Are patient-regarding preferences stable?

Evidence from a laboratory experiment with physicians and medical students from different countries

Jian Wang
Department of Health Management and Health Economics, University of Oslo and Dong Furen Institute of Economic and social development,Wuhan University China

Tor Iversen
Department of Health Management and Health Economics, University of Oslo

Heike Hennig-Schmidt
Department of Economics, University of Bonn and Department of Health Management and Health Economics, University of Oslo

Geir Godager
Department of Health Management and Health Economics, University of Oslo and Health Services Research Unit, Akershus University Hospital, Norway

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK

Working paper 2019: 1

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-265-0

Are patient-regarding preferences stable?

Evidence from a laboratory experiment with physicians and medical students from different countries[☆]

Jian Wang [1,4,*], Tor Iversen [1,*], Heike Hennig-Schmidt[1,3,*], Geir Godager*[1,2,*]

Abstract

We quantify patient-regarding preferences by fitting a bounded rationality model to data from incentivized laboratory experiments, where Chinese medical doctors, German medical students and Chinese medical students participate. We find a remarkable stability in patient-regarding preferences when comparing subject pools and we cannot reject the hypothesis of equal patient regarding preferences in the three groups. The results suggest that health economic experiments can provide knowledge that reach beyond the student subject pool, and that knowledge on preferences of decision-makers in one cultural context can be of relevance for very different cultural contexts.

Keywords: Laboratory experiment, Bounded rationality, Payment mechanism, Physician behavior

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

[☆]Financial support by the Research Council of Norway (IRECOHEX, Project-No. 231776) and by National Nature Science Foundation of China (NSFC, Grant-No. 71373146) is gratefully acknowledged.

[*]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]Health Services Research Unit, Akershus University Hospital, Norway

[3]BonnEconLab, University of Bonn, Germany

[4]Dong Furen Institute of Economic and Social Development, Wuhan University, China

1. Introduction

Laboratory experiments—being a complementary approach to surveys, field studies, randomized control trials, and experiments in the field—have the unique feature of allowing the researcher to investigate the causal effects of changes in the variable of interest on behavior, as laboratory experiments provide the opportunity for ceteris-paribus variations. One may, for example, implement a variation of a payment mechanisms, while keeping all other variables constant (Falk and Fehr, 2003; Falk and Heckman, 2009). Controlled lab experiments also have great potential as a 'test bed' for field experiments, large-scale studies, and policy reforms, before these changes are implemented. They require much less time and financial resources to be implemented and analyzed (see Hennig-Schmidt et al. (2011); Cox et al. (2016)). Finally, laboratory experimentation provides a scalable approach, as it allows for the flexible adaptation of the experimental setting. While the use of laboratory experiments has contributed to new knowledge on provider preferences, critics argue that artificial context and specific or irrelevant subject pools are substantial limitations reducing the external validity of results. Recent efforts to investigate the replicability of laboratory experiments have also documented that results from many laboratory experiments cannot easily be replicated Camerer et al. (2016), Camerer et al. (2018).

Our study addresses the important issues of replicability and validity of experimental results, i.e. whether results have relevance for subject pools that are not represented in the experiment. We focus on physician decision making under different payment mechanisms. We "bring the field to the lab" by recruiting medical doctors to participate in our lab experiment alongside medical students. Our experiment is an extended version of the laboratory experiment by Hennig-Schmidt et al. (2011). While the experimental parameters are the same as in Hennig-Schmidt et al. (2011), subjects in our experiments make treatment choices in both capitation (CAP) and fee-for-service (FFS) payment schemes. We use a medically framed setting in which subjects in the role of physicians make decisions on the provision of medical services. A subject's quantity choice determines his or her own profit and a patient's health benefit. Decisions are incentivized by monetary rewards determined by the payment method in question (FFS or CAP). Under FFS, participants receive a fee for each medical procedure or service they provide to a patient. Under CAP, they receive a fixed payment for each patient they treat, independent of the quantity of medical services they provide. We extend Hennig-Schmidt et al.'s (2011) between-subject design by confronting doctors and medical students sequentially with both FFS and CAP payment schemes, while varying the order of payment schemes across sessions. Each participant in our experiment is assigned a physician's role and joins the experiment only once. A real

patient's health is affected by the participants' treatment decisions.

Doubling the number of decisions and conducting the experiment with a substantially larger sample in China enable identification of differences in patient-regarding preferences across very different subject pools. This paper also contributes to the literature by fitting a model of bounded rationality to the incentivized choice data. To the best of our knowledge, this is the first paper to quantify preference parameters in a bounded rationality model by means of experimental data on medical treatment choices. The large number of choice occasions enables us to quantify the impact of more experienced subjects on the degree of rational decision-making in a model of bounded rationality.

We address three research questions in this paper. Our first research question is whether there is any evidence that the degree of patient-regarding preferences depends on the subject pool. This is an important question concerning external validity, as using students as experimental subjects is widespread, and if preferences of medical students change when they become medical doctors, the external validity of experimental results that rely on students is limited. We are not the first to conduct a laboratory study on payment incentives with real doctors; the other papers we know of are Brosig-Koch et al. (2016; 2018), Fink and Kairies-Schwarz (2017) and Hafner et al. (2017).

Results in the previous literature on differences between physicians and medical students are mixed, and, to the best of our knowledge, no previous studies provide parameter estimates of patient-regarding preferences with a physician sample large enough to provide statistical power in between-subject-pool tests for differences. We estimate preference parameters for physicians from China (N=99), medical students from China (N=178) and medical students from Germany (N=42). We find a remarkable stability in patient-regarding preferences when comparing physicians from China, medical students from China and medical students from Germany, and we cannot reject the hypothesis of equal patient regarding preferences in the three groups.

Our second contribution to the literature is in our second research question, where we ask how accumulating experience in decision-making in the lab affect subject behavior. We find that the subjects behave less random when they become more experienced with making decisions in the lab. Within the context of our model, the interpretation is that experience cause more rational behavior.

Our third research question concerns the validity of results from lab experiments. We ask whether behavior of medical doctors in a particular experimental incentive scenario can be predicted without using the experimental data on doctors' behavior in this particular scenario. We ask whether experimental data on doctors' behavior in CAP combined with

data on student behavior in CAP and FFS are sufficient to provide accurate predictions on how doctors will behave in FFS, and whether doctors' behavior in CAP can be predicted in a similar fashion when data on doctors' behavior in CAP is excluded from the analysis. We find that our out-of-sample-predictions of doctors' behavior closely resembles observed behavior, as the distributions of predicted action probabilities and observed relative frequencies are not statistically different.

We also check the replicability of the results in the original study by Hennig-Schmidt et al. (2011). We investigate whether the main findings of the between-subject analysis reported therein are robust, or whether a substantial enlargement of the subject pool, within-subject-analysis of effects of experimental conditions, as well conducting the experiment in a very different context will change the results. Evidence from our analyses suggests that the findings reported by Hennig-Schmidt et al. (2011) are robust. We find that also in our within-subject design, both doctors and medical students provide fewer medical services under CAP than they do under FFS. As in the original experiment, whether CAP or FFS is beneficial for the patient depends on the patient type.

The remainder of the paper is organized as follows: in Section 2, we relate our study to the literature on physician behavior and payment scheme experiments as well as the literature on bounded rationality and revealed preferences. In section 3, we describe the experimental design, parameters, and procedure. In Section 4 we compare the present experiment with the original study. Section 5 presents an empirical model of bounded rationality, as well as results from maximum likelihood estimation. Section 6 discusses our findings and concludes.

2. Related literature

2.1. Physician preferences

The question of how physicians should be paid in order to promote higher quality health care services while controlling costs has been central in health economics research for decades. Understanding how physicians respond to economic incentives is fundamental when aiming to achieve these goals. The existing theoretical literature and the empirical literature based on field data and from controlled laboratory experiments provide evidence that the design of a payment system for health care providers affects their decisions (see for example Ellis and McGuire, 1986, 1990; Scott, 2000; Gosden et al., 2001; Iversen and Lurås, 2000; Iversen, 2004; Yip et al., 2010; Hennig-Schmidt et al., 2011; Brosig-Koch et al., 2016, 2017). When analyzing the most common forms of physician payment—fee-for-service (FFS) and per-capita payment (CAP) (see, e.g., McGuire, 2000)—a reoccurring

result is that the former promotes activity, and the resulting service volume can be higher than optimal. Likewise, the latter prospective payment system encourages the provision of few services, and the resulting service volume tends to be smaller than optimal (Newhouse, 1996).

Payment systems based on FFS have traditionally been the prevailing payment method for health care providers in many countries around the world. However, rapidly increasing health care expenditures have motivated discussions on payment reform (see the discussion in Hennig-Schmidt et al., 2011; Yip and Hsiao, 2008; Eggleston, 2012). In recent years, policy makers in many countries (e.g. USA, China, Germany and Norway) have implemented health care reforms using prospective payment methods including capitation in order to curb the growth in health expenditures. When implementing a payment reform, policy makers, however, face the challenge of accounting for health care providers' patient-regarding preferences, as the relative size of patient-regarding preferences influence the optimal mixture of fee-for-service and capitation-based payment components. The empirical evidence in the literature on how payment schemes affect physician behavior most often relies on field studies, register- or survey data. These data are characterized by an absence of control, which is necessary in order to provide reliable causal inferences about the effects of incentives. Uncontrolled variations in the field can include, e.g., unobserved characteristics of the patient population or self-selection of providers (Gaynor and Gertler, 1995; Sørensen and Grytten, 2003; Grytten et al., 2009; Devlina and Sarma, 2008).

There are few studies in the experimental literature on physician behavior that investigate the differences between medical students and physicians. The evidence is inconclusive. Among the contributions are Brosig-Koch et al. (2016; 2018). The former study finds that on the one hand, medical students and physicians respond to financial incentives of FFS and CAP in a similar and consistent way. The response differs between subject pools, however, with physicians responding less than students do. In the latter study – analyzing the introduction of performance pay based on a CAP system – the effect on patient-regarding service provision is not significantly different between physicians and medical students.

2.2. Bounded rationality- and revealed preference studies

There is no consensus on best practice when it comes to representing human behavior by models. The assumption that humans (behave as if they) maximize their utility has been a fundamental element in the larger part of economic research. An increasing mass of evidence is indicating that individuals often make choices that are inconsistent with utility maximization by perfectly rational individuals. Within the empirical game theory

literature, much research has documented the lack of support for the hypothesis that rational subjects maximize utility and choose alternatives consistent with a Nash equilibrium. The lacking support for the hypothesis of rational, utility-maximizing individuals is much discussed in the literature on empirical game theory, and Goeree and Holt (2001) give an enlightening overview.

A rich literature exists on the applicability of Samuelson's (1938) *revealed preference principle*, and whether human behavior in a non-strategic environment is compatible with the revealed preference (RP) axioms. The RP axioms (see, for example Cox (1997) and Andreoni and Miller (2002) and the references therein) provide necessary and sufficient conditions for an observed sequence of choices to be consistent with utility maximization, and the Weak Axiom of Revealed (WARP), Strong Axiom of Revealed Preference (SARP) and Generalized Axiom of Revealed Preference (GARP) have been subject to rigorous testing. See for example Afriat (1973) and Varian (1982,1983) for earlier contributions and Cox (1997), Mattei (2000), Février and Visser (2004) for more recent contributions based on the experimental approach.

It has been apparent for decades that behaviors that violates the RP axioms are too frequently observed to be overlooked as minor anomalies without scientific importance. Some might find it reassuring that violations of GARP are more common among children than adults, see Harbaugh et al. (2001). Choice sequences which violate the revealed preference axioms, include violations of the transitivity requirement, such as choosing A over B, next B over C and then C over A. Violations of necessary conditions should normally lead to the rejection of a hypothesis. Here, one might reject *either*, the hypothesis that one can represent stable human preferences by a monotonous utility function, or, the hypothesis that individuals are perfectly rational upon maximizing their utility, or possibly reject both of the above hypotheses. Instead, development of ad-hoc approaches to address deviation from utility maximization became part of the RP literature. An example is the development of tools to measure the *seriousness* of deviations from rational behavior, such as Afriat's (1972) *"cost efficiency index"* and the *"violation index"* developed by **?**. While such efforts in support of revealed preference theory have been questioned for decades (March, 1978), such tools are applied also in recent RP studies (Li et al., 2017; Li, 2018). The scientific literature includes many contributions criticizing revealed preference theory, and notable are the contributions by Tversky and Kahneman (1974; 1979) and Sen (1973; 1977; 1993; 1997).

While there is no disagreement among economists on whether or not a given sequence of choices violates the revealed preference axioms, views differ on how to address the fact that

humans so frequently behave inconsistently and make choices that violates the revealed preference axioms. We distinguish between three different approaches. One approach is to categorize a given choice sequence as either rational or irrational, and thereafter classify the severity of irrational choices. Examples of this approach can be found in Andreoni and Miller (2002), Fisman et al. (2007), Li et al. (2017) and (Li, 2018).

A second approach is to depart from the dichotomy of rational versus irrational behavior, and consider choices to be a result of a probabilistic process, where individuals, who are assumed rational to some *degree*, (behave as if they) maximize a combination of utility and noise. As argued by McFadden et al. (1999) the perfect rationality assumption is unnecessarily strong. An approach relaxing the restrictive assumption of perfect rationality is to model individual behavior by means of a random utility model (RUM). As described by McFadden (2001), in his Nobel lecture on the history of random utility models and choice modelling, substantial achievements in the analysis of economic choices are from contributions that consider choice to be the result of a stochastic process. Within this modelling paradigm there is a positive probability that sub-optimal alternatives are chosen. Alternatives that provide higher utility are more likely to be chosen, however. The bounded rationality model we estimate below is a type of RUM, which provides an internally consistent set of assumptions that allow for degrees of rationality without any need for auxiliary measures such as violation indices to identify preference parameters. The RUM has close links to behavioral models in other fields, and Glimcher, (2011, p72) argues that economic models of random utility can be reduced to psychological models of percept as well as to neurobiological models of biochemical transduction[1].

The third perspective, starting with the work of 1978 Nobel laureate Herbert A. Simon, perceives the notion of bounded rationality in a fundamentally different way. Individuals are neither assumed capable of maximizing a utility function, nor assumed to behave *as if* they are doing so. They use different *processes* to be able to make difficult decisions and solve complex problems. This way of problem solving is not to be viewed as irrational but follows its own specific rules, which can be and have been studied in the laboratory and the field (Simon, 1957). Simon formulated his main concerns in his 1978 Nobel lecture, where he highlighted the need for a descriptive decision theory, which focuses on how decisions are made, and not just on the decision outcomes (Simon, 1979 , p. 498). When giving his Nobel lecture, theories already existed that incorporated the behavioral notion of bounded rationality like, for instance, the need to search for decision alternatives, the replacement of optimization by targets and satisficing goals, and mechanisms of learning and adaptation (Simon, 1979, p. 510). Simon's approach has been developed further in the domains of strategic as well as individual decision making, in particular by Reinhard

Selten (e.g. Selten, 1998b; 1998a; 1997; Selten et al., 1997; see also Ockenfels and Sadrieh, 2010) Werner Güth (e.g Güth et al., 1982; Güth and Kliemt, 2010) and Gerd Gigerenzer (e.g. 1999; 2001)

In this paper, we apply the second approach, which perceives individual behavior as if it were the result of maximizing a combination of utility and noise. We are not the first to study how contextual factors such as experience influence the degree of rationality in a bounded rationality RUM. In a study about forest management, Holmes and Boyle (2005) found that later choices in their stated preference experiment were significantly less influenced by noise than the earlier choices, and suggest that the phenomenon is caused by respondents' learning about the choice task. Olsen et al. (2017) found that time of day affects randomness in behavior in online food choice experiments. The noise term in the approach we use can be interpreted as capturing influencing factors and decision motives not made explicit in the utility function.

The possible relation between experience in laboratory decision making and rationality in strategic decision making is discussed by McKelvey and Palfrey (1995) who analyze the data by Lieberman (1960), and find strong evidence suggesting that the influence of random noise in their quantal response equilibrium model declines systematically as experimental subjects become more experienced in laboratory decision making. In a study of strategic decision-making in the context of oligopolistic competition with varying number of competing opponents, Ge and Godager (2019) find that decision-making is less influenced by randomness in more competitive settings.

3. Experiment

3.1. Experimental design

Basic setup and decision situation

Our experimental design draws on the seminal model by Ellis and McGuire (1986). The physician is assumed to be concerned about her own profit π as well as about the patient benefit B, the latter depending on the quantity of medical services q. The specifics of the experimental design are taken from Hennig-Schmidt et al. (2011). Our experiment differs from theirs, however, in that we apply a within-subject design whereas Hennig-Schmidt et al. employ a between-subject setup.

Each participant in our experiment acts in the role of the physician. The decision task is 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 ($k = A, B, C, D, E$). She is sequentially confronted with the same 15 decisions (patients) in both payment systems FFS and CAP. – with either CAP first and FFS second or vice versa. Patient types reflect the patients' different states of health. 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 ($B_{1k}(q)$, $B_{2k}(q)$, $B_{3k}(q)$). The patient health benefit is measured in monetary terms. A physician's choice of medical services simultaneously determines the patient benefit and her own profit ($\pi_{jk}(q)$). The patient is assumed to be passive and fully insured, accepting each level of medical service provided by the physician. In our experiment, no real patients are present. However, outside the lab, physicians' quantity choices have consequences for a real patient. The money corresponding to patient benefits aggregated over all decisions was transferred to a real patient's in-hospital account (see the Instructions in Appendix B). Thus, participants in our experiment did have an incentive to take the patient benefit into account when making their decisions. We did not inform the participants about the name of the person to whom the money was transferred.

To illustrate the physicians' task, Figure 1a provides the decision screen for patient 1C under CAP whereas Figure 1b shows the decision screen for the same patient under FFS. See also the Chinese decision screens in Appendix D. The physician gets information on her remuneration, costs and profit as well as on the patient's benefit for each quantity from 0 to 10. 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).

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

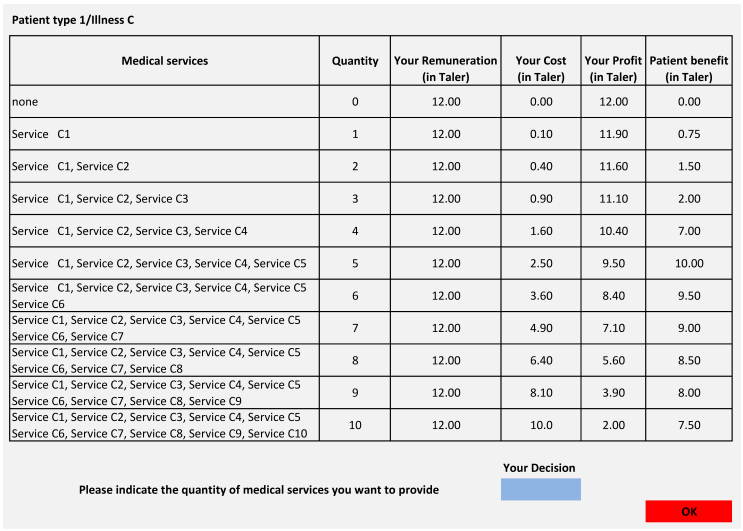
Please indicate the quantity of medical services you want to provide

OK

The first two columns of the screens state the medical services and the corresponding

quantities. Column 3 indicates the physician's remuneration that corresponds to a lump-sum payment per patient in CAP (Figure 1a), whereas under FFS, the remuneration increases in the quantity of medical services (Figure 1b). Column 4 shows the costs of medical services that are constant across patient types in both parts of the experiment. Physician's profit (remuneration minus costs) is given in the fifth column, and the final column comprises the patient benefit.

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

Parameters

Physicians are paid a lump sum of 12 Token per patient under CAP. Under FFS, physicians' remuneration increases in q. Remuneration differs with illnesses, $R_{jA}(q), R_{jB}(q), ..., R_{jE}(q)$. The lump sum paid under CAP is close to the average maximum profit per patient a subject could achieve under FFS. 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. A concave benefit function is applied, the common characteristic of which is a global optimum on the quantity interval [0, 10]. There is a unique quantity q^*_{jk} that yields the highest benefit to patients of type j for illnesses k. 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—and the participants are informed of all values before they make their quantity decision. Patient benefit $B_{jk}(q)$ is shown in panel IV of Table A1. We refer to quantities smaller than q^*_{jk} as underprovision of medical care, while provision of quantities larger than q^*_{jk} is defined as overprovision. 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. Physicians have to bear costs $c_{jk}(q) = 1/10 \times q^2$ under both payment systems. Under CAP, profits are the same for all illnesses. The profit-maximizing quantity $\hat{q}$ is 0 for all patients, jk. Under FFS, pro-

fits 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, $\hat{q} = q^* = 5$.

3.2. Experimental protocol

Applying a within-subject design, each of the 178 Chinese medical students and 99 doctors participating in our experiment was sequentially confronted with the same 15 decisions (patients) in both of the two payment systems FFS and CAP. The subjects were randomly assigned to experimental sessions where either CAP was implemented in Part 1 of the session followed by FFS in Part 2 (condition CF) or in reversed order (FFS in Part 1 followed by CAP in Part 2, condition FC). This 2 x 2 design allows us to compare the behavior of the two subject pools over experimental conditions. Each participant was assigned a physician's role and joined the experiment only once, either in condition CF or in condition FC. Participants were informed at the beginning that the experiment consisted of two parts, but they did not know what the second part would be.

Our experiment was conducted in September 2012 and 2013 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 material distributed to the participants was translated into Chinese by a Chinese native fluent in German from the original German version by using the back translation method (Brislin, 1970). For a translation into English, see Appendix C1. It is important to instruct participants in their native language because the language the experiment is presented in may affect their behavior; see e.g. Costa et al. (2014a; 2014b). Medical students who voluntarily participated in the experiment were recruited via notices posted at the campus and by email invitations. Doctors were recruited through a phone call stating that a research experiment from Shandong University needed volunteers.

The experimental procedure was as follows and was exactly the same for medical students and doctors. After having arrived and before the experiment started, participants were randomly allocated to their workstations. The workstations were numbered and separated from each other by wooden panels and curtains. It was thus guaranteed that they made their decisions in both parts of the experiment in complete anonymity. Then, instructions for Part 1 of the experiment were distributed to participants and read out by a native experimenter. Participants decided under either a CAP or an FFS system. Subjects were given plenty of time to read the instructions and to ask clarifying questions in private, and questions were answered individually. In cases that the content was important for all participants, the question and answer were repeated in public. To check for participants'

understanding of the decision task, they had to answer a set of test questions on remuneration, costs, physician profit and patient benefit for three different quantities of medical services for a patient they were not confronted with in the actual experiment. See Appendix C2 for the English translation of test questions and the respective computer screens. Each participant then went through a sequence of 15 choices (patients) on the quantity of medical services to be provided. The order of patients to be treated was predetermined and kept constant across conditions. After each decision, each participant in both parts of the experiment was informed about his/her profit and the patient benefit generated by the previous choice. At the end of the first part of the experiment, each participant received information about his/her total profit achieved and the total health benefit generated during all 15 quantity decisions. Finally, the participants answered some open-ended questions.

Next, instructions for the second part of the experiment were distributed and read out by the native experimenter. In Part 2, participants decided under the payment system they had not yet been confronted with. Again, each decision-maker received information on his/her total profit achieved and the total health benefit created during all 15 decisions. After the second part of the experiment had been completed, participants were again asked some open-ended questions. The doctors were also asked about socio-demographic variables and professional experience. Finally, participants were informed about their individual total profit and the resulting total benefit aggregated over Parts 1 and 2 of the experiment as well as on their final monetary payoff. Finally, participants were paid in private and dismissed individually.

To ensure that the doctors and medical students trusted the experimenters to actually transfer the money derived from the patient benefit, a certain procedure was applied to ensure trust: 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 decisions of all participants in the respective session. Then, the monitor and an assistant to the experimenters went by taxi to the Shandong Cancer Hospital in Jinan and paid the corresponding amount in cash into the patient's account at the hospital-cashier's desk. This procedure is similar to Eckel and Grossman (1996), Hennig-Schmidt et al. (2011), Godager and Wiesen (2013), Hennig-Schmidt and Wiesen (2014), Godager et al. (2016) and Brosig-Koch et al. (2016, 2017a). 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 that the appropriate monetary amount was paid into the patient's hospital account. All participants in each session received an email stating the amount equaling the aggregate health benefits generated during the respective session.

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 with conditions alternating across sessions. Sessions lasted for about 90 minutes. Based on the decisions in the two conditions, each of the 178 medical students on average earned 28 RMB; 15 RMB (€1.80) in CAP and 13 RMB (€1.56) in FFS plus a show-up fee of 15 RMB (€1.80). Doctors on average earned 160 RMB (86 RMB (€10.32) in CAP and 74 RMB (€8.88) in FFS. Average payoffs for students approximately corresponds to the hourly wage of a student helper at Shandong University of about 30 RMB. For doctors the average hourly wage is about 120 RMB. Based on all 8,310 decisions, a total of 19,814 RMB (€2,377.68) was transferred to the real patient's account; 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 procedure was given by Norwegian Social Science Data Services (reference 44267).

4. Comparing results with the original experiment

We start by describing the subject pools and proceed to testing for differences in aggregate provision behavior between CAP and FFS. Throughout the paper, all statistical tests applied are two-sided. We give a summary of subject characteristics in Table 1. In our experiment, 277 subjects participated. Of these, 178 were medical students of whom 56 % were females. The overall average duration of study was 4.9 semesters. The major of all medical students was Clinical Medicine. The number of participating doctors was 99 with an average age of 40, and 70 % were females. They had on average of 16.23 years of professional experience. The doctors were practicing as general practitioners (75 %), in traditional Chinese medicine (10 %) or in public health (4 %); 11 % of the doctors practiced in all or several of these fields. All doctors were employed at community health centers, where salaries are set according to a fixed salary scheme. Thus, both the medical students and the doctors have in common that they had little or no practical experience with fee-for-service payment or capitation payment systems.

Table 1: Subject characteristics

	Chinese students		Chinese Doctors		German students †	
Female	56 %	N=178	70 %	N=99	62 %	N=42
Age (Mean)	-		40.0	N=89	22.3	N=22
Semester (Mean)	4.9	N=177	-		-	
Years of practice (Mean)	-		16.2	N=88	-	

† The German data were provided by Hennig-Schmidt et al. (2011)

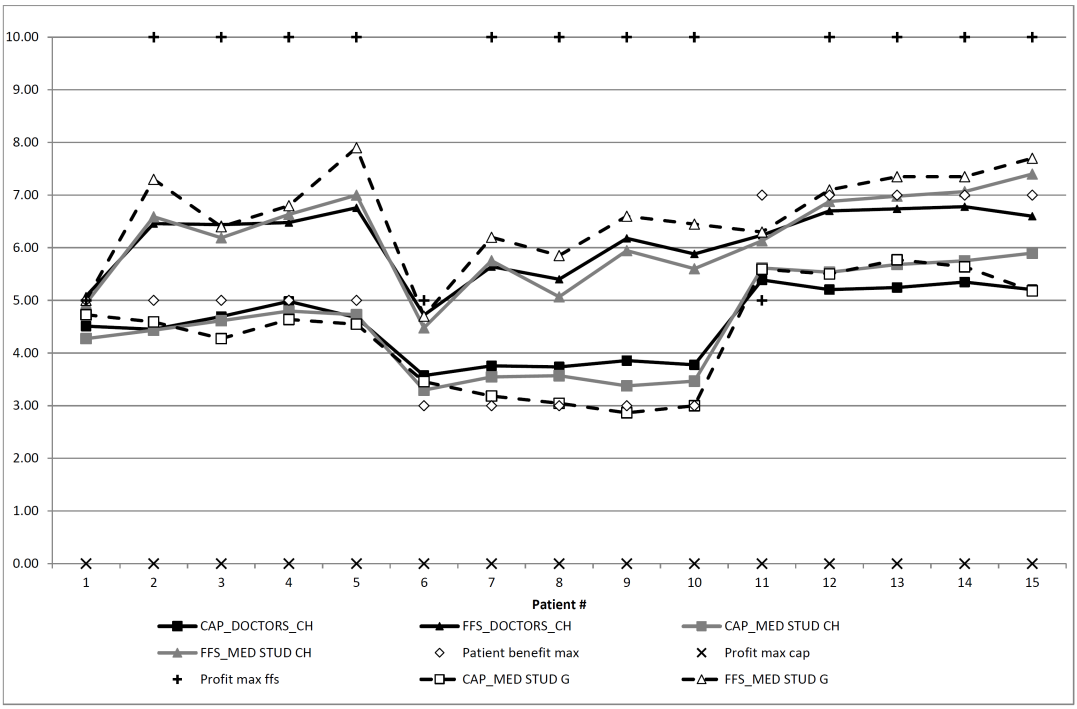
Table 2. Aggregate behavior of Chinese doctors and medical students under CAP and FFS. Mean (Std.Dev) of quantity and patient benefit, and the number of decisions

Payment	Doctors		Medical students		Total	
system	Quantity	#obs	Quantity	#obs	Quantity	#obs
CAP	4.59 (1.78)	1485	4.53 (1.57)	2670	4.55 (1.65)	4155
FFS	6.03 (1.92)	1485	6.16 (1.78)	2670	6.11 (1.83)	4155

Notes:
This table shows descriptive statistics on quantities of service provision over payment systems and subject pools. #obs is the number of decisions under each payment scheme.

The aggregate provision behavior under CAP and FFS is presented in Table 2. We analyze the data pooled over decisions within the two payment schemes and compare doctors and medical students (N=277 subjects; 4155 decisions per payment system). We here also pool data from the same payment scheme, regardless of whether the scheme was implemented first or second in the experiment. In line with earlier studies, we find that our participants respond to the incentives given by the payment systems: average quantities in CAP are lower than in FFS (CAP: 4.55, FFS: 6.11; N=277).

Figure 2. Mean quantity provision for each of the 15 Patients under CAP and FFS differentiated according to subject pools – pooled over both parts of the experiment.



Notes: This figure shows average quantities of service provision as well as patient benefit and profit maxima for payment systems FFS and CAP for Chinese doctors (N=99), and Chinese medical students (N=178), and German medical students (N=42), pooled over both parts of the experiment.

Our within subject design enables us to test whether the amount of service provided to a given type of patient by a given subject, differs between the two payment schemes. We conduct 15 tests on the difference between payment schemes, matching the provided service quantity for a given occasion in FFS to the corresponding patient scenario in CAP. For each test we may reject the null hypothesis that provided service quantity does not differ over payment schemes ($p \leq 0.0001$ in each test, Wilcoxon matched-pairs signed-ranks test, WM in the following). Applying a conservative Bonferroni correction for multiple hypothesis testing gives an adjusted threshold for statistical significance of $p = 0.05/15 = 0.0033$ when tests are applied 15 times. Hence, applying Bonferroni corrections would not influence our conclusions. Over- and underprovision for the three patient types in the present experiment are affected by the payment system in a similar way as in the original experiment, as described in Figure (2). In line with previous empirical and experimental studies (Hennig-Schmidt et al., 2011, Keser et al., (2013), Hennig-Schmidt and Wiesen, (2014) and Brosig et al., 2016, 2017), the incentives of the two payment systems affect medical service provision in that participants provide more services under FFS than under CAP. We conclude that the main findings of Hennig-Schmidt et al. (2011) are confirmed when applying a within-subject configuration of the experiment.

5. Estimating preference parameters of a bounded rationality model.

We refer to the vast choice modelling literature that build on the early work of Luce (1959) and McFadden (1974) when specifying our *random utility model*: We assume that patient-regarding subjects make choices that maximize a (log) linear combination of utility and noise. The inclusion of a noise term implies that a subject who consistently maximizes the objective function can choose different alternatives in two identical choice occasions. Our bounded rationality RUM enables us to depart from the *rational* versus *irrational* dichotomy, and consider rationality to always be present to some degree.

Consider a subject type, indexed by n, choosing treatment alternative, indexed by $j = 0, 1, 2, \ldots 10$ to maximize a Cobb-Douglas function of profit, patient benefit and noise:

$$U_{njt} = \underline{U} B_{jt}^{\alpha_n} \pi_{jt}^{\beta_n} \epsilon_{njt}^{\mu_{nt}} \ , \ \alpha_n \text{ and } \beta_n \in (0,1) \ \forall n \quad . \tag{1}$$

In order to simplify notation, we suppress the index for each of the 30 choice occasions (15 for the German students). The index $t = 1, 2$ indicates whether the choice occasion is in the first payment scheme (t=1) or second payment scheme (t=2) in the experimental session, while n indicates subject type: we let $n = c$ denote Chinese medical student, $n = d$ denotes Chinese medical doctor, and $n = g$ denote German medical student.

Only the relative size of α_n, β_n and μ_{nt} can be identified (Train, 2009), and hence a normalization, such as assuming the relative preference weights sum to unity, is necessary for identification.

Experiment scale and identification of μ_{nt}

In the experimental protocol of Hennig-Schmidt et al. (2011) and in our experiment, the real values of the experimental tokens were set with the aim that hourly payment rates within the experiment are close to subjects' alternative income. For example, the token value for medical doctors were set six times higher than for the Chinese medical students. In the estimations that follow, we use the experimental tokens *as is*, without converting to any real currency. We now show that this does not result in a loss in generality. We let r_n denote the token exchange rate for subject pool n, and rewrite the objective function as

$$U_{njt} = \underline{U} [B_{jt} r_n]^{\alpha_n} [\pi_{jt} r_n]^{(1-\alpha_n)} \epsilon_{njt}^{\mu_{nt}}, \ \alpha_n \in (0,1) \ \forall n \quad . \tag{2}$$

Which can be written:

$$U_{njt} = r_n \underline{U} {B_{jt}}^{\alpha_n} {\pi_{jt}}^{(1-\alpha_n)} {\epsilon_{njt}}^{\mu_{nt}} \quad , \tag{3}$$

or in log-linear form:

$$\tilde{U}_{njt} = ln(\underline{U}) + \alpha_n ln(B_{jt}) + (1-\alpha_n) ln(\pi_{jt}) + ln(r_n) + \mu_{nt}\varepsilon_{njt} \quad , \tag{4}$$

with $\varepsilon_{njt} = ln(\epsilon_{njt})$. We see in (4) that the token exchange rate enters our model as an additive, subject-specific constant which does not change ordinal utility over alternatives, and therefore we cannot identify the effect of token exchange rate. However, we may identify differences in randomness in behavior across subject pools, by means of subject pool dummies. Using the notation for the S-MNL model by Fiebig et al. (2010) we may write our model as:

$$\tilde{U}_{njt} = \sigma_{nt}[ln(\underline{U}) + \alpha_n ln(B_{jt}) + (1-\alpha_n) ln(\pi_{jt})] + a_j + \varepsilon_{njt} \tag{5}$$

where a_j is a vector of alternative specific constants (ASC). While behaviorally equivalent, the unit of measurement differs between Equations 4 and 5. In Equation 4, the unit of measurement is *utility* , whereas the unit of measurement in 5 is that of the error term. Hence, μ_{nt} and σ_{nt} is definitionally linked, and their relation is simply $\mu_{nt} = \sigma_{nt}^{-1}$. Following Fiebig et al. (2010) we do not multiply the alternative specific constants by σ_{nt}. The reason is that alternative specific constants are fundamentally different from observable attributes, and it is reasonable to consider ASCs to be part of the error structure.

In the S-MNL model, σ_{nt} is given by:

$$\sigma_{nt} = exp(\theta z_{nt}) \quad , \tag{6}$$

where z_{nt} is a vector of variables which are constant within each choice occasion, but varies between subject pools. Included in z_{nt} are two dummies equal to 1 for correspondingly medical doctors and German students (meaning that Chinese medical students is the reference category), a dummy equal to 1 in choice occasions where subjects are experienced ($t = 2$), and 17 dummies which indicate the 18 unique choice occasions, 15 in FFS and 3 in CAP. We assume that ε_{njt} is type 1 extreme value distributed, and by implication (5) is a scaled logit model, or S-MNL model in the terminology by Fiebig et al. (2010).

In the experimental design, some available alternatives have either zero profit or zero patient benefit, which complicates the use of logs. This is solved by replacing ln(0) by 0,

and introducing a dummy equal to 1 if either profit or patient benefit is zero. In this way, we are also able to identify the reference utility, $\underline{U}$, which is fixed for all subjects.

After estimating the parameters of the S-MNL model by means of STATA 15 (Gu et al., 2013), we compute the subject type- and occasion specific σ_{nt} estimates by inserting the estimated θ-vector in (6). Next we simply use the definition $\mu_{nt} = \sigma_{nt}^{-1}$ to acquire the estimates of μ_{nt}.

Under the assumption that $\mu_{nt} > 0$, we do not impose strong restrictions on which alternative can be chosen by an individual who maximizes (5). For example, an individual might possibly choose a Pareto-inferior alternative, for example by overproviding services under CAP payment. Also, an individual might choose A rather than B on one occasion, and B rather than A on another, identical occasion. Such behavior would be inconsistent with maximizing (5) with $\mu_{nt} = 0$.

Our application of the S-MNL model relies on the assumption that $\mu_{nt} > 0$, meaning that some degree of randomness in behavior is present. Before we proceed to the estimation, we show that the hypothesis that subject behavior is influenced by randomness can be supported by data directly: In CAP payment scheme, each subject make treatment decisions five times for each patient type without any variation in incentives. Subjects in all three subject pools frequently change their minds, and make different choices across identical scenarios. In Table 3, we describe individuals' choice variation for each of the three patient types in CAP. We see that for patient 1, 146 (49 %) subjects make the same treatment choice in each of the 5 identical choice occasions, whereas 153 subjects (51 %) vary their treatment choice and are observed with more than one unique action. Correspondingly, 115 (38%) and 186 (62%) subjects vary their treatment choice for patients 2 and 3. With this observation in mind, we assume that $\mu_{nt} > 0$ when we estimate the parameters in (5). We present the estimation results in Table (4).

Table 3: Prevalence of choice variation in absence of incentive variation (CAP)

	Subsample			
Patient 1	All	Chinese	Chinese	German
# unique actions		student	doc	student
1 (No variation)	146	99	34	13
2	73	47	22	4
3	41	20	18	3
4	23	7	15	1
5	16	5	10	1
Total	299	178	99	22
Patient 2	All	Chinese	Chinese	German
# unique actions		student	doc	student
1 (No variation)	184	128	43	13
2	56	29	22	5
3	27	6	18	3
4	25	13	11	1
5	7	2	5	0
Total	299	178	99	22
Patient 3	All	Chinese	Chinese	German
# unique actions		student	doc	student
1 (No variation)	113	67	36	10
2	110	70	31	9
3	46	31	14	1
4	18	7	9	2
5	12	3	9	0
Total	299	178	99	22

This table shows the frequency of choice variation when subjects make 5 repeated treatment choices for the same patients (1 ,2 and 3). Sample: 178 Chinese students, 99 Chinese doctors and 22 German students.

Table 4: Results from maximum likelihood estimation

Sample: 178 Chinese students, 99 Chinese doctors, 42 German students.
30 (15) choice occasions for each Chinese (German) subject

		Chinese student		**Chinese doctor**		**German student**	
α_n		0.51 * CI(0.36 -0.66)		0.42* CI(0.29 - 0.55)		0.40* CI(0.23 - 0.58)	
		t=1	t=2	t=1	t=2	t=1	t=2
	FFS_{1A}	0.31	0.19	0.61	0.37	0.23	
	FFS_{1B}	0.37	0.23	0.73	0.45	0.28	
	FFS_{1C}	0.35	0.21	0.68	0.42	0.26	
	FFS_{1D}	0.32	0.20	0.64	0.39	0.24	
	FFS_{1E}	0.41	0.25	0.82	0.50	0.31	
	FFS_{2A}	0.14	0.09	0.28	0.17	0.11	
	FFS_{2B}	0.46	0.28	0.90	0.55	0.34	
	FFS_{2C}	0.29	0.18	0.58	0.35	0.22	
μ_{nt}†	FFS_{2D}	0.39	0.24	0.76	0.47	0.29	(N.A)
	FFS_{2E}	0.57	0.35	1.13	0.69	0.43	
	FFS_{3A}	0.27	0.17	0.54	0.33	0.21	
	FFS_{3B}	0.36	0.22	0.70	0.43	0.27	
	FFS_{3C}	0.20	0.12	0.40	0.24	0.15	
	FFS_{3D}	0.29	0.18	0.58	0.36	0.22	
	FFS_{3E}	0.20	0.12	0.40	0.24	0.15	
	CAP_1	0.55	0.34	1.08	0.66	0.41	
	CAP_2	0.49	0.30	0.96	0.59	0.37	
	CAP_3	0.23	0.14	0.46	0.28	0.18	

Confidence intervals are based on standard errors that are clustered at the level of the individual subject.

*Estimated parameter is significantly different from zero with a p-value < 0.001

† Based on estimated θ parameter, μ_{n1} is significantly different from μ_{n2} with a p-value < 0.001

We see that the confidence intervals of α_c, α_d and α_g in Table (4) have substantial overlap. We test the joint hypothesis $\alpha_c = \alpha_d = \alpha_g$, and we find that this hypothesis cannot be rejected (p-value 0.28, Wald tests). With reference to our first research question, we do not find any evidence suggesting that patient regarding preferences differ between subject pools. Preferences are stable in space, in that preferences of German and Chinese medical students appear similar. Preferences can also be considered as stable over time, noting that medical students and medical doctors in China have closely similar preferences, despite their age difference.

RESULT 1: We do not find any evidence suggesting that patient regarding preferences differ between subject pools.

We find strong evidence that experience causes reductions in μ. Within the context of our theoretical specification of the bounded rationality model, the interpretation is that experience causes more rational behavior. We see that for the Chinese subject pool, with an additional second payment scheme adding 15 choice occasions to the experiment, the influence of noise on decision-making is reduced in occasions where subjects are experienced ($t = 2$) compared to when they are inexperienced ($t = 1$). This implies that subjects are significantly more likely to choose their optimal response when they are more experienced compared to when they have less experience. With reference to our second research question, we find evidence suggesting experience affects the degree of rationality in decision-making. The hypothesis that experience does not affect the degree of rationality can be rejected for both Chinese medical students and medical doctors - the two subject pools who experienced an additional set of 15 choice occasions.

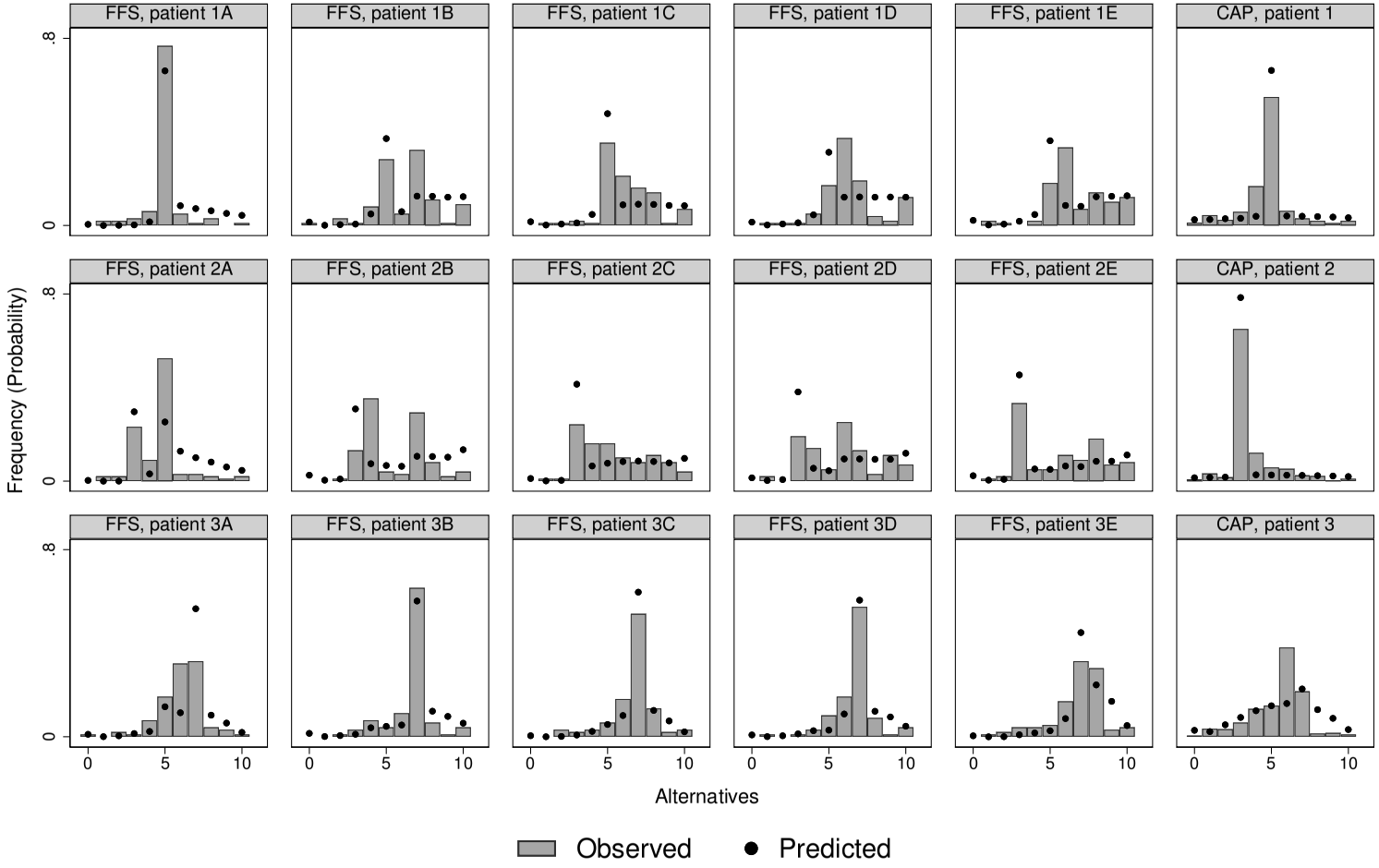
RESULT 2: We find evidence that experience does affect the degree of rationality in decision-making in that subjects are significantly more likely to choose their optimal response when they are more experienced with making decisions in the lab.

5.1. Further about experimental validity.

We now show that the use of student subjects in lab experiments can contribute to knowledge on how medical doctors *would* behave in a similar situation. Based on the result that preferences of students and medical doctors are not statistically different, we refit a restricted version of model (5) constraining preferences to be identical across subject types by assuming $\alpha_n = \alpha \ \ \forall n$. First, we exclude from our estimation sample all data records where doctors make decisions under FFS payment. We use our parameter estimates from this sub-sample, where no doctor choices under FFS are included, to predict out of sample how medical doctors are expected to behave under FFS payment. Next, we repeat the procedure to predict medical doctors' behavior under CAP payment, utilizing only the data where doctor behavior under CAP is excluded. It turns out that based on parameter estimates acquired from data on student behavior in CAP and FFS, and doctor behavior in CAP only, we can predict quite closely the behavior of medical doctors under FFS. Similarly, we can quite closely predict how doctors will behave in CAP without using any data from doctor behavior under CAP.

Figure 3. Out of sample predictions of FFS and CAP behavior of Chinese doctors

Predictions based on estimates from a restricted version of (5), assuming $\alpha_n = \alpha \ \ \forall n$. Estimates acquired while excluding FFS (CAP) choices of Chinese doctors from estimation.



Support is provided in Figure 3, where observed and predicted behavior of Chinese medical doctors in FFS and CAP scenarios is shown. There are in total 198 unique treatment alternatives in the experiment, 165 treatment alternatives for the 15 different choice scenarios in FFS, and 33 treatment alternatives for the 3 different choice scenarios in CAP. For both FFS and CAP we apply statistical tests of matched pairs to test whether the observed frequency distribution differ from the predicted distribution. We cannot reject the null hypothesis that the observed and predicted frequencies for the alternative treatments in FFS and CAP respectively are the same (p-value=0.99 for both FFS and CAP, Fisher-Pitman permutation test for paired replicates).With reference to our third research question, we find that behavior of medical doctors in a particular experimental setting can be predicted without the use of experimental data on doctors behavior in that particular scenario.

RESULT 3: We find evidence that based on behavioral data for doctors from a prevailing payment scheme and experimental data from students in both a prevailing payment scheme and a payment scenario to be introduced in a payment reform allows predicting how doctors would behave after the reform.

6. Discussion and concluding remarks.

We introduce a fully incentivized laboratory experiment, which extends the well-known experiment by Hennig-Schmidt et al. (2011) by including two payment schemes and twice the number of individual level observations. We broaden the set of included subject pools by recruiting Chinese medical doctors as well as Chinese medical students to our experiment. Our results replicate the results by Hennig-Schmidt et al. (2011), even after introducing a larger and more heterogeneous subject pool. The results corroborate the general results in the health economics literature that FFS payment encourages higher service volumes than CAP, and services volumes under FFS can become higher than what is in the best interest of the patient, and vice versa for CAP-systems.

Our results suggest that preferences of subjects from very different subject pools are similar, and hence that the financial incentives of payment systems work in a similar way in the two countries in which our participants are educated and operate. An implication of this finding would be that results from health economic laboratory experiments can provide broad knowledge on expected behavior under cultural and institutional contexts that are different from where the actual experiment is conducted. Further, there is evidence that we are able to provide accurate predictions of doctor behavior based on behavioral data for doctors from a prevailing payment scheme and experimental data from students in both a

prevailing payment scheme and a payment scenario to be introduced in a payment reform. Thus, using existing and experimental behavioral data can provide valid knowledge, which reaches beyond the included subject pools.

In our analysis, we assume individuals are boundedly rational. An interesting question is how individuals would have behaved if they had preferences given by our estimated Cobb-Douglas function and were perfectly rational, such that the influence of noise in the optimization was absent, $\mu_{nt} = 0$. We investigate how behavior in the experiment would have been under these assumptions, and the aggregate quantities of service over subjects and payment schemes can be found in Table B1 in the appendix. Our illustration shows that the scientific approach to understanding economic choices, and whether humans are regarded as perfectly rational, or boundedly rational, have a substantial influence on the predicted behavioral response of a payment reform. In the case of our chosen experimental parameters, the predicted difference in behavior between two payment schemes is exaggerated if one assume perfectly rational individuals who maximize our proposed Cobb-Douglas preference function, while boundedly rational individuals with the same Cobb-Douglas preference function provide a close fit to observed behavior, even when predicting behavior out of sample. Our computation in the Appendix shows that assuming perfect rationality can distort predictions used for policy making: Imagine a policy maker who is in favor of replacing a FFS system by a CAP payment system if the CAP scheme was expected to reduce average service quantity for patients by only 1.6 units. This policy maker might well prefer to prolong the FFS scheme if a quantity reduction of 2.5 units was expected.

Acknowledgements.

We thank Lin Jing both for his assistance in conducting experiments in China and for his very helpful comments and suggestions. We are grateful to Chaoliang Yang for testing the z-tree program as well as for translating all the material, including instructions and the text on the computer screens, into Chinese. We also thank him for his assistance in preparing and conducting the experiments at Shandong University. We are grateful to Jingyi Luo for her assistance in translating material from Chinese into English. We thank Tom Chang for his valuable comments and advice given at ASHECON in Los Angeles, USA, in June 2014. We have also benefited from discussions with Daniel Wiesen, participants at the IHEA-World Congress in Sydney in 2013, ASHECON 2014, and the Second Workshop on Behavioral and Experimental Health Economics in Hamilton, Canada, 2015. Thanks to the authors of Hennig-Schmidt et al. (2011) for providing us with their data and experimental design. We are grateful for financial support from both the Independent

Innovation Fund of Shandong University (grant no. 2012JC038) and the National Natural Science Foundation of China (grant no. 71373146) for funding the experiments conducted at Shandong University. Financial support for Geir Godager (Project-No. 231776), Heike Hennig-Schmidt (Project-No. 231776) and Tor Iversen from the Research Council of Norway is gratefully acknowledged.

Literature

Afriat, S. N. (1972): "Efficiency estimation of production functions," *International Economic Review*, 13, 568–598.

——— (1973): "On a system of inequalities in demand analysis: an extension of the classical method," *International Economic Review*, 14, 460–472.

Andreoni, J. and J. Miller (2002): "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 70, 737–753.

Brislin, R. W. (1970): "Back-translation for cross-cultural research," *Journal of Cross-cultural Psychology*, 1, 185–216.

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

——— (2017): "The effects of introducing mixed payment systems for physicians: Experimental evidence," *Health Economics*, 26, 243–262.

Brosig-Koch, J., H. Hennig-Schmidt, J. Kokot, N. Kairies-Schwarz, and D. Wiesen (2018): "Physician performance pay: Experimental evidence." *Paper presented at AEA/ASSA 2018, Philadelphia, PA, US.*

Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, et al. (2016): "Evaluating replicability of laboratory experiments in economics," *Science*, 351, 1433–1436.

Camerer, C. F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, et al. (2018): "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015," *Nature Human Behaviour*, 2, 637.

Costa, A., A. Foucart, I. Arnon, M. Aparici, and J. Apesteguia (2014a): ""Piensa" twice: On the foreign language effect in decision making," *Cognition*, 130, 236–254.

Costa, A., A. Foucart, S. Hayakawa, M. Aparici, J. Apesteguia, J. Heafner, and B. Keysar (2014b): "Your morals depend on language," *PloS one*, 9, e94842.

Cox, J. C. (1997): "On testing the utility hypothesis," *The Economic Journal*, 107, 1054–1078.

Cox, J. C., E. Green, and H. Hennig-Schmidt (2016): "Experimental and behavioral economics of healthcare," *Journal of Economic Behavior & Organization*, 131, A1–A4.

Devlina, R. and S. Sarma (2008): "Do Physician Remuneration Schemes Matter? The Case of Canadian Family Physicians," *Journal of Health Economics*, 27, 1168–1181.

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

Eggleston, K. (2012): "Health care for 1.3 billion: China's remarkable work in progress," *Milken Institute Review*, 16–27.

Ellis, R. P. and T. G. McGuire (1986): "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, 5, 129–151.

——— (1990): "Optimal Payment Systems for Health Services," *Journal of Health Economics*, 9, 375–396.

Falk, A. and E. Fehr (2003): "Why labour market experiments?" *Labour Economics*, 10, 399–406.

Falk, A. and J. Heckman (2009): "Lab experiments are a major source of knowledge in the social sciences," *Science*, 326, 535–538.

Février, P. and M. Visser (2004): "A study of consumer behavior using laboratory data," *Experimental Economics*, 7, 93–114.

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

Fink, C. and N. Kairies-Schwarz (2017): "Performance pay in hospitals: an experiment on bonuses and fines," *Paper presented at the 2017 Annual meeting of Gesellschaft für experimentelle Wirtschaftsforschung, Kassel, Germany.*

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

Fisman, R., S. Kariv, and D. Markovits (2007): "Individual preferences for giving," *American Economic Review*, 97, 1858–1876.

Gaynor, M. and P. Gertler (1995): "Moral Hazard and Risk Spreading in Partnerships," *Rand Journal of Economics*, 26, 591–613.

Ge, G. and G. Godager (2019): "Predicting behavior in games with vector payoff: An application of a quantal response equilibrium choice model," *Unpublished manuscript currently under review in Proceedings of the National Academy of Sciences*.

Gigerenzer, G. and R. Selten (2001): "Rethinking rationality," in *Bounded rationality: The adaptive toolbox*, Cambridge, MA, London, The MIT Press,.

Gigerenzer, G., P. Todd, A. R. Group, et al. (1999): *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.

Glimcher, P. W. (2011): *Foundations of neuroeconomic analysis*, Oxford: Oxford: Oxford University Press.

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

Godager, G. and D. Wiesen (2013): "Profit or Patients' Health Benefit? Exploring the Heterogeneity in Physician Altruism," *Journal of Health Economics*, 32, 1105–116.

Goeree, J. K. and C. A. Holt (2001): "Ten little treasures of game theory and ten intuitive contradictions," *American Economic Review*, 91, 1402–1422.

Gosden, T., F. Forland, I. Kristiansen, M. Sutton, B. Leese, A. Guiffrida, M. Sergison, and L. Pedersen (2001): "Impact of Payment Method on Behavior of Primary Care Physicians: A Systematic Review," *Journal of Health Services Research and Policy*, 6, 44–54.

Grytten, J., D. Holst, and I. Skau (2009): "Incentives and Remuneration Systems in Dental Services," *International Journal of Health Care Finance and Economics*, 9, 259–278.

Gu, Y., A. R. Hole, and S. Knox (2013): "Fitting the generalized multinomial logit model in Stata," *Stata J*, 13, 382–397.

Güth, W. and H. Kliemt (2010): "(Un) Bounded Rationality in Decision Making and Game Theory–Back to Square One?" *Games*, 1, 53–65.

Güth, W., R. Schmittberger, and B. Schwarze (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3, 367–388.

Hafner, L., S. Reif, and M. Seebauer (2017): "Physician Behavior under Prospective Payment Schemes: Evidence from artefactual field and lab experiments," .

Harbaugh, W. T., K. Krause, and T. R. Berry (2001): "GARP for kids: On the development of rational choice behavior," *American Economic Review*, 91, 1539–1545.

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

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

Holmes, T. P. and K. J. Boyle (2005): "Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions," *Land Economics*, 81, 114–126.

Iversen, T. (2004): "The effects of a patient shortage on general practitioners' future income and list of patients," *Journal of Health Economics*, 23, 673–694.

Iversen, T. and H. Lurås (2000): "The Effect of Capitation on GPs' Referal Decision," *Health Economics*, 9, 199–210.

Kahneman and A. Tversky (1979): "An analysis of decision under risk," *Econometrica*, 47, 263–292.

Keser, Montmarquette, M. S. and C. Schnitzler (2013): "Custom-made healthcare - An experimental investigation,CIRANO Scientific Series 2013s–15," University of Goettingen, cege Discussion Papers No. 218.

Li, J. (2018): "Plastic surgery or primary care? Altruistic preferences and expected specialty choice of US medical students," *Journal of Health Economics*, 62, 45–59.

Li, J., W. H. Dow, and S. Kariv (2017): "Social preferences of future physicians," *Proceedings of the National Academy of Sciences*, 114, E10291–E10300.

Lieberman, B. (1960): "Human behavior in a strictly determined 3× 3 matrix game," *Behavioral Science*, 5, 317–322.

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

March, J. G. (1978): "Bounded rationality, ambiguity, and the engineering of choice," *The Bell Journal of Economics*, 9, 587–608.

Mattei, A. (2000): "Full-scale real tests of consumer behavior using experimental data," *Journal of Economic Behavior & Organization*, 43, 487–497.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. E. Zarembka, Academic Press, New York, 105–142.

——— (2001): "Economic Choices," *American Economic Review*, 91, 351–378.

McFadden, D., M. J. Machina, and J. Baron (1999): "Rationality for economists?" in *Elicitation of preferences*, Springer, 73–110.

McGuire, T. G. (2000): "Physician Agency," in *Handbook of Health Economics, Vol. 1 A*, ed. by Cuyler and Newhouse, North-Holland, Amsterdam (The Netherlands), 461–536.

McKelvey, R. D. and T. R. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 10, 6–38.

Newhouse, J. P. (1996): "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection," *Journal of Economic Literature*, 34, 1236–1263.

Ockenfels, A. and A. Sadrieh, eds. (2010): *The Selten School of Behavioral Economics: A Collection of Essays in Honor of Reinhard Selten*, Springer.

Olsen, S. B., J. Meyerhoff, M. R. Mørkbak, and O. Bonnichsen (2017): "The influence of time of day on decision fatigue in online food choice experiments," *British Food Journal*, 119, 497–510.

Samuelson, P. A. (1938): "A note on the pure theory of consumer's behaviour," *Economica*, 5, 61–71.

Scott, A. (2000): "Economics of General Practice," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, Elsevier, vol. 1, 1175–1200.

Selten, R. (1998a): "Aspiration adaptation theory," *Journal of Mathematical Psychology*, 42, 191–214.

——— (1998b): "Features of experimentally observed bounded rationality," *European Economic Review*, 42, 413–436.

Selten, R., M. Mitzkewitz, and G. R. Uhlich (1997): "Duopoly strategies programmed by experienced players," *Econometrica*, 65, 517–555.

Sen, A. (1973): "Behaviour and the Concept of Preference," *Economica*, 40, 241–259.

——— (1993): "Internal consistency of choice," *Econometrica*, 61, 495–521.

——— (1997): "Maximization and the Act of Choice," *Econometrica*, 65, 745–779.

Sen, A. K. (1977): "Rational fools: A critique of the behavioral foundations of economic theory," *Philosophy & Public Affairs*, 317–344.

Simon, H. A. (1957): *Models of man; social and rational.*, Wiley, New York.

——— (1979): "Rational decision making in business organizations," *The American Economic Review*, 69, 493–513.

Sørensen, R. and J. Grytten (2003): "Service Production and Contract Choice in Primary Physician Services," *Health Policy*, 66, 73–93.

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

Tversky, A. and D. Kahneman (1974): "Judgment under uncertainty: Heuristics and biases," *Science*, 185, 1124–1131.

Varian, H. R. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50, 945–973.

——— (1983): "Non-parametric tests of consumer behaviour," *The Review of Economic Studies*, 50, 99–110.

Yip, W. and W. C. Hsiao (2008): "The Chinese health system at a crossroads," *Health Affairs*, 27, 460–468.

Yip, W. C.-M., W. Hsiao, Q. Meng, W. Chen, and X. Sun (2010): "Realignment of incentives for health-care providers in China," *The Lancet*, 375, 1120–1130.

A. Experimental parameters.

Table A1: 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.

B. Additional Table

Table B1. Illustration of behavior under the perfect rationality assumption: Predicted behavior of Chinese doctors and medical students under CAP and FFS, assuming $\mu = 0$.

Payment system	Doctors Quantity	Medical students Quantity	Total Quantity
CAP	4.33	4.67	4.55
FFS	7.33	6.87	7.03

Notes:
This table describes how aggregate quantities of service provision over payment systems and subject pools would have appeared if randomness in decision-making was absent, $\mu = 0$.

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.}

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:

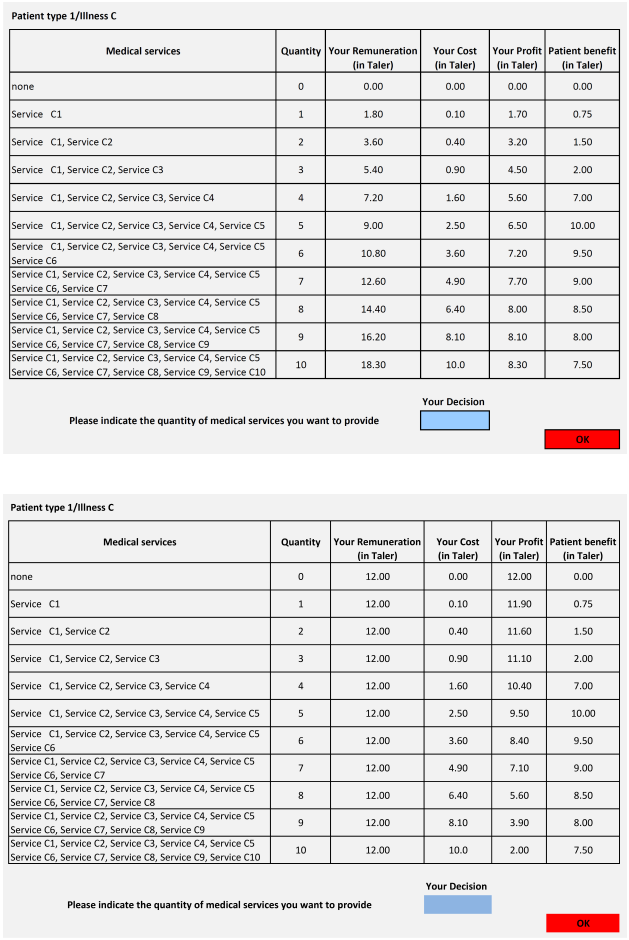
Condition CAP: For each patient you receive a lump-sum payment that is independent of the quantity of medical services.

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}



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

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

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 Provincial Cancer Hospital]. The money will be directly transferred to the patient's account in the hospital, to help him/her with part of the 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 account. To this end the experimenter and a monitor will go together to Shandong Qilu Hospital [Shandong Provincial 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}

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

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

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 Provincial Cancer Hospital]. The money will be directly transferred to the patient's account in the hospital, to help him/her with part

of the 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).

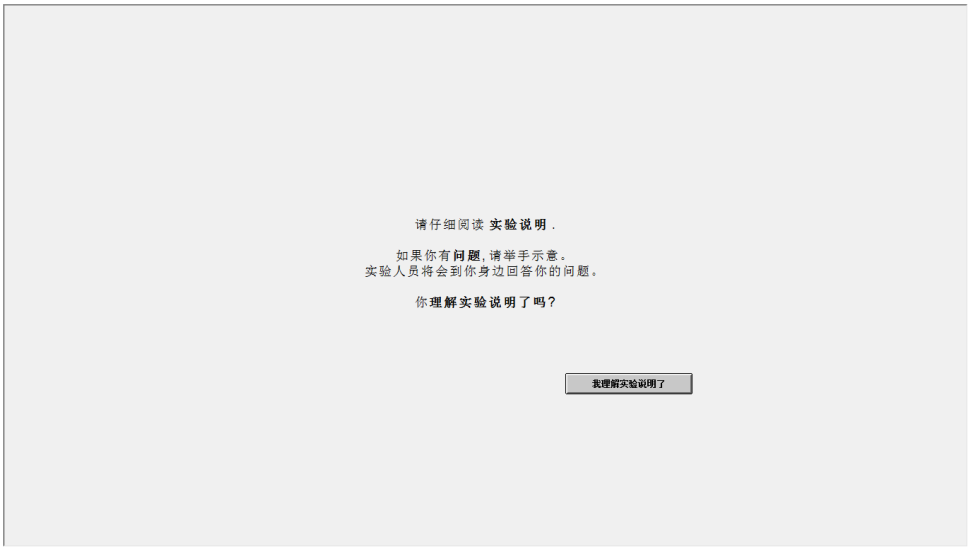
The patient benefit gained by all 15 patients will be converted into RMB at the end of the experiment, too, and will be transferred to the real patient's account. To this end the experimenter and a monitor will go together to Shandong Qilu Hospital [Shandong Provincial Cancer Hospital]. After the transfer, 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. 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: Test questions prior to the experiment

The following example applies to FFS condition. For CAP condition, screens 3 to 5 are similar to Figure C1b in Appendix C4.

Screen 1



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?

Screen 2

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

Screen 3 [4, 5]


患者类型 1 / 临床症状 F

医疗服务	数量	你的诊疗费 (以代币计算)	你的成本 (以代币计算)	你的净收益 (以代币计算)	患者效益 (以代币计算)
不提供	0	0.00	0.00	0.00	0.00
服务 F1	1	0.90	0.10	0.80	0.75
服务 F1 / 服务 F2	2	1.60	0.40	1.20	1.50
服务 F1 / 服务 F2 / 服务 F3	3	5.10	0.90	4.20	2.00
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4	4	5.10	1.60	3.50	7.00
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5	5	5.10	2.50	2.60	10.00
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5 / 服务 F6	6	7.40	3.60	3.80	9.50
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5 / 服务 F6 / 服务 F7	7	7.40	4.90	2.50	9.00
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5 / 服务 F6 / 服务 F7 / 服务 F8	8	11.60	6.40	5.20	8.50
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5 / 服务 F6 / 服务 F7 / 服务 F8 / 服务 F9	9	12.40	8.10	4.30	8.00
服务 F1 / 服务 F2 / 服务 F3 / 服务 F4 / 服务 F5 / 服务 F6 / 服务 F7 / 服务 F8 / 服务 F9 / 服务 F10	10	13.00	10.00	3.00	7.50

假设一位医生准备为上述患者提供数量为0 项的医疗服务。

1 a) 诊疗费是多少?

1 b) 成本是多少?

1 c) 净收益是多少?

1 d) 患者效益是多少?

ok

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?

Screen 6


周期

1 的 1

剩余时间 [秒]: 497

练习题到此结束。

请用鼠标点击按键，开始实验!

ok

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

C3: Questionnaires after Part 1 and Part 2 of the experiment.

[Information in brackets was requested from doctors only.]

{{Numbers/text in double braces refer to Part 2 of the experiment.}}

Please confirm your terminal number on your questionnaire. After you have made all decisions in Part 1 {{2}} of the experiment we would like to ask you to answer the following questions as good as possible. These answers are extremely important for our studies. Thank you for your cooperation.

Please put yourself back into the decision situation of Part 1 {{2}} of the experiment.

- What factors did influence your decision? Why did you decide in this way?
- How did the profit influence your decision?
- How did the patient benefit influence your decision?
- {{Major (faculty / main subject(s)):}}
- {{What is the number of your semester?}}
- {{Your gender: female/male}}
- {{Your nationality: (students only)}}
- {{[Your age]:}}
- {{[How many years of professional experience do you have?]}}
- {{[Your specification
- General Practitioner
- Traditional Chinese Medicine
- Public Health
- Other]}}

C4. Chinese decision screens.

Figure C1a. Illustration of the decision screen for patient 1C under CAP

患者类型 1/临床症状 C

医疗服务	数量	你的诊疗费 (代币)	你的成本 (代币)	净收益 (代币)	患者效益 (代币)
不提供	0	0.00	0.00	0.00	0.00
服务 C1	1	1.80	0.10	1.70	0.75
服务 C1, 服务 C2	2	3.60	0.40	3.20	1.50
服务 C1, 服务 C2, 服务 C3	3	5.40	0.90	4.50	2.00
服务 C1, 服务 C2, 服务 C3, 服务 C4	4	7.20	1.60	5.60	7.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5	5	9.00	2.50	6.50	10.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6	6	10.80	3.60	7.20	9.50
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7	7	12.60	4.90	7.70	9.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8	8	14.40	6.40	8.00	8.50
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8, 服务 C9	9	16.20	8.10	8.10	8.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8, 服务 C9, 服务 C10	10	18.30	10.0	8.30	7.50

你的决策

请填写你要提供的医疗服务的数量

OK

Figure C1b. Illustration of the decision screen for patient 1C under FFS

患者类型 1/临床症状 C

医疗服务	数量	你的诊疗费 (代币)	你的成本 (代币)	净收益 (代币)	患者效益 (代币)
不提供	0	12.00	0.00	12.00	0.00
服务 C1	1	12.00	0.10	11.90	0.75
服务 C1, 服务 C2	2	12.00	0.40	11.60	1.50
服务 C1, 服务 C2, 服务 C3	3	12.00	0.90	11.10	2.00
服务 C1, 服务 C2, 服务 C3, 服务 C4	4	12.00	1.60	10.40	7.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5	5	12.00	2.50	9.50	10.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6	6	12.00	3.60	8.40	9.50
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7	7	12.00	4.90	7.10	9.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8	8	12.00	6.40	5.60	8.50
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8, 服务 C9	9	12.00	8.10	3.90	8.00
服务 C1, 服务 C2, 服务 C3, 服务 C4, 服务 C5 服务 C6, 服务 C7, 服务 C8, 服务 C9, 服务 C10	10	12.00	10.0	2.00	7.50

你的决策

请填写你要提供的医疗服务的数量

OK