Does gender affect medical decisions? Results from a behavioral experiment with physicians and medical students

Geir Godager
Department of Health Management and Health Economics,
University of Oslo,
Norway

Heike Hennig-Schmidt
BonnEconLab,
University of Bonn, Germany

Jing Jing Li
Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

Jian Wang
Dong Fureng Institute of Economic and Social Development,
Wuhan University, China

Fan Yang
Department of Health Management and Health Economics,
University of Oslo, Norway

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK

Working paper 2021:1

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-271-1

Does gender affect medical decisions?

Results from a behavioral experiment with physicians and medical students

Geir Godager[1,5,*], Heike Hennig-Schmidt[2], Jing Jing Li[4,1], Jian Wang[3,1], Fan Yang[1]

Abstract

It is rarely the case that differences in behaviors of females and males are described under a *ceteris paribus* condition, and behaviors can potentially be influenced by the environment in which decisions are made. In the case of medical decisions, physicians are expected to account for patient characteristics as well as observed and unobserved contextual factors, such as whether the patient has a healthy lifestyle. Since one usually do not randomize physicians to context, reported gender differences in medical practice can have several alternative interpretations.

A key question is whether the medical treatment of a *given* patient is expected to depend on the gender of the physician. To address this question, we quantify gender effects using data from an incentivized laboratory experiment, where Chinese medical doctors and Chinese medical students choose medical treatment under different payment schemes. We estimate preference parameters of females and males assuming decision-makers have patient-regarding preferences. We cannot reject the hypothesis that gender differences in treatment choices are absent. Preference parameters of females and males are not statistically different in a log-likelihood ratio test, and there is no evidence that the degree of randomness in choices differs between genders.

The absence of gender effects in the laboratory, where choice context is fixed, provides nuance to previous findings on gender differences, and highlights the general difficulty of separating individuals' behavior from the context they are in.

Keywords: Gender, Laboratory experiment, Bounded rationality, Physician behavior

JEL-Classification: C92, D82, I11, H40, J33

[*]Corresponding author. Email: geir.godager@medisin.uio.no
[1]Institute of Health and Society, Department of Health Management and Health Economics, University of Oslo, Norway
[2]BonnEconLab, University of Bonn, Germany
[3]Dong Fureng Institute of Economic and Social Development, Wuhan University, China
[4]Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, 250021,China
[5]Health Services Research Unit, Akershus University Hospital, Norway

1. Introduction

Differences in behaviors of females and males are rarely considered surprising since examples of observed gender differences exist in fundamental economic choices regarding education, labor market participation and saving, in economic outcomes such as income and in attitudes like risk. While much research has aimed to shed light on the causal mechanisms behind observed gender differences, there is no unanimity in the scientific conclusions. Gender differences in education attainment and labor market participation has shown remarkably little stability over time. This suggests that understanding the differences in the contexts in which individuals make economic decisions is important when aiming to provide new knowledge on the causes of observed gender differences in behavior. As highlighted by Niederle (2016), a key question is whether detected differences between males and females are indeed "true" differences in preferences, or whether they are reflecting the design of the decision environment and the choice architecture that activate psychological attributes in which large gender differences have been found.

In this paper we ask whether the physician's gender affects the choices regarding patient treatment. This is an important research topic as medical care has been characterized by a substantial rise in female labor market participation worldwide due to a considerable increase of women's enrollment in medical schools and residency programs (Levinson and Lurie, 2004). In the United States, for instance, the proportion of female medical students rose from 10% to 50% from 1970 to 2019 (Boyle, 2019), and females now constitute the majority of medical students in many countries (OECD, 2020). If female doctors treat their patients differently than male physicians do, the change in the gender composition in the medical professions would change the supply of medical services and the way patients are treated. This, in particular, would be the case if patient-regarding preferences, measured by the relative weight a physician puts on patient's health benefit, differ between males and females. Knowledge on behavioral gender differences in the medical workforce is therefore of general and political interest. Our paper sheds light on this essential topic.

Research on the effect of increased female participation in the medical profession reports inconclusive results (see Section 2 on related literature). The inconclusive findings may be caused by confounding factors like different payment systems for doctors in the samples studied or differences in the characteristics of the patient populations as well as observable and unobservable differences in patient mix between providers caused by patients choosing actively the provider who suit their preferences. Endogenous matching challenges the identification and quantification of gender effects by use of field data, as it becomes difficult to separate the effect of gender from the effect of the differences in context experienced by female and male providers.

Using surveys, administrative data, questionnaire or vignettes studies do not enable the implementation of ceteris-paribus changes of the decision environment to study causality, and the confounding factors can rarely be controlled in the field. Our paper addresses these challenges by using data from a controlled economic laboratory experiment enabling us to identify and quantify gender differences in preferences and behavior. Our rather "pure" experimental design involves individual decision-making tasks where each male or female participant acts in the role of a physician. Patients and diseases are kept constant and are abstract, which eliminates differences in patient populations, potential gender differences in strategic behavior or discrimination and also complex interactions like endogenous matching of physicians and patients. Participants are Chinese medical students (N = 178: 101 females and 77 males). We also "bring the field to the lab" by recruiting Chinese medical doctors (N = 99: 69 females and 30 males) to analyze whether gender differences can occur with participants from outside of the laboratory and beyond experiments with students. Our experiment applies the same experimental parameters as Hennig-Schmidt et al. (2011), and our study is based on the data of Wang et al. (2020). We study physician decision-making under two different payment mechanisms and use a medically framed setting where subjects' choices determine both a physician's profit and a patient's health benefit. Decisions are incentivized by monetary rewards. Even though in our experiment patients are abstract, a real patient outside the lab is supported by the monetary equivalent of the aggregated patient benefits.

To the best of our knowledge, our experimental study is currently the only one that explicitly analyzes the gender effect in a physician decision-making task[1] involving doctors and medical students. We are also the first to distinguish explicitly between gender differences in observed treatment choices, gender differences in patient-regarding preferences and gender differences in the degree of randomness in medical decision-making.

We address the following three research questions in this paper:

1. Does gender affect observed treatment choice?
2. Do females and males differ in their willingness to sacrifice profit in order to improve patient benefit?
3. Does the degree of randomness in treatment choice depend on gender?

[1] Brosig-Koch et al. (2020) and Attema et al. (2020), for example, control for gender in regression analyses but do not make gender differences the main focus of their studies.

Given our brief surveys on medical field studies and on results of other-regarding behavior in related experimental games (see Section 2) no clear directional hypotheses regarding gender differences in our experiment on physician treatment behavior become apparent. Taking into account our "pure" experimental design where many factors are excluded that affect the context of medical decision making, we expect rather small differences, if at all. Our hypotheses in all three cases, therefore, conform with the null-hypothesis that a gender effect is absent.

The latter two research questions seem unnecessary when perfect rationality is assumed. They are not, however, under the assumption of bounded rationality. The reason is that the existence of a one-to-one relation between preferences and behavior observed in a given context cannot be established when decision-makers are boundedly rational. If the degree of randomness in behavior varies across decision-makers, they may seem heterogeneous in preferences even when they are not (Louviere and Eagle, 2006). For the same reason, differences in the degree of randomness across choice contexts can make individuals' preferences appear context-dependent even when preferences are stable. Swait and Louviere (1993); Louviere and Eagle (2006); Fiebig et al. (2010) argue that the degree of randomness in behavior is unlikely to be constant, as the impact of noise on choices can vary over conditions, contextual circumstances or situations, as well as between decision-makers. If, for example, subjects are learning by doing during the course of a laboratory experiment and the researcher applies an empirical strategy that (silently) assumes the degree of randomness to be constant, the researcher might erroneously conclude that preferences changed during the experiment. Fortunately, data from controlled laboratory experiments provide the most favorable conditions for identifying both preference parameters and the degree of randomness in decision-makers' behavior.

For all three research questions, the empirical analysis of our experimental data does not provide evidence for rejecting the null-hypothesis that gender differences are absent. Thus, in our samples of Chinese medical doctors and medical students, males and females did not show statistically different behavior in service provision, in the willingness to sacrifice profit in order to raise patient benefit, or in the degree of randomness in decision-making.

The paper proceeds as follows. In Section 2, we briefly survey medical field studies and behavioral papers on other-regarding behavior in economic experiments related to our design. Section 3 gives a description of the experimental design and explains the procedure. Results are presented in Section 4 by describing observed behavior, by providing non-parametric test results, and by presenting estimation results from an empirical model of bounded rationality. We discuss our results and conclude in Section 5. The Appendix provides additional analyses as well as further information on experimental parameters, and the experimental material participants were provided with.

2. Related literature

The increase of females in the medical profession has been observed, for instance, in the US, the United Kingdom, Russia, Norway, Canada, Sweden, The Netherlands, and Australia (Kilminster et al., 2007). Female doctors have become common in China as well. The share of licensed female doctors increased from 43% in 2010, to 47% in 2019, and in the hospital sector the percentage of female doctors is even higher with 70% in 2010 and 74% in 2019 (Ministry of Health in People's Republic of China, 2011; China Health Commission, 2020). This trend is likely to continue.

These changes in the gender composition might bring about changes in medical service provision and patient treatment as has been suggested by, for instance, Riska (2001), Boulis and Jacobs (2008), or McKinstry (2008). One example in this regard is effective physician-patient communication; the evidence, however, is inconclusive. Some studies found female doctors to spend more time than male doctors (Jefferson et al., 2013; Roter and Hall, 2004). Others report the opposite (Hampson et al., 1996; Roter et al., 1999) while Bertakis et al. (1995) and Jefferson et al. (2015) found nearly no difference.[2]

Studying gender differences in the medical practice style or in patient treatment of Norwegian general practitioners (GPs) by means of detailed and complete national register data has also lead to inconclusive evidence. Iversen and Lurås (2002) report that female GPs offer shorter waiting time to their patients compared to male GPs. Yet, more patients decide to switch out of the patient lists of female general practitioners (Iversen and Lurås, 2011). Godager et al. (2015) found no significant differences in referrals to hospitals and specialists between female and male GPs, and there was no signifIcant difference between female and male GPs' propensity for working voluntarily for the community health service (Godager and Lurås, 2009). Godager (2012) found support for the hypothesis of endogenous doctor-patient matching in that patients were more likely to prefer a GP of the same gender. This example shows that the potential for endogenous matching challenges identifying and quantifying differences of treatment choices between males and females in general medical practice .

As no previous experimental studies seem to exist that explicitly analyze the gender effect in a physician decision-making scenario, we also survey controlled experimental studies where ceteris-paribus conditions can be created to study the causal effect of gender on behavior. A broad experimental economics literature exists on gender differences in altruism

[2]See also Dacre (2008) and Hedden et al. (2014), for further discussions of the topic.

(other-regarding preferences), the topic we are analyzing in the present paper.[3] Behavioral experiments on the dictator game (Forsythe et al., 1994) and on donations to a charity (Eckel and Grossman, 1996; Grossman and Eckel, 2015) are tools frequently applied when studying altruism. These games involve analyzing distributional preferences in a scenario where one person, the dictator, decides how to distribute money between himself and another person (see also Engel, 2011, for a meta study on dictator games) or a charity. This setup has some similarity with our design, were a physician's treatment choice simultaneously determines his own profit and the patients' health benefit, since benefit of patients can be increased if the physician sacrifices own profit.

The surveys by Eckel and Grossman (2008), Croson and Gneezy (2009) and Niederle (2016) report mixed findings from dictator game experiments in that there is evidence for no gender differences but also for more altruistic and cooperative behavior of women than of men. The latter findings are supported by a recent dictator game study of Brañas-Garza et al. (2018) conducted on a large sample of US residents recruited through Amazon Mechanical Turk. Moreover, both women and men expect women to be more altruistic than men. Falk and Hermle (2018) also report higher altruism in females than in males. Their data include altruism measures for 80,000 individuals in 76 representative country samples. Altruism is quantified by first-mover behavior in a hypothetical dictator game with a charitable organization as hypothetical recipient. Other studies show that behavior of males and females do not differ in the standard dictator game, yet additional experimental features induce behavioral differences. Andreoni and Vesterlund (2001), for instance, find that male and females dictators' transfers do not differ on average. With varying costs of transfers, however, women are more generous than men when giving is relatively expensive, while the opposite holds as the price of giving decreases. The lower sensitivity of females to the price of giving is corroborated by Visser and Roelofs (2011). Boschini et al. (2018) study gender differences in a large random sample of the Swedish population. They find that women are only more altruistic than men when participants are reminded of their gender. These priming effects occur only in mixed-gender pairings. A dictator game study by Klinowski (2018) suggests that non-payoff-related motivations like reluctancy in giving may play a role in producing gender differences in transfers.

3. Experiment

3.1. Basics of the experimental design

Participants in our experiment act in the role of physicians, who are assumed to be concerned about their own profit π as well as about the patient benefit, B, the latter depending on the quantity of medical services $q \in 0, 1, ..., 10$. The participants' task is to choose a quantity of medical services for a given patient whose health benefit is determined by that choice[4]. Each physician decides for three different patient types with five different abstract illnesses, i.e. for 15 patients in total. The combination of patient type and illness characterizes a specific patient. Patient types differ in the health benefit they gain from the medical services. Like many theoretical papers (e.g., Ellis and McGuire, 1986; Ma, 1994; Choné and Ma, 2011) we use a concave patient benefit function that has a global optimum yielding the highest benefit to a patient. We refer to quantities smaller than the optimum as underprovision of medical care, and to quantities larger than the highest patient benefit as overprovision. The three types of patients reflect the patients' different states of health. Patients 1 to 5 of type 1 have an intermediate state of health. Patients 6 to 10 are of type 2 with a good state of health, and patients 11 to 15 are of type 3 suffering from a bad state of health.

A physician's choice of medical services simultaneously determines the patient benefit and the physician's own profit. The patient is assumed to be passive and fully insured, accepting each level of medical service provided by the physician. We apply a within-subject design in that the physician is sequentially confronted with the same 15 patients (choice occasions) in the two payment systems CAP and FFS with either CAP first and FFS second or vice versa. Under FFS, physicians' remuneration increases in the number of medical services provided to a patient. Physicians are paid a lump sum per patient under CAP. The patient health benefit is measured in monetary terms. In our experiment, no real patients are present. However, physicians' quantity choices have consequences for a real patient outside the lab. The money corresponding to the patient benefits aggregated over all decisions was transferred to one real patient's in-hospital account to reduce his out-of-pocket payment for his cancer treatment Thus, subjects have an incentive to care for the patient when making their decisions. We did not inform the participants about the identity of the person to whom the money was transferred.

Before making his or her decision, the physician gets information on her remuneration, costs and profit as well as on the patient's benefit for each quantity that can be chosen. All monetary amounts are in Token, our experimental currency, the exchange rate being 10 Token = 1 RMB for students and 10 Token = 6 RMB for doctors (1 RMB was approximately 0.12 at the time of the experiment).

[3]Other research fields comprise competition (c.f. Niederle and Vesterlund, 2010, Wieland and Sarin, 2012, Cadsby et al., 2013, and Dreber et al., 2014), or risk (c.f. Gong and Yang, 2012, Wieland and Sarin, 2012, and Dreber et al., 2014) where large gender differences have been found (Niederle, 2016).

[4]For the experimental parameters see Table A.1 in Appendix A1. A more detailed description of the experimental design is found in Appendix B.

3.2. Experimental protocol

Our experiment was conducted in September 2012 (medical students) and 2013 (doctors) at the Center for Health Economic Experiments and Public Policy at Shandong University in Jinan, China, and was programmed with z-Tree (Fischbacher, 2007). All experimental material was provided in Chinese; see Appendix C for the English version. Each of the Chinese male and female medical students and doctors participating in our experiment was sequentially confronted with the same 15 patients (choice occasions) in FFS and CAP. The subjects were randomly assigned to experimental sessions where either CAP was implemented in Part 1 followed by FFS in Part 2 (condition CAP-FFS) or in reversed order (condition FFS-CAP). Each participant joined the experiment only once, either in CAP-FFS or in FFS-CAP. Participants were informed at the beginning that the experiment consisted of two parts, but did not know what the second part would be about. The male and female medical students, who voluntarily participated in the experiment, were recruited via notices posted at the campus and by email invitations. Doctors who were working at community health service centers in five districts of Jinan were recruited through a phone call by the respective District Department of Health informing the doctors that a research experiment from Shandong University needed volunteers.

The experimental procedure was exactly the same for all medical students and doctors. Participants were randomly allocated to their workstations separated by wooden panels and curtains to guarantee anonymity of their decisions. Then, instructions for Part 1 of the experiment were distributed and read out by a Chinese experimenter. Subjects got plenty of time to read the instructions and to ask clarifying questions in private that were answered individually. Then they had to answer a set of test questions. Participants decided under either a CAP or a FFS system and went through a sequence of 15 patients (choice occasions) on the quantity of medical services to be provided. The order of patients was predetermined and kept constant across conditions. After each decision, participants were informed about his/her profit and the patient benefit generated by the previous choice. At the end of each part of the experiment, participants received information about their total profit achieved and the total health benefit generated during all 15 quantity decisions. At the end of Part 1, the participants answered some open-ended questions. In Part 2 of the experiment, participants made the same decisons under the payment system they had not yet been confronted with. All participants answered questions on socio-demographic variables, and the doctors also stated their medical speciality and professional experience. Finally, participants were informed about their individual total profit and the total benefit resulting from their decisions in Parts 1 and 2 of the experiment as well as on their final monetary payoff. After having been paid in private they left the laboratory individually.

To ensure that the doctors and medical students trusted the experimenters to actually transfer the money derived from the patient benefit, we implemented a procedure already used in several other experiments[5]. A monitor was randomly selected from the participants in a session. He/she verified the amount of money corresponding to the patient benefits aggregated over all participants' decisions in the respective session. Then, the monitor and an assistant to the experimenters went by taxi to the Shandong University Cancer Hospital in Jinan, and paid the corresponding amount in cash at the hospital-cashier's desk into the patient's account. We took great care to ensure that the monitor did not see the name of the real patient in order to maintain the patient's anonymity. The monitor signed a statement on the appropriate transfer of the monetary amount. After all sessions had been conducted, all participants in each session received an email stating the respective transfer. Each monitor in the medical student subject pool was paid an additional 50 RMB and each doctor 200 RMB.

We conducted four sessions with medical doctors, and six sessions with medical students. Each experimental session comprised one condition (CAP-FFS or FFS-CAP), and lasted for about 90 minutes. A female student on average earned 27 RMB (3.20), while a male student earned 28 RMB (3.40) plus a show-up fee of 15 RMB (1.80). Female doctors on average earned 159 RMB (19.10) and male doctors earned 163 RMB (19.60) plus a show-up fee of 120 RMB (14.40).[6] Based on all 8,310 decisions, a total of 19,814 RMB (2,377.68) was transferred to the real patient's in-hospital-account to be used for reducing his out-of-pocket payment for cancer treatment; 4,751 RMB (570.12) for the sessions with medical students and 15,063 RMB (1,807,56) for the sessions with doctors. Ethical review and approval of the experimental design and procedure was given by Norwegian Social Science Data Services (reference #44267).

4. Empirical results

4.1. Descriptive results

Table 1 provides the gender composition in the two conditions CAP-FFS and FFS-CAP and the respective numbers of doctors and students.[7] We observe that 137 subjects participated in sessions where CAP was followed by FFS, whereas 140 subjects

[5] See, e.g., Hennig-Schmidt et al. (2011), Godager and Wiesen (2013), Hennig-Schmidt and Wiesen (2014), Godager et al. (2016), Brosig-Koch et al. (2016; 2017; 2019), Ge et al. (2021), and Wang et al. (2020)

[6] We adjusted stake sizes according to opportunity costs (Herrmann et al., 2008; Gächter and Schulz, 2016) outside the laboratory, i.e., the hourly wage of a student helper and the average hourly wage of a doctor in the respective employment situation. The average payoff for students approximately corresponded to the hourly wage of a student helper at Shandong University of about 30 RMB. For doctors the average hourly wage was about 120 RMB.

[7] The following analyses are based on the data set of Wang et al. (2020).

participated in sessions where FFS was followed by CAP. The participation in the two conditions is approximately balanced among both medical doctors (49 CAP-FFS, 50 FFS-CAP) and medical students (88 CAP-FFS, 90 FFS-CAP).

Table 1: Gender composition among doctors (N=99) and medical students (N=178) over the two conditions of the experiment

Condition	Gender	No of subjects (Doc.)[Stud.]
CAP-FFS	Female	76 (31)[45]
	Male	61 (18)[43]
	Total	137 (49)[88]
FFS-CAP	Female	94 (38)[56]
	Male	46 (12)[34]
	Total	140 (50)[90]

In addressing our first research question regarding whether females and males differ in their observed choices of service quantity, we compare the quantity choices aggregated for each of the two payment schemes. We further differentiate between male and female doctors and medical students. Table **??** provides the results.

4.2. Quantifying patient-regarding preferences of females and males

Our second research question relates to gender differences in the participants' willingness to sacrifice own profit to improve patients' benefit. Our third research question asks whether males and females differ in the degree of randomness in behavior. To answer these questions, we fit a bounded rationality model to our experimental data. Our choice model builds on early work of Luce (1959), Tversky (1972) and McFadden (1974), as well as the more recent literature on explicitly scaled choice models (Swait and Louviere, 1993; Hole et al., 2006; Fiebig et al., 2010; Bech et al., 2011; Hess and Rose, 2012; Swait and Marley, 2013; Hess and Train, 2017; Wallin et al., 2018; Wang et al., 2020).

We use the index t for the 30 choice occasions (patients) in the experiment, 15 in each of the payment schemes CAP and FFS. The index n denotes the decision maker *type*, here interpreted as female or male, medical student or medical doctor. We use j to index the eleven different treatment alternatives (quantities of service provision, $q \in 0, 1, ..., 10$ that are available for each choice occasion. Our model specification given in Equation (1) is as a scaled logit model with alternative specific constants (ASC), denoted by a_j. The error component in our model is given by: $a_j + \varepsilon_{njt}$. By including ASCs the restrictive independence of irrelevant alternatives assumption (IIA) is relaxed.[8]

$$F_{njt} = R_{nt}[\alpha_n ln(B_{jt}) + (1 - \alpha_n)ln(\pi_{jt})] + (1 - R_{nt})[a_j + \varepsilon_{njt}] \quad \alpha_n \in (0, 1)\ \forall n \ . \tag{1}$$

In textbook applications described by Train (2009), the ε_{njt} terms are commonly assumed to be independent, type 1 extreme value distributed. This a sufficient, but not necessary, condition for ensuring that maximizing decision-makers and the model specification in (1) result in choice probabilities given by the logit formula. Further details on model deduction is provided in Appendix A2.

The specification in (1) reflects the assumption that rationality is present to some *degree*. The elements in the first bracket are the rational part of the individuals' objective, which is the individual's deterministic utility as a function of health benefits B_{jt} and profit π_{jt}. R_{nt} denotes the weight assigned to this part of the objective.[9] The term $1 - R_{nt}$ is a measure of the degree in which the individual's behavior is affected by aspects that are irrelevant to utility, and these irrelevant aspects are captured by the error components. By definition, $R_{nt} \in (0, 1)$ applies for all decision-makers and choice occasions, and we allow R_{nt} to vary between individuals and between choice occasions in the experiment. Hence our empirical specification enables us to test whether the *degree of randomness*, as measured by R_{nt}, differs between females and males.

With the assumption $\alpha_n \in (0, 1)$ we assume a utility function that is homogeneous of degree one (constant returns to scale). The parameter α_n denotes the relative valuation of the health benefit in n's preference function. Specifying physician preferences to comprise a linear combination of profit and patients' health benefit has been common in the health economics literature for more than three decades, see e.g Ellis and McGuire (1986, 1990); Scott (2000); Léger (2008), and this assumption has been shown to fit data from both the field (Godager et al., 2009, 2015) and experiments (Godager and Wiesen, 2013; Wang et al., 2020).

[8]As in Fiebig et al. (2010), ASCs are assumed to be part of the error structure.

[9]It does not appear realistic to account for all aspects of human rationality with one parameter R_{nt}. There are many examples of rational randomness. We therefore use the term *degree of determinism*, rather than *the degree of rationality* in this paper. Preference weights are relative, since the *absolute* weight on utility is in general not identified. Further, the *relative* weights of utility and noise in (1) are identified if, and only if, appropriate functional form restrictions are introduced for the utility function $V(\,)$ (Train, 2009; Fiebig et al., 2010).

The constant returns to scale assumption is not only a convenient assumption in line with mainstream health economics, it also introduces constraints which enable identification of R_{nt}. This identification strategy is applied also by Swait and Marley (2013); Wallin et al. (2018) and Wang et al. (2020), and differs from constraints introduced in so-called *willingness-to-pay space models* by, e.g. Train and Weeks (2005); Scarpa et al. (2008) and Hole and Kolstad (2012). From our log-linear specification it follows that the relative willingness-to-pay (RWTP) is given by:

$$-\frac{d\pi_{jt}}{dB_{jt}}\frac{B_{jt}}{\pi_{jt}} = \frac{\alpha_n}{1-\alpha_n} \quad . \tag{2}$$

The RWTP in (2) is the percentage sacrifice in profit that will render decision-maker's utility unchanged if patient benefit is increased by one percent.[10]

We use the program gmnl in STATA 16, written by Gu et al. (2013), to estimate our model parameters by means of maximum likelihood.

4.3. Results from maximum likelihood estimation

In Table 2, we report results from maximum likelihood estimation of our behavioral model (1). The point estimates of α are 0.46 for females and 0.49 for males, and we see that their confidence intervals overlap. We cannot reject the hypothesis that $\alpha_F = \alpha_M$ (p-value=0.6375, Wald test). The estimated preference parameters have clear economic interpretation. We recall that the interpretation of the RWTP is the percentage reduction in profit which would leave utility unchanged if patient benefit were simultaneously increased by one percent. Using the formula for RWTP given in (2) we find that the point estimates of the two RWTP are 0.86 for females and 0.97 for males.

We also estimated a fully flexible model where each of the four groups, female doctors, male doctors, female students and male students had group-specific α and R parameters. We compared this fully flexible model to a restricted model where preference parameters were constrained to be identical for the four groups. We could not reject the null hypothesis that the most flexible model does not provide better fit to the data than the restricted model (p-value= 0.3109, likelihood-ratio test). We conclude that the more parsimonious model in Table 3 is sufficient for addressing our research questions. The $\theta_{\_FEMALE}$-

Table 2: Results from maximum likelihood estimation of the behavioral model in Equation (1)
Sample: 178 Chinese students and 99 Chinese doctors, 30 decisions for each Chinese subject. Subjects are more experienced when in the second half of the experimental session (EXPERIENCE=1) .

	Estimate	(95% C.I.)	P-value
Preference parameters			
α_F	0.46	(0.33— 0.60)	< 0.001
α_M	0.49	(0.34— 0.64)	< 0.001
Scale heterogeneity†			
$\theta_{\_DOCTOR}$	−0.58	(-1.00— -0.15)	0.008
$\theta_{\_EXPERIENCE}$	0.49	(0.18— 0.80)	0.002
$\theta_{\_FEMALE}$	−0.26	(-0.58— 0.07)	0.125

Note: p-values and C.I. are based on standard errors clustered at the level of each individual. Alternative- and occasion-specific constants not shown. † We report the parameter denoted by θ in Fiebig et al. (2010), as provided by the program of Gu et al. (2013). θ-parameters are marginal effects on the log of the scale parameter: $\frac{\delta ln(\sigma)}{\delta z}$.

parameter is our estimated measure of gender differences in the *degree of randomness*. We observe in Table 3 that $\theta_{\_FEMALE}$ is not statistically significant. We cannot reject the hypothesis that females and males are equally influenced by irrelevant aspects when choosing medical treatments. Additional descriptions and interpretations of differences in the degree of randomness in behavior is provided in the appendix. With reference to our remaining two research questions, we have found:

RESULT 2: We do not find a gender difference in patient-regarding preferences.

RESULT 3: We do not find a gender difference in the degree of randomness in treatment choices.

5. Discussion and conclusion

In this paper we investigate whether females and males differ in their choices of medical treatment, their patient-regarding preferences, and their degree of determinism in behavior. The research questions are motivated by the fact that the share of females employed in the health care sector has risen sharply over recent decades, and, if gender differences exist, they might

[10]The RWTP should not be confused with the *elasticity of substitution*, which in case of the Cobb-Douglas function with constant return to scale, is a given constant equal to one.

bring about changes in the provision of medical care. We apply data from a fully incentivized laboratory experiment based on the experimental design of of Hennig-Schmidt et al. (2011). Our use of data from a controlled laboratory experiment enables identification of gender differences holding decision context fixed. We analyze the data by means of non-parametric and parametric methods. Based on non-parametric tests, we cannot find evidence that gender affects treatment choices. We estimate a scaled choice model to test whether patient-regarding preferences or the degree of randomness in treatment choices differ between females and males. Our measure of patient-regarding preferences is the decision-makers willingness to sacrifice profit (in%) for raising patient benefit with one percent. We cannot find evidence that patient-regarding preferences differ over genders, and we also cannot find evidence that one gender behaves more random then the other.

Our results are obtained in a stylized physician decision-making context that is stripped of many confounding factors like differences in patient populations, strategic behavior, discrimination or complex interactions with patients, insurers or other third-party payers. Our design is reduced to the basic question in a doctor/patient interaction: How much weight does a physician put on the patient's health benefit? In this admittedly rather purified scenario we find that male and female participants behave rather similarly. This is in line with quite some findings reported in the experimental literature on gender effects regarding altruism, see Section 2.

One can argue that the strength of providing a controlled but artificial context facilitating causal inference when using experimental data is, at the same time, the weakness of the experimental method. The decision context in the laboratory will obviously differ from that of any real doctor-patient encounter. We are, however, convinced that analyzing behavior also in a laboratory context is important. Such behavioral studies provide an additional piece of evidence, a broader picture on and a better understanding of gender differences in general and in medical decision-making in particular that are nearly not possible in the field and, therefore, are complementary to field studies. To make participants feel familiar with their professional decision situation and to make the ethical norm of altruistic – in our scenario: patient-regarding – behavior more salient (Eckel and Grossman, 1996; Grossman and Eckel, 2015), we introduced a medical context. Moreover, our participant sample consists of decision makers that are or will be real actors in the field. The doctors had about 16 years of professional experience on average. And also the prospective physicians were not newcomers as their average duration of medical study was about five semesters. Finally, the patient involved is a real person in strong need of financial support for his expensive medical treatment to survive his cancer. Thus, many features of our experiment are of real relevance which also many participants stressed in the open questions about the factors that have influenced their decisions.

We are aware that not finding gender differences in analyzing decisions made in a given medical laboratory context does not preclude existence of gender differences in some real world context. On the other hand, our study does provide an example of a context in which we were unable to find a gender difference in choices of medical treatment. This points to the need of additional future research on when and in which professional and institutional contexts gender differences in choices are exacerbated or reduced.

Acknowledgements

Funding from The Research Council of Norway (Project-No. 231776 and No. 296114) is gratefully acknowledged.

References

Andreoni, J. and Vesterlund, L. (2001), 'Which is the fair sex? gender differences in altruism', *The Quarterly Journal of Economics* **116**(1), 293–312.

Attema, A., Galizzi, M., Groß, M., Hennig-Schmidt, H., Karay, Y., L'Haridon, O. and Wiesen, D. (2020), Physician altruism. mimeo.

Bech, M., Kjær, T. and Lauridsen, J. (2011), 'Does the number of choice sets matter? results from a web survey applying a discrete choice experiment', *Health Economics* **20**(3), 273–286.

Bertakis, K. D., Helms, L. J., Callahan, E. J., Azari, R. and Robbins, J. A. (1995), 'The influence of gender on physician practice style.', *Medical Care* .

Boschini, A., Dreber, A., von Essen, E., Muren, A. and Ranehill, E. (2018), 'Gender and altruism in a random sample', *Journal of Behavioral and Experimental Economics* **77**, 72 – 77.
URL: *http://www.sciencedirect.com/science/article/pii/S2214804318300788*

Boulis, A. K. and Jacobs, J. A. (2008), *The changing face of medicine: women doctors and the evolution of health care in America*, The culture and politics of health care work, ILR Press.

Boyle, P. (2019), 'More women than men are enrolled in medical school'. (Accessed December 14 2020.).
URL: *https://www.aamc.org/news-insights/more-women-men-are-enrolled-medical-school*

Brañas-Garza, P., Capraro, V. and Rascón-Ramírez, E. (2018), 'Gender differences in altruism on mechanical turk: Expectations and actual behaviour', *Economics Letters* **170**, 19 – 23.
URL: *http://www.sciencedirect.com/science/article/pii/S0165176518301952*

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

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2017), 'The effects of introducing mixed payment systems for physicians: Experimental evidence', *Health Economics* **26**(2), 243–262.

Brosig-Koch, J., Hennig-Schmidt, H., Kokot, J., Kairies-Schwarz, N. and Wiesen, D. (2019), 'Physician performance pay: Experimental evidence.', *Discussion paper, SSRN-id3467583* .

Brosig-Koch, J., Hennig-Schmidt, H., Kokot, J., Kairies-Schwarz, N. and Wiesen, D. (2020), Physician performance pay: Experimental evidence., Technical report, University of Oslo, Health Economics Research Network.

Cadsby, C. B., Servátka, M. and Song, F. (2013), 'How competitive are female professionals? a tale of identity conflict', *Journal of Economic Behavior & Organization* **92**, 284–303.

China Health Commission, ed. (2020), *China Health Statistic Yearbook*, Peking Union Medical College Press, Beijing.

Chone, P. and Ma, C. A. (2011), 'Optimal Health Care Contract under Physician Agency', *Annales d'Economie et de Statistique* **101/202**, 229–256.

Croson, R. and Gneezy, U. (2009), 'Gender differences in preferences', *Journal of Economic literature* **47**(2), 448–74.

Dacre, J. (2008), 'Are there too many female medical graduates? no', *BMJ* **336**(7647), 749–749.

Dagsvik, J. K. (1995), 'How large is the class of generalized extreme value random utility models?', *Journal of Mathematical Psychology* **39**(1), 90–98.

Dagsvik, J. K. (2008), 'Axiomatization of stochastic models for choice under uncertainty', *Mathematical Social Sciences* **55**(3), 341–370.

Dagsvik, J. K. (2016), 'What independent random utility representations are equivalent to the iia assumption?', *Theory and Decision* **80**(3), 495–499.

Dagsvik, J. K. (2018), 'Invariance axioms and functional form restrictions in structural models', *Mathematical Social Sciences* **91**, 85–95.

Dreber, A., von Essen, E. and Ranehill, E. (2014), 'Gender and competition in adolescence: task matters', *Experimental Economics* **17**(1), 154–172.

Eckel, C. C. and Grossman, P. J. (2008), Differences in the economic decisions of men and women: Experimental evidence, Vol. 1 of *Handbook of Experimental Economics Results, Chapter 57*, Elsevier, pp. 509 – 519.
URL: *http://www.sciencedirect.com/science/article/pii/S1574072207000571*

Eckel, C. and Grossman, P. (1996), 'Altruism in Anonymous Dictator Games', *Games and Economic Behavior* **16**, 181–191.

Ellis, R. P. and McGuire, T. G. (1986), 'Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply', *Journal of Health Economics* **5**, 129–151.

Ellis, R. P. and McGuire, T. G. (1990), 'Optimal payment systems for health services', *Journal of Health Economics* **9**, 375–396.

Engel, C. (2011), 'Dictator games: A meta study', *Experimental Economics* **14**(4), 583–610.

Erlander, S. (1998), 'Efficiency and the logit model', *Annals of Operations Research* **82**, 203–218.

Falk, A. and Hermle, J. (2018), 'Relationship of gender differences in preferences to economic development and gender equality', *Science* **362**(6412).

Fiebig, D. G., Keane, M. P., Louviere, J. and Wasi, N. (2010), 'The generalized multinomial logit model: accounting for scale and coefficient heterogeneity', *Marketing Science* **29**(3), 393–421.

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

Forsythe, R., Horowitz, J. L., Savin, N. E. and Sefton, M. (1994), 'Fairness in simple bargaining experiments', *Games and Economic Behavior* **6**(3), 347–369.

Gächter, S. and Schulz, J. F. (2016), 'Intrinsic honesty and the prevalence of rule violations across societies', *Nature* **531**(7595), 496.

Ge, G., Godager, G. and Wang, J. (2021), 'Exploring physician agency under demand-side cost sharing — an experimental approach', *HERO Working Paper Series* **2021**.

Godager, G. (2012), 'Birds of a feather flock together: A study of doctor—patient matching', *Journal of Health Economics* **31**(1), 296 – 305.
URL: *http://www.sciencedirect.com/science/article/pii/S0167629611001640*

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

Godager, G., Iversen, T. and Ma, C. A. (2009), 'Service motives and profit incentives among physicians', *International journal of health care finance and economics* **9**(1), 39–57.

Godager, G., Iversen, T. and Ma, C. A. (2015), 'Competition, gatekeeping, and health care access', *Journal of Health Economics* **39**, 159–170.

Godager, G. and Lurås, H. (2009), 'Dual job holding general practitioners: the effect of patient shortage', *Health Economics* **18**(10), 1133–1145.

Godager, G. and Wiesen, D. (2013), 'Profit or patients' health benefit? exploring the heterogeneity in physician altruism', *Journal of Health Economics* **32**, 1105–116.

Gong, B. and Yang, C.-L. (2012), 'Gender differences in risk attitudes: Field experiments on the matrilineal mosuo and the patriarchal yi', *Journal of Economic Behavior & Organization* **83**(1), 59–65.

Grossman, P. J. and Eckel, C. C. (2015), 'Giving versus taking for a cause', *Economics Letters* **132**, 28–30.

Gu, Y., Hole, A. R. and Knox, S. (2013), 'Fitting the generalized multinomial logit model in stata', *Stata Journal* **13**(2), 382–397.

Hampson, S. E., McKay, H. G. and Glasgow, R. E. (1996), 'Patient-physician interactions in diabetes management: consistencies and variation in the structure and content of two consultations', *Patient Education and Counseling* **29**(1), 49–58.

Hedden, L., Barer, M. L., Cardiff, K., McGrail, K. M., Law, M. R. and Bourgeault, I. L. (2014), 'The implications of the feminization of the primary care physician workforce on service supply: a systematic review', *Human Resources for Health* **12**(1), 32.

Hennig-Schmidt, H., Selten, R. and Wiesen, D. (2011), 'How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation', *Journal of Health Economics* **30**, 637–646.

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

Herrmann, B., Thöni, C. and Gächter, S. (2008), 'Antisocial punishment across societies', *Science* **319**(5868), 1362–1367.

Hess, S. and Rose, J. M. (2012), 'Can scale and coefficient heterogeneity be separated in random coefficients models?', *Transportation* **39**(6), 1225–1239.

Hess, S. and Train, K. (2017), 'Correlation and scale in mixed logit models', *Journal of Choice Modelling* **23**, 1–8.

Hole, A. R. and Kolstad, J. R. (2012), 'Mixed logit estimation of willingness to pay distributions: A comparison of models in preference and wtp space using data from a health-related choice experiment', *Empirical Economics* **42**(2), 445–469.

Hole, A. R. et al. (2006), 'Small-sample properties of tests for heteroscedasticity in the conditional logit model', *Economics Bulletin* **3**(18), 1–14.

Iversen, T. and Lurås, H. (2002), 'Waiting time as a competitive device: an example from general medical practice', *International Journal of Health Care Finance and Economics* **2**(3), 189–204.

Iversen, T. and Lurås, H. (2011), 'Patient switching in general practice', *Journal of health economics* **30**(5), 894–903.

Jefferson, L., Bloor, K., Birks, Y., Hewitt, C. and Bland, M. (2013), 'Effect of physicians' gender on communication and consultation length: a systematic review and meta-analysis', *Journal of Health Services Research & Policy* **18**(4), 242–248.

Jefferson, L., Bloor, K. and Hewitt, C. (2015), 'The effect of physician gender on length of patient consultations: observational findings from the uk hospital setting and synthesis with existing studies', *Journal of the Royal Society of Medicine* **108**(4), 136–141.

Kilminster, S., Downes, J., Gough, B., Murdoch-Eaton, D. and Roberts, T. (2007), 'Women in medicine- is there a problem? a literature review of the changing gender composition, structures and occupational cultures in medicine', *Medical Education* **41**(1), 39–49.

Klinowski, D. (2018), 'Gender differences in giving in the dictator game: the role of reluctant altruism', *Journal of the Economic Science Association* **4**(2), 110–122.

Léger, P. T. (2008), Physician Payment Mechanisms, *in* M. Lu and E. Jonsson, eds, 'Financing Health Care: New Ideas for a Changing Society', Wiley-VCH Press, Weinheim (Germany), pp. 149–176.

Levinson, W. and Lurie, N. (2004), 'When most doctors are women: what lies ahead?', *Annals of Internal Medicine* **141**(6), 471–474.

Louviere, J. J. and Eagle, T. (2006), Confound it! that pesky little scale constant messes up our convenient assumptions, *in* 'Sawtooth Software Conference', Sawtooth Software Inc.

Luce, R. D. (1959), *Individual Choice Behavior a Theoretical Analysis*, Oxford, England: John Wiley.

Ma, C. A. (1994), 'Health Care Payment Systems: Cost and Quality Incentives', *Journal of Economics and Management Strategy* **3**, 93–112.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. E. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–142.

McKinstry, B. (2008), 'Are there too many female medical graduates? yes', *bmj* **336**(7647), 748–748.

Ministry of Health in People's Republic of China, ed. (2011), *China Health Statistic Yearbook*, Peking Union Medical College Press, Beijing.

Niederle, M. (2016), Gender, *in* J. H. Kagel and A. E. Roth, eds, 'Handbook of Experimental Economics', Vol. 2, Princeton university press, pp. 481 – 553.

Niederle, M. and Vesterlund, L. (2010), 'Explaining the gender gap in math test scores: The role of competition', *Journal of Economic Perspectives* **24**(2), 129–44.

OECD (2020), *OECD Health Statistics 2020*, OECD. (Accessed December 14 2020.).
URL: *https://www.oecd.org/health/health-data.htm*

Riska, E. (2001), 'Towards gender balance: but will women physicians have an impact on medicine?', *Social Science & Medicine* **52**(2), 179–187.

Roter, D. L., Geller, G., Bernhardt, B. A., Larson, S. M. and Doksum, T. (1999), 'Effects of obstetrician gender on communication and patient satisfaction', *Obstetrics & Gynecology* **93**(5), 635–641.

Roter, D. L. and Hall, J. A. (2004), 'Physician gender and patient-centered communication: a critical review of empirical research', *Annu. Rev. Public Health* **25**, 497–519.

Scarpa, R., Thiene, M. and Train, K. (2008), 'Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the alps', *American Journal of Agricultural Economics* **90**(4), 994–1010.

Scott, A. (2000), Economics of general practice, *in* A. J. Culyer and J. P. Newhouse, eds, 'Handbook of Health Economics', Vol. 1, Elsevier, pp. 1175–1200.

Swait, J. and Louviere, J. (1993), 'The role of the scale parameter in the estimation and comparison of multinomial logit models', *Journal of Marketing Research* **30**(3), 305–314.

Swait, J. and Marley, A. A. (2013), 'Probabilistic choice (models) as a result of balancing multiple goals', *Journal of Mathematical Psychology* **57**(1-2), 1–14.

Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge (UK).

Train, K. E. and Weeks, M. (2005), Discrete choice models in preference space and willingness-to-pay space, *in* R. Scarpa and A. Alberini, eds, 'Applications of simulation methods in environmental and resource economics', Springer, pp. 1–16.

Tversky, A. (1972), 'Choice by elimination', *Journal of Mathematical Psychology* **9**(4), 341–367.

Visser, M. S. and Roelofs, M. R. (2011), 'Heterogeneous preferences for altruism: Gender and personality, social status, giving and taking', *Experimental Economics* **14**(4), 490–506.

Wallin, A., Swait, J. and Marley, A. (2018), 'Not just noise: A goal pursuit interpretation of stochastic choice.', *Decision* **5**(4), 253.

Wang, J., Iversen, T., Hennig-Schmidt, H. and Godager, G. (2020), 'Are patient-regarding preferences stable? evidence from a laboratory experiment with physicians and medical students from different countries', *European Economic Review* p. 103411.

Wieland, A. and Sarin, R. (2012), 'Domain specificity of sex differences in competition', *Journal of Economic Behavior & Organization* **83**(1), 151–157.

Appendix A1. Experimental parameters

Table A.1: Experimental parameters

	Payment	Var	0	1	2	3	4	5	6	7	8	9	10
I	FFS	R_{jA}(q)	0.00	1.70	3.40	5.10	5.80	10.50	11.00	12.10	13.50	14.90	16.60
		$R_{jB}(q)$	0.00	1.00	2.40	3.50	8.00	8.40	9.40	16.00	18.00	20.00	22.50
		$R_{jC}(q)$	0.00	1.80	3.60	5.40	7.20	9.00	10.80	12.60	14.40	16.20	18.30
		$R_{jD}(q)$	0.00	2.00	4.00	6.00	8.00	8.00	15.00	16.90	18.90	21.30	23.60
		$R_{jE}(q)$	0.00	1.00	2.00	6.00	6.70	7.60	11.00	12.30	18.00	20.50	23.00
	CAP	$R(q)$	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00
II	FFS,CAP	c(q)	0.00	0.10	0.40	0.90	1.60	2.50	3.60	4.90	6.40	8.10	10.00
III	FFS	$\pi_{jA}(q)$	0.00	1.60	3.00	4.20	4.20	8.00	7.40	7.20	7.10	6.80	6.60
		$\pi_{jB(q)}$	0.00	0.90	2.00	2.60	6.40	5.90	5.80	11.10	11.60	11.90	12.50
		$\pi_{jC(q)}$	0.00	1.70	3.20	4.50	5.60	6.50	7.20	7.70	8.00	8.10	8.30
		$\pi_{jD(q)}$	0.00	1.90	3.60	5.10	6.40	5.50	11.40	12.00	12.50	13.20	13.60
		$\pi_{jE(q)}$	0.00	0.90	1.60	5.10	5.10	5.10	7.40	7.40	11.60	12.40	13.00
	CAP	$\pi(q)$	12.00	11.90	11.60	11.10	10.40	9.50	8.40	7.10	5.60	3.90	2.00
IV	FFS,CAP	B1k(q)	0.00	0.75	1.50	2.00	7.00	10.00	9.50	9.00	8.50	8.00	7.50
		$B_{2k}(q)$	0.00	1.00	1.50	10.00	9.50	9.00	8.50	8.00	7.50	7.00	6.50
		$B_{3k}(q)$	0.00	0.75	2.20	4.05	6.00	7.75	9.00	9.45	8.80	6.75	3.00

Note: This table shows all experimental parameters. $R_{jk}(q)$ denotes physicians' payment for patient type j and illness k. Under FFS, $R_{jk}(q)$ varies with illnesses k and increases in q, whereas under CAP, $R_{jk}(q)$ remains constant. The costs for providing medical services $c_{jk}(q)$ increase in q and are the same under all experimental conditions. The physicians' profit $\pi_{jk}(q)$ is equal to $R_{jk}(q) - c_{jk}(q)$. $B_{jk}(q)$ denotes the patient benefit for the three patient types $j = 1, 2, 3$ held constant across conditions.

Appendix A2. Methodological Appendix

The conventional way of deriving a choice model as described by Train (2009), is to assume individuals who maximize random utility, and let random utility be the sum of a deterministic utility term and a random term. As highlighted by Hess and Rose (2012) and Hess and Train (2017), the model we apply in this paper is in practice the same as traditional choice models in textbooks. The motivation and interpretation differ, however, as we highlight bounded rationality as a source of randomness in behavior rather than explaining randomness in behavior as driven by factors that are unobservable to the researcher, as in McFadden (1974). The argument for motivating the source of randomness differently, is that factors being unobservable to the researcher are less plausible when data are from a controlled laboratory experiment. We assume boundedly rational decision makers and allow for individuals' behavior to be influenced by factors that are irrelevant to their utility. Boundedly rational individuals are assumed to maximize F_{njt}, which is a linear combination of utility, $V(B_{jt}, \pi_{jt})$, and noise, ϵ_{njt}:

$$F_{njt} = R_{nt}V(B_{jt}, \pi_{jt}) + (1 - R_{nt})\epsilon_{njt} \quad , \tag{.1}$$

The specification in (.1) reflects the assumption that rationality is present to some *degree*: The rational part of the individuals' objective, $V(B_{jt}, \pi_{jt})$, is assumed to be a function of health benefits B_{jt} and profit π_{jt}, and R_{nt} denotes the weight assigned to this part of the objective. The term $1 - R_{nt}$ is a measure of the degree in which the individual's behavior is affected by aspects that are irrelevant to utility, and these irrelevant aspects are captured by the noise term ϵ_{njt}. By definition, $R_{nt} \in (0, 1)$ applies for all decision-makers and choice occasions, and we allow R_{nt} vary between individuals and between choice occasions in the experiment. Hence our empirical specification enables us to test whether the *degree of randomness*, as measured by R_{nt}, differ between females and males.

We assume a utility function that is homogeneous of degree one (constant returns to scale) and implement this assumption by a log-linear Cobb-Douglas specification:

$$V(B_{jt}, \pi_{jt}) = \alpha_n ln(B_{jt}) + (1 - \alpha_n) ln(\pi_{jt}) \quad , \quad \alpha_n \in (0, 1)\ \forall n \quad , \tag{.2}$$

where the parameter α_n denotes the relative valuation of health benefit in n's preference function. Specifying physician preferences to comprise a linear combination of profit and patients' health benefit has been common in the health economic literature for more than three decades, see e.g Ellis and McGuire (1986, 1990); Scott (2000); Léger (2008) and this assumption has been shown to fit data from both field (Godager et al., 2009, 2015) and experiments (Godager and Wiesen, 2013; Wang et al., 2020). Inserting for ϵ_{njt} and $V(B_{jt}, \pi_{jt})$ in Equation (.1), Our model specification becomes:

$$F_{njt} = R_{nt}[\alpha_n ln(B_{jt}) + (1 - \alpha_n) ln(\pi_{jt})] + (1 - R_{nt})[a_j + \varepsilon_{njt}] \quad , \tag{.3}$$

which is the model presented above in Equation (1) Assuming that ε_{njt} terms are independently, type 1 extreme value distributed is a sufficient, but not necessary, condition for ensuring that maximizing decision-makers and the model specification in (1) result in choice probabilities given by the logit formula:

$$P_{nit} = \frac{\exp\left(\frac{R_{nt}}{1-R_{nt}} V(B_{it}, \pi_{it})\right)}{\sum_{j=1}^{J} \exp\left(\frac{R_{nt}}{1-R_{nt}} V(B_{jt}, \pi_{jt})\right)} \quad . \tag{.4}$$

Choice probabilities given by (.4) can be derived axiomatically under weaker assumptions (Dagsvik, 1995; Erlander, 1998; Dagsvik, 2008, 2018, 2016), hence the specification of logit models to characterize human choices does not rely on strong assumptions.

Appendix A3.

Additional empirical results and robustness checks

Table A.2: P-values from Mann-Whitney-U tests for 30 choice occasions. H_0: gender difference is zero

OCCASION #	FFS	CAP
1	0.0595	0.4662
2	0.1622	0.5252
3	0.7929	0.3034
4	0.4965	0.2268
5	0.2815	0.0418
6	0.1248	0.2888
7	0.4714	0.4450
8	0.9595	0.2056
9	0.4672	0.7065
10	0.7533	0.6945
11	0.1144	0.8197
12	0.6046	0.9518
13	0.9304	0.4206
14	0.1447	0.4220
15	0.7325	0.3139

Table A.3: Estimation results from a linear regression model with random effects. N=277 subjects, T=30 occasions.

Dependent variable: chosen q		
	Estimate	Robust Std. Err.†
FEMALE	-0.07	(0.07)
DOCTOR	-0.03	(0.08)
EXPERIENCE	-0.07	(0.10)
OCCASION		
FFS_2	1.56	(0.10) ***
FFS_3	1.26	(0.09) ***
FFS_4	1.55	(0.10) ***
FFS_5	1.90	(0.11) ***
FFS_6	-0.44	(0.08) ***
FFS_7	0.61	(0.13) ***
FFS_8	0.23	(0.13)
FFS_9	0.74	(0.14) ***
FFS_{10}	0.52	(0.17) ***
FFS_{11}	1.24	(0.08) ***
FFS_{12}	1.86	(0.08) ***
FFS_{13}	1.92	(0.08) ***
FFS_{14}	1.94	(0.08) ***
FFS_{15}	2.26	(0.09) ***
CAP_1	-0.39	(0.09) ***
CAP_2	-0.41	(0.10) ***
CAP_3	-0.25	(0.10) **
CAP_4	-0.14	(0.08)
CAP_5	-0.23	(0.09)
CAP_6	-1.68	(0.09) ***
CAP_7	-1.59	(0.08) ***
CAP_8	-1.56	(0.10) ***
CAP_9	-1.59	(0.10) ***
CAP_{10}	-1.58	(0.10) ***
CAP_{11}	0.58	(0.11) ***
CAP_{12}	0.55	(0.11) ***
CAP_{13}	0.61	(0.11) ***
CAP_{14}	0.66	(0.11) ***
CAP_{11}	0.70	(0.10) ***
constant	5.06	(0.09) ***
rho = 0.13	(fraction of variance due to u_i)	

† Clustered at the level of the individual decision-maker.
***(**)[*] indicate statistically significant parameter,
with p-value <0.0001(<0.001)[<0.01]

Table A.4: Results from maximum likelihood estimation of an ordinal regression model with random effects. N=277 subjects, T=30 occasions.

Dependent variable: chosen q

	FFS		CAP	
	Estimate	Robust Std. Err.†	Estimate	Robust Std. Err.†
FEMALE	-0.16	0.15	0.02	0.12
DOCTOR	-0.10	0.15	0.03	0.12
EXPERIENCE	-0.09	0.07	-0.10	0.07
OCCASION#				
2	7.65	0.15 ***	6.91	0.15 ***
3	7.19	0.15 ***	7.12	0.15 ***
4	7.51	0.15 ***	7.21	0.15 ***
5	8.06	0.16 ***	7.13	0.15 ***
6	5.17	0.13 ***	5.04	0.14 ***
7	6.45	0.15 ***	5.11	0.14 ***
8	5.99	0.14 ***	5.09	0.14 ***
9	6.56	0.15 ***	5.03	0.14 ***
10	6.48	0.16 ***	5.04	0.14 ***
11	7.26	0.16 ***	8.64	0.16 ***
12	8.09	0.15 ***	8.59	0.16 ***
13	8.18	0.15 ***	8.71	0.16 ***
14	8.16	0.15 ***	8.74	0.16 ***
15	8.77	0.16 ***	8.83	0.16 ***
Cutoffs				
cut_1	3.00	0.14 ***	2.93	0.12 ***
cut_2	3.08	0.14 ***	3.27	0.13 ***
cut_3	3.18	0.15 ***	3.51	0.13 ***
cut_4	4.18	0.15 ***	5.68	0.14 ***
cut_5	4.62	0.16 ***	6.40	0.15 ***
cut_6	6.13	0.16 ***	8.41	0.16 ***
cut_7	7.12	0.17 ***	9.90	0.17 ***
cut_8	9.03	0.18 ***	11.88	0.20 ***
cut_9	10.23	0.19 ***	12.46	0.21 ***
cut_{10}	10.89	0.19 ***	13.03	0.24 ***
var($u_{i\text{FFS}}$)	= 1.22			
var($u_{i\text{CAP}}$)	= 0.78			
cov($u_{i\text{FFS}}$, $u_{i\text{CAP}}$)	= -0.43			
Log likelihood	=-16254.871			

† Clustered at the level of the individual decision-maker.
***(**)[*] indicate statistically significant parameter with p-value <0.0001(<0.001)[<0.01] .
The model is estimated by means the gsem module in STATA 16.

Interpreting effects on the degree of randomness in Table 3.

While the degree of randomness in treatment choices does not differ significantly between genders, it might be interesting to give an interpretation of the absolute values of the point estimates of $\theta_{\text{\_DOCTOR}}$, $\theta_{\text{\_EXPERIENCE}}$ and $\theta_{\text{\_FEMALE}}$. The interpretation is that the gender difference is smaller than the difference caused by having experienced 15 additional decisions in the laboratory. It is interesting to note that the standard error of $\theta_{\text{\_FEMALE}}$ is smaller, and the confidence narrower, than the corresponding estimates for $\theta_{\text{\_DOCTOR}}$ and $\theta_{\text{\_EXPERIENCE}}$. In order to illustrate how a gender difference in θ would have translated to differences in R_{nt}, we compute the R_{nt}-estimates for doctors and medical students in the two parts of the experiment, $t = 1$ and $t = 2$, and present the results in Table A.5. We use the formula $R = \frac{\lambda}{\lambda+1}$, where λ is the scale parameter given by $\exp(Z \times \theta)$ in Gu et al. (2013). This specification is frequently reported in the literature. Yet, the procedure can be criticized for ignoring the noise captured by alternative specific constant, as R becomes slightly exaggerated overall. Fortunately, the *relative* differences in R over EXPERIENCE and between subject types are accurate anyway.

When comparing the differences in R between students and doctors, and over decision-maker's level of experience, we keep in mind that the interpretation of the point estimates of the three θ parameters is that the gender difference in R_{nt}, is smaller than the change in R_{nt} caused by having more experience. We remember that R represents the *degree of determinism*, reflecting the degree in which choices are driven by utility differences, while $1 - R$ measures the degree in which choices are driven by factors that are irrelevant to utility. We observe in Table A.5 that for decisions made in the second half of the experiment, treatment choices are more influenced by utility differences, and less influenced by irrelevant aspects: In the first line of Table A.5, we observe that for doctors deciding for patient 1 in FFS, R equals 0.63 if the decision maker is *less* experienced ($t = 1$, decision occurs in sessions where FFS precedes CAP). R rises to 0.73 when the decision maker is *more*

Table A.5: Variation in the *degree of determinism* over occasions and type of subject: Estimates of R_{nt} from the behavioral model in Equation (1)

	Doctors		Students	
Occasion	t=1	t=2	t=1	t=2
FFS$_1$	0.63	0.73	0.75	0.83
FFS$_2$	0.59	0.70	0.72	0.81
FFS$_3$	0.64	0.74	0.76	0.84
FFS$_4$	0.64	0.75	0.76	0.84
FFS$_5$	0.57	0.68	0.70	0.79
FFS$_6$	0.80	0.87	0.88	0.92
FFS$_7$	0.54	0.66	0.68	0.78
FFS$_8$	0.62	0.73	0.75	0.83
FFS$_9$	0.57	0.69	0.70	0.80
FFS$_{10}$	0.46	0.58	0.60	0.71
FFS$_{11}$	0.67	0.77	0.78	0.85
FFS$_{12}$	0.62	0.72	0.74	0.82
FFS$_{13}$	0.75	0.83	0.84	0.90
FFS$_{14}$	0.66	0.76	0.78	0.85
FFS$_{15}$	0.75	0.83	0.84	0.90
CAP$_{1\text{-}5}$	0.56	0.67	0.69	0.78
CAP$_{6\text{-}10}$	0.58	0.69	0.71	0.80
CAP$_{11\text{-}15}$	0.72	0.81	0.82	0.88

Note:
$t = 1(t = 2)$ refer to first (second) half of a session, and capture difference in EXPERIENCE

experienced ($t = 2$, decision occurs in sessions where CAP precedes FFS). The same result also applies to decisions made by students, and we observe that when students decide for patient 1 in FFS, R rises from 0.75 for the less experienced student ($t = 1$) to 0.83 for the more experienced student ($t = 2$).[11] The variation in R over choice occasions is substantial, ranging from the lowest estimate of 0.46 for patient 10 in FFS, to the highest estimates, 0.90, which are found for patients 13 and 15 in FFS.

Appendix B: Experimental design

B1: Decision situation

Each participant in our experiment acts in the role of a physician and is assumed to be concerned about her own profit π as well as about the patient benefit B. He/she has to choose a quantity of medical services for a given patient whose health benefit is determined by that choice. Each physician i decides on the quantity of medical services $q \in 0, 1, ..., 10$ for three patient types ($j = 1, 2, 3$) with five abstract illnesses ($k = A, B, C, D, E$). The combination of patient type and illness characterizes a specific patient $1A, 1B, 1C, ..., 3D, 3E$. Patient types differ in the health benefit they gain from the medical services ($B_{1k}(q)$, $B_{2k}(q)$, $B_{3k}(q)$). A common characteristic of $B_{jk}(q)$ is a global optimum q^*_{jk} on the quantity interval [0,10] that yields the highest benefit to patients of type j for illnesses k. The level of health benefit patients receive from optimal treatment is nearly the same for all three patient types, only the quantity of medical services differs to get there. The three types of patients reflect the patients' different states of health (intermediate, good, bad).

To illustrate the physicians' task, Figure 1a provides the decision screens for patient 1C under FFS and whereas Figure 1b shows the decision screen for the same patient under CAP. Columns 1 to 6 of the screen, respectively, indicate: (1-2) medical services and the corresponding quantities; (3) physician's remuneration, increasing in the quantity of medical services under FFS (Figure 1a), whereas under CAP the remuneration corresponds to a lump-sum payment per patient (Figure 1b); (4) costs of medical services that are constant across patient types in both parts of the experiment; (5) physician's profit (remuneration minus costs); (6) patient benefit.

B2: Parameters

Under FFS, physicians' remuneration increases in q, and remuneration differs with illnesses, $R_{jA}(q), R_{jB}(q), \ldots, R_{jE}(q)$. Physicians are paid a lump sum of 12 Token per patient under CAP, which was set close to the mean of the maximum profits a subject could achieve under FFS when averaging over patients. For an overview of all payment parameters, see panel I in Table A1 in Appendix A.

The patient benefit $B_{jk}(q)$ varies across patient types. The quantities that maximize patient benefit are $q^*_{1k} = 5$, $q^*_{2k} = 3$ and $q^*_{3k} = 7$ for patient types 1, 2, and 3, respectively with the highest level of health benefit from optimal treatment being nearly the same for all three patient types. Patient benefit $B_{jk}(q)$ is shown in panel IV of Table A1.

[11] An interesting finding which Wang et al. (2020) did not report in their paper, is that R_{nt} also rises as decision makers acquire experience with the current payment scheme. This can most easily be seen in CAP where R rises from one choice occasion to the next without exceptions.

Further parameters relevant for physicians' decisions are costs $c_{jk}(q)$ and, particularly, profit $\pi_{jk}(q)$; see panels II and III of Table A.1. Under both payment systems, physicians have to bear costs $c_{jk}(q) = 1/10 \times q^2$. Under CAP, profits are the same for all illnesses, and the profit-maximizing quantity, $\hat{q}$, is 0 for all patients, jk. Under FFS, profits vary across illnesses because remuneration differs while costs are kept constant. The profit-maximizing quantity, $\hat{q}$, is 10 for all patients, jk, except for those with illness A, (i.e., patients 1A, 2A and 3A) as $\hat{q}_{jA} = 5$. For patient 1A, $\hat{q} = q^* = 5$. For the sake of simplicity, the patients are numbered from 1 to 15.

Figure 1a: Decision screen for patient 1C under FFS

Patient type 1/Illness C

Medical services	Quantity	Your Remuneration (in Taler)	Your Cost (in Taler)	Your Profit (in Taler)	Patient benefit (in Taler)
none	0	0.00	0.00	0.00	0.00
Service C1	1	1.80	0.10	1.70	0.75
Service C1, Service C2	2	3.60	0.40	3.20	1.50
Service C1, Service C2, Service C3	3	5.40	0.90	4.50	2.00
Service C1, Service C2, Service C3, Service C4	4	7.20	1.60	5.60	7.00
Service C1, Service C2, Service C3, Service C4, Service C5	5	9.00	2.50	6.50	10.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6	6	10.80	3.60	7.20	9.50
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7	7	12.60	4.90	7.70	9.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8	8	14.40	6.40	8.00	8.50
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8, Service C9	9	16.20	8.10	8.10	8.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8, Service C9, Service C10	10	18.30	10.0	8.30	7.50

Your Decision

Please indicate the quantity of medical services you want to provide

OK

Figure 1b: Decision screen for patient 1C under CAP

Patient type 1/Illness C

Medical services	Quantity	Your Remuneration (in Taler)	Your Cost (in Taler)	Your Profit (in Taler)	Patient benefit (in Taler)
none	0	12.00	0.00	12.00	0.00
Service C1	1	12.00	0.10	11.90	0.75
Service C1, Service C2	2	12.00	0.40	11.60	1.50
Service C1, Service C2, Service C3	3	12.00	0.90	11.10	2.00
Service C1, Service C2, Service C3, Service C4	4	12.00	1.60	10.40	7.00
Service C1, Service C2, Service C3, Service C4, Service C5	5	12.00	2.50	9.50	10.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6	6	12.00	3.60	8.40	9.50
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7	7	12.00	4.90	7.10	9.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8	8	12.00	6.40	5.60	8.50
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8, Service C9	9	12.00	8.10	3.90	8.00
Service C1, Service C2, Service C3, Service C4, Service C5 Service C6, Service C7, Service C8, Service C9, Service C10	10	12.00	10.0	2.00	7.50

Your Decision

Please indicate the quantity of medical services you want to provide

OK

Appendix C: Experiment material

C1: Instructions of the experiment

[Numbers/text in brackets refer to the conditions where doctors participate.]

{Sentences/decision screens in braces are inserted into the instructions either in condition FFS or in condition CAP.}

[[Text in double brackets refer to explanatory notes.]]

Instructions Part 1
General Information

In the following experiment, you will make a couple of decisions. Following the instructions and depending on your decisions, you can earn money. It is therefore very important that you read the instructions carefully.

You take your decisions anonymously on your computer screen. During the experiment, you are not allowed to talk to any other participant. Whenever you have a question, please raise your hand. The experimenter will answer your question in private in your cubicle. If you disregard these rules, you can be excluded from the experiment without receiving any payment. All amounts of money in the experiment are stated in Token. At the end of the experiment, your earnings will be converted into RMB at an exchange rate of 10 Token = 1 [6] RMB and paid to you in cash.

The experiment consists of two parts. We we will inform you now on the decision situation in Part 1. We will provide you with the instructions for Part 2 as soon as Part 1 has ended. Please note that your decisions in Part 1 have no influence on your decisions in Part 2 and vice versa.

Your decisions in Part 1 of the experiment
During the experiment, you are in the role of a physician. You have to make 15 decisions regarding the treatment of patients. All participants of this experiment take their decisions in the role of physicians. You decide on the quantity of medical services you want to provide for given clinical symptoms of a patient.

You decide on your computer screen where five different kinds of clinical symptoms – A, B, C, D, and E – of three different patient types – 1, 2, and 3 – will be shown one after another. For each patient you can provide 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 medical services.

Your remuneration is as follows:

{Condition CAP: For each patient you receive a lump-sum payment that is independent of the quantity of medical services.}

{Condition FFS: A different payment is assigned to each quantity of medical services. The payment increases in the quantity of medical services.}

While deciding on the quantity of medical services, in addition to your payment you determine the costs you incur when providing these services. Costs increase with increasing quantity provided. Your profit in Token is calculated by subtracting your costs from your payment.

A certain benefit for the patient is assigned to each quantity of medical services, the patient benefit that the patient gains from your provision of services (treatment). Therefore, your decision on the quantity of medical services not only determines your own profit, but also the patient benefit. An example for a decision situation is given on the following screen.

{Decision screen for patient 1C under FFS and CAP}

[[NOTE: The same screens as in Figures 1a and 1b in Appendix B1.]]

You decide on the quantity of medical services on your computer screen by typing an integer between 0 and 10 into the box labeled ”Your Decision”.

After all participants have taken their decisions for the respective patient you will proceed to the next patient. There are no real, but abstract patients participating in this experiment. Yet, the patient benefit, which an abstract patient receives by your providing medical services, will be beneficial for a real patient. The total amount of patient benefit determined by your 15 decisions will be provided to a patient with cancer treated in Shandong Qilu Hospital [Shandong University Cancer Hospital]. The money will be directly transferred to the patient's in-hospital account to finance part of his/her treatment fee.

Each time you make a decision on the quantity of medical services you will be informed on your profit and the patient benefit. After you have made your 15 decisions in Part 1 of the experiment you will get to know your total profit and the corresponding total patient benefit.

Earnings in Part 1 of the experiment
After you have made your decisions in Part 1 of the experiment, your overall earnings will be calculated by summing up your profits from providing medical services to the 15 patients. This amount will be converted from Token into RMB. Your earnings of Part 1 of the experiment together with the earnings of Part 2 will be paid to you in cash at the end of the experiment (rounded to 1 Yuan).

The patient benefit gained by all 15 patients will be converted into RMB at the end of the experiment, too, and will be transferred to the real patient's in-hospital account. To this end the experimenter and a monitor will go together to Shandong Qilu Hospital [Shandong University Cancer Hospital]. After the transfer, the signed receipt will be scanned into electronic form and will be sent to all the participants via e-mail in order to ensure the authenticity of the above process. Personal information will be blinded black to respect the patient's privacy.

After the end of Part 2 of the experiment, one participant is randomly assigned the role of the monitor. The monitor receives a payment of 50 [200] RMB in addition to the payment from the experiment. In the end, the monitor signs a form to verify that the procedure described above was actually carried out. This form will be sent to all participants together with the receipt via e-mail.

Next, please answer some questions familiarizing you with the decision situation. After your 15 decisions, please answer some further questions on your screen.

Instructions Part 2

The experiment will now be repeated including one change. Like in Part 1 you will make 15 decisions. After these 15 decisions the experiment will end.

The General Information from Part 1 also applies for Part 2 of the experiment.

Your decisions in Part 2 of the experiment

Also in Part 2 of the experiment, you are in the role of a physician and you have to make 15 decisions regarding the treatment of patients. All participants take their decisions in the role of physicians. You decide on the quantity of medical services you want to provide for given clinical symptoms of a patient.

Like in Part 1 you decide on your computer screen where five different kinds of clinical symptoms A, B, C, D, and E of three different patient types (1, 2, and 3) will be shown one after another. For each patient you can provide 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 medical services.

Your remuneration is as follows:
{Condition CAP: For each patient you receive a lump-sum payment that is independent of the quantity of medical services.}
{Condition FFS: A different payment is assigned to each quantity of medical services. The payment increases in the quantity of medical services.}

As in Part 1, while deciding on the quantity of medical services, in addition to your payment you determine the costs you incur when providing these services. Costs increase with increasing quantity provided. Your profit in Token is calculated by subtracting your costs from your payment.

A certain benefit for the patient is assigned to each quantity of medical services, the patient benefit that the patient gains from your provision of services (treatment). Therefore, your decision on the quantity of medical services not only determines your own profit, but also the patient benefit. An example for a decision situation is given on the following screen.

{Decision screen for patient 1C under FFS and CAP}

[[NOTE: The same screens as in Part 1. Yet, participants who saw the screen for FFS in Part 1, now see the screen for CAP – and vice versa]]

You decide on the quantity of medical services on your computer screen by typing an integer between 0 and 10 into the box labeled "Your Decision".

After all participants have taken their decisions for the respective patient you will proceed to the next patient.

Also in this part of the experiment there are no real, but abstract patients participating in this experiment. Yet, the patient benefit, which an abstract patient receives by your providing medical services, will be beneficial for a real patient. Also in the second part of the experiment the total amount of patient benefit determined by your 15 decisions will be provided to a patient with cancer treated in Shandong Qilu Hospital [Shandong University Cancer Hospital]. The money will be directly transferred to the patient's in-hospital account to finance part of his/her treatment fee.

Each time you made a decision on the quantity of medical services you will be informed on your profit and the patient benefit. After you have made your 15 decisions in Part 2 of the experiment you will get to know your total profit and the corresponding total patient benefit.

Earnings in Part 2 of the experiment
After you have made your decisions in Part 2 of the experiment, your overall earnings will be calculated by summing up your profits from providing medical services to the 15 patients. This amount will be converted from Token into RMB at the end of the experiment and will be paid to you in cash together with the earnings of Part 1 of the experiment (rounded to 1 Yuan).

The patient benefit gained by all 15 patients will be converted into RMB at the end of the experiment, too, and will be transferred to the real patient's in-hospital account. To this end the experimenter and a monitor will go together to Shandong Qilu Hospital [Shandong University University Hospital]. After the transfer, the signed receipt will be scanned into electronic form and will be sent to all the participants via e-mail in order to ensure the authenticity of the above process. Personal information will be blinded black to respect the patient's privacy. Information about the procedure has been given in Part 1 of the experiment.

Next, please answer some questions in this part of the experiment that will familiarize you with the present decision situation. After your 15 decisions, please answer some further questions on your screen.

C2: Comprehension questions prior to the experiment

Please read the instructions carefully. If you have a question, please raise your hand. The experimenter will come to you and answer your question. Have you understood the instructions?

To familiarize you with the decision situation we first ask you to answer 3 questions. We will inform you when the actual experiment starts.

Assume a physician wants to provide the quantity of 0 [10, 4] medical services for the patient above.
1 [2, 3] a) What is the remuneration?
1 [2, 3] b) What are the costs?
1 [2, 3] c) What is the profit?
1 [2, 3] d) What is the patient benefit?

The test questions are now completed. When you click on the button the experiment will start.