



Physician behavior and conditional altruism: the effects of payment system and uncertain health benefit

Peter Martinsson^{1,2} · Emil Persson¹

Published online: 29 June 2019
© The Author(s) 2019

Abstract

This paper experimentally investigates the altruistic behavior of physicians and whether this behavior is affected by payment system and uncertainty in health outcome. Subjects in the experiment take on the role of physicians and decide on the provision of medical care for different types of patients, who are identical in all respects other than the degree to which a given level of medical treatment affects their health. We investigate physician altruism from the perspective of ethical principles, by categorizing physicians according to how well their treatment decisions align with different principles for priority setting. The experiment shows that many physicians are altruistic toward their patients but also that the degree of altruism varies across patients with different medical needs. We find a strong effect of payment system that is overall unaffected by the introduction of risk and ambiguity in patients' health outcomes. There is, however, substantial heterogeneity across individuals, in particular under the capitation payment system where physicians' responses to the introduction of uncertainty in patient health are modulated by their own generic risk and ambiguity preferences.

Keywords Physician behavior · Experiment · Altruism · Uncertainty · Principles for priority setting

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11238-019-09714-7>) contains supplementary material, which is available to authorized users.

✉ Emil Persson
emil.persson@liu.se

Peter Martinsson
peter.martinsson@economics.gu.se

¹ Department of Management and Engineering, Division of Economics, Linköping University, 581 83 Linköping, Sweden

² Department of Economics, University of Gothenburg, Box 640, 40530 Gothenburg, Sweden

1 Introduction

Uncertainty in health and altruistic behavior by doctors are two fundamental aspects in the organization of health care (Arrow 1963). Individuals are uncertain about the incidence of and recovery from diseases and illnesses, and about the availability of treatments and their effectiveness for health improvements. Naturally, physicians have greater knowledge about the alternative treatments and their consequences for recovery. This information asymmetry between patients and physicians is a traditional principal–agent problem. Physicians will to a large extent, if not fully, influence the quantity and quality of the health care provided, and their effort is influenced by their preferences for own income as well as altruism toward the patient, which are together affected by the prevailing payment system (e.g., McGuire 2000, for a general overview of payment systems). Due to budget constraints, a central issue in the provision of health care is allocation. Many countries have implemented explicit prioritization rules based on general ethical principles, such as equal access, cost and clinical effectiveness, and severity-graded valuation of health gains. Physicians are then incentivized to provide medical treatment that is in each patient’s best interest given these rules. As physicians are the gatekeepers to the health care system, it is important for policy makers to understand how physicians trade-off own income and altruism and how their behavior in this way corresponds to general principles for priority setting under different payment systems and in the presence of uncertainty about health outcomes. Our objective is to investigate whether and to what extent a physician’s altruistic behavior depends on the payment system and uncertainty in health outcome. To this end, we conduct an experiment building on the novel framework introduced by Hennig-Schmidt et al. (2011). We contribute to filling important gaps in the literature concerning the heterogeneity of physician altruism and the effects of introducing risk and ambiguity in health outcomes for patients.¹

Existing evidence concerning the heterogeneity of physician altruism is relatively scarce.² Hennig-Schmidt et al. (2011) found that altruism was an important component of physicians’ choices, and Godager and Wiesen (2013) and Brosig-Koch et al. (2016, 2017) showed that the level of altruism varied across both physicians and patients. We extend these results both conceptually and empirically by categorizing physicians according to how well their treatment decisions align with three different ethical principles for priority setting: severity of illness, capacity to benefit, and ex post equality. We find that many physicians are altruistic toward their patients and that the degree of altruism varies across patients with different medical needs. Interestingly, the type classification of physicians based on ethical principles is unaffected by payment system; the common categorization is that physician altruism is guided by severity of illness, both under capitation and fee-for-service. We replicate the previous finding that patients are undertreated in capitation payment systems and overtreated in fee-for-service systems, and we show that this result is independent of whether patient’s health benefit is deterministic, risky or ambiguous. This is our main contribution and a

¹ We use the term uncertainty to include both risk (known probabilities) and ambiguity (unknown probabilities).

² For an overview on behavioural health economics see Galizzi and Wiesen (2018).

robustness result for the particular methodology introduced by Hennig-Schmidt et al. (2011) indicating that results from previous studies relying solely on deterministic patient's health benefits can be extended to risky and ambiguous situations. In more detailed analyses, we find substantial individual heterogeneity in physicians' responses to the introduction of risk and ambiguity in patient health. This heterogeneity partly depends on payment system because there is a significant correlation between physicians' behavior and their generic risk and ambiguity preferences under capitation but not under fee-for-service.

2 Background

Medical costs are increasing over time and this is largely driven by an aging population and rapid technological growth in the health care sector (e.g., Newhouse 1992; Cutler 2002). This poses a challenge for health care systems around the world and raises important questions regarding how to organize the health care system, with efficiency and priority setting often being the most important issues to consider (Culyer and Wagstaff 1993; Zweifel et al. 2009). A central element is, of course, that physicians provide medical treatment that is in patients' best interest, given existing rules and regulations for the health care sector. Thus, an important question is how to design physician payment systems to achieve the best possible outcome for patients. The optimal design of these payment systems will naturally depend on how physicians trade-off own income and altruism when making their decisions (e.g., Woodward and Warren-Boulton 1984; Ellis and McGuire 1986; Chalkley and Malcomson 1998; Ma and Riordan 2002; Galizzi et al. 2015).

One common way to understand physicians' behavior is based on Ellis and McGuire (1986), who model physicians as motivated by a combination of altruism and a preference for own income. In this context, 'altruism' (toward the patient) captures the extent to which a physician values the patient's health. Under capitation, the physician receives a fixed payment related to the people she is assumed to be responsible to care for, and then makes an effort to improve their health. Thus, the emphasis is on patient health and the altruism of physicians, who must give up own income to improve the health of the patients. In contrast, under fee-for-service, physicians are paid for each amount of health care provided. Thus, these two systems consider the physician's preference for own income differently. In the extreme case of a physician motivated exclusively by own income maximization, the two systems provide completely different amounts of treatment. Under capitation, there will not be any treatment and hence the patient will be undertreated, while under fee-for-service the patient will receive the maximum treatment, or until the point where the marginal utility of effort equals the utility of income provided by that unit of effort for the physician. In a fee-for-service system, this could result in harmful treatment levels, which relates to the concept of supplier-induced demand (e.g., Fuchs 1978; Dranove and Wehner 1994; Gruber and Owings 1996). In the other extreme case, where the physician is only affected by altruism, both payment systems result in optimal provision of health care for the patient. More likely, however, is that we have intermediate cases where the physician derives utility both from own income and altruism, as discussed by for example Arrow

(1963). The relative degree of physicians' preference for own income in relation to their altruism influences which of the two remuneration systems that is preferred from a societal perspective.

Despite the increased reliance on monetary incentives in health care, little is known empirically about the effects on physicians' decision making at patient level. A key reason is the difficulty to find exogenous variation in physician payment system using naturally occurring data. It is very challenging to conduct field experiments, not least from an ethical perspective since the physical health of the participants would be affected (without their consent), and naturally only a few studies exist.³

An emerging literature in health economics has used laboratory experiments to facilitate detailed investigation of the causal impact of payment system on physician behavior. Hennig-Schmidt et al. (2011) introduced a novel framework where subjects take on the role of physicians and decide on the provision of medical care for different types of patients, who are identical in all respects other than the degree to which a given level of medical treatment affects their health. One of the key benefits of using a lab experiment is that there exists a theoretically known and unique optimal quantity of medical care that maximizes the health benefit in each patient, and furthermore that it facilitates *ceteris paribus* variation in physician payment systems in a controlled environment. A robust finding in this experimental literature is that physicians provide more medical care under fee-for-service than under capitation, to the extent that patients are overtreated in fee-for-service and undertreated in capitation (Hennig-Schmidt et al. 2011; Brosig-Koch et al. 2013; Keser et al. 2014; Hennig-Schmidt and Wiesen 2014; Brosig-Koch et al. 2015, 2017). Moreover, Brosig-Koch et al. (2013) and Brosig-Koch et al. (2015, 2017) investigate whether physicians provide better medical treatment under capitation or fee-for-service, and capitation seems to be the marginally better system in this respect. Using the data in Hennig-Schmidt et al. (2011), Godager and Wiesen (2013) estimate the relative weight attached to own income and altruism. They find that most subjects assign a weight to both components, but there is large heterogeneity in the weights assigned. An interesting possibility concerning physician altruism is that it might also vary across patients with different illness conditions both in terms of severity and benefit from treatments, as discussed by Brosig-Koch et al. (2017).

Physicians might be more altruistic toward patients with more severe illnesses, or they might be more altruistic toward patients with a high capacity to benefit from treatment. These are two different principles for allocation of health care based on medical need: The capacity to benefit principle implies a goal to maximize the overall level of health in the population whereas the severity of illness principle implies that health maximization should be weighted by medical condition (e.g., Dolan and Olsen 2001; Dolan et al. 2005). In practice, capacity to benefit has been the overarching goal

³ Krasnik et al. (1990) find that a change from capitation to fee-for-service for Danish General Practitioners led to higher provision of medical services and lower referral rates. Gaynor and Gertler (1995) analyze survey data on medical group practice and find strong correlation between physician effort and the degree to which payment is linked to productivity. Devlin and Sarma (2008) investigate how physicians in Canada select into different payment schemes and how their provision behavior is affected by the incentive structure. They find that the number of patient visits per week is significantly higher among physicians under fee-for-service. Li et al. (2014) exploit a natural experiment in Canada and find that the introduction of pay-for-performance incentives had a relatively modest effect on the provision of targeted primary care services.

in for example the UK, using maximization of quality-adjusted life years (QALYs) as an explicit criterion, whereas severity of illness has been important in Norwegian, Swedish, and German health policy (Shah 2009). How physicians' behavior corresponds to different principles for priority setting, therefore, has implications for how well the quality and quantity of health care provided correspond with the existing rules and guidelines established for the health care sector. Different principles for allocation according to medical need yield different behavior, which can be traced back to different rules for priority setting (Culyer and Wagstaff 1993; Olsen 1997; Cookson and Dolan 2000). Based on the taxonomy in Williams and Cookson (2000), we can identify three general principles for provision according to medical needs (two of which have already been mentioned). First, the severity of illness principle gives precedence to the patient with the worst no-treatment profile, i.e., the one with the worst outlook in the absence of treatment. Justifications for this principle can be made on grounds of equality (e.g., Williams 1962) and on a "rule of rescue" (Cookson and Dolan 2000). Second, a capacity to benefit principle places the greatest weight on patients who will benefit the most from treatment, justified on grounds of maximizing the overall level of population health. Finally, one could aim to equalize ex post health, justified by for example the "fair innings" argument, which stresses everyone's right to a similarly normal span of life years in good health (Williams 1997).

For a systematic and detailed investigation of the influence of patients' medical needs on physician decision making, an experimental approach is suitable since it facilitates controlled *ceteris paribus* variation in patient characteristics and medical conditions, something that would be extremely difficult to accomplish using naturally occurring data (see, e.g., Falk and Heckman 2009 for a general discussion on causality and experimental and survey methods). Based on an extension of Hennig-Schmidt et al. (2011), we introduce a type classification of physicians in terms of conditional altruism, capturing how well their treatment decisions in the experiment align with the different principles for priority setting discussed above.

Uncertainty is prevalent in almost every situation involving clinical decision making, since it is difficult to pin down the exact probability of a successful treatment outcome due to, for example, the complexity of many illnesses and expert disagreement regarding the effect of many treatments (for a discussion, see, e.g., Berger et al. 2013). The health outcome for the patient can differ across conditions and interventions even when the expected health gain for a single patient is quite similar. In some rare cases, the outcome for the patient is deterministic, i.e., it is known for sure, but in other cases there is a risk in the outcome with known probabilities, for example, based on a large sample of previous treatment outcomes. In other cases, however, there is ambiguity in treatment outcomes, meaning that the probability of a success of a treatment is unknown. Previous experimental research has shown that individuals make different choices under risk than under ambiguity, and that they tend to prefer situations characterized by risk rather than ambiguity even when one takes their subjective beliefs about probabilities into account (Ellsberg 1961; for a review see, e.g., Camerer and Weber 1992). Previous studies that investigated risk taking on behalf of others using neutral or non-medical frames have indicated mixed results compared to risk taking for oneself (e.g., Andersson et al. 2016; Chakravarty et al. 2011; Heinrich and Mayrhofer 2018; Vieider et al. 2016), while similar comparison

for decisions under ambiguity is scarce (König-Kersting and Trautmann 2016). Arrieta et al. (2017) investigated attitudes toward risk taking for others in different medical contexts and found that subjects on average were risk averse but also that the magnitude of risk aversion varied across different contexts. Since previous experimental studies on physicians' decision making abstract away from risk and ambiguity in the effect of medical treatment on patient's health, we introduce these features in our experiment as a robustness check for the particular methodology developed by Hennig-Schmidt et al. (2011). This allows us to investigate if uncertainty could affect physician behavior and altruism toward the patient.

3 Experiment

The experimental design builds on the framework introduced by Hennig-Schmidt et al. (2011) and extended by Brosig-Koch et al. (2016). We stay close to the latter version. Subjects in the role of physicians decide on the quantity of medical treatment $q \in \{0, 1, \dots, 10\}$ for different types of patients. Physicians incur (convex) costs when treating a patient and they also receive a payment, either a fixed sum up front (capitation) or a sum that varies with the level of medical treatment they provide (fee-for-service).

Physicians met patients with five different health profiles (A–E), each characterized by a health benefit function mapping the amount of medical treatment (q) into health benefit. Figure 1 displays the benefit functions for the health profiles used in the experiment. For instance, if the physician gives a patient with health profile A five units of medical treatment ($q = 5$), the resulting patient's health benefit is eight Taler (the experimental currency used, later translated into euros at a 1:1 conversion rate). For each patient, there is a unique interior level of medical treatment where the health benefit is maximized, and in either payment system the physician is faced with a trade-off between patient health and own payoff maximization. Under capitation, physician profit (π) is given by $\pi(q) = R - c(q)$, where R is the capitation payment set to $R = 10$ and $c(q) = q^2/10$ is the convex cost function. Under fee-for-service, physician profit is instead given by $\pi(q) = 2q - q^2/10$, where $2q$ is the fee-for-service payment. Experiment parameters are, thus, set such that a selfish physician would choose $q = 0$ under capitation and $q = 10$ under fee-for-service. A physician who is also motivated by altruism would choose closer to the level of treatment that is optimal for the patient, and if sufficiently altruistic would provide precisely at the optimum. To facilitate comparison across capitation and fee-for-service, the absolute values of marginal profits and the maximum profits are equivalent under both systems. Moreover, the benefit functions are symmetric, which is a simplification of reality but it ensures that the marginal effects (for the patient) of undertreatment and overtreatment are identical (Brosig-Koch et al. 2017).

There are no real patients present in the lab, but the amounts resulting from physicians' decisions are transferred to a charity funding medical treatments. Thus, the health benefits in the experiment were expressed in monetary terms, and in the experiment they were labeled Talers. These features were key parts in Hennig-Schmidt et al. (2011) and have since then been adopted by subsequent studies building on that design.

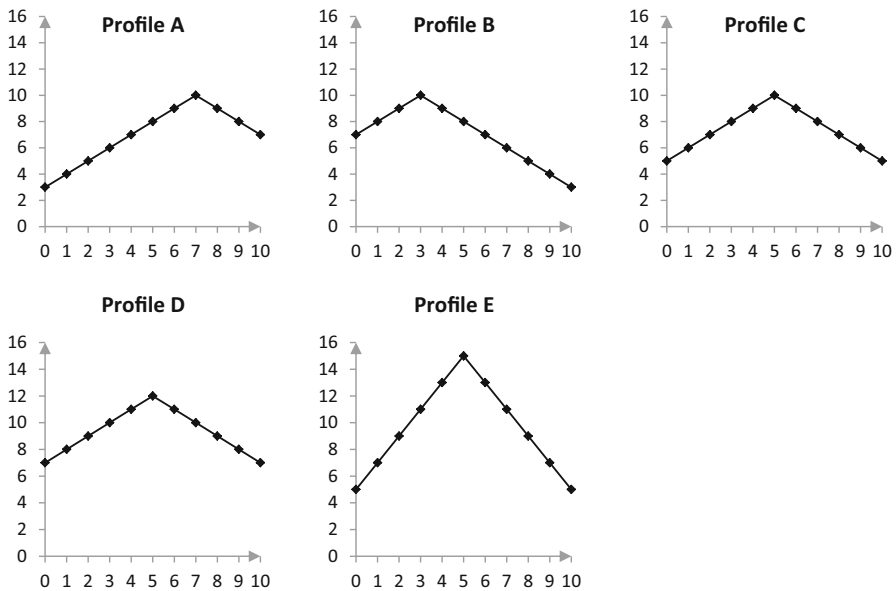


Fig. 1 Patient benefit functions (medical treatment (q) on the horizontal axes and health benefit in monetary terms on the vertical axes). Note: the health benefit (on the vertical axes) for patients is expressed in Taler, the experimental currency that was translated into euros at the end of the experiment

The donation is crucial since it means that decisions taken for (abstract) patients in the experiment have consequences for real patients outside the lab. To ensure participants that the donations would in fact be made after the experiment, they were informed that receipts of each transfer would be published on the experimental laboratory's blackboard on a given date.⁴

To meet our research goals, we extend the design in Brosig-Koch et al. (2016) along two dimensions: (i) allowing for explicit tests for principles for priority settings and (ii) uncertainty. First, our design enables us to investigate physician behavior and altruism toward the patients based on systematic pairwise comparison of different types of benefit functions. Properties of the benefit functions capture how patients differ from each other in terms of medical need, i.e., with respect to severity of health problem (intercept), capacity to benefit from medical treatment (slope), and optimal level of medical treatment. Health profiles C and E are carefully designed to facilitate pairwise comparison in this respect. Thus, C patients are worse off without medical treatment (severity of illness principle) and they require treatment closer to the level where their health benefit is maximized to equalize health ex post (equality principle). Then, we develop a classification of physicians based on how well their treatment decisions for C vs. E align with different principles for priority setting.⁵ These profiles are similar

⁴ The charity used was Christoffel Blindenmission Deutschland e.V., an organization that funds treatment of patients with eye cataracts, and the same organization was used as a recipient in previous studies using this design.

⁵ The analyses of priority setting in this paper focuses on physician behavior in individual treatment decisions and how this behavior may depend on patients' medical needs. For example, a physician may

Table 1 Classification of physicians based on priority principles

| Principle | Criteria | Explanation |
|---------------------|--|---|
| Equality ex post | C is given medical treatment that is closer to the level of treatment where health benefit is maximized ($q = 5$). | C patients have a flatter benefit function, and thus need medical treatment that is closer to the maximizer of the benefit function when the goal is to equalize health benefit ex post. For example, two units of treatment to E and four to C would equalize their health benefit |
| Capacity to benefit | E is given medical treatment that is closer to the level of treatment where health benefit is maximized ($q = 5$). | E patients should be given priority since the marginal effect of medical treatment on health benefit is higher compared to C patients |
| Severity of illness | Provide C and E with the same level of medical treatment, but not at the profit-maximizing level (selfish physicians) or at the level that is optimal for the patient (altruistic physicians) ^a | C patients and E patients have the same no-treatment profile, i.e., identical health without medical treatment |

^aWe classify physicians as selfish if they provide at the profit-maximizing level ($q = 0$ in capitation and $q = 10$ in fee-for-service) for both C and E, and as altruistic if they provide at the level where the benefit for the patient is maximized ($q = 5$) for both C and E

in that the health benefit is maximized at the same level of medical treatment ($q = 5$) and that the functions have the same intercept, but the slope of the benefit function is steeper for E than C. The basis for this classification is summarized in Table 1. We can see in the table that the equality principle requires that treatment to C is closer to the level where the health benefit is maximized, since compared to E the benefit function of C is flatter. On the contrary, by the capacity to benefit principle, E patients should be provided treatment that is closer to the maximizer of the benefit function, since the marginal effect of medical treatment is higher for E than C. Finally, since C and E have the same no-treatment profile, we classify physicians by the severity of illness principle if they provide C and E with equal amounts of medical treatment, but at the same time not at the profit-maximizing level (purely selfish) or at the level that is optimal for the patient (purely altruistic).

The second novelty of our design is that we allow for risk and ambiguity in the outcome of medical treatment. Previous studies only considered deterministic outcomes. We introduced risk as follows: for each deterministic health benefit function, a given level of medical treatment (q) yields a good outcome for the patient in terms of health benefit with a 0.5 probability and a bad outcome for the patient with a 0.5 probability. The good outcome was always one Taler above the corresponding outcome in the

Footnote 5 continued

choose to provide more time and effort on patients with more severe illnesses compared to patients with less urgent medical conditions. This is a form of implicit priority setting that is linked to our overall focus on physician payment systems, as an important channel through which policy makers can influence how the overall provision of medical care aligns with existing rules and guidelines based on general ethical principles. For an example on allocation decisions in a context with explicit capacity constraints, see, e.g., the experiment by Ahlert et al. (2012) or the discussion in Doctor et al. (2009).

deterministic case, and the bad outcome was always one Taler below the corresponding deterministic outcome.⁶ Ambiguity was introduced in a similar manner, but in contrast to risk the probabilities of the outcomes were unknown. Risk was represented by an opaque bag consisting of five black and five white balls, and ambiguity by an opaque bag consisting of ten black and white balls with an unknown composition. At the end of the experiment, a draw from each of these bags determined how risk and ambiguity were resolved at the level of the patient, i.e., whether the patient's benefit resulting from a physician's treatment decision would be the good outcome or the bad outcome (see below for more details about "Experiment" procedures).⁷ Generally speaking, given the health benefit functions and assuming risk aversion, we expect physicians in the experiment to on average provide medical treatment closer to patients' optimal treatment levels when the treatment outcome is risky or ambiguous than when it is certain.⁸ The reason is that physicians cannot offset risk for the patient by choosing a safer treatment, but they can offset the downside risk for each level of medical treatment by providing a little closer to the patient optimum than they thought was ideal when patient's health was deterministic. Of course, this effect depends on how physicians weigh their own profits and patient benefits; the effect is likely strongest for intermediate levels of altruism. Physicians who care mostly about own profits will probably provide close to the profit-maximizing level regardless of risk and ambiguity in patient's health. Similarly, physicians who care mostly about patient's health will always provide at treatment levels that are close to optimal for the patient.

We conducted two types of conditions, with the only difference that a capitation payment system was used in CAP condition and a fee-for-service payment system in FFS condition. In each condition, physicians first made five treatment decisions in the case of deterministic patient health and each patient belonged to a different health profile A–E; then followed another five treatment decisions when patient health was risky; and the last five treatment decisions concerned ambiguous patient health.⁹ Table 2 shows the timeline for the experiment investigating the behavior of physicians and describes each condition in more detail. It is important to ensure that subjects

⁶ The choice of probability at exactly 50–50 is made to facilitate understanding. Other combinations would be interesting as well and plausible from a real-world perspective, such as a 0.9 probability of a modest gain (relative to the deterministic treatment) and a 0.1 probability of a very serious outcome.

⁷ At the end of the experiment, one out of the 15 decisions was randomly chosen to be paid out to subjects and this was determined in two steps: (i) determining if the treatment outcome for each subject individually is good or bad in the risk and ambiguous condition, respectively, and (ii) randomly deciding which of the 15 profiles to be payoff relevant. The outcome in the risky and ambiguous case needs to be resolved in a time efficient way given that the experiment lasted approximately 1.5 h. One randomly selected individual drew a ball from each of the bags representing risk and ambiguity, respectively, and the outcome of the draw was binding for all individuals for whom a risk or ambiguity patient was later randomly drawn to be the payoff relevant patient. Before the drawing, however, all subjects had to guess (in private) the color of the ball drawn from each bag and a correct guess resulted in a good outcome and an incorrect guess resulted in a bad outcome for the payoff-relevant decision of that particular individual. Thus, between subjects the risks were uncorrelated.

⁸ If individuals are ambiguity averse, we expect the treatment level to be closer to the optimal level for the patient when there is ambiguity compared to when there is risk.

⁹ In the instructions (see Online Resource), we don't speak about health profiles that are repeated throughout the experiment but rather about decisions to be made on behalf of different patients that differ in two dimensions, i.e., (i) the shape of the benefit function and (ii) how certain their health benefits from treatment are.

Table 2 Conditions and timeline for the experiment investigating physician behavior

| Condition | Health benefit is deterministic | Health benefit is risky | Health benefit is ambiguous | Participants |
|--------------------|---------------------------------|-------------------------|-----------------------------|--------------|
| CAP | | | | |
| Decision (patient) | 1,2,3,4,5 (A–E) | 6,7,8,9,10 (A–E) | 11,12,13,14,15 (A–E) | 66 |
| FFS | | | | |
| Decision (patient) | 1,2,3,4,5 (A–E) | 6,7,8,9,10 (A–E) | 11,12,13,14,15 (A–E) | 64 |

Health profile in parentheses (Patient 1 has health profile A, Patient 2 has profile B, and so on)

understand the experiment and do not learn during the course of the experiment. To consider the familiarity with the decisions in all treatments, we decided to explain all three treatments before the subjects made any decisions. In addition, we included extensive comprehension questions for the different treatments, where the subjects were asked questions about costs and benefits if certain quantities of medical treatments were provided.¹⁰ We made the comprehension questions deliberately tricky by asking open-ended questions and the subjects could only pass a set of questions if all were answered correctly. If at least one question was incorrect, they were not informed which question(s) that was answered incorrectly. A subject could not begin the experiment before all comprehension questions were correct and all subjects passed several sets of comprehension questions. Thus, when the experiment began, the subjects had information about all treatments and passed several extensive comprehension tests. Given our instructions including the comprehension questions, we do not think that the order of treatments should influence the decisions. Thus, we kept the order of the treatments fixed (deterministic, risky, and ambiguous), which was the same order we introduced them in the introductions.¹¹

The experiment was conducted at MELESSA, University of Munich. It was computerized with z-Tree (Fischbacher 2007) and participants were recruited using ORSEE (Greiner 2015). The choice of subject pool needs to be carefully thought of in relation to the research question. Galizzi and Wiesen (2018) conclude their discussion on subject pool on page 10 as: “Taken together, the experimental studies that systematically account for potential differences in the subject pools indicate that the direction of a treatment effect does not differ between (medical or non-medical) students and medical professionals. Importantly, however, the intensity of a behavioral effect might vary across subject pools”. Our key research question focuses on if there is a difference in treatments, if the health outcome is deterministic, risky or ambiguous, i.e., we are primarily interested in the direction of treatments on behavior rather than the intensity per se. To sum up, our subject pool serves the purpose to allow us to investigate our research question.¹² Subjects were randomized into cubicle workstations and given plenty of time to read the instructions (see Online Resource) and ask questions (in

¹⁰ For example, one of the sets of the comprehension questions for CAP treatment with ambiguity when physician provide 2 quantities of medical treatment: What is the capitation? What is the cost? What is the profit? What is the patient benefit if good outcome? What is the patient benefit if bad outcome?

¹¹ This fixed logical order also reduces the cognitive burden for participants, similar to, e.g., Vieider et al. (2015) who conduct risk experiments followed by ambiguity experiments.

¹² Risk attitudes might differ between medical and non-medical students.

private) before the experiment began. They had to answer several control questions, and the experiment did not start until everyone had answered all questions correctly.

In the experiment that investigates the behavior of physicians, subjects made 15 decisions in total, on how much medical treatment to provide different patients. At the end of the experiment, one of these decisions was randomly drawn and paid out to subjects. Risk was represented by an opaque bag consisting of five black and five white balls, and ambiguity by an opaque bag consisting of ten black and white balls with an unknown composition. At the end of the experiment, one randomly selected individual drew a ball from each bag (and another individual announced the outcome). Before the drawing, all subjects had to guess (in private) the color of the ball drawn from each bag and a correct guess resulted in a good outcome and an incorrect guess resulted in a bad outcome for the corresponding decision of that particular individual. Subjects knew that they could verify the results from the public drawing (and check the composition of each bag) at the end of the experiment.

In a second part of the experiment, subjects' individual risk and ambiguity preferences were elicited in a generic setting. We followed Sutter et al. (2013) and used incentivized choice lists where subjects decided between a fixed prospect and a sure amount of money. The fixed prospect consisted of a 50–50 lottery where the outcome was either 0 or 5 Taler.¹³ Risk was represented by an opaque bag consisting of five black and five white balls. We also test for ambiguity preferences using a choice list where the ambiguous outcome was either 0 or 5 Taler and the payoff was represented by an opaque bag consisting of ten black and white balls, exactly as in the first part of the experiment where subjects made decisions in the role of physicians. One of the choices was randomly selected and played out for real. The money won added to the subjects' total earnings in the experiment. In addition, subjects answered a few general questions related to health care and medical decision making and a questionnaire on socio-economic characteristics. In total, 64 undergraduate students participated in the FFS condition and 66 in CAP. We evaluate the randomization into the two conditions using non-parametric tests. The differences in age, gender, years at the university, and individual risk and ambiguity preferences are insignificant at conventional levels of significance, indicating that the randomization was successful.

4 Results

We begin the analysis by investigating the average quantities of medical treatment chosen for each patient and the optimal quantity required by each patient (Fig. 2). We can see a strong effect of payment system. For example, patient 1, where optimal treatment is 7 units, is on average provided with 8.03 units of medical treatment in FFS and 2.73 units of medical treatment in CAP. The experimental results are the same for

¹³ The subject is asked to make repeated choices between a sure amount and the fixed prospect. The sure amount in the first choice is 0.25 Taler and it is then increased by increments of 0.25 Taler until it reaches 5 Taler. For example, a risk-neutral subject would choose the fixed prospect when the sure amount is below 2.50 Taler, which corresponds to the expected value of the lottery, and the sure amount when this amount is 2.50 Taler or higher. Similarly, a risk-averse subject would choose the fixed prospect fewer times than a risk-neutral subject, while the opposite is true for a risk-loving individual.

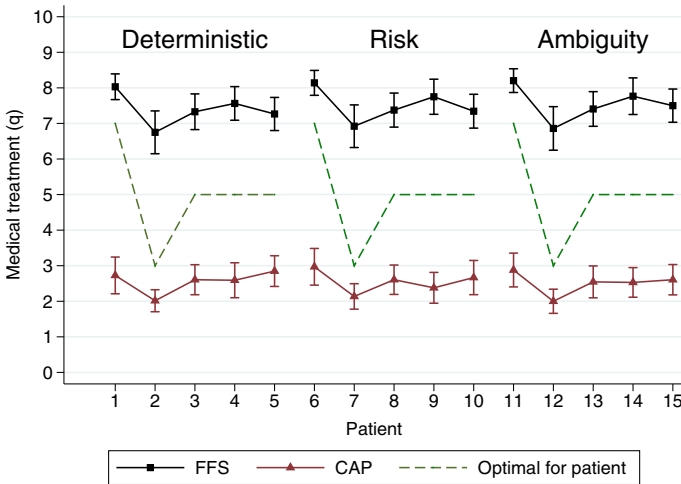


Fig. 2 Average quantity of medical treatment per patient. FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$). Error bars represent 95% confidence intervals calculated from independent sample t tests at the level of each patient (x -axis)

the other patients. These findings are in line with results from previous experiments. Interestingly, the pattern prevails under risk and ambiguity. For example, patients with health profile A (patients 1, 6, and 11) are provided close to 8 units of medical care in FFS irrespective of whether patient health is deterministic, risky, or ambiguous. Taken together, physicians do not change the level of medical treatment for risky and ambiguous health outcomes compared with the case of treating the same types of patients with deterministic health outcomes.

Table 3 shows the effect of payment system on the average deviation from optimal medical treatment. On average, patients are overtreated by 2.48 units in FFS and undertreated by 2.46 units in CAP (Mann–Whitney U test; $P < 0.01$, $N = 130$). Since benefit functions are symmetric, the absolute deviation from optimal treatment is what influences the realized health benefit for the patient. We can see in the table that in this respect there is no significant difference across the two payment systems ($P = 0.82$). Overall, patients are treated equally well (or badly) in either payment system. The results are similar when disaggregated by health profile (A and B vs. C, D, and E) (see Table 9 in the Appendix). The final three columns of Table 3 show the results disaggregated by deterministic, risky, and ambiguous health benefits. We can see that there is a small increase in the deviation from optimal treatment when the outcome in patient health is risky or ambiguous instead of deterministic, but the differences are generally insignificant, except in FFS when health benefits are ambiguous.¹⁴ Taken together, the introduction of risk and ambiguity in the effect of medical treatment on patient health does not affect physicians' provision behavior in the experiment.

Result 1 *The overall pattern of provision behavior is unaffected by the introduction of risk and ambiguity in the effect of medical treatment on patient's health.*

¹⁴ See Tables 10 and 11 in the Appendix for a disaggregation by health profile, which yields similar results.

Table 3 Effect of payment system on medical treatment

| | Average deviation from optimal treatment | Average absolute deviation from optimal treatment | Average absolute deviation from optimal treatment (disaggregated) | | |
|--|--|---|---|--------------------|--------------------|
| | All patients | All patients | Deterministic patients | Risk patients | Ambiguity patients |
| FFS ($N = 64$) | 2.48 (1.79) | 2.53 (1.73) | 2.45 (1.76) | 2.55 (1.73) | 2.58 (1.77) |
| CAP ($N = 66$) | - 2.46 (1.55) | 2.56 (1.43) | 2.52 (1.52) | 2.58 (1.44) | 2.58 (1.46) |
| Null hypothesis | | | | <i>P</i> value | |
| All patients: FFS = CAP (av. deviation) | | | | <0.01 ^a | |
| All patients: FFS = CAP (av. absolute deviation) | | | | 0.82 ^a | |
| FFS: deterministic = risk | | | | 0.15 ^b | |
| FFS: deterministic = ambiguity | | | | 0.04 ^b | |
| CAP: deterministic = risk | | | | 0.91 ^b | |
| CAP: deterministic = ambiguity | | | | 0.79 ^b | |

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)
^aMann–Whitney U test. ^b Wilcoxon signed-rank test. Standard deviation in parentheses. Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit

The effects of risk and ambiguity are investigated in more detail using subjects' generic risk and ambiguity preferences, which were elicited in a later (second) part of the experiment. Table 4 compares deterministic and risky health benefits: medical treatment measured as the absolute deviation from optimal treatment is regressed on a dummy for the five patients with risky benefit functions, on the measure of individual risk preference, and on the interaction between these two variables and several control variables (including age, gender, and health status). Standard errors in the regressions are clustered on individuals. In FFS (columns 1 and 2), there is a small effect (significant at the 10% level) of the introduction of risk for patients, such that physicians on average increase their overtreatment by 0.105 units of medical care; yet there is no effect of physicians' risk preferences. In contrast, in CAP, physicians' risk preferences are correlated with both how well they treat their patients on average (column 3) and how they change their behavior when patient health is risky rather than deterministic (column 4). As indicated by the positive and significant interaction term, introducing risk in the effect of medical treatment on patient's health makes physicians who are more risk averse provide medical treatment closer to patients' optimal treatment levels. This effect is significantly stronger in CAP than in FFS (Appendix Table 12).

Table 5 shows the results from the corresponding analysis of ambiguity in the effect of medical treatment on patient health. We can see an overall and marginally significant effect of ambiguity on physicians' provision behavior in FFS (columns 1 and 2), similar to the effect of risk investigated in Table 4. In CAP, the effect is correlated with physicians' generic ambiguity preferences (column 4) and this effect is similar to the one observed when patient health is risky.

Table 4 Generic risk preference and medical treatment (linear regression)

| Dependent variable: absolute deviation from optimal treatment | FFS | | CAP | |
|--|-----------------|-----------------|-------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| Risky patient | 0.105 (0.058)* | 0.258 (0.214) | 0.013 (0.079) | - 0.719 (0.307)** |
| Risk preference | - 0.089 (0.087) | - 0.081 (0.089) | - 0.127 (0.055)** | - 0.169 (0.057)*** |
| Risky patient × risk preference | - | - 0.02 (0.023) | - | 0.083 (0.039)** |
| Control variables | Yes | Yes | Yes | Yes |
| Observations | 630 | 630 | 640 | 640 |
| Clusters | 63 | 63 | 64 | 64 |

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$. Errors are clustered on individuals (within parentheses). Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit. Risky patient is a dummy for patients for whom the outcome of medical treatment is risky (and the base category is patients for whom the outcome of medical treatment is deterministic). Risk preference is the first row where the subject chose the certain amount of money (over the risky lottery) in the individual choice task, where a higher number indicates less risk aversion. Control variables: age, gender, whether the subject smokes, tries to keep a healthy diet, number of visits to the doctor in the last 3 and 12 months, health status and political views on a scale from left to right. A total of three subjects (in FFS) switched multiple times in at least one of the two individual choice tasks and are, therefore, removed from the analysis. FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

Table 5 Generic ambiguity preference and medical treatment (linear regression)

| Dependent variable: absolute deviation from optimal treatment | FFS | | CAP | |
|---|-----------------|-----------------|-----------------|------------------|
| | (1) | (2) | (3) | (4) |
| Ambiguity patient | 0.143 (0.075)* | - 0.067 (0.260) | 0.05 (0.091) | - 0.345 (0.184)* |
| Ambiguity preference | - 0.091 (0.067) | - 0.096 (0.067) | - 0.036 (0.073) | - 0.060 (0.078) |
| Ambiguity patient × Ambiguity preference | - | 0.009 (0.032) | - | 0.048 (0.022)** |
| Control variables | Yes | Yes | Yes | Yes |
| Observations | 630 | 630 | 640 | 640 |
| Clusters | 63 | 63 | 64 | 64 |

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$. Errors are clustered on individuals (in parentheses). Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit. Ambiguity patient is a dummy for patients for whom the outcome of medical treatment is ambiguous (and the base category is patients for whom the outcome of medical treatment is deterministic). Ambiguity preference is the first row where the subject chose the certain amount of money (over the ambiguity lottery) in the individual choice task, where a higher number indicates less ambiguity aversion. Control variables: age, gender, whether the subject smokes, tries to keep a healthy diet, number of visits to the doctor in the last 3 and 12 months, health status and political views on a scale from left to right. A total of three subjects (in FFS) switched multiple times in at least one of the two individual choice tasks and are, therefore, removed from the analysis. FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

Result 2 *Physicians' generic risk and ambiguity preferences affect their reactions to risk and ambiguity in the effect of medical treatment on patient health under CAP. In particular, more risk-averse physicians provide treatment closer to patients' optimal treatment levels following the introduction of risk in patient health; and more ambiguity-averse physicians provide treatment closer to patient optimum following the introduction of ambiguity in patient health.*

Next, we turn to the analysis of altruism and conditional altruism of physicians. Table 6 displays the proportion of subjects who made a selfish decision at least once and the proportion who were selfish in all their decisions. To make the presentation easier to overview, we restrict attention to the case of deterministic benefit functions given small differences compared to risky and ambiguous benefits. In FFS, 20.3% of physicians always made a selfish decision and 31.3% did so at least once. By comparison, in CAP, 10.6% were always selfish and 27.3% were selfish at least once. It is interesting that physicians who were always selfish when deciding for C–E patients were also selfish in their decisions for patients with profiles A and B, even though it is indirectly less costly (in terms of foregone profit) to provide A (in FFS) and B (in CAP) with optimal treatment.¹⁵ Table shows a similar analysis for the proportion of physicians who made optimal decisions for their patients: 46.9% in FFS and 53.0% in CAP decided optimally at least once, and 10.9% in FFS and 3.0% in CAP provided optimal treatment for all their patients. This indicates that many physicians are altruistic but also that the degree of altruism varies across different types of patients and payment system. This is further explored in Table 7, where physicians' treatment of patients from health profiles C and D is compared. The comparison is interesting since we know that C patients are in equal or greater medical need than D patients. While C and D patients' capacity to benefit from medical care is identical, C patients are worse off without treatment (severity of illness principle) and they require treatment closer to the level where their health benefit is maximized to equalize health ex post (equality principle). We can see in the table that C patients indeed receive better medical treatment than D patients do, and that the difference is most pronounced under fee-for-service.

Result 3 *Many physicians are altruistic toward their patients. The degree of altruism depends on the medical need of the patient: patients in greater need of medical care receive better treatment, i.e., they are given treatment closer to the level where their health benefit is maximized.*

The type classification of physicians based on how their behavior and altruism vary across patients with different medical needs is based on a comparison of medical treatment provided to patients with health profiles C and E. We can identify three broad categories of physicians: (1) purely selfish physicians who always choose so as to maximize own income, (2) purely altruistic physicians who always choose optimally for the patient, and (3) physicians who choose a mix between full altruism and own income maximization. Table 8 shows the distribution of physicians over these three broad categories, based on their decisions when health outcomes are deterministic.

¹⁵ Optimal medical treatment is 7 units for patients from A, 3 units for patients from B, and 5 units for patients from C, D, or E. Thus, in CAP physicians earn a higher profit when B is provided with optimal treatment compared to C, D, or E, FFS physicians earn a higher profit when A is provided with optimal medical treatment.

Table 6 Proportion of selfish and altruistic physicians when patient's health is deterministic, by treatment and health profile

| | Selfish decision at least once (%) | Selfish decision always (%) | Optimal decision at least once (%) | Optimal decision always (%) |
|-------------------------|------------------------------------|-----------------------------|------------------------------------|-----------------------------|
| FFS ($N = 64$) | | | | |
| Profiles: A, B, C, D, E | 31.3 | 20.3 | 46.9 | 10.9 |
| Profiles: A, B | 31.3 | 21.9 | 45.3 | 12.5 |
| Profiles: C, D, E | 29.7 | 20.3 | 32.8 | 17.2 |
| CAP ($N = 66$) | | | | |
| Profiles: A, B, C, D, E | 27.3 | 10.6 | 53.0 | 3.0 |
| Profiles: A, B | 24.2 | 15.2 | 51.5 | 3.0 |
| Profiles: C, D, E | 22.7 | 10.6 | 27.2 | 10.6 |

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$), *Optimal* the level of medical treatment that maximizes the patient's health benefit

Table 7 Provision of medical care to patients from health profiles C and D

| Dependent variable: absolute deviation from optimal treatment | Profile C | Profile D | P value ^a |
|---|-------------|-------------|------------------------|
| FFS ($N = 64$) | | | |
| Deterministic | 2.39 (1.93) | 2.59 (1.85) | 0.05 |
| Risk | 2.41 (1.87) | 2.75 (1.98) | <0.01 |
| Ambiguity | 2.41 (1.95) | 2.83 (1.97) | <0.01 |
| CAP ($N = 66$) | | | |
| Deterministic | 2.45 (1.63) | 2.56 (1.80) | 0.22 |
| Risk | 2.45 (1.59) | 2.68 (1.67) | 0.05 |
| Ambiguity | 2.55 (1.69) | 2.53 (1.60) | 0.82 |

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

^aWilcoxon signed-rank test. Standard deviation in parentheses. Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit

Interestingly, there are more purely selfish physicians but also more purely altruistic physicians in FFS than in CAP, which thus has a greater share of physicians motivated both by own income and altruism. The difference is not statistically significant, however (Chi-square test; $P = 0.13$, $N = 139$).

We compare treatment decisions for patients with health profiles C and E, to categorize subjects based on how their treatment decisions align with different principles for priority setting (explained in Table 1). The distribution of physician types based on this classification is shown in the bottom panel of Table 8. Among those who choose a mix between full altruism and own income maximization, the altruistic part is mostly described by concerns for severity of illness (38.5% and 47.1% of FFS and CAP subjects, respectively). This principle implies priority to patients with worse no-treatment

Table 8 Distribution of physician types

| | Full sample | | Only mix | |
|---------------------------------|------------------|------------------|------------------|------------------|
| | FFS ($N = 64$) | CAP ($N = 66$) | FFS ($N = 39$) | CAP ($N = 50$) |
| 1. Selfish | 20.3% | 12.1% | – | – |
| 2. Altruistic (optimal C and E) | 18.8% | 10.6% | – | – |
| 3. Mix | | | | |
| Principle | | | | |
| Equality ex post (%) | 15.6 | 10.6 | 25.6 | 13.7 |
| Capacity to benefit (%) | 21.9 | 30.3 | 35.9 | 39.2 |
| Severity of illness (%) | 23.4 | 36.4 | 38.5 | 47.1 |
| P value ^a | 0.19 | | 0.35 | |

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

^aChi-square test for equality of proportions between FFS and CAP. The type distribution is based on physician decisions for patients with deterministic benefit functions from health profiles C and E: *selfish* = always profit-maximizing choice ($q = 0$ in CAP and $q = 10$ in FFS); *altruistic* = optimal care (max. health benefit) to both C and E; *equality ex post* = closer to the maximizer of the benefit function ($q = 5$) for C than for E; *capacity to benefit* = closer to the maximizer of the benefit function for E than C; and *severity of illness* = provide C and E with equal amounts of care but not at the profit-maximizing level or at the level that is optimal for the patient. See Table 1 for more details and explanations

profiles, even if their capacity to benefit from medical treatment is comparatively low. The capacity to benefit principle follows quite closely (35.9% of subjects in FFS and 39.2% in CAP), and only a minority of subjects behave in accordance with the principle of ex post equality. We can also see that there are some differences across the two payment systems, in particular that the distribution is more balanced in FFS than in CAP, but the difference is not significant using a Chi-square test.

Result 4 *There are more pure altruists but also more pure selfish physicians in FFS than in CAP, but the more detailed distribution of physician types according to how well their treatment decisions correspond to the different principles for priority setting is not affected by payment system. In both CAP and FFS, altruistic behavior toward the patients is mostly guided by severity of illness.*

5 Discussion and conclusion

We use a lab experiment to investigate the altruistic behavior of physicians and whether this behavior is affected by payment system, and risk and ambiguity in health outcome. The experiment shows that many physicians behave altruistically toward their patients but also that the degree of altruism varies across patients with different medical needs. This variation across patients is best described by concerns for severity of illness. There are more selfish physicians but also more purely altruistic physicians under fee-for-service than under capitation. Interestingly, the type classification of physicians based on ethical principle is unaffected by payment system. Moreover, we replicate the strong effect of payment system on physicians' provision behavior found by previous studies using a similar design, and show that this result is robust to the introduction of risk and

ambiguity in patient health. There is, however, substantial variation across individuals, and this effect partly depends on payment system. Under capitation, greater generic risk aversion means that subjects on average make larger adjustments toward patients' optimal treatment levels when deciding for patients with risky health benefits. There is a similar effect for ambiguity aversion, but neither of these effects is present under fee-for-service payments. Although the overall differences between treatments were not statistically different, a natural extension is to investigate the same research question in a natural setting to investigate the differences in intensity.¹⁶

Medical need is a key aspect in selecting socially preferred allocations of scarce medical resources in both normative analyses and in health policies. Discussing how physician behavior and altruism correspond to different principles for priority setting based on medical need, the capacity to benefit principle can be seen as the benchmark since it is directly connected to the marginal health benefit of the patient.¹⁷ It is interesting to note that a majority of physicians in the experiment do not behave in accordance with this principle, but rather seem to factor in other aspects of medical need as well, such as the severity of illness. This is in line with the public view in many countries, that health improvements are more valuable for people with more severe illnesses (e.g., Dolan et al. 2005; Shah 2009; Nord and Johansen 2014). This is also reflected in health policy. For example, the NICE Citizen Council in the UK has concluded that severity of illness should be taken into account when making decisions, alongside the already established criteria based on cost and clinical effectiveness (NICE 2008).¹⁸ Similarly, in the early 1990s, the State of Oregon increased the coverage of the publicly funded Medicaid health program. Prioritization based strictly on cost effectiveness was emphasized to contain costs, but the public reactions were negative and the policy was later adjusted so that other factors, including the severity of illness, were also allowed to influence prioritization (Ham 1998; Tinghög 2011).

For health policy, the main takeaway message from our experimental findings is that capitation and fee-for-service payment systems seem to affect neither the distribution of medical care nor the overall physician response to risk and ambiguity in patient health for reasonably large probabilities. However, our results, as well as results from most experiments, are based on a narrow set of parameters. For example, many cases in medical decision making involves cases with small probability for a very adverse outcome, and empirical studies on decision making under uncertainty indicate that subjects overestimate losses with small probabilities (e.g., Abdellaoui 2000). Moreover, real-life decisions might also be sensitive to contextual factors, where case-based studies would be needed to investigate such an issue. These important extensions of our work, we leave for future research. One aspect in favor of the capitation system is

¹⁶ For a discussion on the choice of subject pool see Galizzi and Wiesen (2018), who conclude that medical and non-medical students do not differ in the direction of a treatment effect but in the intensity.

¹⁷ For example, in Ellis and McGuire (1986), the marginal rate of substitution between own income and health benefit for the patient determines how much medical treatment the physician wants to provide.

¹⁸ The National Institute for Health and Clinical Excellence (NICE) provides advice and guidance for clinical and public health professionals in the UK. A citizen council was established by NICE in 2002 in order to "provide advice about the social values that should underpin the Institute's guidance" (NICE 2008). The council has met and discussed several different issues and in 2008 they met to discuss the case of severity of illness.

that it resulted in fewer cases of physicians only interested in own income. This payment system also generated a greater proportion of physicians motivated by a mixture of own income and altruism corresponding to one of the three principles for priority setting introduced by Culyer and Wagstaff (1993), and these physicians were generally responsive to patients' medical need in their treatment decisions. However, physicians in the capitation system were also more heterogeneous in how they responded to the introduction of risk and ambiguity in patient health, where their responses seemed to be modulated by their own risk and ambiguity preferences. We reasoned beforehand that introducing risk in patient health should push treatment levels toward the patient optimum if physicians were risk averse, and this is consistent with the effect that we see for risk-averse physicians under capitation, but not under fee-for-service. One interpretation of this difference between payment systems could be that monetary incentives under fee-for-service are more salient to physicians, thus leaving less room for decision making modulated by individual risk preference and other factors. The heterogeneity within each payment system regarding how physicians' respond to patient risk and how their behavior corresponds to general principles for priority setting calls for careful consideration when designing payment systems in the future. In the future, it is also important for research to compare actual behavior with different principles for priority setting, and of course to fine-tune and test for different definitions of principles for priority settings (e.g., Williams and Cookson 2000). An important aspect in this respect could be to develop clearer clinical guidelines, something that has been emphasized, e.g., in the UK. Another suggestion is to test mixed or performance-based payment systems (e.g., Brosig-Koch et al. 2013, 2017), which could be further investigated in future research.

Acknowledgements We are grateful to Håkan J. Holm, Gustav Kjellsson, Martin Kocher, Jeannette Brosig-Koch, Olof Johansson-Stenman, Daniel Wiesen, and two anonymous reviewers for valuable comments and suggestions. We wish to thank seminar and conference participants at the University of Gothenburg and the 4th Workshop in Behavioral and Experimental Health Economics in Cologne for helpful comments and discussion. Financial support from Formas through the program Human Cooperation to Manage Natural Resources (COMMONs) program is gratefully acknowledged, and we kindly thank MELESSA at the University of Munich for providing laboratory resources.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Disaggregation of results by health profile

See Tables 9, 10, 11, 12.

Table 9 Effect of payment system on medical treatment (by health profile)

| | Average deviation from optimal treatment | Average absolute deviation from optimal treatment |
|-------------------|--|---|
| Profiles: A, B | | |
| FFS ($N = 64$) | 2.48 (1.76) | 2.56 (1.69) |
| CAP ($N = 66$) | - 2.55 (1.51) | 2.64 (1.40) |
| <i>P</i> value | <0.01 | 0.59 |
| Profiles: C, D, E | | |
| FFS ($N = 64$) | 2.48 (1.83) | 2.51 (1.79) |
| CAP ($N = 66$) | - 2.40 (1.64) | 2.50 (1.49) |
| <i>P</i> value | <0.01 | 0.99 |

Mann–Whitney *U* test. Standard deviation in parentheses. Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

Table 10 Effect of risk and ambiguity on medical treatment in FFS (by health profile)

| Health profile | Treatment | | |
|----------------|---------------|-------------|--------------|
| | Deterministic | Risk | Ambiguity |
| A | 1.22 (1.29) | 1.30 (1.26) | 1.30 (1.24) |
| B | 3.75 (2.41) | 3.92 (2.40) | 3.86 (2.45) |
| C | 2.39 (1.93) | 2.41 (1.87) | 2.41 (1.95) |
| D | 2.59 (1.85) | 2.75 (1.98) | 2.83 (1.97)* |
| E | 2.30 (1.82) | 2.38 (1.86) | 2.50 (1.87)* |

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$. Wilcoxon signed-rank tests on the difference from Deterministic for each health profile and treatment (Risk and Ambiguity). Standard deviation in parentheses. Dependent variable is the absolute deviation from optimal treatment (the absolute difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit)

FFS condition with fee-for-service payment ($N = 64$)

Table 11 Effect of risk and ambiguity on medical treatment in CAP (by health profile)

| Health profile | Treatment | | |
|----------------|---------------|---------------|--------------|
| | Deterministic | Risk | Ambiguity |
| A | 4.27 (2.10) | 4.06 (2.04) | 4.12 (1.93)* |
| B | 1.05 (1.21) | 1.17 (1.22) | 1.21 (1.20) |
| C | 2.45 (1.63) | 2.45 (1.59) | 2.55 (1.69) |
| D | 2.56 (1.80) | 2.68 (1.67) | 2.53 (1.60) |
| E | 2.24 (1.64) | 2.55 (1.67)** | 2.49 (1.60)* |

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$. Wilcoxon signed-rank tests on the difference from Deterministic for each health profile and treatment (Risk and Ambiguity). Standard deviation in parentheses. Dependent variable is the absolute deviation from optimal treatment (the absolute difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit)

CAP condition with capitation payment ($N = 66$)

Table 12 Generic risk/ambiguity preference and medical treatment, difference between payment systems (linear regression)

| Dependent variable: absolute deviation from optimal treatment | (1) Risk | (2) Ambiguity |
|---|-------------------|-------------------|
| CAP | 0.384 (0.977) | - 0.566 (0.839) |
| Risky patient | 0.258 (0.212) | |
| Risk preference | - 0.106 (0.086) | |
| Risky patient × risk preference | - 0.016 (0.023) | |
| Risky patient × CAP | - 0.978 (0.371)** | |
| Risk preference × CAP | - 0.041 (0.102) | |
| Risky patient × Risk preference × CAP | 0.099 (0.045)** | |
| Amb. patient | | 0.067 (0.259) |
| Amb. preference | | - 0.125 (0.059)** |
| Amb. patient × Amb. preference | | 0.009 (0.032) |
| Amb. patient × CAP | | -0.412 (0.317) |
| Amb. preference × CAP | | 0.074 (0.090) |
| Amb. patient × Amb. preference × CAP | | 0.040 (0.038) |
| Control variables | Yes | Yes |
| Observations | 1270 | 1270 |
| Clusters | 127 | 127 |

* $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$. Errors are clustered on individuals (within parentheses). Deviation from optimal treatment is the difference between chosen level of medical treatment and the level of medical treatment that maximizes the patient's health benefit. CAP is a dummy for the CAP condition (FFS is the base category). Risky patient is a dummy for patients for whom the outcome of medical treatment is risky (and the base category is patients for whom the outcome of medical treatment is deterministic). Risk preference is the first row where the subject chose the certain amount of money (over the risky lottery) in the individual choice task (10 = risk neutrality). Amb. patient is a dummy for patients for whom the outcome of medical treatment is ambiguous (and the base category is patients for whom the outcome of medical treatment is deterministic). Amb. preference is the first row where the subject chose the certain amount of money (over the ambiguity lottery) in the individual choice task (10 = ambiguity neutrality). Control variables: age, gender, whether the subject smokes, tries to keep a healthy diet, number of visits to the doctor in the last 3 and 12 months, health status and political views on a scale from left to right. A total of three subjects (in FFS) switched multiple times in at least one of the two individual choice tasks and are, therefore, removed from the analysis

FFS condition with fee-for-service payment ($N = 64$), CAP condition with capitation payment ($N = 66$)

References

- Abdellaoui, M. (2000). Parameter-free elicitation of utilities and probability weighting functions. *Management Science*, 46, 1497–1512.
- Ahlert, M., Felder, S., & Vogt, B. (2012). Which patients do I treat? An experimental study with economists and physicians. *Health Economics Review*, 2, 1–25.
- Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2016). Deciding for others reduces loss aversion. *Management Science*, 62, 29–36.
- Arrieta, A., García-Prado, A., González, P., & Pinto-Prades, L. (2017). Risk attitudes in medical decisions for others: an experimental approach. *Health Economics*, 26(S3), 97–113.
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53, 941–973.

- Berger, L., Bleichrodt, H., & Eeckhoudt, L. (2013). Treatment decisions under ambiguity. *Journal of Health Economics*, 32, 559–569.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., & Wiesen, D. (2013). *How effective are pay-for-performance incentives for physicians? A laboratory experiment* (p. 413). Duisburg: Ruhr Economic Papers.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization*, 131, 17–23.
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: experimental evidence. *Health Economics*, 26, 243–262.
- Brosig-Koch, J., Kairies-Schwarz, N., & Kokot, J. (2015). *Sorting into physician payment schemes—a laboratory experiment* (p. 529). Ruhr Economic Papers: Duisburg.
- Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5, 325–370.
- Chakravarty, S., Harrison, G. W., Haruvy, E. E., & Rutström, E. (2011). Are you risk averse over other people's money? *Southern Economic Journal*, 77, 901–913.
- Chalkley, M., & Malcomson, J. (1998). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17, 1–19.
- Cookson, R., & Dolan, P. (2000). Principles of justice in health care rationing. *Journal of Medical Ethics*, 26, 323–329.
- Culyer, A. J., & Wagstaff, A. (1993). Equity and equality in health and health care. *Journal of Health Economics*, 12, 431–457.
- Cutler, D. (2002). Equality, efficiency, and market fundamentals: the dynamics of international medical-care reform. *Journal of Economic Literature*, 40, 881–906.
- Devlin, R. A., & Sarma, S. (2008). Do physician remuneration schemes matter? The case of Canadian family physicians. *Journal of Health Economics*, 27, 1168–1181.
- Doctor, J. N., Miyamoto, J., & Bleichrodt, H. (2009). When are person tradeoffs valid? *Journal of Health Economics*, 28, 1018–1027.
- Dolan, P., & Olsen, J. A. (2001). Equity in health: the importance of different health streams. *Journal of Health Economics*, 20, 823–834.
- Dolan, P., Shaw, R., Tsuchiya, A., & Williams, A. (2005). QALY maximization and people's preferences: a methodological review of the literature. *Health Economics*, 14, 197–208.
- Dranove, D., & Wehner, P. (1994). Physician-induced demand for childbirths. *Journal of Health Economics*, 13, 61–73.
- Ellis, R., & McGuire, T. (1986). Provider behavior under prospective reimbursement: cost sharing and supply. *Journal of Health Economics*, 5, 129–151.
- Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75, 643–669.
- Falk, A., & Heckman, J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326, 535–538.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, 171–178.
- Fuchs, V. (1978). The supply of surgeons and the demand for operations. *Journal of Human Resources*, 13, 35–56.
- Galizzi, M., Tammi, T., Godager, G., Linnosmaa, I., & Wiesen, D. (2015). Provider altruism in health economics. National Institute for Health and Welfare Discussion Paper 4/2015.
- Galizzi, M. M., & Wiesen, D. (2018). Behavioral experiments in health economics. In J. H. Hamilton, A. Dixit, S. Edwards, & K. Judd (Eds.), *Oxford research encyclopedia of economics and finance*. Oxford research encyclopedias. Oxford: Oxford University Press.
- Gaynor, M., & Gertler, P. (1995). Moral hazard and risk spreading in partnerships. *RAND Journal of Economics*, 26, 591–613.
- Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32, 1105–1116.
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114–125.
- Gruber, J., & Owings, M. (1996). Physician financial incentives and cesarean section delivery. *RAND Journal of Economics*, 27, 99–123.

- Ham, C. (1998). Retracing the Oregon trail: The experience of rationing and the Oregon health plan. *British Medical Journal*, *316*, 1965–1969.
- Heinrich, T., & Mayrhofer, T. (2018). Higher-order risk preferences in social settings. *Experimental Economics*, *21*, 434–456.
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behavior—an experimental investigation. *Journal of Health Economics*, *30*, 637–646.
- Hennig-Schmidt, H., & Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science and Medicine*, *108*, 156–165.
- Keser, C., Peterle, E., & Schnitzler, C. (2014). Money talks: Paying physicians for performance. Center for European Governance and Economic Development Research Discussion Paper 173.
- König-Kersting, C., & Trautmann, S. T. (2016). Ambiguity attitudes in decisions for others. *Economic Letters*, *146*, 126–129.
- Krasnik, A., Groenewegen, P. P., Pedersen, P. A., Scholten, P., Mooney, G., Gottschau, A., et al. (1990). Changing remuneration systems: Effects on activity in general practice. *British Medical Journal*, *300*, 1698–1701.
- Li, J., Hurlley, J., Decicca, P., & Buckley, G. (2014). Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, *23*, 962–978.
- Ma, C. A., & Riordan, M. H. (2002). Health insurance, moral hazard and managed care. *Journal of Health Economics and Management Strategy*, *11*, 81–107.
- McGuire, T. (2000). Physician agency. In A. Culyer & J. Newhouse (Eds.), *Handbook of Health Economics*. Amsterdam: North Holland.
- Newhouse, J. (1992). Medical care costs: How much welfare loss? *Journal of Economic Perspectives*, *6*, 3–21.
- NICE. (2008). Quality adjusted life years (QALYs) and the severity of illness. NICE Citizens Council Meeting Report 10. <http://www.nice.org.uk/Media/Default/Get-involved/Citizens-Council/Reports/CCReport10QALYSeverity.pdf> Accessed 20 Dec 2016.
- Nord, E., & Johansen, R. (2014). Concerns for severity in priority setting in health care: a review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. *Health Policy*, *116*, 281–288.
- Olsen, J. A. (1997). Theories of justice and their implications for priority setting in health care. *Journal of Health Economics*, *16*, 625–639.
- Shah, K. K. (2009). Severity of illness and priority setting in healthcare: A review of the literature. *Health Policy*, *93*, 77–84.
- Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, *103*, 510–531.
- Tinghög, G. (2011). The Art of Saying No: The Economics and Ethics of Health Care Rationing. PhD thesis, Linköping University.
- Veider, F., Lefebvre, M., Bouchoucha, R., Chmura, T., Hakimov, R., Krawczyk, M., et al. (2015). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, *13*, 421–452.
- Veider, F. M., Villegas-Palacio, C., Martinsson, P., & Mejia M. (2016). Risk taking for others and oneself: A structural model C. approach. *Economic Inquiry*, *54*, 879–894.
- Williams, B. (1962). The idea of equality. In P. Laslett & W. G. Runciman (Eds.), *Philosophy, politics and society*. Oxford: Basil Blackwell.
- Williams, A. (1997). Intergenerational equity: An exploration of the 'fair innings' argument. *Health Economics*, *6*, 117–132.
- Williams, A., & Cookson, R. (2000). Equity in health. In A. Culyer & J. Newhouse (Eds.), *Handbook of health economics*. Amsterdam: North Holland.
- Woodward, R., & Warren-Boulton, F. (1984). Considering the effect of financial incentives and professional ethics on 'appropriate' medical care. *Journal of Health Economics*, *3*, 223–237.
- Zweifel, P., Breyer, F., & Kifmann, M. (2009). *Health economics*. Berlin: Springer.