of the confirmation of the receipt of payment from the Christoffel Blindenmission will be sent from the Department of Business Administration and Health Care Management via e-mail to all participants.

References

- ALLARD, M., I. JELOVAC, AND P. LÉGER (2011): “Treatment and Referral Decisions under Different Physician Payment Mechanisms,” *Journal of Health Economics*, 30, 880–893.
- ANDREONI, J. (1989): “Giving with impure altruism: applications to charity and Ricardian equivalence,” *Journal of Political Economy*, 1447–1458.
- ANDREONI, J. AND J. MILLER (2002): “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism,” *Econometrica*, 70, 737–753.
- ARROW, K. J. (1963): “Uncertainty and the Welfare Economics of Medical Care,” *American Economic Review*, 53, 941–969.
- BARROS, P. P. AND P. OLIVELLA (2005): “Waiting Lists and Patient Selection,” *Journal of Economics & Management Strategy*, 14, 623–646.
- BIGLAISER, G. AND C.-T. A. MA (2007): “Moonlighting: public service and private practice,” *RAND Journal of Economics*, 38, 1113–1133.
- BREKKE, K. R., R. NUSCHELER, AND O. R. STRAUME (2007): “Gatekeeping in health care,” *Journal of Health Economics*, 26, 149–170.
- BREKKE, K. R., L. SICILIANI, AND O. R. STRAUME (2012): “Quality competition with profit constraints,” *Journal of Economic Behavior & Organization*, 84, 642–659.
- BROSIG-KOCH, J., H. HENNIG-SCHMIDT, N. KAIRIES, AND D. WIESEN (2013): “How Effective are Pay-for-performance Incentives for Physicians? A Laboratory Experiment,” *Ruhr Economic Papers*.
- BROSIG-KOCH, J., H. HENNIG-SCHMIDT, N. KAIRIES-SCHWARZ, AND D. WIESEN (forthcominga): “The Effects of Introducing Mixed Payment Systems for Physicians: Experimental Evidence,” *University of Duisburg-Essen - Health Economics*.

- (forthcomingb): “Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision,” *University of Duisburg-Essen - Journal of Economic Behavior & Organization*.
- CAMERER, C. (2003): *Behavioral Game Theory*, Princeton (NJ), Princeton University Press.
- CHALKLEY, M. AND J. M. MALCOMSON (1998): “Contracting for Health Services when Patient Demand Does not Reflect Quality,” *Journal of Health Economics*, 17, 1–19.
- CHANDRA, A. AND J. SKINNER (2012): “Technology Growth and Expenditure Growth in Health Care,” *Journal of Economic Literature*, 50, 645–80.
- CHARNESS, G. AND M. RABIN (2002): “Understanding social preferences with simple tests,” *Quarterly journal of Economics*, 817–869.
- CHONÉ, P. AND C.-T. A. MA (2011): “Optimal Health Care Contract under Physician Agency,” *Annales d’Economie et de Statistique*, 101/202, 229–256.
- DECAMP, M. AND L. S. LEHMANN (2015): “Guiding Choice Ethically Influencing Referrals in ACOs,” *New England Journal of Medicine*, 372, 205–207.
- DELLAVIGNA, S., J. A. LIST, AND U. MALMENDIER (2012): “Testing for Altruism and Social Pressure in Charitable Giving,” *Quarterly Journal of Economics*, 127, 1–56.
- ELLIS, R. P. AND T. G. MCGUIRE (1986): “Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply,” *Journal of Health Economics*, 5, 129–151.
- FALK, A. AND J. J. HECKMAN (2009): “Lab Experiments Are a Major Source of Knowledge in Social Sciences,” *Science*, 326, 535–538.
- FARLEY, P. J. (1986): “Theories of the Price and Quantity of Physician Services : A synthesis and critique,” *Journal of Health Economics*, 5, 315–333.

- FISCHBACHER, U. (2007): “Z-tree: Zurich Toolbox for Readymade Economic Experiments – Experimenter’s Manual,” *Experimental Economics*, 10, 171–178.
- GARBER, A. M. AND J. SKINNER (2008): “Is American Health Care Uniquely Inefficient?” *Journal of Economic Perspectives*, 22, 27–50.
- GARCIA-MARIÑOSO, B. AND I. JELOVAC (2003): “GPs’ payment contracts and their referral practice,” *Journal of Health Economics*, 22, 617–635.
- GARICANO, L. AND T. SANTOS (2004): “Referrals,” *American Economic Review*, 94, 499–525.
- GODAGER, G., T. IVERSEN, AND C.-T. A. MA (2015): “Competition, gate-keeping, and health care access,” *Journal of health economics*, 39, 159–170.
- GODAGER, G., C.-T. A. MA, AND D. WIESEN (2016): “Competition and Altruism,” Working paper, Boston University.
- GODAGER, G. AND D. WIESEN (2013): “Profit or Patients’ Health Benefit? Exploring the Heterogeneity in Physician Altruism,” *Journal of Health Economics*, 32, 1105–1116.
- GONZÁLEZ, P. (2010): “Gatekeeping versus direct-access when patient information matters,” *Health Economics*, 19, 730–754.
- GRASSI, S. AND C.-T. A. MA (forthcoming): “Information Acquisition, Referral, and Organization,” *University of Lausanne - RAND Journal of Economics*.
- GREINER, B. (2004): “An Online Recruitment System for Economic Experiments,” in *Forschung und wissenschaftliches Rechnen : Beiträge zum Heinz-Billing-Preis 2003*, ed. by K. Kremer and V. Macho, Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, 79–93.
- HENNIG-SCHMIDT, H., R. SELTEN, AND D. WIESEN (2011): “How Payment

- Systems Affect Physicians' Provision Behavior – An Experimental Investigation,” *Journal of Health Economics*, 30, 637–646.
- HENNIG-SCHMIDT, H. AND D. WIESEN (2014): “Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students,” *Social Science & Medicine*, 108, 156 – 165.
- INDERST, R. AND M. OTTAVIANI (2012): “Competition through Commissions and Kickbacks,” *American Economic Review*, 102, 780–809.
- KESER, C., C. MONTMARQUETTE, M. SCHMIDT, AND C. SCHNITZLER (2014a): “Custom-Made Healthcare - An Experimental Investigation,” University of Goettingen, cege Discussion Papers No. 218.
- KESER, C., E. PETERLE, AND C. SCHNITZLER (2014b): “Money talks: Paying physicians for performance,” University of Goettingen, cege Discussion Papers No. 218.
- KESTERNICH, I., H. SCHUMACHER, AND J. WINTER (2015): “Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus?” *Journal of Public Economics*, 131, 1 – 11.
- LAGARDE, M. AND D. BLAAUW (2015a): “Carrot, stick or competition? Experimental evidence of performance contracts in a multitasking setting,” Working paper, London School of Hygiene and Tropical Medicine.
- (2015b): “Studying physicians’ altruism and responses to incentives: a medically framed real effort experiment,” Working paper, London School of Hygiene and Tropical Medicine.
- LEE, H.-H., E. J. PINKER, AND R. A. SHUMSKY (2012): “Outsourcing a two-level service process,” *Management Science*, 58, 1569–1584.
- MALCOMSON, J. M. (2004): “Health Service Gatekeepers,” *RAND Journal of Economics*, 35, 401–421.

- MCCLELLAN, M., A. N. MCKETHAN, J. L. LEWIS, J. ROSKI, AND E. S. FISHER (2010): “A National Strategy To Put Accountable Care Into Practice,” *Health Affairs*, 29, 982–990.
- MCGUIRE, T. G. (2000): “Physician Agency,” in *Handbook of Health Economics, Vol. 1 A*, ed. by Cuyler and Newhouse, North-Holland, Amsterdam (The Netherlands), 461–536.
- MORREIM, E. (1989): “Conflicts of interest: Profits and problems in physician referrals,” *Journal of the American Medical Association*, 262, 390–394.
- OWEN, B. M. (1977): “Kickbacks, specialization, price fixing, and efficiency in residential real estate markets,” *Stanford Law Review*, 931–967.
- PAULY, M. V. (1979): “The Ethics and Economics of Kickbacks and Fee Splitting,” *The Bell Journal of Economics*, 10, 344–352.
- SCOTT, A. (2000): “Economics of general practice,” *Handbook of health economics*, 1, 1175–1200.
- SHUMSKY, R. A. AND E. J. PINKER (2003): “Gatekeepers and referrals in services,” *Management Science*, 49, 839–856.
- SONG, Z., T. D. SEQUIST, AND M. L. BARNETT (2014): “Patient referrals: A linchpin for increasing the value of care,” *Journal of the American Medical Association*, 312, 597–598.
- WILENSKY, G. R., N. WOLTER, AND M. M. FISCHER (2007): “Gain sharing: a good concept getting a bad name?” *Health Affairs*, 26, w58–w67.
- ZHANG, Z. G., H. P. LUH, AND C.-H. WANG (2011): “Modeling security-check queues,” *Management Science*, 57, 1979–1995.