UiO : University of Oslo

Ge Ge

Essays in experimental health economics

Thesis submitted for the degree of Philosophiae Doctor

Department of Health Management and Health Economics
Faculty of Medicine

2020

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-841-0

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Summary

In order to implement effective health policy reforms, knowledge of physician preferences, and hence, their responses to policy reforms is desirable. This thesis consists of three papers aiming to address fundamental research questions on physician behavior. Paper I investigates whether the change of information scheme affects physicians' prescribing behavior. The results suggest that preannouncing a mystery shopper audit reduces physicians' probability of prescribing drugs to the pseudopatients. Paper II explores physicians' response to cost-sharing borne by the patients and finds that future physicians are concerned about the influences of their medical treatment choices on patients' consumption opportunities after co-payment. Paper III introduces a strategic decision scenario and studies physician treatment decisions under competition. The results indicate that the substantial difference in behavior between markets may be attributed to changes in individuals' scale parameter. The scale parameter rises as markets become more competitive, implying a higher degree of determinism in behavior. The data of all three papers are collected from experiments. Under the framework of stochastic choice theory, three special cases of a generalized multinomial logit model are employed in the data analysis.

Acknowledgments

The pursuit of a Ph.D. has been a challenging yet rewarding journey. I am grateful for the support and guidance I received from many others.

First and foremost, I would like to thank my main supervisor, Geir Godager. Thank you for your excellent guidance, patience, and encouragement. Your curiosity in science and enormous diligence have deeply inspired me. Thank you for sharing your time generously whenever I had questions. Thank you for all the opportunities I was given in the past years.

This thesis would not have become possible without the support and inspiration of many people. Thanks to Julie Riise, Arne Risa Hole, Tor Iversen, and Eline Aas for all the insightful discussions and patient advice. Thanks to John Dagsvik, who was the supervisor of my master's thesis, for continuing to share your expertise, humor, and encouragement with me. Thanks to Oddvar Kaarbøe for the research opportunity in registry data analysis and policy evaluation that certainly broadens my skill set. Thanks to my co-authors for their valuable input and assistance in conducting the experiments.

I acknowledge the Research Council of Norway for providing financial support for this PhD through the project of International Research Collaborative for Health Economic Experiments. I would like to thank the Department of Health Management and Health Economics at the University of Oslo for providing an interdisciplinary research environment with excellent IT and admin support. Thank you to my colleagues at the department for the help in various ways and all the fun we had at lunch and coffee breaks. Thanks to Martin Helsø for kindly sharing this beautiful LaTeX thesis template.

Last but not least, I would like to thank my friends who brought fun in this lonely endeavor, and thank you to Yannick, Kaya, and my parents for all the unconditional support and love.

Ge Ge
Oslo, November 2020

List of Papers

Paper I

The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment
Cheo, R., Ge, G., Godager, G. *et al*. *Health Economics Review* **10** (2020).
DOI: https://doi.org/10.1186/s13561-020-00290-z.

Paper II

Exploring physician agency under demand-side cost sharing — An experimental approach
Ge, G., Godager, G., and Wang, J.
Revise & Resubmit at *Health Economics*

Paper III

Predicting strategic medical choices: An application of a quantal response equilibrium choice model
Ge, G. and Godager, G.
Revise & Resubmit at *Journal of Choice Modelling*
Appendix E attached to this paper is submitted to and currently under review at *Data in Brief*.

Contents

Summary i

Acknowledgments iii

List of Papers v

Contents vii

1 Introduction 1

2 Background 3
2.1 Physician behavior . . . 3
2.1.1 Physician objective . . . 3
2.1.2 Physician competition . . . 7
2.2 Stochastic choices . . . 9
2.2.1 Interpretations of stochastic choices . . . 10
2.2.2 Stochastic choices in strategic games . . . 12
2.2.3 Sources of preference data . . . 13

3 Thesis objectives 15

4 Experiments and data 17
4.1 Economic experiments . . . 17
4.1.1 Experimental methods in economics . . . 17
4.1.2 Purposes and advantages of experiments . . . 20
4.1.3 Control in experiments . . . 21
4.2 Experimental design . . . 23
4.2.1 Variables in experiments . . . 24
4.2.2 The perceived experiment . . . 28
4.2.3 Summary of the three experiments . . . 29

4.3 Data structure . . . 31
4.4 Ethics . . . 32

5 Models and estimation methods **37**
5.1 Paper I . . . 39
5.2 Paper II . . . 40
5.3 Paper III . . . 42

6 Summary of results **47**
6.1 Paper I . . . 47
6.2 Paper II . . . 47
6.3 Paper III . . . 48

7 Discussion and future research **49**
7.1 Discussion of the results . . . 49
7.2 Discussion of methodology and future research . . . 51
7.2.1 A couple of issues on quantal response equilibrium . . 51
7.2.2 Diversity of economic experimental methods . . . 53
7.2.3 Structural models and experimental economics . . . 54

8 Conclusion **57**

References **58**

Papers **80**

I Paper I **81**

II Paper II **103**

III Paper III **139**

1 | Introduction

The aim of this thesis is to address fundamental research questions on physician behavior by combining experimental economic methods and structural stochastic choice models.

Physician behavior is at the heart of health economics. Physicians are the "captain of the team" and make decisions in many aspects of health care that determine costs and outcomes (Fuchs, 2011). Peculiarities in the health care market, such as information asymmetry and physicians' privileged social role, challenge fundamental issues on physician decision-making (Arrow, 1963). In order to implement effective health policy reforms, knowledge of physician preferences, and hence, their responses to policy reforms is desirable. This thesis contributes to this literature by acquiring knowledge on physicians' preferences when making treatment decisions. Specifically, Paper I provides evidence on the effect of preannouncement of an audit on physicians' prescribing behavior. Paper II contributes to the demand-side cost sharing literature by investigating physicians' response to cost sharing borne by patients. Paper III introduces a strategic decision scenario and studies physician treatment decisions under competition.

We address these issues by means of experiments. The experimental approach has been employed in a large body of health economic research. Controlled experiments facilitate exogenous changes of the variables of interest and thus *ceteris paribus* inference (Friedman et al., 1994). Moreover, choice scenarios that do not exist in the real market can be constructed in experiments, and trade-offs unobserved from the real world can therefore be studied (Louviere et al., 2000). All three experiments in this thesis study some choices of medical treatments. The experiment in Paper I is conducted in the field with physicians, and a randomized control trials design is employed. The experiments in Papers II and III are performed in the laboratory with medical framing. The former incorporates real incentives into a discrete choice experimental design, and the latter applies a design from strategic games.

The validity of experiments can be improved by combining the experimental data with theory-based structural models (Low and Meghir, 2017). While the experiments generate exogenous variations that help identify economic effects, the theory-based structural models describe the mechanisms through which effects operate and thus provide the framework for an interpretation of the experimental results and analysis of counterfactuals (Attanasio et al., 2012; Low and Meghir, 2017). Treatment decisions in the three experiments are discrete economic choices which can be analyzed by a well-developed class of stochastic choice modeling methods that builds on McFadden (1974). The theory of stochastic choice acknowledges the inconsistency in human behavior and provides a structure of the choice probabilities that can be justified from behavioral arguments (McFadden et al., 1999). Under the framework of stochastic choice, in the papers, we employ three choice models that can be described as special cases of a generalized multinomial logit model (Fiebig et al., 2010).

This thesis is structured as follows. Chapter 2 provides a general background on the research of physician behavior and stochastic choices and describes how my papers may contribute to the literature. Chapter 3 summarizes the objectives of each paper. Chapter 4 provides an overview of the economic experimental methods with examples from the three papers, describes the data structure, and discusses ethical considerations. Chapter 5 presents the general multinomial logit model and its special cases that we employ in the three papers. It is followed by a description of the estimation methods for each model. Chapter 6 summarizes the results from each individual paper. Chapter 7 discusses the results and methods and provides suggestions for future research. Chapter 8 concludes.

2 | Background

2.1 Physician behavior

In order to implement effective health policy reforms, knowledge of physician preferences and their responses to policy reforms is desirable. In the health care market, where the products and services provided are mostly credence goods, understanding the drivers of physicians' treatment decisions has been a central issue in health economics.

In his seminal paper, Arrow (1963) proclaimed that information asymmetry is one important characteristic of medical care that distinguishes it from many commodity goods. Due to the complexity of medical knowledge, physicians undoubtedly hold greater information about treatment alternatives and their potential effects than do patients. According to this argument, physicians might exercise market power to influence the demand for health services to maximize their profit. At the same time, as Arrow (1963) remarked, "It is clear from everyday observation that the behavior expected of sellers of medical care is different from that of businessmen in general." Indeed, physicians are specially trained and highly educated professionals who take the oath to serve as patients' agents. It is thus apparent that physicians' motives are guided by professional norms and hence different from profit maximization and are subject to the concern for patients (see for example Feldstein, 1970; Ellis and McGuire, 1986). These peculiarities in the health care market naturally challenge fundamental issues of physician incentives when making a treatment decision and how physicians act in the role of a patient's agent. In addition, the presence of health insurance, price regulation and reimbursement, and the uncertainty of the product quality add to the challenge.

2.1.1 Physician objective

Following Ellis and McGuire (1990), we consider a simple model of the physician's medical treatment decision. In this model, the physician chooses a treatment alternative x that maximizes his utility. Physician's utility $W(x)$ is a function of three elements: the profit $\pi(x)$, the patient's benefit $B(x)$, and other factors Z. For simplicity, we assume the utility is linear in these three elements and given by:

$$W(x) = \beta_1 \pi(x) + \beta_2 B(x) + Z. \tag{2.1}$$

The coefficient β_1 captures the physician's relative utility weight on profit and other factors, and the ratio of β_1 and β_2 denotes the physician's relative utility weight on profit and patient benefit. Within this simple framework, a physician prefers one treatment over another for several reasons. First, the chosen treatment might generate high profit for the physician. Second, the chosen treatment might provide more health benefit to the patient. Third, other factors, for example, social norms, reputation, and information disclosure, can also affect treatment choices. Lastly, the weight the physician puts on profit, patient benefit, and other factors might differ in different contexts, for example, competition levels of the market. The β parameters can also vary across individual physicians. In the following, I first discuss the elements in this simple framework of physician objective under the assumption that physicians' decisions are independent.[1] In the next section, I briefly introduce competition under which physicians' behaviors have impact on others.

Payment schemes

There is extensive literature examining the impact of financial incentives on physicians' treatment behavior (for example, Ellis and McGuire, 1990; Barnum et al., 1995; McGuire, 2000; Allard et al., 2011b; Hennig-Schmidt et al., 2011). One example of both theoretical and empirical importance is the discussion of the effect of salary, fee-for-service (FFS) payment, and capitation (CAP) payment. Salary-paid physicians receive a fixed salary regardless of the number of patients, quantity, or types of services provided. Under FFS, a physician receives a fee for each service he provides, and hence his profit is directly related to the type or intensity of services provided. Under CAP, a physician receives a payment for every patient he provides services for, regardless of the volume of the services. Theoretical analyses show that FFS provides incentives for volume and can lead to a problem in which physicians provide services that exceed the optimal volume for the patient or society. In contrast, CAP provides incentives that discourage volume and may result in underprovision, effort minimizing, and cream-skimming of low-cost patients (Ellis and McGuire, 1986; Blomqvist, 1991; Newhouse, 1996; Iversen and Lurås, 2000; Barros, 2003). There are also mixed-payment systems combining salary, FFS and CAP. Léger (2008) shows that a mixed payment combining

[1] Chandra et al. (2011) provide an excellent discussion of supply-side drivers of clinical decisions in general.

FFS and CAP can encourage an efficient level of care.

Empirical results on the effect of payment systems are, to some extent, mixed. A large volume of studies support the theoretical hypotheses that FFS-component in the payment system incentivize the provision of more services and services of higher intensity, while CAP-component works in the opposite direction (e.g., Krasnik et al., 1990; Coulam and Gaumer, 1992; Sørensen and Grytten, 2003; Devlin and Sarma, 2008). At the same time, some evidence also suggests small or even no impact of payment systems on medical services provision (e.g., Hurley and Labelle, 1995; Hutchison et al., 1996). The biases from self-selection into different payment schemes challenge the identification of behavioral responses to changes in payment in studies using registered data. To overcome this identification challenge and complement observational studies, Hennig-Schmidt et al. (2011) performed a controlled laboratory experiment and presented results in line with the theoretical predictions.

Trade-off between profits and patients

While it is apparent that physicians respond to financial incentives, it has been recognized that sources of motivation other than income are also important. One approach to modify the model is to include patient health benefit (B) as an argument in the physician's utility function (Ellis and McGuire, 1986; Léger, 2008). In these models, physicians are assumed to be concerned not only for their own profits, but also some benefits the patients receive from the treatment. The ratio between coefficient β_1 and β_2 in the utility function (2.1) explicitly models the trade-off or the marginal rate of substitution (MRS) of profit for patient benefit.

The adoption of physicians' other-regarding motivation in their utility function is a natural practice following Arrow's (1963) argument that differentiates physicians from purely profit-maximizing agents. Among many interpretations of the MRS, such as professionalism, ethical constraint, and the degree of physician agency, one widely received interpretation of the MRS is physician altruism. Despite the importance of its role and its implications in health care from a theoretical perspective (Ma, 1994; Ellis and McGuire, 1986; Chalkley and Malcomson, 1998; Siciliani, 2009; Allard et al., 2011a), empirical research on physician altruism is scarce (Galizzi et al., 2015). Data from surveys, interviews, and prescription records (Hellerstein, 1998; Lundin, 2000; Allaby, 2003) have been utilized to find evidence of physician altruism in an indirect manner. More recently, several incentivized lab experiments have contributed to the investigation of physician altruism and its distribution. These studies found significant heterogeneity in physician altruism (Hennig-Schmidt et al., 2011; Godager and Wiesen, 2013; Brosig-Koch et al., 2017b; Wang et al., 2020).

The empirical results and implications of physician patient-regarding motives depend on the specification of the physicians' utility function. Parallel to models that incorporate patient health benefit into physicians' objectives, some include patients' utility or welfare (Farley, 1986; Lerner et al., 1994). Physicians in the latter models are assumed to be concerned about patients' consumption opportunities after paying for medical treatment. In the case of full insurance, when the patients' consumption opportunities are not affected by the medical treatment, there is no loss in generality from specifying the physician objective that excludes patient consumption (Ellis and McGuire, 1990). However, in the majority of health care systems worldwide, demand-side cost sharing, i.e., when a patient pays partly out of pocket for the medical treatment, is present. Therefore, it is highly relevant to examine whether the patient's consumption opportunities influence the physician's choice of medical treatment.

In Paper II in this thesis, we provide new evidence on physicians' concern for patient welfare under demand-side cost sharing. Theoretically, we show that the optimal calibration of physician payment mechanisms depends on whether or not physicians ignore the influence their medical decisions have on patients' consumption opportunities. Specifically, under demand-side cost sharing, when a physician is concerned about patient utility, the optimally calibrated payment mechanism has a smaller fixed-payment component and a larger fee-for-service component, compared to the optimal calibration when physicians ignore patient consumption. Empirically, we contribute to the scarce literature on the influence of patient co-payment on physician behavior (Hellerstein, 1998; Lundin, 2000; Lu, 2014; Hu et al., 2017). We employ an incentivized lab experiment approach, and find strong evidence that future physicians care about their patients' consumption opportunities. This result also opens a new discussion about the interpretation of results from the RAND Health Insurance Experiment (Newhouse, 1974) and implies that the actual response to demand-side cost sharing from this study might come from both patients and physicians.

Other sources of motivation

Factors other than profit and patient benefit might motivate physician behavior. Some are independent of income and patient welfare, and some interact with them or might even indirectly determine physicians' income and patient utility. While here I discuss a couple of examples of Z in function (2.1), their relevance is highly contextual.

A non-pecuniary element in the physician's objective is what Bénabou and Tirole (2006) refer to as "recognition by others" or "social norms." Due to the information asymmetry, physicians might not make choices that are in patients' best interest. For physicians who value social norms, it is plausible that revealing their performance could encourage them to make better decisions on behalf of the agents since such disclosure reduces information asymmetry. One way to disclose the performance is to routinely collect and report data on physicians' activity and performance.[2] This approach relies on advances in the infrastructure of health registers. Another commonly used form of information disclosure is *auditing* which records performance data only over a specified period of time (Ivers et al., 2012). Most studies focus on the effects of auditing when combined with other measures, for instance, reminders (Eccles et al., 2001; Östervall, 2017), feedback (Baker et al., 2003; Godager et al., 2016), and education (Kerse et al., 1999).

Identifying the effect of information disclosure may contribute to the development of valuable policies. However, few studies are able to disentangle the effect of change in information regimes from the effect of financial incentives or other measures. This motivated us to design the field experiment in Paper I. We investigate whether preannouncement of an audit affects physicians' prescribing behavior.

Examples of other elements that might influence physicians' treatment choices are physician's specialty (Fowler Jr et al., 2000; Chandra et al., 2011), geographic region (Collins et al., 2002), physician's location of training (Lucas et al., 2010), and autonomy (Lerner et al., 1994). Some also affect physician utility indirectly through interaction with physician financial or patient-regarding incentives. For example, physicians' wages can vary substantially across specialties (Lucas et al., 2010). At the same time, physicians with different specialities choosing different treatment even for the same sickness or same patient might reflect both their genuine beliefs in the chosen treatment in terms of the health benefit and the true benefits as a result of their training and experience in the specific field of medicine (Chandra and Staiger, 2007).

2.1.2 Physician competition

Much of the analysis of physician behavior discussed so far is formulated under the assumption that each physician's decision is independent. In spite of being interesting and important, physician competition has drawn relatively little attention from health economists (Gaynor and Vogt, 2000; Gaynor and Town, 2011). Traditionally, health

[2] In some health care systems, the performance data is also linked to financial incentives for the purpose of quality improvement. For example, the Quality and Outcomes Framework in the United Kingdom (Gillam et al., 2012).

markets were characterized with little concentration and large product differentiation, and hence, meaningful and efficient competition was argued to be almost implausible. However, there are reasons to consider the essential role of competition in physician health service markets. In some institutional settings, using the US as an example, there might be substantial costs of entry to the market because physicians must be members of an insurer's network to serve the enrollees (Gaynor and Town, 2011). In Australia, general practitioners are not subject to government regulation of fees, so that they can compete on prices. In particular, they can bulk-bill a large fraction of patients and charge no co-payment (Gravelle et al., 2016). Recently, patient choice reforms, for example, in the UK and Norway, have encouraged competition driven by patient choice among private physicians and hospitals (Cooper et al., 2011).

Recent theoretical work that takes into consideration physician competition provides mixed implications on its impact. Allard et al. (2009) show that in a dynamic framework with certain conditions, competition plays a socially beneficial role in inducing sufficient incentives for physicians to provide services consistent with their patients' desire. On the other hand, Dulleck and Kerschbamer (2009) argue that competition does not necessarily enhance welfare in the case when experts (who can give both diagnosis and treatment) reduce their effort in performing diagnosis to prevent consumers from switching to a discounter (who only provides treatments) at a lower price.

A large body of empirical studies has focused on the effect of competition in hospital or insurance markets, while the literature on physician competition is much smaller. A main challenge in the empirical study of competition is that market structure is endogenous. In other words, the observed effect of competition could be biased as a result of unobserved characteristics that affect the degree of competition and outcome variables, such as service provision and quality of care. While the instrumental variable approach can, in principle, mitigate the problem, the lack of data adds to the challenge. However, several papers deal with this endogeneity problem and provide compelling evidence on the causal relationship between physician competition and behavior. Similar to Kessler and McClellan (2000), Dunn and Shapiro (2014) construct predicted measures of competition, "fixed-travel-time HHI," to mitigate endogeneity due to higher-quality physicians attracting a higher proportion of patients. They find that physicians in more concentrated markets charge higher prices. A paper by Gravelle et al. (2016) also investigates the effect of competition on general practitioners' consultation prices in Australia. They measure degree of competition by the distance between GPs, and they use variations within areas to account for the endogeneity of GP location decisions. Their results show that GPs with more distant competitors charge higher prices, and a smaller proportion of their patients make no out-of-pocket payment. A recent paper by Brekke et al. (2019) focuses on the impact of competition on physicians' service provision. They address endogeneity issues nicely by exploiting physicians who work in both their own practice (with competition) and the emergency center (without competition). They conclude that physicians are more likely to certify sick leave at their own practice than at the emergency center, and further, the competition effect is reinforced by physicians' financial incentives.

Addressing the peculiarities of physician market competition represents an important avenue of research. With the challenge of a lack of data and an endogenous market structure, more innovative approaches to collecting data and designing studies are called for (Gaynor and Town, 2011). The experimental method potentially serves as a good complement to studies using registered data. However, experimental health economics is at its very infancy in the topic of physician competition. Paper III in the thesis adds to this small literature by exploiting an experimental design that facilitates exogenous changes in market structure.

2.2 Stochastic choices

Consider an individual who chooses one alternative from a finite set of mutually exclusive alternatives. In neoclassical economics, it is assumed that the individual has a deterministic utility function and perfect processing capacity to rank the alternatives in a consistent way. These assumptions lead to the prediction that the individual makes the same choice in repeated situations. However, McFadden et al. (1999) noted that this assumption of human behavior is highly restrictive. As Tversky (1972b) noted, "When faced with a choice among several alternatives, people often experience uncertainty and exhibit inconsistency. That is, people are often not sure which alternative they should select, nor do they always make the same choice under seemingly identical situations." Thus, models for analyzing stochastic choices were developed to accommodate the observed inconsistency in behavior. Under the framework of stochastic choice, an alternative is not chosen with certainty, rather with some *probability*. The theory of stochastic choice provides a structure of the probabilities that can be justified from behavioral arguments.

Before the 1960s, economists had already started to consider individual idiosyncrasies in theoretical work. However, when the theory was applied empirically, this complication was mostly neglected. Instead, the concept of a representative agent, who represents the mean behavior of the population, was largely employed in empirical studies of market demand (McFadden, 2001). A representative consumer was often modeled to maximize a deterministic utility function of a vector of different goods at various levels contingent on a budget constraint. Any deviation from the implied behavior by the representative agent was, thus, formulated as an additive error term. The deviation was often attributed to measurement errors in the observed data.

While rooted in psychology, the stochastic choice models have seen rapid development in economics since the 1960s, accelerated by the availability of data on individual behavior and computational advances. Instead of treating them as ad hoc disturbances, economists started to model and interpret the heterogeneous behaviors of a population and varying preferences of an individual. The stochastic choice models provide a structure of the choice probabilities that can be justified from a behavioral perspective. One prevalent type of stochastic choice models is the *Random Utility Maximization* (RUM) model, named by Marschak (1960). He introduced Thurstone's (1927) psychophysical model into economics and presented the model with an individual's utility that contained a random term. The deterministic part of the utility was assumed to be a function of observable variables, such as individual characteristics and properties of possibly individual specific set of choice alternatives. Later, McFadden (1974) developed and popularized the application of an econometric presentation of the random utility model. The most famous application probably was McFadden and his colleagues' work on predicting the number of people who would ride the new BART train in the San Francisco Bay Area (McFadden et al., 1977). The official government prediction was 15%. McFadden's model predicted 6.3%, which was much closer to the actual number of 6.2%. Thereafter, stochastic choice models have seen a rapid development in theoretical and empirical literature in many fields of economics, especially in transportation, marketing, environmental valuation, and labor economics.

2.2.1 Interpretations of stochastic choices

A large body of models has been developed for analyzing stochastic choices. Historically, they were developed by both psychologists and economists, so they differ in the interpretation of the random mechanism that determines the stochasticity of the choices. These models are conventionally categorized into two families according to the random mechanism (Anderson et al., 1992; Kjær, 2005).

The first family assumes that the *decision rule is stochastic*, whereas the utility is deterministic, hence attributing a stochastic element to the decision process (e.g., Luce, 1959; Tversky, 1972a,b). The first presentation of such a model was by Luce (1959), who applied an axiomatic approach and showed that when some axioms are satisfied, the probabilities can be derived from the deterministic values that are defined over alternatives. Another model in this family was proposed by Tversky (1972a,b), and the choice of an alternative is analyzed as a stochastic process of successive elimination by aspects.

For the other family, the decision process is deterministic, whereas the *utility is stochastic*; hence attributing a stochastic element to the utility (e.g., Thurstone, 1927; Manski, 1977). The first model was explained by Thurstone (1927), based on his interpretations of results from psychological experiments. He found inconsistent responses from individuals when they were asked to rank objects in terms of weights or tones in terms of loudness. He interpreted this as variation resulted from comparison of the realization of random variables assigned to the alternatives. Later, economists (Marschak, 1960; McFadden, 1974; Manski, 1977) embraced such random utility model with an econometric representation. However, the point of departure is conceptually different. From an econometrician's point of view, in line with neoclassical consumer theory, the utility is deterministic to the individual. The stochastic feature of utility is thus introduced not to capture uncertainty of the individual but to reflect the lack of available information on individuals and alternatives to the researcher.

In practice, one can adopt either of the two interpretations or both because identical mathematical presentations can be derived for both (Marschak, 1960; McFadden, 1974). However, it is worth noting that the "error term" in the stochastic choice models is not only a micro-econometric representation, but also has its theoretical root in psychological approaches to decision-making.

So far, the two classes of models discussed above take the point of departure that the stochasticity in choices occurs involuntarily. Alternatively, stochastic choices can also be derived from a third possible interpretation. Models in this third class postulate that the stochasticity in behavior is a deliberate choice of the individuals as they desire to randomize. In other words, in some contexts, individuals might *choose* to make different decisions even from the same menu. Swait and Marley (2013), following a similar line as Machina (1985), conceptualize this stochasticity in choice as the result of pursuing multiple goals simultaneously. They later illustrate this motivation by an example of individuals' vegetable consumption behavior and suggest that individuals have an underlying propensity for variety-seeking in addition to the systematic component of utility (Wallin et al., 2018). Some experimental evidence supports such deliberate stochasticity. For example, Agranov and Ortoleva (2017) conduct an experiment in which subjects are asked the same questions repeatedly multiple times, and even when they are aware of the repetition, a majority of them seem to explicitly decide to report different answers to the questions. A recent contribution from Cerreia-Vioglio et al. (2019) formalizes the intuition and axiomatically develops a general model of stochastic choice over lotteries as the outcome of a deliberate desire to randomize.

The studies in this thesis employ the models derived from the first two interpretations. Even though the "deliberately stochastic" interpretation has its appeal in some contexts, for example, choices of consumer products, lotteries, and investment portfolios, it appears less plausible in the context of medical decision making. Nevertheless, exploring the nature of stochastic choices remains an interesting and important avenue of research.

2.2.2 Stochastic choices in strategic games

In a strategic scenario, the payoff to a decision-maker not only depends on his own action, but also his opponent(s)'. The game theory literature has seen a great endeavor in explaining and predicting strategic behaviors. The Nash equilibrium (NE) has undoubtedly been pivotal to the development of theoretical predictions of games. However, in empirical applications, the strict assumption of NE often does not fit with the observed behavior (Goeree and Holt, 2001).

In their seminal contribution, McKelvey and Palfrey (1995) proved the existence of a quantal response equilibrium (QRE). It generalizes the Nash equilibrium by incorporating perspectives from stochastic choice theories that allow for errors in the decision making process. Unlike NE, in which decision-makers are assumed to be perfect maximizers of the expected payoff, QRE assumes that decision-makers alter their decisions in anticipation of their own and others' mistakes and maximize a linear combination of the expected utility and noise. As a result, the alternative that maximizes the expected payoff is not chosen with probability of one, and the sub-optimal alternatives are chosen with non-zero probabilities. In addition, the probability of any alternative being chosen is positively related to the payoff from that strategy. A QRE is the statistical generalization of a NE (Camerer, 2011b) since it converges to a subset of NE in the limit as the noise weight diminishes.

QRE has been applied in many classic types of games and demonstrated great capability of predicting behaviors (Anderson et al., 2001; Goeree et al., 2005; Goeree and Holt, 2005; Goeree et al., 2010; Matějka and McKay, 2015; Wright and Leyton-Brown, 2017). However, 25 years after McKelvey and Palfrey (1995), most applications of QRE are still focusing on scenarios where the choice alternatives are characterized by scalar payoffs. Generalizations to payoffs with multiple elements are straight forward, and the practical toolbox of a choice modeler can contribute greatly to analyzing strategic behavioral data from such games. In Paper III, we utilize an existing choice modeling software module to analyze data from strategic games with vector payoffs. We propose a simple two-step estimator in estimating the preference of multiple attributes and variance of the noise. Not only is this two-step estimator a convenient approach, the results from our Monte Carlo simulations show that it is also accurate even with a moderate sample size.

2.2.3 Sources of preference data

Two common sources of preference data that have been used in choice studies are revealed preference (RP) data and stated preference (SP) data.[3] While RP data in general refer to the observation of actual choices from real markets, SP data contain hypothetical decisions from controlled experiments and surveys (Louviere et al., 2000).

There are advantages and disadvantages to both sources of data.[4] Although RP reflects real choices and contributes to external validity, some obvious challenges of such data include: the key explanatory variables lack variation and can be highly collinear; the choice sets, attributes of alternatives and individual characteristics are difficult to reverse-engineer; and new products are not traded in the market. In contrast, data from carefully designed and conducted SP choice experiments can address some of these issues. In a choice experiment, for instance, a decision-maker is asked to rank or choose from more than one alternative. The alternatives are characterized by several attributes whose levels are varied. Since the choice scenarios are hypothetical, new features and non-existing combinations of features and levels can be included and tested. In addition, the alternatives are constructed in such a way that the decision-maker often needs to make a trade-off between the attributes, which greatly contributes to the richness of the collected information. Nevertheless, the reliability of SP data is constantly debated for the lack of realism and consequent hypothetical biases in some cases (Harrison, 2006).

There is growing literature on combining RP and SP data to correct certain deficiencies in the data source and hence produce a forecasting model to predict future scenarios (Morikawa, 1989; Ben-Akiva and Morikawa, 1990; Hensher and Bradley, 1993; Louviere et al., 2000). In particular, RP data are collected for the information of current market equilibrium, and SP data contribute most with the information of attribute trade-offs in a wider range of market settings. The practice of combining RP and SP data, in principle, improves the richness and validity of the data, with one implicit assumption that the SP data generation process is well designed (Louviere et al., 2000). A potential

[3] Stated preference is sometimes referred to as stated choice (SC), for example in Louviere et al. (2000).
[4] See Kjær (2005) for a review and comparison of the methods used for collecting RP and SP data.

solution to the hypothetical biases generated from SP experiments closely relates to the salience condition in the induced value theory I discuss in Chapter 4. If real outcomes are attached to the choice alternatives in a clear and salient manner, the subjects will be incentivized to perform an action that is consistent with their latent preferences. For this reason, making choice studies incentive compatible is a natural next step (Harrison, 2006).

3 | Thesis objectives

The general objective of this thesis is to address fundamental research questions on physician behavior by combining structural stochastic choice models and the experimental economic methods. More specifically, the objectives of the papers in this thesis are:

Paper I: The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment

Health care systems in many countries are characterized by the limited availability of provider performance data that can be used to design and implement welfare improving reforms in the health sector. The objective of this study is to investigate whether preannouncement of a performance audit can be an effective measure to reduce overprescribing behavior in primary care in such settings.

Paper II: Exploring physician agency under demand-side cost sharing — An experimental approach

The objective of this study is to contribute to new knowledge of physician behavior in the context of demand-side cost sharing. Specifically, we investigate whether physicians are concerned about their patients' consumption level after the co-payment. If they are, the optimal calibration of physician payment will depend on the level of demand-side cost sharing. It also contributes to the small literature on the identification of physician response to demand-side cost sharing.

Paper III: Predicting strategic medical choices: An application of a quantal response equilibrium choice model

The objective of this paper is to contribute to new knowledge on how market competition affects behavior in a medical setting. We study behavior in a designed competition experiment and apply recent advances in empirical game theory.

4 | Experiments and data

4.1 Economic experiments

4.1.1 Experimental methods in economics

Economics has seen a revolutionary change in the view of the experimental method in the past 40 years. Unlike physics, a traditional example of experimental science, economics was believed to be a non-experimental science and the development of economic methods was exactly an adaptation to the infeasibility of controlled experiments (Bardsley et al., 2010). As reflected in what Milton Friedman (1953) wrote in 1953, empirical studies in economics mostly relied on observational data collected from naturally occurring occasions in the field:

> Unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences. Generally, we must rely on evidence cast up by the "experiments" that happen to occur.

Despite this general perception of methodology in economics, a few landmark contributions initiated and strongly influenced the growth of contemporary experimental economics[5]. One example is L.L. Thurstone's (1931) investigation in a "very old problem that overlaps economic theory and psychophysical experimentation." In his report, Thurstone, a leading psychologist of his time, attempted to elicit individuals' indifference curve from their response to binary choices in an experiment. He described the use of experimental methods in determining the economic concept that previously had no direct empirical grounding. Even though the report was published in a psychology journal and received critical reviews from economists (for example, Wallis and Friedman, 1942), its innovative use of experiment sparked subsequent experimental investigations in economic behavior and preferences (Mosteller and Nogee, 1951;

[5] Roth and Kagel (1995, Chapter 1) provide a detailed description of this.

Allais, 1953; Davidson and Marschak, 1959). Starting with this strand of individual choice experiments and the seminal publication of *Theory of Games and Economic Behavior* by Neumann and Morgenstern (1944), another series of experiments involving interactive behavior was brought under focus. An early example is the experiment conducted in 1950 at RAND Corporation, in which the prisoner's dilemma game was formulated (Flood, 1958). The discussion of this experiment among economists stimulated collaboration between game theorists and experimenters and inspired work that looked at experimental design issues in more general interactive games (for example, Kalisch et al., 1954). Since then, the appearance of "strictly planned experiments" was much appreciated (Morgenstern, 1954), and the concern of the rule of the games was brought under considerable attention. Another contribution that has influenced modern experiments focuses on industrial organization and is represented by Edward Chamberlain's (1948) attempt to construct experimental markets and test its efficiency. Since then. his technique has been employed by many and stimulated later investigations concerning duopoly and oligopoly market behavior (Sauermann and Selten, 1960; Siegel and Fouraker, 1960). In particular, it is exactly his experiments that impressed a participant back then, Vernon Smith, and sparked Smith's endeavor in experimental economics.

Even with these landmark contributions, it was not until the 1980s that the majority of economists were convinced of the validity and value of experiments. Experimental economics started to be seen as "an exciting new development" (Samuelson and Nordhaus, 1992), even among those who had most doubts about it. Economics has since seen a rapid growth in the use of experimental approaches, and its significance received public recognition when the Nobel Prize in Economic Sciences was awarded to experimental contributions: Vernon L. Smith "for having established laboratory experiments as a tool in empirical economic analysis" in 2002 (NobelPrize.org., 2002), and Abhijit Banerjee, Esther Duflo, Michael Kremer "for their experimental approach to alleviating global poverty" in 2019 (NobelPrize.org., 2019). Nowadays, experimental economics is assuredly an important part of the discipline and a major source of knowledge in the social sciences (Falk and Heckman, 2009).

The experimental approach is employed as a tool in a very large body of successful economic research. For the purpose of this thesis, it is of great interest and relevance to provide some insights into the development and applications of a broad range of experimental approaches in health economics, especially those that deeply inspired our experimental studies.

The most famous and influential field experiment utilizing the randomized control trial (RCT) method in health economics is the RAND Health Insurance Experiment (HIE) on health care costs, utilization, and outcomes in the United States (Newhouse, 1974). In the experiment, 5809 people were randomly assigned to insurance plans with no cost sharing, 25%, 50%, or 95% co-insurance with a ceiling of an annual payment of 1000 US dollars. The main results indicate a significant effect of demand-side cost sharing on the utilization of health care services, but the effect on health status varies upon types of medical services and socioeconomic status (Keeler et al., 1985; Lohr et al., 1986; Manning et al., 1987; Lurie et al., 1989; Newhouse, 2004). RAND HIE has been referred to as the "gold standard" in research on effects of health insurance, and the research question in Paper II is partially motivated by this landmark experiment. The RCT type of field experiments has since been employed in many investigations of various health-related questions. To name a few, RCT experiments on the effect of public health insurance on healthcare utilization (Finkelstein et al., 2012; Taubman et al., 2014) and clinical outcomes (Baicker et al., 2013), effects of auditing combined with other measures on provision and quality of health services (Feder et al., 1995; Baker et al., 1997; O'Connell et al., 1999; Eccles et al., 2001; Kiefe et al., 2001; Currie et al., 2011, 2014; Lu, 2014; Östervall, 2017), and determinants of individual health behavior (Charness and Gneezy, 2009; Volpp et al., 2009; Milkman et al., 2011; Bronchetti et al., 2015; Carrera et al., 2018; Halpern et al., 2015; List and Samek, 2015; Belot et al., 2016). The experimental design in Paper I benefits from the RCT design and investigates the effect of the announcement of auditing on prescribing behavior.

Controlled laboratory experiments got off to a late start in health economic studies, only about a decade ago, (Cox et al., 2016) despite the fact that leading health economists have advocated its value in complementing traditional methods in healthcare research (Fuchs, 2000; Frank, 2007). One early application is Hennig-Schmidt et al.'s (2011) investigation of physicians' responses toward different payment schemes. They compared medical students' supply of medical services under a fee-for-service (FFS) and a capitation (CAP) payment and found physicians provided significantly more services under FFS than CAP. Although the subjects were affected by the payment system, the results indicated that patients' health benefits also played an important role in future physicians' decision-making. The novelty of this experiment is the salience of the "patient benefit" accumulated in the lab, and it is achieved by linking it to a charity donation to treatments for real cataract patients outside the lab. Subsequent lab experimental studies on physician altruistic behavior largely built on Hennig-Schmidt et al.'s design and extended it for analyses of various aspects of physician motivation and behavior (Godager and Wiesen, 2013; Hennig-Schmidt and Wiesen, 2014; Godager et al., 2016;

Brosig-Koch et al., 2017a; Wang et al., 2020). The two lab experiments (Papers II and III) in this thesis build on the protocol in Hennig-Schmidt et al. (2011) and update with some innovative adjustments to suit the specific research questions of interest. Controlled lab experiments have seen a growth in applications addressing other topics in health as well, such as patient-physician interaction (Huck et al., 2016), choice of health insurance (Schram and Sonnemans, 2011; Buckley et al., 2012; Kairies-Schwarz et al., 2017), determinants of the provision of health services (Mimra et al., 2016), and time and risk preferences in health behavior (Anderson and Mellor, 2008; Arrieta et al., 2017).

Another strand of experiments that has been increasingly advocated and widely employed in eliciting preferences of health care providers and patients is discrete choice experiments (DCEs). DCEs involve asking individuals to select preferred alternatives over hypothetical and specially constructed choice scenarios. Choice alternatives in DCEs are described by several attributes, and individuals' responses reveal whether preferences are affected by the attributes and their relative importance. The use of this method dates back to the 1970s in transportation (Train, 1978) and marketing (Train, 1986; Train et al., 1987) research, and the data analysis is grounded on a well-established theoretical basis for discrete choices (McFadden, 1974). In the past couple of decades, DCE has seen its growing popularity in the health domain and is employed in a wide range of topics, such as valuing health outcomes and experiences, eliciting health care providers' preferences and job choices, and developing priority setting frameworks (see Ryan and Gerard, 2003; de Bekker-Grob et al., 2012; Clark et al., 2014; Soekhai et al., 2019, for reviews of the applications). Despite its usefulness and advantages, DCE has received criticism for its lack of real incentives and, thus, real behavior (Galizzi and Wiesen, 2018). To address this concern, in the DCE-inspired experiment in Paper II, we attach separate monetary incentives to all three choice attributes to ensure the salience. In this way, medical students' choices in the lab have real consequences for a real patient in the hospital in terms of consumption opportunity and treatment payment. To the best of our knowledge, this is the first laboratory experiment of its kind.

4.1.2 Purposes and advantages of experiments

Economic experiments serve many purposes for which different designs are employed. Alvin Roth (1986) classified the purposes as the following:

1. Speaking to theorists: testing and modifying formal theories;

2. Whispering into the ears of princes: providing inputs for policymakers;

3. Searching for facts: detecting interesting phenomena and unanticipated regularities.

Even though his classification was made for lab experiments, it seems suited for experiments in general. In addition to these purposes, there is growing interest in eliciting individual preferences (such as willingness to pay for public goods) and measuring behavioral parameters (such as risk or rationality parameters). Economists also find it useful to use lab experiments or small-scale field experiments as "test bed" before introducing policies or interventions in the field. Last but not least, the early economic experiment recorded by Chamberlin (1948) had an important pedagogical purpose. As with the evolution of experimental economics, it is a natural trend to use experimental demonstrations to illustrate economic propositions in schools (Friedman et al., 1994).

One may have noticed that the purposes of experimental studies in economics are not radically different from observational studies utilizing happenstance data collected from the field.[6] Why do economists propose controlled experiments?[7] The answer relates to the challenges in observational studies, one of which is a matter of identification. Consider a set of happenstance data of which the data generating process is uncontrolled. If an outcome variable Y is always associated with a variable X, without control, one cannot make a confident causal conclusion. This is because the observed correlation could be due to the direct effect of X on Y, or some unobserved factor Z that affects both X and Y. This is a well-known identification challenge faced by economists in empirical studies utilizing naturally occurring data. Thus, their main aim is to determine the set of assumptions that best describe the unknown data generating process and therefore identify the causal effects of the treatment. Experiments, on the contrary, allow the experimenter to decide on some elements in the data generating process in accordance with the research question of interest. In the aforementioned example, if individuals' characteristics and the properties of the institution where individuals act are well controlled in an experiment, one can conclude with confidence that any change of X is *exogenous* and thus *causes* changes in Y. In other words, controlled experiments facilitate *ceteris paribus* inferences. A controlled experiment can also make it possible to measure and therefore eliminate the confounding effect of elements in Z.

4.1.3 Control in experiments

A controlled environment is essential for achieving the objective of an experiment. Here, the *environment* is often referred to as individual *subject* and an *institution* through

[6] See Fig. 1.2 (p.4) in Friedman et al. (1994) for examples of data sources.

[7] Jacquemet and l'Haridon (2018), in Chapter 3 of their book, provide a comprehensive illustration of some identification challenges in empirical economics that can be addressed by controlled experiments.

which subjects act (Friedman et al., 1994). An institution specifies the framing and rules of the experiment, such as the type of game, possible actions and corresponding outcomes for the subjects, sequence of the actions, information conditions. Control over the institution in principle can be achieved by a clear explanation of the experiment to the subjects and strict enforcement of the rules.

Meanwhile, controlling for the subjects' characteristics is more challenging. Subjects usually have their own characteristics that might not be in accordance with the presumed specifications of the experiment, for example, initial resource endowment and access to technology. In the lab experimental context, these can sometimes be held to a level that is compatible with the experiment. However, if it is not feasible, other techniques, such as randomization, need to be put in place to minimize the confounding from subjects' innate characteristics. Subjects' preferences are latent but essential to control. In Vernon Smith's seminal paper on induced value theory (Smith, 1976), he identifies sufficient conditions to achieve a valid control of subjects' preferences. He points out that under several conditions, the experimenter can induce subjects' preference by proper use of a reward medium, typically monetary payment. In other words, the subjects are incentivized to act consistently with their latent preferences. Smith (1976) provides practical advice which has long guided the experimental design in economics.

We now summarize conditions identified by the induced value theory following Smith (1976), his extended discussion (Smith, 1982), and presentation in Friedman et al. (1994) (p.12-15):[8]

Monotonicity and nonsatiation

Subjects prefer more reward to less. If $V(m,z)$ is the subject's unobserved preference over the reward m and everything else z, the condition implies

$$\frac{\partial V(m,z)}{\partial m} > 0, \text{ for every feasible combination of } (m,z).$$

Smith (1976) suggests using local currency to achieve this condition. This is employed in Papers II and III.

Salience

The reward received by the subject, Δm, is associated directly with the alternative chosen by the subject (and other subjects in a strategic game). The relation between the choice alternatives and the reward is clearly stated and explained to the subjects. This idea of

[8] Friedman et al. (1994) illustrate how fulfilment of these conditions enables us to induce subjects' preferences using a simple model.

salience distinguishes controlled incentivized economics experiments from other methods, such as surveys or interviews, and potentially increases the validity of the results. Using the experiment in Paper II as an example, the reward is directly linked to each subject's decision. Based on a randomly drawn decision by the subject, the amount equal to "Your profit" was paid to the subject; the money corresponding to the "Health benefit for the patient" was transferred to the patient's in-hospital-account for medical treatment use, and the amount of cash equal to "Money available to the patient" was given to the same patient for his own disposal. In contrast, the 25 Yuan (3.77 USD) fixed show-up fee is not salient. The purpose is to barely compensate subjects' time for participating.

Dominance
The reward, m, dominates any other influences, z, that might change subjects' utility. Since z captures everything else and in almost all cases is latent, this condition is most challenging. First of all, researchers can in practice increase the amount of the reward so the utility from other influences is negligible. In addition, dominance is more plausible once the most obvious other influences are held fixed. In general, researchers can apply different measures to mitigate the effects of z on the subjects' utility. For instance, researchers can hold private information of each subject's reward if it is assumed that subjects' decisions are influenced by rewards earned by others (and if it is not the research question of interest). Another practical approach that suits most of the experimental settings is to neutralize the experimental description to avoid any researcher-induced behavior. To satisfy the dominance condition, specific procedures were employed in my papers. In Paper II, for example, patient's initial endowment was fixed for each decision scenario to alleviate its influence on the subjects' utility. To avoid the influence of others' decisions, in both Papers II and III, the information regarding subjects' choices and rewards was kept private, and in Paper III, the matching of players was random and kept confidential from the subjects.

4.2 Experimental design

The extent to which a study permits causal inferences is often referred to as the internal validity. For an experimental study, internal validity requires proper experimental design and data analysis. In this section, we describe some best and common practices in experimental design and discuss their rationale. Chapter 5 takes up the issue regarding data analysis and empirical models.

Two main aspects of a good experimental design facilitate identification (Jacquemet and l'Haridon, 2018). The first is a well-controlled decision environment, i.e., individual subjects and the institution through which subjects act. In other words, the parameters that characterize the participating subjects and institution need to be carefully controlled to sharpen the actual effect of treatment variable on the behavioral outcome. However, this is not enough, as the behavior generated in an experiment results from the subjects' perception of the experiment. For example, if the payment procedure information is not well-conveyed to the subjects, the reward might fail to incentivize behavior that reflects the latent preferences. As a result, controlling how subjects perceive the experiment is an equally essential part of the experimental design.

It is worth emphasizing that the choice of an experimental design and the trade-offs of aspects should be made to serve the specific research question. In other words, some variations are less likely to confound the effect than others, depending on the questions to be addressed. For example, if the investigation is about the effect of patient cost sharing on physicians' medical treatment decisions, the color of a patient's clothes is clearly less important to control for than, for instance, physician's profit and patient health benefit from the treatment. The list of unobservable confounders is endless; thus, testing all one after the other is pointless. As a result, there is never a perfectly controlled experiment, nor a universal experimental design everyone should comply with. The best practice is therefore to carefully choose the design that sharpens the proper identification of the relevant variables and minimizes confounding due to other variables of little or no interest (Chap.3 in Friedman et al. (1994)).

4.2.1 Variables in experiments

Variables that we are interested in the effects of are often called *treatment* or *focus* variables, while other variables that might or might not confound the main effect are referred to as *control* or *nuisance* variables (Friedman et al., 1994). Identification of treatment effect is achieved if the nuisance variables are uncorrelated with the treatment variables or the confounding effects from any nuisance variable are measured and eliminated. In the following, we describe some common design techniques (summarized from Friedman et al., 1994; Louviere et al., 2000; Moffatt, 2015; Jacquemet and l'Haridon, 2018) and how they are employed to achieve identification, with examples from the three essays.

Direct control

The simplest way to control a variable directly is to hold it constant at a reasonable level throughout the whole experiment. It is the most straightforward approach to generating experimental data and thus providing direct control over the experimental environment.

In general, variables that can be directly controlled, to name a few, are the exchange rate of the experimental currency, the type of the subject pool, the values of parameters in the game, and the rules of interaction among subjects. In the context of the competition game in Paper III, the market demand function is predetermined and kept constant in the experiment. The demand function provides a clear description of competition in the market by specifying the effect of a physician's decision on his demand of patients and thus his payoffs (physician profit and patient health benefit). Outside the lab, physicians' payoffs are often private information and unobservable. Direct control in the lab in this context, to the contrary, allows us to include payoffs of both chosen and non-chosen alternatives and hence facilitates the identification of treatment effects.

It is a trade-off between keeping a variable at a constant level (i.e., considering it as a nuisance variable) and varying it (i.e., considering it as a treatment variable). The more variables one holds constant, the cheaper and simpler the experiment becomes, but less knowledge is gained on the direct effects and interactions of the variables. The decision should be made based on consideration of several aspects, among which the research purposes, the potential correlation between nuisance and treatment variable, the number of observations needed to achieve statistical power are important. The research question we address in Paper II is whether physicians are concerned about how their choices of medical treatment affect their patients' consumption opportunities. We are also interested in quantifying the relative weights of patient consumption compared to physician profit and patient health benefit in physician's utility, and these three variables might correlate. Hence, these three variables are taken as treatment variables in the experiments while other nuisances are held at a constant level, for example, patients' initial endowment. One may argue for the relevance of the patient's wealth as a treatment variable, but we resisted this temptation due to the increased requirement of number of observations and the boredom or fatigue that might result from a more complicated and lengthy experiment.

Randomization

Not all the nuisance variables can be controlled or even observed by the researchers. The simplest technique to avoid confounding problems is a completely randomized design in which subjects are assigned to groups of treatments and control at random. The implementation is referred to as "between-subject" design in some experimental designs. By definition, this design breaks down the correlation between treatment variable and subject characteristics. However, it does not rule out the possibility of noise in the data and imprecision in the statistical analysis if variation of nuisance is large and the distributions of noise in treatment and control are not identical. For this reason, this type of randomization is demanding in terms of sample size.[9]

In the field experiment in Paper I, we employed this design and randomly assigned clinics to a treatment or control group. The mystery shopper intervention was only implemented in the treatment group. Hence, the chance that the intervention is correlated with clinic- or physician- specific characteristics, such as the location and size of the clinics, and the age and gender of the physicians, was minimized. The sample size was chosen to have the power sufficient for detecting a 30-percentage-point effect size. The randomization technique was also applied in the assignment of pseudo patients to clinics, such that characteristics of the patients do not confound the intervention effect. For similar reasons, subjects in Papers II and III were randomly assigned a seat (by drawing a seat number card upon entering the experiment) such that they do not choose the decision booklet, where to sit, and whom their neighbors are. In Paper III, subjects were also randomly matched to groups of two and four in duopoly and quadropoly markets, respectively.

Blocking

When variation in nuisance variables is inevitable, an alternative strategy is to divide subjects into blocks by their characteristics. In this way, the nuisance is held constant within the block, and since it no longer varies, it is no longer confounding. Thereafter, randomization can be performed within the blocks. One typical example is to block on gender, as it can be an important source of variation in the outcome. By blocking on it, we increase the precision of the analysis of the main intervention effect.

Another specific case of blocking is a "within-subject" design in which the same experimental subject experiences more than one treatment, one after the other. The unobserved individual heterogeneity is, in this way, held unchanged across interventions, and hence its confounding effect is eliminated. This design is employed in Paper III as every subject is asked to make medical decisions in all three market environments, namely, monopoly, duopoly, and quadropoly. To avoid the confounding effects due to the order in which the market environments are implemented,[10] the subjects were randomized into different orders.

[9] List et al. (2011) provide a detailed discussion on this issue.
[10] Known as the "order effects."

Multiple treatment variables and optimal design

In many cases, researchers are interested in the effects of more than one treatment, each with several discrete values or levels. Consider an example with three treatment variables; each takes two levels. In total, there are $2 \times 2 \times 2 = 8$ possible combinations of these levels across all treatment variables.[11] *Full factorial design* is a design in which a complete enumeration of these combinations is represented (Louviere et al., 2000). Full factorial design guarantees that the effects of all variables are truly independent. However, it requires a large sample size of experimental subjects or tedious decision-making of all treatment combinations for each subject.

As the number of variables and levels increases, it is necessary to reduce the number of combinations implemented in the experiment. This is often done by *fractional factorial design*, which contains a subset of the complete set of combinations. Rather than random selection from the complete factorial, statistical sampling methods have been developed for selecting fractional designs. Nevertheless, in general, fractional design involves some loss of information, and sometimes it limits the ability to study effects of interactions between two or more attributes (Louviere et al., 2000).

Researchers often consider the *efficiency* of the experimental design. From a statistical perspective, optimal or efficient design means a design that gives maximal precision in the estimation of the parameters (Moffatt, 2015). In a typical choice experiment, there are three steps of experimental design in which efficiency should be taken into consideration (Kjær, 2005). The first step is the selection of alternatives (i.e., combinations of levels of variables) if a full factorial design is not feasible. The second step involves pairing the alternatives into choice sets. It should be done in a way that there exists some trade-off between alternatives (i.e., no dominant alternatives), and the differences in attribute levels for each choice set are not correlated. The last step involves dividing choice sets into multiple blocks. In the situation when the subjects are not presented with all the choice sets, the choice sets need to be split into blocks that retain their statistical properties.

For a pre-specified model, an efficiency criterion results in minimizing the generalized variance of the parameter estimates, corresponding to maximizing the information.[12] D-efficiency criterion, among others, is a common efficiency measure used to optimize

[11] Also referred to as experimental conditions, experimental treatments, runs, or trials (Friedman et al., 1994).

[12] This is because common measures of efficiency are based on the information matrix, and the variance-covariance matrix of the vector of parameter estimates is proportional to the reciprocal of the information matrix. Kuhfeld (2005) provides a pedagogical illustration of efficiency designs.

the design. It maximizes the determinant of the information matrix. There has been an increased interest in using D-optimal design in choice experiments due to the availability of software modules such as `dcreate` (Hole, 2017) in Stata, and its robustness even with biased priors (Carlsson and Martinsson, 2003).

Consider the experiment in Paper II; we are interested in effects of three variables: physician profit, patient health benefit, and patient consumption, and each takes eight levels. A full factorial design generates $8 \times 8 \times 8 = 512$ treatment combinations (choice alternatives in our application). We chose to have two alternatives in a choice set. We utilized `dcreate` to obtain a D-efficient fractional design with four blocks and 23 decision scenarios in each block.

4.2.2 The perceived experiment

Regardless of the experimental design, behavior of the subjects only reflects what they think they are involved in. In explaining the researchers' idea of design to the subjects, it is essential to have the experiment's instructions in place. These instructions are the primary information source of the experimental procedure, and they are usually presented on a printed sheet of paper distributed to the subjects at the beginning of the session. The way the instructions is written and communicated has two goals: each and every subject understands the experiment well, and their understanding of the experiment is homogeneous (Jacquemet and l'Haridon, 2018). The former ensures the consistency between the experiment designed by the researchers and the one perceived by the subjects, and the latter mitigates any noise that would be introduced to the outcome behavior from subjects' various understanding of the experiment.

These two goals have implications on practical matters in communicating experimental design to the subjects. First of all, the instructions should be formulated in a simple and clear manner so they are accessible to the least-talented subject one can imagine. In practice, for example, jargon should be avoided, and sentence structures should be simple. Secondly, providing examples assists in getting across the experimental design through their concrete applications (see Paper II for an example). Another rule of thumb is to use control questions which are asked to be answered by the subjects and checked and commented on by the experimenter before the start of the experiment (see Papers II and III). Lastly, to ensure subjects' homogeneous understanding of the experiment, it is wise to write the instructions using "neutral" words, and the same instruction is communicated and distributed to each and every subject.

4.2.3 Summary of the three experiments

Paper I: The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment

Using a randomized treatment-control design, we conducted a field experiment in primary care clinics in a Chinese city. We investigate whether informing physicians of a forthcoming mystery shopper audit influences their prescribing behavior. We sampled for-profit clinics with one practicing physician, and randomly assigned 48 clinics to the control group, and 48 clinics to the treatment group.

Two audits were performed in each clinic. For the treatment group, an intervention of announcing a forthcoming mystery shopper audit was conducted before the second audit. The goal of the first audit was to check randomization, and the second audit was to examine the intervention effect. In the audits, pseudopatients presented symptoms of a minor common cold to the physician according to a script. They allowed the physician to measure their temperature and visually inspect their throat and were instructed to refuse any other treatment or diagnostic test by the physician. After the visit, the pseudopatients filled out a data collection sheet, including data on the prescribed drugs and characteristics of the physicians and the clinics.

The intervention of announcing a forthcoming mystery shopper audit was conducted before the second audit. A representative of the research project visited the clinics in the treatment group one by one to announce the mystery shopper audit. The announcement was made in person by presenting a letter containing information about a quality evaluation of primary care services in Jinan, particularly service, professionalism, and adequacy of treatment. The clinics were informed that an anonymous patient would visit the clinics and collect the quality information.

Paper II: Exploring physician agency under demand-side cost sharing — An experimental approach

In this incentivized laboratory experiment, the participating medical students played the role of a physician, and each made 23 treatment decisions for patients. Decisions were made independently and anonymously. A decision task was to choose from two treatment alternatives for a patient who had an endowment of 50 Chinese Yuan (7.55 USD). The patient did not have full insurance and was passive, which means he had to accept the treatment chosen by the physician. The choice of treatment simultaneously determined the physician's profit, the patient's health benefit, and the patient's consumption level after the co-payment.

There were no real patients participating in the experiment. To induce patient-regarding motives, a procedure similar to that of Wang et al. (2020) was applied. The choices made in the experiment had consequences for a real hospital patient who was randomly chosen from a list of patients. One out of 23 decisions were randomly chosen to determine each participant's payment and the amount of money to be transferred to the patient.[13] The money corresponding to the sum of health benefits provided by all participants in one of the 23 occasions in the experiment was transferred to the hospital account of the patient, ensuring that the money can only be used for medical treatment. At the same time, physicians' choices also determined the co-payment and the amount of money available for patient consumption, and the latter was given in cash to the same hospital patient.

Paper III: Predicting strategic medical choices: An application of a quantal response equilibrium choice model

This incentivized laboratory experiment had a medical framing. The participants were instructed to play the role of a physician and choose medical treatment for eight types of patients in each of three different market settings: monopoly, duopoly, and quadropoly. The treatment choices of participants determined their own profit and patients' health benefit. The profit and patient benefit accrued in the laboratory were converted into monetary transfers to the participants and a charity dedicated to providing surgeries for ophthalmic patients. This element of our protocol, which is identical to Hennig-Schmidt et al. (2011), motivated participants' patient-regarding behavior in the laboratory.

In each market, there is a fixed market demand of 100 patients who seek medical care. Upon each choice occasion, the decision-maker is asked to pick one treatment alternative out of 11 available alternatives. In the monopoly dictator game, the physician serves the whole market. Alternatives that provide more per-patient-benefit result in less per-patient-profit. In the strategic scenarios, duopoly and quadropoly, the payoff matrices reflect a positive demand response from providing benefits to the patients. Under competition, physicians who provide more health benefit to patients (for a given health benefit provided by physicians' opponents) obtain a larger market share. On the other hand, providing more health benefit per patient reduces physicians' profit margins.

In duopoly and quadropoly, participants were randomly matched to groups of two and four, respectively. The joint decisions by the matched group determined the profit

[13] A recent paper (Charness et al., 2016) surveys theoretical predictions and empirical evidence of different payment approaches, and suggests that both methods are effective, in spite of their pros and cons.

and benefit. To facilitate non-cooperative decision-making, the random match was dissolved after the completion of all tasks in each market setting, and the participants were not informed about the outcomes before all tasks in a market were completed. Decision-makers could use a "calculator" in the z-Tree program (Fischbacher, 2007) to see how a combination of players' choices determines the payoffs.[14] In other words, subjects could inspect each cell of the payoff matrix. The participants played in the three markets by different orders.

4.3 Data structure

Table 4.1 presents a summary of data from three experiments. The data contain participants' choices and their characteristics.

In Paper I, each of 94 physicians made choices on prescribing or not prescribing on four types of drugs. In total, 376 decisions were made. Of the participants, 46.8% were male, and 53.2% were female; 48.9% were younger than 40 years, and 51.1% were older than 40 years.

In Paper II, 202 medical students participated in the experiment. They made binary choices on treatment alternatives. Each choice set consisted of two alternatives, and each participant was presented with 23 choice sets. In total, 4645 decisions were made.[15] More females than males participated in the experiments; 63.9% were female, and 35.6% were male. Since all the participants were students, 99% of them were younger than 40 years, and two (around 1%) of them did not provide age information.

In Paper III, 136 university students were recruited to take part in the experiment. They made choice decisions on treatment alternatives. Each choice set contains 11 alternatives, and each participant made decisions on 24 choice sets. Female (49%) and male (51%) represented half of the sample roughly. The majority of the participants (97.8%) were younger than 40 years, while 2.2% were older.

[14] Requate and Waichman (2011) find that the use of a profit table or a profit calculator yields indistinguishable performance.

[15] $23 \times 202 = 4646$. One decision is missing from one participant.

Table 4.1: Summary of data by paper

	Paper I[a]	Paper II	Paper III
Choices			
Type of choice	Binary choice of prescribing or not prescribing	Binary choice of treatment alternatives	Discrete choice of treatment alternatives
No. of alternatives	2	2	11
No. of choice occasions per individual	4	23	24
No. of decisions made	376	4645[b]	3264
Participants			
Type of participants	Physicians	Medical students	University Students
No. of participants	94	202	136
Gender			
Male	46.8%	35.6%	51%
Female	53.2%	63.9%	49%
Unknown	0%	0.5%	0%
Age[c]			
≤ 40	48.9%	99.0%	97.8%
> 40	51.1%	0%	2.2%
Unknown	0%	1.0%	0%

a Only data from the second audit is presented in the table because it was used for the primary analyses of the intervention effect.
b One decision is missing from one participant.
c The age data is primarily used in Paper I, so for the purpose of summarizing the data we follow the categorization in that paper.

4.4 Ethics

Experimental economics is subject to a broad range of general ethical principles, which include the agreement with the laws, rules and regulations, honesty and replicability of the study, appropriate credit, and disclosure of conflicts of interest. In addition, an important area of ethical consideration in economic studies using experimental methods stems from the employment of human subjects. In Norway, studies using personal data with identifiable personal information have to notify the Norwegian Centre for

Research Data (Norwegian: Norsk senter for forskningsdata, NSD) about the project.[16] Beyond the data protection and privacy policy, there are other ethical concerns related to interactions between researchers and human subjects in experimental economics (Ifcher and Zarghamee, 2016). In the following, we summarize and discuss these challenges following the common source of guidelines in the context of our projects.[17]

Protection of participants and their rights

For the lab experiments, the participants are recruited voluntarily before the experiment and are informed that they can withdraw from the project any time before or during the experiment for whatever reason, and without explanation or penalty (Papers II and III). That means participants are free to leave the room whenever they wish to do so. Before the experiment starts, full information regarding the experiment procedure, payment methods, potential benefits and risks, and usage of the data should be provided to all participants. In the experiments in Papers II and III, not only was the information fully disclosed and explained to the participants but also, sufficient time was provided for clarifying questions.

In advance of and during the experiment, all participants and research assistants must be fully informed and protected from physical, psychological, social, legal, and economic harm at all times. This was carefully taken care of in the mystery shopper experiment in Paper I. All participating students, both pseudopatients and accompanying students, underwent a sufficient amount of training and rehearsal to ensure adherence to the protocol. The pseudopatients, recruited from the medical school, had at least one semester of basic medical training and were instructed to refuse any treatment and diagnostic test by the physician except for temperature measurement and visual inspection of the throat. In addition, they were always accompanied by a fellow student during the experiment and were instructed to report with full information should any adverse event occur.

[16] In this thesis, the review and approval was provided by the NSD, with reference number 44243 for Paper I, 53301 for Paper II, and 43709 for Paper III. The Regional Committees for Medical and Health Research Ethics (REC) was also consulted before all three experiments, but they concluded that the cases were not relevant.

[17] For example, the Belmont Report (1979) written by the National Commission for the Protection of Human Services of Biomedical and Behavioral Research in the US (retrieved from www.hhs.gov), and Ethical Principles for the European Economist (2017) drafted by the European Economic Association Ethics Committee (retrieved from www.eeassoc.org).

Deception and informed consent

One of the most-discussed ethical issues in designing and conducting economic experiments is the use of deception. Deceiving the subjects can contaminate the reputation of the lab among students and create a difficult environment for other researchers. Furthermore, the prohibition of deception enhances the validity of the inferences from experimental observations since it ensures that subjects make decisions induced by the monetary incentive instead of the psychological feelings of manipulation (Davis and Holt, 1993). That said, there is no apparent agreement on classifying an experiment as deception or not within social science research (Wilson, 2014). Informed consent is favored out of respect for the subjects' autonomy and to protect their safety. The subjects are usually informed of the nature of the experiment and their right to withdraw from the experiment at any time.

While there is no doubt about the experiments in Papers II and III in this regard, whether the field experiment in Paper I poses ethical concerns is debatable. Firstly, the experiment lacks consent from the audited physicians. It is not uncommon in field experiments for the experimental subjects to be unaware that they are participating in an experiment. List (2008) argues that it is exactly this design of natural field experiments that "combines the most attractive elements of the laboratory and of naturally occurring data: randomization and realism." According to Wilson (2014), in practice, not providing the full information regarding the experiment itself is insufficient to be considered unethical. Therefore, the waiver of consent is grounded to ensure scientific validity. Secondly, during the audit, the pseudopatients strictly follow the script, which does not reflect their true health status. This can be considered as a violation of honesty. Nevertheless, following Rhodes and Miller's (2012) ethical analysis, we argue that it can be justified because the confidentiality of research data is protected, risks or burdens to the subjects are minimal, and the research has potential value to human knowledge. In addition, the field experiment also contributed positively to the revenues of the clinics in the study sample since physicians gained profit by selling prescribed medications.

Privacy policy

To protect participants' privacy, they took part in the experiment anonymously, and their behavioral data was only used for analyses agreed in the protocols approved by the ethical committees. In the experiments in Papers II and III, participants received the payment in private at the end of the experiments. The real patient who received the benefits from the experiment in Paper II was randomly chosen from a short-list of patients provided by Qilu Hospital, and any private information of the patient was strictly kept confidential, even from the researchers. Data and unique ID assigned to the participants are detached at the end of all three experiments.

5 | Models and estimation methods

We start with a generalized multinomial logit model (GMNL) (Fiebig et al., 2010), and proceed to its special cases, which we employ in the three papers.

Consider a sample of N individuals who are faced with a choice set, C, of discrete alternatives. In each of the T choice occasion, each individual chooses one alternative from C. The utility of individual n from choosing choice alternative j at choice occasion t is given by:

$$\begin{gathered} U_{njt} = \boldsymbol{\beta}'_n \boldsymbol{x}_{njt} + \varepsilon_{njt} \\ n = 1,...,N; j = 1,...,J; t = 1,...,T, \end{gathered} \tag{5.1}$$

where $\boldsymbol{x}_{njt}$ denotes the vector of attributes of alternatives, and $\boldsymbol{\beta}_n$ is a vector of individual specific parameters. ε_{njt} are idiosyncratic errors and are assumed independently, identically distributed over individual, alternative, and choice occasion. We assume further that the errors follow a Type I extreme value distribution which has a cumulative distribution function as:

$$F(\varepsilon_{njt}) = e^{-e^{-\varepsilon_{njt}}}. \tag{5.2}$$

The extreme value distribution leads to the closed-form logit formula which is readily interpretable (Train, 2009).[18] It then follows that the probability of individual n choosing alternative j at scenario t takes the following logit formula (see, for example, Train, 2009, pp. 74-75):

$$P_{njt} = \frac{\exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{njt})}{\sum_{k \in C} \exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{nkt})}. \tag{5.3}$$

[18] Luce and Suppes (1965) showed that the extreme value distribution leads to the logit formula, and McFadden (1974) showed the converse that the logit formula for choice probabilities implies the unobserved utility is extreme value distributed.

The vector of individual specific parameters, $\boldsymbol{\beta}_n$, captures the variations in the behaviors of individuals. How to model it and interpret it is an essential part of choice modeling analysis. One interpretation is that individuals have heterogeneous preferences or tastes regarding choice attributes (McFadden and Train, 2000). In other words, some individuals value particular choice attributes more than others. This explanation is represented by a class of models that allows for random parameters. One popular example within the family of logit models is the mixed logit model (MIXL).[19]

Some researchers argue that the preference heterogeneity may be better explained as scale heterogeneity (see, for example, Swait and Louviere, 1993; Louviere et al., 1999, 2002; Louviere and Eagle, 2006; Louviere et al., 2008). In this argument, the scale is inversely related to the variance of the error term. This implies that, with preference or taste fixed, individuals with smaller scale behave more randomly than those with larger scale. This interpretation has motivated models which explicitly model the scale of individuals, such as the scale heterogeneity multinomial logit model (SMNL). Compared to the mixed logit model specification, where the distributions of one or more coefficient parameters need to be estimated, the SMNL model gives a much more parsimonious description of the data (Fiebig et al., 2010).

Fiebig et al. (2010) noted that random coefficient and scale heterogeneity models are not fundamentally different. They are different ways of specifying the distribution of coefficient heterogeneity. To incorporate both sources of heterogeneity, namely preference and scale, Fiebig et al. (2010) developed the generalized multinomial logit model (GMNL). Following the GMNL specification, the vector of individual specific parameters, $\boldsymbol{\beta}_n$, is defined as

$$\boldsymbol{\beta}_n = \lambda_n \boldsymbol{\beta} + [\gamma + \lambda_n(1-\gamma)]\boldsymbol{\eta}_n. \tag{5.4}$$

λ_n is the individual specific scale of the idiosyncratic error. $\boldsymbol{\beta}$ is the constant vector of preference parameters, capturing the mean attribute weights in the population. $\boldsymbol{\eta}_n$ is the vector of individual n's deviation from the mean preferences. γ is a parameter that captures how residual preference heterogeneity varies with the scale.[20]

Despite the fact that disentangling scale heterogeneity from other sources of heterogeneity still requires explicit restrictions (Hess and Train, 2017), GMNL provides an

[19]Other names for this type of model are the random parameters logit model (RPL) (for example Greene, 2018, p.845), the mixed multinomial logit or random coefficients multinomial logit model (MMNL) (for example in McFadden and Train, 2000).

[20]See Fiebig et al. (2010, p.7) for a description of γ in detail. They show how, using the GMNL model, MIXL and SMNL can be nested in two ways when γ takes the value of zero and one, respectively.

alternative mechanism in modeling variability in utility than a mixed logit model with full coefficient covariance. The different models applied in the three papers in this thesis can be described as special cases of GMNL:

1. Multinomial logit model (MNL)/Conditional logit model (CL): $\boldsymbol{\beta}_n = \boldsymbol{\beta}$.
   This is when we assume there is no scale heterogeneity, and the scale is normalized to be one $\lambda_n = 1$, and no preference heterogeneity that the variance of residual preference is assumed to be zero $\text{var}(\boldsymbol{\eta}_n) = 0$.

2. Mixed logit model (MIXL): $\boldsymbol{\beta}_n = \boldsymbol{\beta} + \boldsymbol{\eta}_n$.
   This is when we assume there is no scale heterogeneity, and the scale is normalized to be one $\lambda_n = 1$.

3. Scaled multinomial logit model (SMNL): $\boldsymbol{\beta}_n = \lambda_n \boldsymbol{\beta}$.
   This is when we assume there is no preference heterogeneity that the variance of residual preference is assumed to be zero $\text{var}(\boldsymbol{\eta}_n) = 0$.

In the following, I present these three models with extensions or restrictions suitable for the analyses in my three papers and the corresponding estimation methods.

5.1 Paper I

For the analysis of this paper, we assume homogeneous preference and scale among individuals. Each individual makes decisions to "prescribe" or "not prescribe" ($j = 1, 2.$) on four types of drugs one by one ($t = 1, 2, 3, 4.$). The deterministic part of the utility of an individual is specified as:

$$V_{njt} = \sum_{t=1}^{4} \alpha_{jt} D_t + \beta_j I_n, \tag{5.5}$$

where $\alpha_{jt} D_t$ is the mean marginal utility of choosing j without the intervention for each drug type t, I_n indicates whether the individual is in the intervention group, and β_j captures the mean intervention effect. We also allowed the intervention effect β_j to vary over the different types of drugs.

According to the definitions in Hoffman and Duncan (1988), MNL focuses on the individual as the unit of analysis and uses the characteristics of the individual as explanatory variables, as opposed to CL, which focuses on the attributes of alternatives for each individual and uses the characteristics of alternatives as explanatory variables. When the utility is dependent on both attributes of alternatives and characteristics of the individuals, such models are referred to as a general form of the multinomial logit model.[21] As we note here in our study, given the drug type, V_{njt} is only dependent on characteristics of the individual. So it is a special case of the general MNL.

Maximum likelihood estimation

The parameters of a conditional logit model or multinomial logit model are estimated by maximum likelihood estimation method. Assuming that the sample observations are drawn independently and randomly from the population, maximum likelihood estimation finds the value of parameters that generates the observed sample most often (Ben-Akiva and Lerman, 1985). The likelihood of each individual in the sample choosing the alternative that is actually chosen is:

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}(P_{njt})^{y_{njt}}, \tag{5.6}$$

where P_{njt} is the conditional probability in (5.3) and $y_{njt} = 1$ if individual n chose j at occasion t and zero otherwise. Conventionally, the logarithm of $L(\boldsymbol{\beta})$ is maximized instead of $L(\boldsymbol{\beta})$ itself due to the fact that the log function is strictly monotonically increasing. The log-likelihood function is then presented as follows:

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} y_{njt}\ln P_{njt}. \tag{5.7}$$

McFadden (1974) has shown that the log-likelihood function with logit choice probabilities is globally concave in parameters when the utility is specified as linear-in-parameter. In general, maximum likelihood estimators are consistent, asymptotically normal, and asymptotically efficient.[22] To solve for the value of $\boldsymbol{\beta}$ that maximizes this function, many software packages are available, such as SAS, Stata, Nlogit, R, and Python.

5.2 Paper II

In this study, we allow for individual heterogeneity in preference, while assuming there is no scale heterogeneity, and the scale is normalized to be one. The vector of individual specific coefficients, $\boldsymbol{\beta}_n$, varies stochastically over individuals with density $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta})$ and of which $\boldsymbol{\theta}$ denotes the parameters of the distribution.[23] Although there is no

[21] See Greene (2018, pp.828-831) for a full exposition of it.

[22] See, for example, Ben-Akiva and Lerman (1985) Chap.2.5 for a discussion of the properties.

[23] The specification of the mixed logit model can be derived from different motivations or interpretations. Train (2009) discussed random coefficients and error components (see pp.137-141). These two are formally equivalent. Here, we take the most used and straightforward perspective of random coefficients.

restriction on the types of coefficient distribution, in practice, the multivariate normal distribution and the log-normal distribution are widely adopted in most applications in health economics (e.g., Hole, 2008; Hole and Kolstad, 2012; Godager and Wiesen, 2013; Song et al., 2015; Gutacker et al., 2016).[24]

The mixed logit model takes other names, such as the random parameters logit model (RPL) (for example in Greene, 2018, p.845), the mixed multinomial logit, or random coefficients multinomial logit model (MMNL) (for example, in McFadden and Train, 2000). The MIXL is a flexible model that, with adequate mixing specification, it can approximate any random utility model well (McFadden and Train, 2000).

Maximum simulated likelihood estimation

In MIXL, $\boldsymbol{\beta}_n$ is assumed to vary across individuals and follow a density function $f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta})$ with θ referring to the parameters of the distribution. The probabilities of an individual choosing a sequence of choices conditional on $\boldsymbol{\beta}_n$ are denoted as $S_n(\boldsymbol{\beta}_n)$:

$$S_n(\boldsymbol{\beta}_n) = \prod_{t=1}^{T}\prod_{j=1}^{J}(P_{njt}(\boldsymbol{\beta}_n))^{y_{njt}}. \tag{5.8}$$

The unconditional probability is then the integral of the conditional probability over the distribution of $\boldsymbol{\beta}_n$:

$$P_n(\boldsymbol{\theta}) = \int_{\boldsymbol{\beta}_n} S_n(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n \mid \boldsymbol{\theta}) d\boldsymbol{\beta}_n. \tag{5.9}$$

Since the probabilities cannot be solved analytically, they are approximated using simulation (Train, 2009, p.144).[25] The simulated probabilities $\hat{P}_n(\boldsymbol{\theta})$ are then defined as the average of the conditional probabilities evaluated at R random draws from the distribution of $\boldsymbol{\beta}_n$:

$$\hat{P}_n(\boldsymbol{\theta}) = \frac{1}{R}\sum_{r=1}^{R} S_n(\boldsymbol{\beta}_n^r). \tag{5.10}$$

Here, $\boldsymbol{\beta}_n^r$ is the r-th random draw.[26] Inserting $\hat{P}_n(\boldsymbol{\theta})$ into the log-likelihood, the corresponding simulated log-likelihood is thus:

$$SLL(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln \hat{P}_n(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln \frac{1}{R}\sum_{r=1}^{R} S_n(\boldsymbol{\beta}_n^r). \tag{5.11}$$

[24] When a coefficient is assumed to be positive or negative for all individuals, the log-normal distribution is often used.

[25] In addition to the maximum simulated likelihood, Train (2009, Chap.10) discussed advantages and limitations of another two methods of estimation: method of simulated moments and method of simulated scores. In this thesis, the estimation of the mixed logit was performed by the widely used maximum simulated likelihood method.

[26] Different software packages employ different methods of random draw. In this thesis, we use Stata SE15.1 package `mixlogit` (Hole, 2007) with Halton draws.

It is shown in (Hajivassiliou and Ruud, 1994) that the maximum simulated likelihood estimator is consistent, asymptotically normal and efficient, and equivalent to the maximum likelihood estimator when the number of draws R increases faster than the square root of the sample size $\sqrt{N}$.

5.3 Paper III

In this paper, we assume there is no preference heterogeneity among individuals. The variance in behavior is attributed to the differences in scale. However, as opposed to a standard SMNL model which allows the scale to vary across individuals, we assume the scale varies between different market settings but not individuals. This is motivated by Louviere and Eagle's (2006) argument that scale is highly unlikely to be constant, as the impact of noise on choices can vary over conditions, contexts, circumstances, or situations, as well as between decision-makers. We further extend this special case of SMNL to a strategic game scenario, where the individual's payoff of a choice alternative not only depends on his own choice but also his opponent(s)' choice.

Consider a symmetric normal form game with two individual players, a row player and a column player. Both players are asked to choose from a choice set, C, comprising J alternatives that are characterized by a vector of attributes. We report the payoffs of the row player instead of both because in a symmetric game, the payoff of choosing j when the opponent chooses s, denoted as $\boldsymbol{x}_{j|s}$, is identical to both players, $j,s \in C$. We assume the utility of the row player choosing j *conditional on* his opponent choosing s is linear in parameters with coefficient vector $\boldsymbol{\beta}$. The attributes of alternatives are denoted as $\boldsymbol{x}_{j|s}$. The utility can be expressed as:

$$v_{j|s} = \boldsymbol{\beta}' \boldsymbol{x}_{j|s}. \tag{5.12}$$

Note again that the utility of the row player choosing j is uncertain, as it depends on the opponent's choice.

The opponent's action is unknown ex ante. The probability that the opponent chooses alternative s is denoted by P_s. It follows that when the row player chooses alternative j, his *expected utility*, $V_j(\boldsymbol{P})$, is given by:

$$V_j(\boldsymbol{P}) = \sum_{s \in C} P_s v_{j|s}, \tag{5.13}$$

where $\boldsymbol{P} = (P_1, ..., P_J)$ denotes the vector of the opponent's choice probabilities for all alternatives.

We further assume the observed choices to be the result of individuals maximizing a linear combination of the expected utility and a noise. This leads to a small and natural augmentation of the standard SMNL, and we get:

$$\lambda_m V_j(\boldsymbol{P}) + \varepsilon_j, \tag{5.14}$$

where the scale parameter λ_m is kept constant among individuals, but allowed to vary between market settings.

Since there is no individual heterogeneity in preference and scale, the probability of choosing j is the same for each player. Assume ε_j is Type I extreme value distributed, it follows then:

$$P_j = \frac{\exp\left(\lambda_m V_j(\boldsymbol{P})\right)}{\sum_{r\in C}\exp\left(\lambda_m V_r(\boldsymbol{P})\right)} = \frac{\exp\left(\lambda_m \boldsymbol{\beta}' \sum_{s\in C} P_s \boldsymbol{x}_{j|s}\right)}{\sum_{r\in C}\exp\left(\lambda_m \boldsymbol{\beta}' \sum_{s\in C} P_s \boldsymbol{x}_{r|s}\right)}, \tag{5.15}$$

where the $\sum_{s\in C} P_s \boldsymbol{x}_{j|s}$ is the expected attribute vector. The key feature that distinguishes this model from a standard SMNL model is that the expected attribute vector on the right-hand side of the equation is unobservable because the probabilities of each alternative being chosen, P_s, are unknown.

It is essential to note that λ and $\boldsymbol{\beta}$ cannot be identified individually without additional restrictions, although the product can be estimated. The identification issue is less a concern if one is interested in estimating the marginal rates of substitution between two attributes, for example, willingness to pay (Train and Weeks, 2005; Hole and Kolstad, 2012; Godager and Wiesen, 2013; Scott et al., 2013), since λ cancels out when the ratio of two coefficients is estimated.

Quantal response equilibrium

In their seminal contribution, McKelvey and Palfrey (1995) proved that a quantal response equilibrium (QRE) always exists for finite games. At QRE, each player's belief of the opponent's choice probabilities $\boldsymbol{P}$ is identical to the equilibrium probabilities $\boldsymbol{P}^*$. So to find the quantal response equilibrium probabilities, one needs to solve for the fixed point of:

$$p_j = \frac{\exp(\lambda_m V_j(\boldsymbol{P}^*))}{\sum_{r\in C}\exp(\lambda_m V_r(\boldsymbol{P}^*))}. \tag{5.16}$$

QRE can be described as a statistical version of the Nash equilibrium (NE) (Camerer, 2011b). In the limit, when the scale parameter λ approaches infinity, the players make no errors, and the behavior becomes deterministic. In other words, in the limit, the players are perfectly responsive to the differences in expected utility across alternatives, and the behavior converges to a subset of NE. On the other extreme, when the λ approaches zero, the behavior becomes purely random, and the player plays each pure strategy with equal probability. In the behavioral game theory literature inspired by McKelvey and Palfrey (1995), the scale parameter is often referred to as the "rationality parameter."

Two-step and full information maximum likelihood estimation

The log-likelihood function given our model specification in (5.15) can be written as:

$$LL(\lambda_m, \boldsymbol{\beta})|_{\boldsymbol{P}} = \sum_n \sum_j y_{nj} ln \frac{\exp(\lambda_m V_j(\boldsymbol{P}, \boldsymbol{\beta}))}{\sum_{r \in C} \exp(\lambda_m V_r(\boldsymbol{P}, \boldsymbol{\beta}))}. \tag{5.17}$$

There are at least two methods for estimating the scale parameter λ and the preference parameter $\boldsymbol{\beta}$, with suitable restrictions to achieve identification.

The first one is a two-step maximum likelihood estimation. In the first step, we compute the expected attribute vector. Since the player's belief of the opponent's probabilities of choices, $\boldsymbol{P}$, is unknown, it is replaced by realized relative frequencies, $\boldsymbol{f}$. Then, in the second step, the log-likelihood function below is maximized in estimating the parameters:

$$LL(\lambda_m, \boldsymbol{\beta})|_{\boldsymbol{f}} = \sum_n \sum_j y_{nj} ln \frac{\exp(\lambda_m V_j(\boldsymbol{f}, \boldsymbol{\beta}))}{\sum_{r \in C} \exp(\lambda_m V_r(\boldsymbol{f}, \boldsymbol{\beta}))}. \tag{5.18}$$

Since $\boldsymbol{f}$ is a consistent estimator of $\boldsymbol{P}$, log-likelihood (5.18) converges in probability to (5.17) as the number of individuals increases toward infinity. It follows that the estimators $\hat{\lambda}$ and $\hat{\boldsymbol{\beta}}$ which maximize equation (5.18) are consistent. The two-step estimators are a straightforward choice as they can be estimated by existing software modules, such as `gmnl` in Stata (Fiebig et al., 2010; Gu et al., 2013).

The second possibility is to fit the full model at once. This approach relies on the assumption of QRE which is the fixed point of (5.16). The log-likelihood function is then:

$$LL(\lambda_m, \boldsymbol{\beta})|_{\boldsymbol{P}} = \sum_n \sum_j y_{nj} ln \frac{\exp(\lambda_m V_j(\boldsymbol{P}^*, \boldsymbol{\beta}))}{\sum_{r \in C} \exp(\lambda_m V_r(\boldsymbol{P}^*, \boldsymbol{\beta}))}. \tag{5.19}$$

Using a full information maximum likelihood estimation means estimating λ and $\boldsymbol{\beta}$ and computing $\boldsymbol{P}^*$ simultaneously. This procedure is computationally complicated due to the fact that QRE has to be calculated at each stage in the parameter search in order to evaluate the likelihood function at each stage. One drawback is that no software modules exist that can perform the full information maximum likelihood estimation for such model. Thus, researchers have to program on their own, which makes it costly and less attractive to use, especially for analyzing games with more than two players.[27]

[27] See Moffatt (2015, pp.398-400) for an example of Stata code for the full information maximum likelihood estimation. The complete code contains the mata optimize program for finding the equilibrium probabilities, and the ml program for estimating two parameters: a scale parameter λ and a risk parameter.

6 | Summary of results

6.1 Paper I

The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment

We categorize physicians' prescribing, upon seeing a patient with minor symptoms of a common cold, into four types: antibiotics, other prescription drugs (Other Rx), over-the-counter drugs (OTC), and alternative and nonpharmacological treatments (Alternatives).

The estimated average intervention effect is negative and statistically significant, implying that the mystery shopper intervention caused a reduction in physicians' mean marginal utility of prescribing in general and thus reduced their probability of prescribing drugs to the pseudopatient. Compared to the control group, the mystery shopper intervention reduced the odds of prescribing by 19.2%. When we account for the possibility of substitution by allowing for between-drug variation in the intervention effect, we find that the announcement of a mystery shopper audit led to a reduction in prescribing of Other Rx and OTC. Compared to the control group, the mystery shopper intervention reduced the odds of prescribing Other Rx by 61.0% and OTC by 42.7%.

6.2 Paper II

Exploring physician agency under demand-side cost sharing — An experimental approach

The results suggest that future physicians are concerned about how their choices of medical treatment affect patients' consumption opportunities, and this main finding is robust across specifications. Our results show evidence of significant preference heterogeneity among individuals. We find that all three marginal utilities (own profit, patient health benefit, and patient consumption opportunity after co-payment) at the median are positive and declining, showing a diminishing marginal utility at the median level. Further, the analysis provides evidence that, on average, profit is considered as a quantity complement to both patient benefit and patient consumption and that marginal utility of the patient health benefit is unaffected by the level of patient consumption.

6.3 Paper III

Predicting strategic medical choices: An application of a quantal response equilibrium choice model

Fitting a choice model with an assumption that the subjects' preferences toward profit and health benefit remain fixed during the experiment, the substantial difference in behavior between markets can be attributed to changes in individuals' scale parameter. We find that the scale parameter rises as markets become more competitive, implying a higher degree of determinism in behavior. One possible intuitive explanation is that competition triggers decision-makers' attention. Another possible explanation is that competition positively affects the perceived return from cognitive effort.

The model captures the substantial differences in behavior across games and market settings, and the out-of-sample predictions are quite similar to observed behavior. Further, results from Monte Carlo simulations show that the two-step maximum likelihood estimator produces accurate and precise estimates, even with a moderate sample size.

7 | Discussion and future research

7.1 Discussion of the results

Paper I

In Paper I, we find that the preannouncement of a mystery shopper audit significantly decreases the probability of prescribing any drug in general. More specifically, when we account for the possibility of substitution by allowing for between-drug variation in the intervention effect, we find that the intervention led to a reduction in prescribing of other prescription drugs and over-the-counter drugs, while there was no effect on the prescribing of antibiotics and alternative nonpharmacological treatments. The findings contribute new knowledge by identifying a separate effect of auditing rather than a combined effect of auditing and other measures, such as reminders, feedback, or education. It is promising that such straightforward intervention has the potential to be implemented for those health care systems challenged by limited access to performance data and overwhelming unnecessary health expenditure to improve provider performance.

That said, the results from this paper shall be understood and generalized with caution. First of all, our effect estimates are acquired in a specific context and with a specific set of presented symptoms. Since we cannot rule out the possibility that the physicians might respond differently to the intervention in the context of illness other than the common cold, the results should be interpreted in the context of the common cold, where the issue of overprescribing is highly relevant. Secondly, our findings do not indicate a significant effect on antibiotic prescribing, which is consistent with the findings of Östervall (2017). The literature provides evidence of several factors that drive inappropriate prescribing of antibiotics, such as the lack of knowledge on antibiotic misuse (Grigoryan et al., 2007; Togoobaatar et al., 2010; Pan et al., 2012; Yu et al., 2014), the diagnostic uncertainty

(Arnold et al., 2005; Kotwani et al., 2010; Xue et al., 2019), and patients' expectations (Reynolds and McKee, 2009; Jin et al., 2011). Understanding the relevant institutional and cultural background of the health system in interest might point to a useful direction of policy design and provide reasonable interpretations of the results.

Paper II

In Paper II, we find that future physicians care about patients' consumption opportunities alongside patients' health benefit and their own profit. In other words, they prefer, ceteris paribus, treatment alternatives with less demand-side cost sharing. The result is intuitive and consistent with conclusions drawn from many survey studies (e.g., Reichert et al., 2000; Khan et al., 2008). The added value of our study to the literature is to provide a quantitative economic measure of the importance of patients' co-payment in physicians' decision-making obtained from a controlled experiment.

The result that physicians include patients' co-payment in their objective means that demand-side cost sharing can influence the equilibrium of health care utilization through two channels. When the behavior of both patients and physicians is affected by demand-side cost sharing, it becomes challenging to compute demand elasticities using conventional empirical approaches with field data. Furthermore, this result provides implications for policy design. For a health care system in which the physicians have knowledge of patients' insurance coverage, a given level of demand-side cost sharing might, on average, have a stronger effect on service utilization than expected when the policy makers only consider the demand responses from the patient's perspective. With regards to the optimal design of physician payment, our result implies that the optimal level of supply-side cost sharing is negatively related to the optimal level of demand-side cost sharing, as both channels influence supply-side incentives in the same direction. This means that the variation in demand-side cost sharing across patients challenges the implementation of optimal physician payment.

Paper III

In Paper III, we observe that competition is beneficial to the patients since the subjects provide significantly more total health benefits as the markets become more competitive. A possible interpretation is that "competition raises moral values" or "competition crowds in pro-social motivation." However, we provide an alternative interpretation with reference to economic theory and the assumption of the fixed preferences. The substantial difference in behavior between markets can then be attributed to changes in individuals' scale parameter. The scale parameter rises as markets become more competitive, implying a higher degree of determinism in behavior. It might be interpreted as competition triggers decision-makers' attention or competition positively affects the perceived return from cognitive effort. Our approach combines a theory-based structural model with experimental data. It enables us to better understand the mechanism through which the effect operates and to predict behavior out-of-sample and hence improve the validity externally (Low and Meghir, 2017). I discuss more on this methodological approach in Section 7.2.3 below.

7.2 Discussion of methodology and future research

7.2.1 A couple of issues on quantal response equilibrium

Quantal response equilibrium has been obtaining significant attention due to its widely documented ability to explain strategic behavior in cases where Nash equilibrium cannot (see, for example, reviews by Goeree et al., 2010; Palfrey et al., 2016). Built on a stochastic choice modeling framework (Harsanyi, 1973; McFadden, 1974), the notion of QRE has contributed as a bridge between individual choice modeling and game theory. It can be interpreted as an application of standard random utility models of discrete choice to strategic scenarios (the angle taken throughout this thesis) or as a statistical generalization of Nash equilibrium that allows for "noisy" players (Goeree et al., 2010).

The importance of its contribution has been highly praised by some. As suggested by Camerer et al. (2004), "QRE almost always explains the direction of deviations from Nash and should replace Nash as the static benchmark to which other models are routinely compared." Later in the book, Camerer (2011b) expressed, "I used to say in classes and seminars that, if John Nash had been a statistician rather than a mathematician, he might have discovered QRE rather than Nash equilibrium."[28] The usefulness of QRE undoubtedly opens up the opportunity to revisit previous experiments from a statistical perspective. Nevertheless, there have also been discussions of limitations of QRE and cautions when applying it. I discuss two issues in the following.

Empirical content of QRE

First of all, the structural QRE has been questioned for the empirical implications in its applications in different games/datasets. Haile et al. (2008) state that QRE imposes no falsifiable restriction, and hence it can rationalize any distribution of behavior in any normal form game. While it is true that, without restrictions on the disturbances, QRE

[28] See p.34 in Camerer (2011b) for the complete anecdote.

can fit any data, Goeree et al. (2010) argue the quantal response function generating the QRE is only determined when the probability distribution of the random payoff disturbances and subjects' preferences are specified. Therefore, to generate falsifiable restrictions and comparative predictions, Goeree et al. (2010) suggest holding the same structural assumptions across different games or datasets. Haile et al.'s (2008) concern raises a note of caution for interpreting results from different games, as it makes little sense to compare the magnitudes of λ (the scale parameter) estimated from different games without imposing the same restrictions of the distribution of subjects' preferences and the payoff disturbances.

In Paper III, we follow suggestions from Goeree et al. (2010) and perform the analysis under the assumption that there is no variation in the subjects' preferences and the payoff disturbances across games. The assumption is reasonable in our study because the sample of subjects and the experimental environments are carefully controlled across games. Fitting the model with this assumption, the hypothesis of a fixed scale across markets is then rejected. In my view, the discussion on the empirical content of QRE motivates future experimental research applying QRE to carefully design the experiments so that the restrictions are likely satisfied, and reasonable inferences can be generated by comparing estimated parameters from different experiments or datasets.

Multiple equilibria

In QRE, each player's belief of the opponent's choice probabilities is accurate and identical to the equilibrium probabilities (McKelvey and Palfrey, 1995). To find the QRE probabilities, one needs to solve for the fixed point of a system of quantal response functions. McKelvey and Palfrey (1995) prove that a QRE always exists for finite games. However, the uniqueness is never guaranteed, even though they claim that uniqueness of the logit QRE applies for "almost all games."

In the related social interaction (also called peer effect or social networks) literature, where individuals' preferences are assumed to be influenced by the aggregated behavior of others, multiple equilibria have been discussed in different settings (Becker, 1974; Manski, 2000; Brock and Durlauf, 2007; Ge, 2015). In the context of consumer choices, under certain conditions when the peer effect is strong enough, there may exist multiple levels of demand that correspond to a same price level (Becker, 1974). Thus, a given price change can cause a dramatic demand response as it jumps to another equilibrium level.[29] In the context of QRE, when multiple equilibria exist, a given vector of

[29]Known as the "tipping" point in Schelling (1971).

parameters (e.g., preference parameters and scale parameter) corresponds to more than one vector of choice probabilities. This clearly complicates the analysis of behavior. In his recent paper, Dagsvik (2020) extends the techniques used in studies of social interaction to the case of QRE with multinomial choices for two players. He establishes necessary and sufficient conditions for the existence of single and multiple QRE in symmetric logit QRE models and provides useful algorithms for concluding on the number of stable QRE. We have applied the algorithms to examine the uniqueness of the equilibrium in the eight games in duopoly (in Paper III in this thesis) and found that the equilibrium is unique for all eight games in duopoly. Generalization of the conditions to asymmetric QRE with more than two players is naturally an interesting and important topic of future research.

7.2.2 Diversity of economic experimental methods

The field of experimental economics has seen the development and contributions of diverse methods, ranging from lab to field experiments and a variety of methods in between.[30] Despite the debate on the validity of experiments,[31] the view seems to converge in the recognition that various experimental methods have different sets of strengths and weaknesses, and they serve as strong complements and can together improve the state of knowledge in science (Harrison and List, 2004; Reiley and List, 2007; Falk and Heckman, 2009; Czibor et al., 2019; Harrison et al., 2015; Kessler and Vesterlund, 2015).

Sparked by this discussion, the experimental design in my papers takes advantage of different types of experimental approaches. For instance, the experiment in Paper II is designed following the principles of a discrete choice experiment that the choice sets are carefully and efficiently constructed and combined. In addition, we set up the experiment in a lab so the environment is carefully controlled. Most importantly, to reduce the hypothetical bias that the choice experiments are often criticized for, we attach real incentives to all the attributes of the decisions (Harrison, 2006). Compared to the commonly used approach in which the health benefits generated from the lab are transferred to a charity (e.g., Hennig-Schmidt et al., 2011), we transfer the money to treat a real patient at the university hospital nearby. We believe this approach enhances the saliency and thus the validity.

[30] In spite of the confusion generated by the various use of terminologies, some researchers propose two differentiate types of experiments. For example, Harrison and List (2004, p.1013) propose definitions and terminologies of different experimental methods, and List (2007) (Figure 2 on p.7) categorizes experiments by the degree of control.

[31] See for example, Levitt and List (2007, 2008) and Camerer (2011a).

This methodological discussion has led to the development of new experimental approaches that potentially integrate strengths of different approaches and strengthen both internal and external validity (Harrison et al., 2015). One example is virtual reality experiments (VX), in which participants make decisions in a real-time computer-generated world that can be interactively experienced through sensory stimuli (Fiore et al., 2009; Innocenti, 2017). Fiore et al. (2009) propose this method and argue that VX takes advantage of the combined strengths of both lab and field experiments since it creates a controlled lab-like environment with contextual cues mimicking those occurring in the field. Patterson et al. (2017) incorporate virtual reality into a discrete choice experiment of neighborhood choice by visualizing the choice alternatives.[32] VX has the potential to add unique insights to our understanding of behavior in many fields of economics, especially where field experiments are rare and time-consuming and where it is difficult to understand the alternatives or tasks in the lab or DC experiments (Fiore et al., 2009; Patterson et al., 2017). In my view, applying VX in experimental health economics can be a useful and interesting path for future research. An interactive virtual environment can facilitate a description of health status that is closer to reality and thus generates better understanding of the value of health and other health-related research questions.

7.2.3 Structural models and experimental economics

Randomized experiments are at the heart of the economics of policy design and analysis. As discussed in Section 4.1, the experimental approaches have their appeal in that they provide causal relations without needing to refer to a specific economic model or structure. Low and Meghir (2017) in their recent paper, argue that the usefulness of experiments is limited when they are not analyzed within a framework that can improve validity externally. Thus, they encourage the use of a combination of experiments and theory-based structural models to enhance analysis that cannot be achieved with either approach alone. While the experiments generate exogenous variations that help to identify economic effects, the theory-based structural models describe the mechanisms through which effects operate and thus provide the framework for an interpretation of the experimental results and analysis of counterfactuals (Attanasio et al., 2012; Low and Meghir, 2017).

Indeed, without referring to a structural model, the observation from Paper III in this thesis that competition brings about more patient benefits may lead to the implication that competition raises moral values. However, we could interpret the results with

[32] They find that preferences estimated this way are based on the displayed images of outcomes rather than individuals' mental images in a conventional text-only version of the choice experiment.

reference to a structural model based on economic theory and provide a coherent way to understand the mechanism. This approach not only enhances the internal validity by allowing us to predict behavior within the experiment but also adds the potential to understand how this particular mechanism of behavior under competition may translate in different environments outside the experiment.

The recommendation to combine experiments and structural modeling dates back at least to Orcutt and Orcutt (1968) and Burtless and Hausman (1978), and this approach is growing in influence (Blundell et al., 1998; Attanasio et al., 2012; Duflo et al., 2012; Huck et al., 2015). Thanks to the progress of both computational power and numerical methods, the use of structural models in experiments, together with other methodological approaches, will continuously add important insights and improve the state of knowledge in economics (Low and Meghir, 2017; Czibor et al., 2019).

8 | Conclusion

This thesis addresses fundamental research questions on physician behavior by combining experimental economic methods and the structural stochastic choice models. The three papers contribute to the growing literature in experimental health economics and add knowledge on physician choices when making treatment decisions. Each paper focuses on a specific aspect that is found to influence physician behavior: Paper I on change of information scheme, Paper II on demand-side cost sharing, and Paper III on physician competition. Specifically, in Paper I, we found that the preannouncement of a mystery shopper audit reduced physicians' probability of prescribing drugs to the pseudopatients. In Paper II, the results suggest that future physicians are concerned about how their choices of medical treatment affect the patients' consumption opportunities. In Paper III, we found that the scale parameter rises as markets become more competitive, implying a higher degree of determinism in behavior. The evidence we found in the papers has implications on health policy designs for the purpose of quality improvement and efficient resource allocation in health care.

The novelty of this thesis is the incorporation of experimental data and theory-based structural models. The thesis benefits from advantages from both approaches: The experiments generate exogenous variations that help to identify economic effects, while the theory-based structural models describe the mechanisms through which effects operate and thus provide the framework for an interpretation of the experimental results and analysis of counterfactuals. Under the framework of stochastic choice, three special cases of a generalized multinomial logit model were employed in the data analysis.

The challenges ahead are numerous due to the peculiarities of the health care and human minds. As McFadden (2001) acknowledged, "Science is a cooperative enterprise." The literature points to the importance of collaboration of different data-collecting and analyzing approaches. Both health economics and decision theory can benefit from collaboration between discrete choice econometrics, experimental economics, psychology, behavioral game theory, cognitive science, and artificial intelligence.

References

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Allaby, M. A. (2003). Doctors for the poor in urban nepal. *Tropical doctor*, 33(2):83–85.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, pages 503–546.

Allard, M., Jelovac, I., and Léger, P. (2011a). Treatment and Referral Decisions under Different Physician Payment Mechanisms. *Journal of Health Economics*, 30:880–893.

Allard, M., Jelovac, I., and Leger, P. T. (2011b). Treatment and referral decisions under different physician payment mechanisms. *Journal of Health Economics*, 30(5):880–893.

Allard, M., Thomas Léger, P., and Rochaix, L. (2009). Provider competition in a dynamic setting. *Journal of Economics & Management Strategy*, 18(2):457–486.

Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, 27(5):1260–1274.

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation*. MIT press.

Anderson, S. P., Goeree, J. K., and Holt, C. A. (2001). Minimum-effort coordination games: Stochastic potential and logit equilibrium. *Games and Economic Behavior*, 34(2):177–199.

Arnold, S. R., To, T., McIsaac, W. J., and Wang, E. E. (2005). Antibiotic prescribing for upper respiratory tract infection: the importance of diagnostic uncertainty. *The Journal of pediatrics*, 146(2):222–226.

Arrieta, A., García-Prado, A., González, P., and Pinto-Prades, J. L. (2017). Risk attitudes in medical decisions for others: an experimental approach. *Health economics*, 26:97–113.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53:941–969.

Attanasio, O. P., Meghir, C., and Santiago, A. (2012). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. *The Review of Economic Studies*, 79(1):37–66.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., Schneider, E. C., Wright, B. J., Zaslavsky, A. M., and Finkelstein, A. N. (2013). The oregon experiment—effects of medicaid on clinical outcomes. *New England Journal of Medicine*, 368(18):1713–1722.

Baker, R., Farooqi, A., Tait, C., and Walsh, S. (1997). Randomised controlled trial of reminders to enhance the impact of audit in general practice on management of patients who use benzodiazepines. *BMJ Quality & Safety*, 6(1):14–18.

Baker, R., Fraser, R. C., Stone, M., Lambert, P., Stevenson, K., and Shiels, C. (2003). Randomised controlled trial of the impact of guidelines, prioritized review criteria and feedback on implementation of recommendations for angina and asthma. *British Journal of General Practice*, 53(489):284–291.

Bardsley, N., Cubitt, R., Loomes, G., Moffat, P., Starmer, C., and Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton University Press.

Barnum, H., Kutzin, J., and Saxenian, H. (1995). Incentives and provider payment methods. *The International journal of health planning and management*, 10(1):23–45.

Barros, P. P. (2003). Cream-skimming, incentives for efficiency and payment system. *Journal of health economics*, 22(3):419–443.

Becker, G. S. (1974). A theory of social interactions. *Journal of political economy*, 82(6):1063–1093.

Belot, M., James, J., and Nolen, P. (2016). Incentives and children's dietary choices: A field experiment in primary schools. *Journal of health economics*, 50:213–229.

Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6):485–495.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.

Blomqvist, Å. (1991). The doctor as double agent: Information asymmetry, health insurance, and medical care. *Journal of Health Economics*, 10(4):411–432.

Blundell, R., Duncan, A., and Meghir, C. (1998). Estimating labor supply responses using tax reforms. *Econometrica*, pages 827–861.

Brekke, K. R., Holmås, T. H., Monstad, K., and Straume, O. R. (2019). Competition and physician behaviour: Does the competitive environment affect the propensity to issue sickness certificates? *Journal of health economics*, 66:117–135.

Brock, W. A. and Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1):52–75.

Bronchetti, E. T., Huffman, D. B., and Magenheim, E. (2015). Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination. *Journal of Economic Behavior & Organization*, 116:270–291.

Brosig-Koch, J., Hehenkamp, B., and Kokot, J. (2017a). The effects of competition on medical service provision. *Health Economics*, 26:6–20.

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

Buckley, N. J., Cuff, K., Hurley, J., McLeod, L., Mestelman, S., and Cameron, D. (2012). An experimental investigation of mixed systems of public and private health care finance. *Journal of Economic Behavior & Organization*, 84(3):713–729.

Burtless, G. and Hausman, J. A. (1978). The effect of taxation on labor supply: Evaluating the gary negative income tax experiment. *Journal of political Economy*, 86(6):1103–1130.

Camerer, C. (2011a). The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list. *Available at SSRN 1977749*.

Camerer, C. F. (2011b). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., Ho, T.-H., and Chong, J. K. (2004). Behavioural game theory: thinking, learning and teaching. In *Advances in understanding strategic behaviour*, pages 120–180. Springer.

Carlsson, F. and Martinsson, P. (2003). Design techniques for stated preference methods in health economics. *Health economics*, 12(4):281–294.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2018). The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise. *Journal of health economics*, 62:95–104.

Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., and Riella, G. (2019). Deliberately stochastic. *American Economic Review*, 109(7):2425–45.

Chalkley, M. and Malcomson, J. M. (1998). Contracting for Health Services when Patient Demand Does not Reflect Quality. *Journal of Health Economics*, 17:1–19.

Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, 56(2):95–108.

Chandra, A., Cutler, D., and Song, Z. (2011). Who ordered that? the economics of treatment choices in medical care. In *Handbook of health economics*, volume 2, pages 397–432. Elsevier.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140.

Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Clark, M. D., Determann, D., Petrou, S., Moro, D., and de Bekker-Grob, E. W. (2014). Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*, 32(9):883–902.

Collins, M. M., Barry, M. J., Zietman, A., Albertsen, P. C., Talcott, J. A., Corkery, E. W., Elliott, D. B., and Fowler Jr, F. J. (2002). United states radiation oncologists' and urologists' opinions about screening and treatment of prostate cancer vary by region. *Urology*, 60(4):628–633.

Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2011). Does hospital competition save lives? evidence from the english nhs patient choice reforms. *The Economic Journal*, 121(554):F228–F260.

Coulam, R. F. and Gaumer, G. L. (1992). Medicare's prospective payment system: a critical appraisal. *Health Care Financing Review*, 1991(Suppl):45.

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

Currie, J., Lin, W., and Meng, J. (2014). Addressing antibiotic abuse in china: an experimental audit study. *Journal of development economics*, 110:39–51.

Currie, J., Lin, W., and Zhang, W. (2011). Patient knowledge and antibiotic abuse: Evidence from an audit study in china. *Journal of health economics*, 30(5):933–949.

Czibor, E., Jimenez-Gomez, D., and List, J. A. (2019). The dozen things experimental economists should do (more of). *Southern Economic Journal*, 86(2):371–432.

Dagsvik, J. K. (2020). On the equilibria in logit models of social interaction and quantal response equilibrium. *HERO On line Working Paper Series*.

Davidson, D. and Marschak, J. (1959). Experimental tests of a stochastic decision theory. *Measurement: Definitions and theories*, 17:274.

Davis, D. D. and Holt, C. A. (1993). *Experimental economics*. Princeton university press.

de Bekker-Grob, E. W., Ryan, M., and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health economics*, 21(2):145–172.

Devlin, R. A. and Sarma, S. (2008). Do physician remuneration schemes matter? the case of canadian family physicians. *Journal of health economics*, 27(5):1168–1181.

Duflo, E., Hanna, R., and Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.

Dulleck, U. and Kerschbamer, R. (2009). Experts vs. discounters: Consumer free-riding and experts withholding advice in markets for credence goods. *International Journal of Industrial Organization*, 27(1):15–23.

Dunn, A. and Shapiro, A. H. (2014). Do physicians possess market power? *The Journal of Law and Economics*, 57(1):159–193.

Eccles, M., Steen, N., Grimshaw, J., Thomas, L., McNamee, P., Soutter, J., Wilsdon, J., Matowe, L., Needham, G., Gilbert, F., et al. (2001). Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *The Lancet*, 357(9266):1406–1409.

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

Ellis, R. P. and McGuire, T. G. (1990). Optimal payment systems for health services. *Journal of Health Economics*, 9:375–396.

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *science*, 326(5952):535–538.

Farley, P. J. (1986). Theories of the price and quantity of physician services: A synthesis and critique. *Journal of Health Economics*, 5:315–333.

Feder, G., Griffiths, C., Highton, C., Eldridge, S., Spence, M., and Southgate, L. (1995). Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? a randomised controlled trial in general practices in east london. *Bmj*, 311(7018):1473–1478.

Feldstein, M. S. (1970). The rising price of physician's services. *The Review of Economics and Statistics*, pages 121–133.

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

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., et al. (2012). The oregon health insurance experiment: Evidence from the first year*. *The Quarterly Journal of Economics*, 127(3):1057–1106.

Fiore, S. M., Harrison, G. W., Hughes, C. E., and Rutström, E. E. (2009). Virtual experiments and environmental policy. *Journal of Environmental Economics and Management*, 57(1):65–86.

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

Flood, M. M. (1958). Some experimental games. *Management Science*, 5(1):5–26.

Fowler Jr, F. J., Collins, M. M., Albertsen, P. C., Zietman, A., Elliott, D. B., and Barry, M. J. (2000). Comparison of recommendations by urologists and radiation oncologists for treatment of clinically localized prostate cancer. *Jama*, 283(24):3217–3222.

Frank, R. G. (2007). Behavioral economics and health economics. In Diamond, P. and Vartiainen, H., editors, *Behavioral economics and its applications*, chapter 6, pages 195–234. Princeton University Press.

Friedman, M. (1953). The methodology of positive economics.

Friedman, S., Friedman, D., and Sunder, S. (1994). *Experimental methods: A primer for economists*. Cambridge University Press.

Fuchs, V. R. (2000). The future of health economics. *Journal of Health Economics*, 19(2):141–157.

Fuchs, V. R. (2011). *Who shall live?: health, economics and social choice*. World Scientific.

Galizzi, M. M., Tammi, T., Godager, G., Linnosmaa, I., and Wiesen, D. (2015). Provider altruism in health economics.

Galizzi, M. M. and Wiesen, D. (2018). Behavioral experiments in health economics. In *Oxford Research Encyclopedia of Economics and Finance*.

Gaynor, M. and Town, R. J. (2011). Competition in health care markets. In *Handbook of health economics*, volume 2, pages 499–637. Elsevier.

Gaynor, M. and Vogt, W. B. (2000). Antitrust and competition in health care markets. *Handbook of health economics*, 1:1405–1487.

Ge, G. (2015). Individual Choice Behavior with Social Interaction: An Empirical Analysis of Choice among Mobile Network Operators. Master's thesis, Department of Economics, University of Oslo, https://www.duo.uio.no/handle/10852/49108.

Gillam, S. J., Siriwardena, A. N., and Steel, N. (2012). Pay-for-performance in the united kingdom: impact of the quality and outcomes framework—a systematic review. *The Annals of Family Medicine*, 10(5):461–468.

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

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

Goeree, J. K. and Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, 91(5):1402–1422.

Goeree, J. K. and Holt, C. A. (2005). An experimental study of costly coordination. *Games and Economic Behavior*, 51(2):349–364.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2005). Regular quantal response equilibrium. *Experimental Economics*, 8(4):347–367.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2010). Quantal response equilibria. In *Behavioural and Experimental Economics*, pages 234–242. Springer.

Gravelle, H., Scott, A., Sivey, P., and Yong, J. (2016). Competition, prices and quality in the market for physician consultations. *The Journal of Industrial Economics*, 64(1):135–169.

Greene, W. (2018). Econometric analysis (eight ed.).

Grigoryan, L., Burgerhof, J. G., Degener, J. E., Deschepper, R., Lundborg, C. S., Monnet, D. L., Scicluna, E. A., Birkin, J., and Haaijer-Ruskamp, F. M. (2007). Attitudes, beliefs and knowledge concerning antibiotic use and self-medication: a comparative european study. *Pharmacoepidemiology and drug safety*, 16(11):1234–1243.

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

Gutacker, N., Siciliani, L., Moscelli, G., and Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of health economics*, 50:230–246.

Haile, P. A., Hortaçsu, A., and Kosenok, G. (2008). On the empirical content of quantal response equilibrium. *American Economic Review*, 98(1):180–200.

Hajivassiliou, V. A. and Ruud, P. A. (1994). Classical estimation methods for ldv models using simulation. *Handbook of econometrics*, 4:2383–2441.

Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern, J., Loewenstein, G., Brennan, T. A., Asch, D. A., and Volpp, K. G. (2015). Randomized trial of four financial-incentive programs for smoking cessation. *N Engl J Med*, 372:2108–2117.

Harrison, G., Lau, M., and Rutström, E. (2015). Theory, experimental design and econometrics are complementary (and so are lab and field experiments). In *Handbook of experimental economic methodology*, pages 296–338. Oxford University Press Oxford, UK.

Harrison, G. W. (2006). Making choice studies incentive compatible. In *Valuing environmental amenities using stated choice studies*, pages 67–110. Springer.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42:1009–1055.

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23.

Hellerstein, J. K. (1998). The importance of the physician in the generic versus trade-name prescription decision. *The Rand journal of economics*, pages 108–136.

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

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

Hensher, D. A. and Bradley, M. (1993). Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters*, 4(2):139–151.

Hess, S. and Train, K. (2017). Correlation and scale in mixed logit models. *Journal of choice modelling*, 23:1–8.

Hoffman, S. D. and Duncan, G. J. (1988). Multinomial and conditional logit discrete-choice models in demography. *Demography*, 25(3):415–427.

Hole, A. (2017). Dcreate: Stata module to create efficient designs for discrete choice experiments. *Statistical Software Components from Boston College Department of Economics*.

Hole, A. R. (2007). Fitting mixed logit models using maximum simulated likelihood. *Stata Journal*, 7:388–401.

Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of health economics*, 27(4):1078–1094.

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

Hu, T., Decker, S. L., and Chou, S.-Y. (2017). The impact of health insurance expansion on physician treatment choice: Medicare part d and physician prescribing. *International journal of health economics and management*, 17(3):333–358.

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

Huck, S., Rasul, I., and Shephard, A. (2015). Comparing charitable fundraising schemes: Evidence from a natural field experiment and a structural model. *American Economic Journal: Economic Policy*, 7(2):326–69.

Hurley, J. and Labelle, R. (1995). Relative fees and the utilization of physicians' services in canada. *Health Economics*, 4(6):419–438.

Hutchison, B., Birch, S., Hurley, J., Lomas, J., and Stratford-Devai, F. (1996). Do physician-payment mechanisms affect hospital utilization? a study of health service organizations in ontario. *CMAJ: Canadian Medical Association Journal*, 154(5):653.

Ifcher, J. and Zarghamee, H. (2016). Ethics and experimental economics. In *Practicing Professional Ethics in Economics and Public Policy*, pages 195–205. Springer.

Innocenti, A. (2017). Virtual reality experiments in economics. *Journal of behavioral and experimental economics*, 69:71–77.

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M. A., Johansen, M., Grimshaw, J., and Oxman, A. D. (2012). Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane database of systematic reviews*, (6).

Iversen, T. and Lurås, H. (2000). The effect of capitation on gps' referral decisions. *Health economics*, 9(3):199–210.

Jacquemet, N. and l'Haridon, O. (2018). *Experimental Economics*. Cambridge University Press.

Jin, C., Ely, A., Fang, L., and Liang, X. (2011). Framing a global health risk from the bottom-up: user perceptions and practices around antibiotics in four villages in china. *Health, risk & society*, 13(5):433–449.

Kairies-Schwarz, N., Kokot, J., Vomhof, M., and Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory–an experiment. *Journal of Economic Behavior & Organization*, 137:374–397.

Kalisch, G., Milnor, J., Nash, J., and Nering, E. (1954). Some experimental n-person games.

Keeler, E. B., Brook, R. H., Goldberg, G. A., Kamberg, C. J., and Newhouse, J. P. (1985). How free care reduced hypertension in the health insurance experiment. *Jama*, 254(14):1926–1931.

Kerse, N. M., Flicker, L., Jolley, D., Arroll, B., and Young, D. (1999). Improving the health behaviours of elderly people: randomised controlled trial of a general practice education programme. *Bmj*, 319(7211):683–687.

Kessler, D. P. and McClellan, M. B. (2000). Is hospital competition socially wasteful? *The Quarterly Journal of Economics*, 115(2):577–615.

Kessler, J. and Vesterlund, L. (2015). The external validity of laboratory experiments: The misleading emphasis on quantitative effects. In *Handbook of experimental economic methodology*, pages 391–406. Oxford University Press Oxford, UK.

Khan, S., Sylvester, R., Scott, D., and Pitts, B. (2008). Physicians' opinions about responsibility for patient out-of-pocket costs and formulary prescribing in two midwestern states. *Journal of managed care Pharmacy*, 14(8):780–789.

Kiefe, C. I., Allison, J. J., Williams, O. D., Person, S. D., Weaver, M. T., and Weissman, N. W. (2001). Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *Jama*, 285(22):2871–2879.

Kjær, T. (2005). A review of the discrete choice experiment-with emphasis on its application in health care.

Kotwani, A., Wattal, C., Katewa, S., Joshi, P., and Holloway, K. (2010). Factors influencing primary care physicians to prescribe antibiotics in delhi india. *Family practice*, 27(6):684–690.

Krasnik, A., Groenewegen, P. P., Pedersen, P. A., von Scholten, P., Mooney, G., Gottschau, A., Flierman, H. A., and Damsgaard, M. T. (1990). Changing remuneration systems: effects on activity in general practice. *British Medical Journal*, 300(6741):1698–1701.

Kuhfeld, W. F. (2005). Experimental design, efficiency, coding, and choice designs. *Marketing research methods in sas: Experimental design, choice, conjoint, and graphical techniques*, pages 47–97.

Léger, P. T. (2008). Physician Payment Mechanisms. In Lu, M. and Jonsson, E., editors, *Financing Health Care: New Ideas for a Changing Society*, pages 149–176. Wiley-VCH Press, Weinheim (Germany).

Lerner, C., Claxton, K., et al. (1994). Modelling the behaviour of general practitioners: a theoretical foundation for studies of fundholding. Technical report.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2):153–174.

Levitt, S. D. and List, J. A. (2008). Homo economicus evolves. *Science*, 319(5865):909–910.

List, J. (2008). Homo experimentalis evolves. *Science*, 321(5886):207–208.

List, J. A. (2007). Field experiments: a bridge between lab and naturally occurring data. *The BE Journal of Economic Analysis & Policy*, 5(2).

List, J. A., Sadoff, S., and Wagner, M. (2011). So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4):439.

List, J. A. and Samek, A. S. (2015). The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption. *Journal of health economics*, 39:135–146.

Lohr, K. N., Brook, R. H., Kamberg, C. J., Goldberg, G. A., Leibowitz, A., Keesey, J., Reboussin, D., and Newhouse, J. P. (1986). Use of medical care in the rand health insurance experiment: diagnosis-and service-specific analyses in a randomized controlled trial. *Medical care*, 24(9):S1–S87.

Louviere, J., Street, D., Carson, R., Ainslie, A., Deshazo, J., Cameron, T., Hensher, D., Kohn, R., and Marley, T. (2002). Dissecting the random component of utility. *Marketing letters*, 13(3):177–193.

Louviere, J. J. and Eagle, T. (2006). Confound it! that pesky little scale constant messes up our convenient assumptions. In *Sawtooth Software Conference*. Sawtooth Software Inc.

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated choice methods: analysis and applications*. Cambridge university press.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D., and Irwin, J. (1999). Combining sources of preference data for modeling complex decision processes. *Marketing Letters*, 10(3):205–217.

Louviere, J. J., Street, D., Burgess, L., Wasi, N., Islam, T., and Marley, A. A. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of choice modelling*, 1(1):128–164.

Low, H. and Meghir, C. (2017). The use of structural models in econometrics. *Journal of Economic Perspectives*, 31(2):33–58.

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

Lucas, F. L., Sirovich, B. E., Gallagher, P. M., Siewers, A. E., and Wennberg, D. E. (2010). Variation in cardiologists' propensity to test and treat: is it associated with regional variation in utilization? *Circulation: Cardiovascular Quality and Outcomes*, 3(3):253–260.

Luce, D. and Suppes, P. (1965). Preference, utility, and subjective probability. *Handbook of Mathematical Psychology*, 3:252–410.

Luce, R. D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons.

Lundin, D. (2000). Moral hazard in physician prescription behavior. *Journal of health economics*, 19(5):639–662.

Lurie, N., Kamberg, C. J., Brook, R. H., Keeler, E. B., and Newhouse, J. P. (1989). How free care improved vision in the health insurance experiment. *American journal of public health*, 79(5):640–642.

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

Machina, M. J. (1985). Stochastic choice functions generated from deterministic preferences over lotteries. *The economic journal*, 95(379):575–594.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., and Leibowitz, A. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *The American economic review*, pages 251–277.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229–254.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of economic perspectives*, 14(3):115–136.

Marschak, J. (1960). Binary choice constraints and random utility indicators. In Arrow, K., editor, *Stanford symposium on mathematical methods in the social sciences*, pages 312–329. Stanford University Press.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

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

McFadden, D. (2001). Economic choices. *American Economic Review*, 91(3):351–378.

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

McFadden, D., Talvitie, A., Cosslett, S., Hasan, I., Johnson, M., Reid, F., and Train, K. (1977). Demand model estimation and validation. *Urban Travel Demand Forecasting Project, Phase*, 1.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470.

McGuire, T. G. (2000). Physician Agency. In Cuyler and Newhouse, editors, *Handbook of Health Economics, Vol. 1 A*, pages 461–536. North-Holland, Amsterdam (The Netherlands).

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108(26):10415–10420.

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

Moffatt, P. G. (2015). *Experimetrics: Econometrics for experimental economics*. Macmillan International Higher Education.

Morgenstern, O. (1954). Experiment and large scale computation in economics. *Economic activity analysis*, pages 484–549.

Morikawa, T. (1989). *Incorporating stated preference data in travel demand analysis*. PhD thesis, Massachusetts Institute of Technology.

Mosteller, F. and Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59(5):371–404.

Neumann, J. v. and Morgenstern, O. (1944). Theory of games and economic behavior. 1944.

Newhouse, J. P. (1974). A design for a health insurance experiment. *Inquiry*, 11(1):5–27.

Newhouse, J. P. (1996). Reimbursing health plans and health providers: efficiency in production versus selection. *Journal of economic literature*, 34(3):1236–1263.

Newhouse, J. P. (2004). Consumer-directed health plans and the rand health insurance experiment. *Health Affairs*, 23(6):107–113.

NobelPrize.org. (2002). The sveriges riksbank prize in economic sciences in memory of alfred nobel 2002. https://www.nobelprize.org/prizes/economic-sciences/2002/summary/.

NobelPrize.org. (2019). The sveriges riksbank prize in economic sciences in memory of alfred nobel 2019. https://www.nobelprize.org/prizes/economic-sciences/2019/summary/.

O'Connell, D. L., Henry, D., and Tomlins, R. (1999). Randomised controlled trial of effect of feedback on general practitioners' prescribing in australia. *Bmj*, 318(7182):507–511.

Orcutt, G. H. and Orcutt, A. G. (1968). Incentive and disincentive experimentation for income maintenance policy purposes. *The American Economic Review*, 58(4):754–772.

Östervall, L. W. (2017). Nudging to prudence? the effect of reminders on antibiotics prescriptions. *Journal of Economic Behavior & Organization*, 135:39–52.

Palfrey, T. R., Holt, C. A., and Goeree, J. (2016). Quantal response equilibrium-a stochastic theory of games.

Pan, H., Cui, B., Zhang, D., Farrar, J., Law, F., and Ba-Thein, W. (2012). Prior knowledge, older age, and higher allowance are risk factors for self-medication with antibiotics among university students in southern china. *PloS one*, 7(7):e41314.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J., and Yazdizadeh, A. (2017). Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning*, 157:63–74.

Reichert, S., Simon, T., and Halm, E. A. (2000). Physicians' attitudes about prescribing and knowledge of the costs of common medications. *Archives of internal medicine*, 160(18):2799–2803.

Reiley, D. H. and List, J. A. (2007). Field experiments in economics. *The New Palgrave Dictionary of Economics*.

Requate, T. and Waichman, I. (2011). "a profit table or a profit calculator?" a note on the design of cournot oligopoly experiments. *Experimental Economics*, 14(1):36–46.

Reynolds, L. and McKee, M. (2009). Factors influencing antibiotic prescribing in china: An exploratory analysis. *Health Policy*, 90(1):32–36.

Rhodes, K. V. and Miller, F. G. (2012). Simulated patient studies: an ethical analysis. *The milbank quarterly*, 90(4):706–724.

Roth, A. E. (1986). Laboratory experimentation in economics. *Economics & Philosophy*, 2(2):245–273.

Roth, A. E. and Kagel, J. H. (1995). *The handbook of experimental economics*, volume 1. Princeton university press, Princeton.

Ryan, M. and Gerard, K. (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied health economics and health policy*, 2(1):55–64.

Samuelson, P. and Nordhaus, W. (1992). Economics. 14th ed.

Sauermann, H. and Selten, R. (1960). An experiment in oligopoly. *General Systems, Yearbook of the Society for General Systems Research*, 5:85–114.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schram, A. and Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review*, 55(6):799–819.

Scott, A., Witt, J., Humphreys, J., Joyce, C., Kalb, G., Jeon, S.-H., and McGrail, M. (2013). Getting doctors into the bush: General practitioners' preferences for rural location. *Social science & medicine*, 96:33–44.

Siciliani, L. (2009). Paying for Performance and Motivation Crowding Out. *Economics Letters*, 103(2):68—71.

Siegel, S. and Fouraker, L. E. (1960). Bargaining and group decision making: Experiments in bilateral monopoly.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, 72(5):923–955.

Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., and Vass, C. M. (2019). Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics*, 37(2):201–226.

Song, K., Scott, A., Sivey, P., and Meng, Q. (2015). Improving chinese primary care providers' recruitment and retention: a discrete choice experiment. *Health policy and planning*, 30(1):68–77.

Sørensen, R. J. and Grytten, J. (2003). Service production and contract choice in primary physician services. *Health Policy*, 66(1):73–93.

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

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

Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., and Finkelstein, A. N. (2014). Medicaid increases emergency-department use: evidence from oregon's health insurance experiment. *Science*, 343(6168):263–268.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

Thurstone, L. L. (1931). The indifference function. *The Journal of Social Psychology*, 2(2):139–167.

Togoobaatar, G., Ikeda, N., Ali, M., Sonomjamts, M., Dashdemberel, S., Mori, R., and Shibuya, K. (2010). Survey of non-prescribed use of antibiotics for children in an urban community in mongolia. *Bulletin of the World Health Organization*, 88(12):930–936.

Train, K. (1978). A validation test of a disaggregate mode choice model. *Transportation Research*, 12(3):167–174.

Train, K. (1986). *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*, volume 10. MIT press.

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

Train, K. E., McFadden, D. L., and Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, pages 109–123.

Train, K. E. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In Scarpa, R. and Alberini, A., editors, *Applications of Simulation Methods in Environmental and Resource Economics*, volume 6 of *The Economics of Non-Market Goods and Resources*, pages 1–16. Springer, Netherlands.

Tversky, A. (1972a). Choice by elimination. *Journal of mathematical psychology*, 9(4):341–367.

Tversky, A. (1972b). Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., et al. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *N Engl J Med*, 360:699–709.

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

Wallis, W. A. and Friedman, M. (1942). The empirical derivation of indifference functions. *Studies in mathematical economics and econometrics in memory of Henry Schultz*, pages 175–189.

Wang, J., Iversen, T., Hennig-Schmidt, H., and Godager, G. (2020). Are patient-regarding preferences stable? evidence from a laboratory experiment with physicians and medical students from different countries. *European Economic Review*, page 103411.

Wilson, B. J. (2014). The meaning of deceive in experimental economic science. In *The Oxford Handbook of Professional Economic Ethics*. Oxford University Press New York, NY.

Wright, J. R. and Leyton-Brown, K. (2017). Predicting human behavior in unrepeated, simultaneous-move games. *Games and Economic Behavior*, 106:16–37.

Xue, H., Shi, Y., Huang, L., Yi, H., Zhou, H., Zhou, C., Kotb, S., Tucker, J. D., and Sylvia, S. Y. (2019). Diagnostic ability and inappropriate antibiotic prescriptions: a quasi-experimental study of primary care providers in rural china. *Journal of Antimicrobial Chemotherapy*, 74(1):256–263.

Yu, M., Zhao, G., Lundborg, C. S., Zhu, Y., Zhao, Q., and Xu, B. (2014). Knowledge, attitudes, and practices of parents in rural china on the use of antibiotics in children: a cross-sectional study. *BMC infectious diseases*, 14(1):112.

Papers

I

I | Paper I

Cheo *et al. Health Economics Review* (2020) 10:33
https://doi.org/10.1186/s13561-020-00290-z

Health Economics Review

RESEARCH Open Access

The effect of a mystery shopper scheme on prescribing behavior in primary care: Results from a field experiment

Roland Cheo[1], Ge Ge[2]*, Geir Godager[2,3], Rugang Liu[4,5], Jian Wang[6,7] and Qiqi Wang[8]

Abstract

Background: Health care systems in many countries are characterized by limited availability of provider performance data that can be used to design and implement welfare improving reforms in the health sector. We question whether a simple mystery shopper scheme can be an effective measure to improve primary care quality in such settings.

Methods: Using a randomized treatment-control design, we conducted a field experiment in primary care clinics in a Chinese city. We investigate whether informing physicians of a forthcoming mystery shopper audit influences their prescribing behavior. The intervention effects are estimated using conditional fixed-effects logistic regression. The estimated coefficients are interpreted as marginal utilities in a choice model.

Results: Our findings suggest that the mystery shopper intervention reduced the probability of prescribing overall. Moreover, the intervention had heterogeneous effects on different types of drugs.

Conclusions: This study provides new evidence suggesting that announced performance auditing of primary care providers could directly affect physician behavior even when it is not combined with pay-for-performance, or measures such as reminders, feedback or educational interventions.

Keywords: Field experiment, Primary care, Prescription, Information and product quality, Social responsibility

JEL-Classification: C93; I11; I18; L15; M14

Background

As noted by Arrow [1], asymmetric information about product quality is a fundamental characteristic of the medical care market. The providers of health services are experts who typically hold information that is superior to that of the patients and the payers of the services. When the presence of asymmetric information limits provider quality assurance, it affects the providers' incentive for quality delivery. Recent health reforms in many countries are designed to encourage quality improvements by linking financial incentives to observable indicators of quality. When feasible, policymakers often take advantage of advances in information and communication technology in developing of policy measures, such as by designing mechanisms for provider payment based on routinely collected data on provider activity and performance. The Quality and Outcomes Framework (QOF) in the United Kingdom is an example of an extensive pay-for-performance program that relies on advanced infrastructure in the form of health registers and patient lists when measuring provider performance.

Many health care systems are still characterized by limited availability of provider performance data and patient registers. Without routinely collected performance data, the implementation of an advanced pay-for-performance system is not feasible in all countries. In the presence

*Correspondence: gege@medision.uio.no
[2]Department of Health Management and Health Economics, University of Oslo, P.O. Box 1089 Blindern, 0317 Oslo Norway
Full list of author information is available at the end of the article

of asymmetric information on service quality, the degree of asymmetry can be influenced by introducing simple auditing schemes that do not rely on routinely collected register data on every provider. Such performance auditing is often designed to improve the quality of services by evaluating the quality against standards and can be implemented without necessarily linking financial incentives to performance. As described by Dranove [2], health plans and hospitals frequently contribute to quality assurance mechanisms by collecting and voluntarily disclosing quality information. While knowledge of hospital performance is a necessity in modern hospital management, auditing primary care physicians more likely requires an external initiative. As reviewed by Ivers and Oxman [3], most intervention studies on auditing focus on the effect of auditing when combined with other measures, such as *reminders* [4, 5], *feedback* [5–11] or *educational interventions* [12, 13]. In a recent study by Östervall [14], however, the effect of auditing primary care physicians' practice in Sweden is separated from the effect of reminding physicians and patients about the inappropriate use of antibiotics. The reminders are found to have a substantial effect on prescribing, whereas introducing audits does not significantly influence physician prescribing behavior. Our study relates to the study by [14] in that we aim to quantify the effect of announced auditing on prescribing behavior.

We question whether announced auditing in the form of a mystery shopper scheme can be an effective measure to improve health care quality in primary care markets where routinely collected performance data is not available, and we propose to identify this effect by applying the method of mystery shopping in a randomized treatment-control design. Mystery shopping is frequently used for performance measurement to reduce the asymmetry of information in industries organized as chains. Mystery shoppers interact with product or service providers following specific scripts of tasks and report back detailed information on the experience. A mystery shopper scheme thus enables decision makers to acquire performance information on subdivisions of an organization, which can be used for pure monitoring purposes as well as performance-based payment [15]. Mystery shopper schemes can be customized to suit different purposes, and using mystery shoppers to collect information for research purposes has become more common in recent years. The key element of a mystery shopper is that parties that are audited are not informed about the mystery shopper's identity and when audits will occur. Decades ago, the mystery shopping approach was adopted in the health domain to study provider behavior, and it has been proved valuable to society [16]. In a health context, mystery shoppers are commonly referred to as *pseudopatients*, *simulated patients*, *standardized patients* or *surrogate patients*. Using pseudopatients involves an element of deception, which generally involves careful ethical considerations, especially in the health research domain. Application of this method can be ethically justified, however, as long as individuals' confidentiality is protected, risks to the research subjects are minimal and the research is potentially valuable in furthering our knowledge on the subject [17]. This project was subject to ethical assessment and was approved by the Data Protection Official for Privacy in Research, Norwegian Social Science Data Services, which serves as the institutional review board for the University of Oslo.[1]

The quality measure applied in our study is the physician's prescribing behavior when the patient presents a specific set of symptoms. The symptoms presented by the pseudopatients in this study are symptoms of a mild common cold. As reviewed by Simasek and Blandino [18] and Allan and Arroll [19], medical studies on various treatments for the common cold do not show clear benefits, and adverse side effects from inappropriate treatment can potentially harm patients. In addition, financial costs paid by patients when purchasing medications contribute negatively to patients' overall welfare. Hence, whether or not medication is prescribed is an observable and convenient quality measure in our specific study setting. In general, prescribing behavior in primary care is a highly relevant quality aspect, as inappropriate prescribing of medication has become a global public health challenge. According to the World Health Organization [20], more than half of medical prescriptions worldwide are inappropriate, causing not only adverse health outcomes but also increasing health expenditures. A typical example is the overprescribing of antibiotics. This practice is common in many countries, leading to widespread resistance against medications used for treatable bacterial infections [21–24]. Governments are increasingly implementing guidelines and regulations to curb such misuse of medications. The literature reveals, however, that antibiotics are prescribed too often, even in the presence of guidelines and gatekeeping [25–27].

We conducted a field experiment on physicians from small private clinics in Jinan, China. The majority of the physicians in our sample are owners or co-owners of the clinics. The profit from medication sales is often their main source of income, as they most often do not charge consultation fees. We randomized clinics into either a treatment or control group. We applied a similar audit methodology and script as Currie et al. [26, 27] and announced a forthcoming mystery shopper audit only to clinics in the treatment group. Physicians' prescribing behavior was categorized into four types, corresponding to the inclusion of antibiotics, other prescription drugs

[1] Case number: 44243.

(Other Rx), over-the-counter drugs (OTC), and alternative and nonpharmacological treatments (Alternatives) in the prescription. We found that the mystery shopper intervention unambiguously reduced the mean marginal utility of prescribing drugs and thereby the probability of prescribing overall. Moreover, the average reduction in prescribing was mostly driven by reductions in Other Rx and OTC.

This paper contributes to the literature using field experiments to acquire knowledge on key mechanisms in health service delivery. To our knowledge, this is the first paper to examine whether providers change behavior in response to preannouncement of a mystery shopper audit. In addition to this innovation, a strength of the paper is the use of a randomized treatment-control design to identify the intervention effect. This paper provides new evidence suggesting that auditing primary care providers can directly affect physician behavior, even when it is not combined with pay-for-performance, or other measures such as reminders, feedback or educational interventions.

Theoretical background and hypotheses

The patient-physician relationship is commonly described as a case of (imperfect) agency [28]. The patient (principal) consults the physician (agent), who is an expert with superior information regarding health and expected treatment effects.[2] Under perfect physician agency, the optimal treatment for the patient will coincide with the optimal treatment option for the physician. In our study setting, income from selling medications comprises a substantial share of physicians' income. Financial incentives to prescribe drugs result in conflicting objectives between patients and physicians, as it becomes costly to always behave as a perfect agent on behalf of the patient.

We studied the case of a patient with a common cold, where prescribed medication is not expected to contribute to positive health benefits. When the patient needs to pay out-of-pocket for medication, one may argue that a rational patient would refrain from drug purchase if the patient and physician were equally well informed. Upon seeing a patient with minor symptoms of a common cold, the physician decides whether or not to prescribe medication.

We assume that the patient passively accepts the physician's treatment recommendation and indicate the prescribing choice by a, where $a = 1$ if the physician chooses to prescribe, and $a = 0$ otherwise. We assume that the physician's net profit, π, from prescribing is positive. The physician's choice affects patient's net benefit, $V(a)$, defined by health benefit measured in money minus cost of medication. In the case of the common cold, prescribing reduces the patient's net benefit, $V(1) < V(0)$, since prescribed medication is not expected to provide positive health benefits, and the patient incurs costs.

We assume that physicians are partly altruistic, and, similar to Farley [29], we include the physician's concern for the patient's overall well-being when specifying the physician's objective. When the physicians are informed of a forthcoming mystery shopper audit, it implies that their service quality and professionalism can be acknowledged by a relevant institution. We propose that the alternative *not prescribe*, being medically appropriate and beneficial to the patient while yielding low physician profit, can become more rewarding after receiving information of a forthcoming mystery shopper audit: In the presence of a mystery shopper scheme, information on medical decisions will reach a broader audience than what is the case in a conventional physician-patient encounter. As described by Bénabou and Tirole [30], the physician's objective might include other elements, such as "recognition by others" or "social stigma" in conjunction with profit motive and concern for patients, and therefore, they may behave differently when a mystery shopper scheme is introduced.

We indicate the existence of a mystery shopper scheme by T, where $T = 1$ when a mystery shopper scheme exists and $T = 0$ otherwise. The element of "recognition by others" or "social stigma" can be included additively in the physician objective as a function $S(a; T)$, which introduces a stigma effect from prescribing in the context of a mystery shopper scheme. We assume that in the absence of a mystery shopper scheme ($T = 0$), stigma does not affect the provider objective, i.e., $S(1; 0) = S(0; 0)$. In the case of mystery shopping ($T = 1$), however, prescribing unnecessary medication results in a negative stigma effect: $S(1; 1) < S(0; 1)$. The objective for a physician who cares about social stigma besides profit and patients' net benefit can be expressed as:

$$U(a; T) = \pi a + bV(a) + cS(a; T) \qquad (1)$$

where the preference parameter, $b > 0$, indicates the weight the physician attaches to the patient's net benefit, and $c \geq 0$ indicates the preference weight of social stigma in the physician's objective function. We assume that physicians behave as if they are maximizing (1).

In the absence of a mystery shopper scheme ($T = 0$) where $S(1; 0) = S(0; 0)$, a physician would prescribe if $U(1; 0) > U(0; 0)$, where $U(1; 0) = \pi + bV(1) + cS(1; 0)$ and $U(0; 0) = bV(0) + cS(0; 0)$. Under the assumption that physicians maximize (1), physicians with low altruism, $b < \frac{\pi}{V(0)-V(1)}$, will prescribe; those with a high altruism, $b > \frac{\pi}{V(0)-V(1)}$, will not prescribe; and physicians with $b = \frac{\pi}{V(0)-V(1)}$ will be indifferent to prescribing choices. In the case of preference heterogeneity in the population of physicians, preference variation will cause practice variations in terms of heterogeneous prescribing choice for a given patient.

[2] There is no third party payer in our study.

In the presence of a mystery shopper scheme ($T = 1$), a physician's decision depends on the sign of $U(1;1) - U(0;1)$, where $U(1;1) = \pi + bV(1) + cS(1;1)$ and $U(0;1) = bV(0) + cS(0;1)$. It can be shown that in a population of physicians that maximize (1) with varying b, introducing a mystery shopping scheme will cause a change in behavior for a subset of physicians.

The result can be illustrated by studying the optimal choice for the physician who is indifferent to prescribing in the absence of mystery shopping, with the altruism parameter given by $b^0 = \frac{\pi}{V(0)-V(1)}$. Introduction of a mystery shopper scheme will cause this physician to strictly prefer the alternative *not prescribe*, since $U(1;1) - U(0;1) = c(S(1;1) - S(0;1)) < 0$. The result is illustrated in Fig. 1. The two lines represent incremental utility from prescribing, with and without a mystery shopper scheme. Under the assumption that physicians maximize (1), physicians choose *prescribe* whenever $U(1;T) - U(0;T) > 0$ and *not prescribe* whenever $U(1;T) - U(0;T) < 0$. We see that in the absence of mystery shopping, the physician's incremental utility from choosing to prescribe is negative for physicians with $b > b^0$. Introducing mystery shopping shifts the incremental utility curve downwards, and now indifference in the prescribing decision occurs for a lower level of altruism $b = b^1$, implying that a mystery shopper scheme will cause a change in behavior for a subset of physicians with altruism parameters $b \in \left(b^1, b^0\right)$.

Based on the model results, we specify our main hypothesis:

The probability of physicians prescribing medication to patients with symptoms of a minor common cold will be reduced by announcing a mystery shopper scheme.

A plausible extension of the model is to allow for heterogeneous stigma effects over different types of prescribed medications. Therefore, a secondary hypothesis can be specified:

The effects of announcing a mystery shopper scheme are heterogeneous over different types of prescribed medications.

We test our hypotheses in a setting where primary care physicians earn a net profit from selling their prescribed drugs and the patients pay the full price out-of-pocket.

Methods

Experimental design and procedure

The literature reveals that Chinese physicians prescribe medication, especially antibiotics, when they should not [25–27]. An important cause of medication overprescribing in China is the financial incentives. Revenues from selling medication have become more important to hospitals since the early 1980s, when the government began to reduce financial support to hospitals [31]. For physicians in private clinics, profit from medication sales is often the main source of income, as they most often do not charge

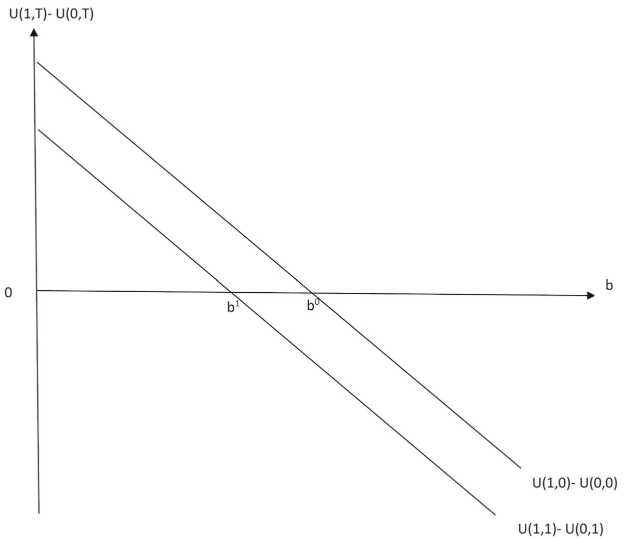


Fig. 1 Incremental utility from prescribing, with and without mystery shopping

consultation fees. To mitigate incentives for overprescribing in China, various reforms have been implemented by the Chinese government since 2009. In general, most of the regulation and reforms target private and public hospitals rather than private clinics. In 2010, the Health Ministry separated doctors' pay from prescription drug sales to curb the widespread prescription of antibiotics in hospitals [32]. In 2011, the Health Ministry also regulated antibiotic prescription for hospitalized patients and outpatients and set targets at less than 60% and 20% of all prescriptions. In addition, antibiotic utilization in hospitalized patients were set at less than 40 daily defined doses per 100 patient days [33]. However, these reforms have not proven effective [34]. We conducted a randomized field experiment in private clinics in China to investigate if pre-announcement of a mystery shopper audit could improve the quality of primary health care services.

Sample and randomization

Our field experiment was performed in Jinan, the capital city of Shandong province in China. By performing the experiment among small walk-in private clinics where no patient ID is required and no patient records are kept, we could randomly assign pseudopatients to clinic visits. It might be more challenging to conduct a similar field experiment in a country where durable physician-patient relations, often formalized as patient list systems, are common. We received support from the School of Public Health at Shandong University and Qilu Health Service Center, which is affiliated with the largest public hospital in Jinan (Qilu Hospital); and this support added substantial credibility to the mystery shopper intervention.

From official Chinese registers in the Health and Family Planning Commission of Jinan Municipality, we identified 118 primary care clinics in Jinan based on these criteria: the clinic is for-profit with only one practicing physician, is located within the five districts of Jinan city,[3] has a valid license on the date of the experiment, and provides general medicine.[4] From the list of suitable clinics, we then randomly assigned 48 clinics to the control group, 48 clinics to the treatment group, and the remaining 22 clinics served as backups. In case any visited clinic was permanently closed, one random clinic from the 22 backups could replace the closed one. According to our prior information on prescribing in primary care, we expected that medications would be prescribed in a majority of consultations. We aimed to assess whether the intervention could generate a substantial reduction in inappropriate prescribing. Our sample size was based on power calculations. With a sample size of 96, the likelihood of correctly rejecting the null-hypothesis (the intervention has no effect) in a Pearson's χ^2 test, given an effect size of 30 percentage points, is 80% when significance level is set at the conventional level of 5%.

Mystery shopper audit

Following Moriarty et al. [35] and Bisgaier and Rhodes [36], we carried out two mystery shopper audits on all 96 clinics in November and December 2015. A time-line of the field experiment is provided in Table 1. Throughout the first audit, we collected baseline data on the characteristics of the clinics and the practicing physicians and their prescribing behavior. Based on the second audit, we compared differences in prescribing behavior between the treatment and control groups.

In both audits, pseudopatients presented symptoms of the common cold to the physician according to a script (see Appendix C) and a protocol (see Appendix D). They described their symptoms as "feel fatigued, have a low grade fever, slight dizziness, a sore throat and a poor appetite", and they told the physician that their body temperature was 37 °C in the morning. The pseudopatients were explicitly instructed not to say to the physician that they have a cold. They allowed the physician to measure their temperature and/or visually inspect their throat. The pseudopatients were strictly instructed to refuse any other treatment or diagnostic test by the physician. If the physician prescribed any medication, the patient was instructed to memorize the names and the pharmaceutical companies of all the medications prescribed. The patient was then to ask for the price of the prescription. The budget for drug purchasing was set at 20 Yuan. The cost for a one-time clinic visit due to a mild common cold would typically be lower. It is important to note that this budget was never revealed to the physician, and the patient's purchasing decision was announced after the drugs were prescribed. Hence, physicians' prescribing behavior was measured by drugs prescribed, not drugs purchased.

A pseudopatient was always accompanied by a fellow student during the audits. The fellow students observed the number of additional patients in the waiting room, the number of additional physicians and patients in the office, the gender and age of the practicing physician and helped the pseudopatient memorize the medication names. The pseudopatient and the accompanying student completed a data collection sheet together after leaving the clinic.

Mystery shopper intervention

The intervention of announcing a forthcoming mystery shopper audit was conducted three weeks before the second audit. A representative of the research project visited

Table 1 Timeline of the field experiment

	Dates
First audit	30th November, 1st December and 2nd December 2015
Intervention	7th December, 8th December and 9th December 2015
Second audit	28th December, 29th December and 30th December 2015

[3]Other districts or counties are too far away.

[4]We excluded dentistry and clinics providing only Chinese medicine because they do not suit our scripted audit scenario.

the clinics in the treatment group one by one to announce the mystery shopper audit. The announcement was made in person by presenting a letter containing information about a current project at Shandong University (see Fig. 3 for an English translation of the project description letter in Appendix E). The project is about quality evaluation of primary care services in Jinan, particularly service, professionalism, and adequacy of treatment. The clinics were informed that an anonymous patient would visit the clinics and collect information about the treatment decision and then evaluate the quality of care. To enhance the credibility of the research project, we offered the clinics three ways to receive feedback from the quality assessment: publicly available feedback (results published on the Shandong University website), feedback in private (results only received by the clinic) or no feedback.[5] The representative read the project description with the physician and ensured that the physician understood the project. In addition, Qilu Health Science Center, affiliated with Shandong University and one of the largest public hospitals (Qilu Hospital) in Jinan, provided an endorsement letter to support the project (see Fig. 4 for an English translation of the endorsement letter in Appendix E). The representative presented the endorsement letter to the physician and left both the stamped project description and the endorsement letter at the clinic.

Training of the pseudopatients

The audits were performed by 12 healthy pseudopatients, each accompanied by a fellow student, recruited from the School of Public Health, Shandong University.[6] Each pseudopatient visited 8 clinics in both audits. Each pair of students (the pseudopatient and the accompanying student) underwent 10 hours of training in total on the 10th and 11th of October 2015. The purpose of the extensive training of the pseudopatients and the accompanying students was to ensure adherence to the script and protocol in order to reduce data variations due to subjective interpretations by the pseudopatients and to enhance the credibility of the pseudopatients so that the physicians are not able to identify them. On the first day, they went through a review of the types of antibiotics and cold medicines on the market. They also had to rehearse and role play using the script. At the end, they practiced filling out the information sheet. Training on the second day involved practice visits to clinics that were not in the 118 identified clinics. To further ensure adherence to the script, the data collection sheets and the physician-patient dialogues from the practice visits were discussed. The teams of pseudopatients were randomly assigned to clinics. They did not visit any clinic twice, and they were not informed about whether the clinics were in the treatment or control group.

Ethical considerations

The mystery shopper audit has been used in the health care domain for decades and has been developed into a scientifically sound experimental method that provides unique and valuable knowledge to society in both developing and developed countries (see for example [16, 35, 53, 54]). The use of deception is controversial in science, and there is no unanimous classification across disciplines. The main ethical dilemma in our study is that the healthy pseudopatients provide incorrect information to the physician when describing their state of health. However, following the ethical analysis of Rhodes and Miller [17], it can be ethically justified as long as confidentiality of research subjects is ensured, risks to the research subjects are minimal and the research is potentially valuable to human knowledge.

To ensure the safety of the pseudopatients, they were always accompanied by a fellow student, so a team of two students always traveled together. Furthermore, the pseudopatients, being students of the School of Public Health, had at least one semester of basic medical training and were specifically instructed to refuse any treatment and/or diagnostic test by the physician except for temperature measuring and visual inspection of the throat. To protect the physicians'/clinics' privacy, we generated a unique series of ID numbers identifying each clinic. The sheet of paper linking ID numbers with clinic addresses was destroyed after the visits, so data from the clinics could not be traced to a particular clinic or physician, even by the researchers. In addition, the field experiment also contributed positively to the revenues of the clinics in the study sample, since physicians gained profit by selling prescribed medications.

Data

The 96 clinics were randomized into the treatment and control groups. The map (see Fig. 2) indicating the locations of the clinics in the treatment and control groups provides a rough impression that the treatment and control clinics were randomly scattered throughout Jinan city. Table 2 reports the inclusion of treatment and control clinics over the five districts in the city. There is no significant difference in representation of treatment and control clinics over the districts (p-value= 0.359, χ^2 test).

Table 3 presents summary statistics from our sample at the clinic level. We collected data on the size of the clinics, measured by the number of additional physicians

[5] After the experiment, feedback was indeed provided to those physicians who had opted in. The intention of providing physicians with feedback options is to enhance the credibility of the intervention, not to attempt to draw any causal relation between feedback choice and prescribing behavior. The reason is that feedback choices made by the physicians are endogenously decided, not exogenously assigned to clinics.

[6] In total, 13 pseudopatients and 13 accompanying students were recruited, allowing for one pair of students to serve as a backup pair.

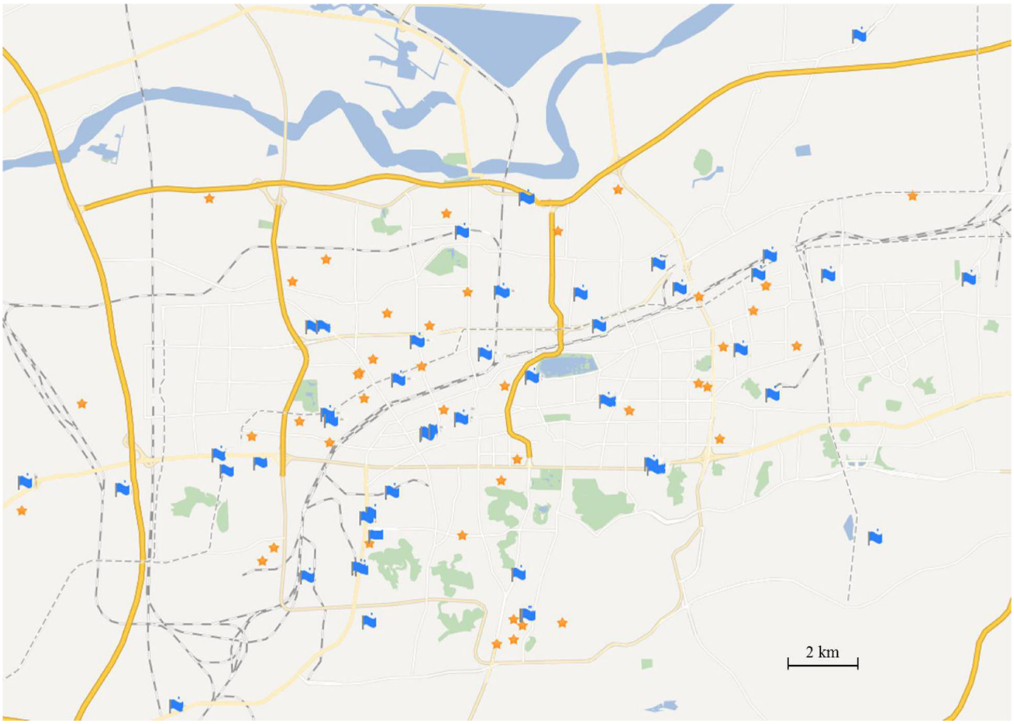


Fig. 2 Map of locations of sampled 96 clinics Stars: the control group; Flags: the treatment group

and patients[7] in a physician's office, and the number of additional patients in the waiting room. Based on the results from Mann-Whitney-Wilcoxon (MWW) tests, there are no significant differences in observed characteristics between clinics in the treatment and control groups in either audit. In the second audit, one clinic in the control group had become a drug store, and one clinic in the treatment group was closed. Therefore, we removed these two clinics from our sample, and data from 94 clinics was used in our study of the second audit. During the experiment, it was discovered that many of the clinics registered as a single-physician unit had more than one physician employed. Due to the design and confidentiality of individual physicians, we cannot ensure a one-to-one match between physicians in the first and second audit.

Table 4 describes the characteristics of the practicing physicians for both audits. We can see that the number of male and female physicians in the control and treatment groups are similar for the first audit, while there are more females in the control and more males in the treatment group in audit two. Physicians' age was observed and categorized into four groups. Physicians' characteristics could potentially influence their prescribing patterns, and the intervention might therefore have varying impacts on physicians of different genders and ages. For this reason, we control for physician characteristics when we analyze the intervention effect on physician prescribing behavior.

Based on the prescribing data and reviews of treatments for the common cold [18, 19], we categorize physicians' prescribing into four prescription types: antibiotics, other prescription drugs (Other Rx), over-the-counter drugs (OTC), and alternative and nonpharmacological treatments (Alternatives). Despite the fact that the common cold is a viral illness, antibiotics are frequently prescribed for this illness in China. Medical studies have recorded

Table 2 Locations of sampled 96 clinics

	District 1	District 2	District 3	District 4	District 5	Total
Control	10	11	3	12	12	48
Treatment	12	5	7	14	10	48

[7] It is not uncommon in China that patients wait in the physician's office, especially for small clinics.

Table 3 Clinic characteristics

Variables	Control Mean	Sd.	N	Treatment Mean	Sd.	N	MWW *p*-value
Audit 1							
# of additional physicians in the office	0.333	0.808	48	0.354	0.758	48	0.792
# of additional patients in physician's office	0.979	1.436	48	0.938	1.359	48	0.865
# of additional patients in the waiting room	0.250	0.636	48	0.375	0.672	48	0.182
Audit 2							
# of additional physicians in the office	0.617	1.054	47	0.447	0.880	47	0.451
# of additional patients in physician's office	1.191	1.313	47	1.511	1.932	47	0.653
# of additional patients in the waiting room	0.234	0.560	47	0.213	0.508	47	0.826

evidence that antibiotics provide no benefit and can potentially cause harm by increasing bacterial resistance [37, 38]. Other prescription drugs are not recommended when only mild cold symptoms are presented due to risks of adverse effects and unclear benefits, especially in our experiment where no other diagnostic tests were ordered other than a visual inspection of the patient's throat and a measurement of the temperature. While there is no cure for the common cold, most over-the-counter drugs are directed at relieving certain symptoms. Examples are paracetamol (acetaminophen), ibuprofen or other pain relievers for body aches or a headache and decongestant nasal sprays. Considering the side effects, they are in general not recommended given the absence of symptoms. In general, OTCs include a wide range of medicines of which the benefits are unclear but likely small in adults. Alternative and nonpharmacological treatments include, for example, vitamin C supplements and cough drops, and the benefits are likely absent [19]. In addition to unclear health benefits, prescriptions for any medications increase patients' financial costs.

Table 5 summarizes physicians' prescribing behavior from the control and treatment groups in both audits. The large majority of physicians prescribed at least one type of drug to the patients in both audits. In the second audit, where all the physicians in the control group provided some medication to the pseudopatients, significantly fewer physicians (χ^2 test p-value=0.022) in the treatment group (89.4%) provided medication. There were only a few physicians who did not prescribe any drug at all: 3 in the control group and 6 in the treatment group in the first audit, and 0 in the control group and 5 in the treatment group in the second audit. From the first audit data, we can see that OTCs were the most prescribed alternative, with more than 80% of physicians choosing to prescribe them. OTCs were followed by antibiotics, which were prescribed by around two-thirds of physicians. This observation clearly confirms the prevalence of antibiotic overprescribing in China in the case of the common cold and is similar as the previously reported rate in experimental studies [26]. The less commonly prescribed treatments were Other RX and Alternatives, provided by approximately 12% and 3% of physicians, respectively. From the second audit data, we observed higher prescription rates of antibiotics and lower rates of Other Rx, OTC and Alternatives in the treatment group compared to the control group. Overall, the qualitative prescribing pattern described by the ranking of prescribing rates of the four

Table 4 Physician characteristics

Variables		Audit 1 Control Freq.	N	Treatment Freq.	N	Audit 2 Control Freq.	N	Treatment Freq.	N
Gender	Male	24	48	23	48	19	47	25	47
	Female	24	48	25	48	28	47	22	47
Age	≤30	2	48	2	48	2	47	2	47
	[31,40]	24	48	26	48	16	47	26	47
	[41,50]	12	48	18	48	21	47	12	47
	≥51	10	48	2	48	8	47	7	47

Age is categorized into four levels: younger than or equal to 30 years old; between 31 to 40 years old; between 41 to 50 years old; older than or equal to 51 years old

Table 5 Physician prescribing behavior

	Control				Treatment			
Variables	**Mean**	**Sd.**	**Freq.**	**N**	**Mean**	**Sd.**	**Freq.**	**N**
Audit 1								
Prescribe any drug	93.8%	0.245	45	48	87.5%	0.334	42	48
Antibiotics	62.5%	0.489	30	48	66.7%	0.476	32	48
Other Rx	12.5%	0.334	6	48	12.5%	0.334	6	48
OTC	85.4%	0.357	41	48	81.3%	0.394	39	48
Alternatives	2.1%	0.144	1	48	4.2%	0.202	2	48
Audit 2								
Prescribe any drug	100%	0	47	47	89.4%	0.312	42	47
Antibiotics	57.4%	0.500	27	47	68.1%	0.471	32	47
Other Rx	23.4%	0.428	11	47	10.6%	0.312	5	47
OTC	85.1%	0.360	40	47	76.6%	0.428	36	47
Alternatives	8.5%	0.282	4	47	6.4%	0.247	3	47

Prescribe any drug: prescribe at least one type of drug; Other Rx: Other prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments

types of drugs is the same in both audits and for both groups.

Empirical strategy

The decision to prescribe a drug to a patient is a standard discrete economic choice [39], and the choice modelling literature comprises a rich toolbox for analyzing how individuals' choice combinations are affected by the characteristics of the available alternatives, as well as differences in context [40]. Choice models are now commonly used in studies applying experimental data (see for example [11, 55–58]). We examine and quantify the intervention effect on prescribing choices of the individual physicians. The prescribing choice can, without loss of generality, be split into a sequence of choices, where the physician first decides whether or not to prescribe antibiotics, and then they decide whether or not to include other types of drugs, one by one, until a complete prescription is chosen.

We estimate the intervention effects using a standard conditional fixed-effects logit model, which allows us to quantify the observed heterogeneity of prescribing patterns across different categories of drugs with and without the intervention. The physician's prescribing decision is indicated by y_{it}, where we use the indices $i = 1, 2, ..., N$ for physician, and $t = 1, 2, 3, 4$ for the types of drug that physician i decides to include or exclude in the medical treatment of the patient. The physician's prescribing decision for each drug type is a binary choice variable such that $y_{it} = 1$ if the physician prescribes drug t, and $y_{it} = 0$ otherwise. Let the mean marginal utility for physician i of prescribing drug t be denoted by v^*_{it}. We allow v^*_{it} to depend on whether or not physician i is in the treatment group, by defining it as $v^*_{it} = v_t\,[1 + \gamma_t I_i]$, where v_t denotes the mean marginal utility of prescribing drug t for physicians without the intervention. The potential effects of the intervention are captured by the inclusion of the intervention dummy I_i. The intervention effect γ_t is allowed to vary over the different types of drugs. In the special case where the intervention effects, γ_t, are all zero, we have $v^*_{it} = v_t$ for physicians in both the treatment and control groups. Letting α_{it} be any unobservable heterogeneity that is fixed for physician i when deciding on whether to prescribe drug t, the conditional logit probability of physician i prescribing drug t is given by:

$$Pr\,(y_{it} = 1) = \frac{\exp\left(\alpha_{it} + v^*_{it}\right)}{\exp\,(\alpha_{it}) + \exp\,(\alpha_{it} + v_{it}*)} \tag{2}$$

From Eq. (2) and the definition of v^*_{it}, we see that when the intervention does not have any effect, i.e., $\gamma_t = 0$, we have $v^*_{it} = v_t$. This means that the marginal utility of prescribing drug t, and thus the probability of prescribing, do not differ between the treatment and the control groups. The γ_t parameter captures the causal effect of the intervention on the marginal utility of prescribing. When γ_t is positive (negative), the interpretation is that the probability that the physicians' treatment recommendation includes drug t is positively (negatively) affected by the intervention.

A convenient feature of the conditional fixed-effects logit model is that the fixed effects α_{it} are conditioned out of the likelihood function, since Eq. (2) reduces to $\frac{\exp(v^*_{it})}{1+\exp(v^*_{it})}$ [41]. By means of a conditional logit model, we may therefore acquire robust estimates of the mean marginal utilities without the intervention, v_t, and the intervention effect, γ_t. Extending from single item choices to choices of bundles is trivial, and a clear deduction is provided by Hole [42]. Applying a robust method that enables analysis of how the intervention affects both the probability of prescribing and the composition of the prescribed drugs is a key feature of the empirical analysis. We estimate the conditional fixed-effects logit models by means of the `clogit` module in Stata 16. The same models are applied to data from the first and second audit, respectively. The intervention effects presented in Table 6 are estimated using data from the second audit. In Appendix A, we present model estimates based on the first audit, providing evidence that physicians in the treatment and control groups did not behave significantly different prior to the intervention.

Table 6 Intervention effects on physician prescribing

	Model 1	Model 2
Panel A: Prescribing pattern (ν_t)		
Antibiotics	0.630***	0.300
	(0.197)	(0.269)
Other Rx	-1.481***	-1.186***
	(0.294)	(0.367)
OTC	1.551***	1.743***
	(0.269)	(0.327)
Alternatives	-2.418***	-2.375***
	(0.531)	(0.688)
Panel B: Average intervention effect (γ)		
	-0.214***	
	(0.064)	
Panel C: Heterogeneous intervention effects (γ_t)		
Antibiotics		0.458
		(0.304)
Other Rx		-0.943**
		(0.434)
OTC		-0.557**
		(0.264)
Alternatives		-0.311
		(0.519)
Number of observations	752	752
Log-Likelihood	-175.4	-173.2
Pseudo R^2	0.327	0.336
AIC	360.9	362.4
BIC	384.0	399.3

Other Rx: Other prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments. Marginal utilities are presented with standard errors in parentheses. The standard errors are adjusted for clustering on groups of physicians by gender and age

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Results

Estimation results from Model 1 and Model 2 are reported in Table 6. The average intervention effect is quantified in Model 1 by assuming the intervention effects on marginal utility of prescribing are fixed over drug types ($\gamma_t = \gamma$, $\forall t \in \{1, 2, 3, 4\}$). The less restrictive Model 2 allows for heterogeneous intervention effects on marginal utilities for the four types of drugs.

In panel A in Table 6, we present the estimates of the mean marginal utilities for each of the four drugs without the intervention, ν_t, with robust standard errors in parentheses. In panel B we report the average intervention effect (γ in Model 1), while heterogeneous intervention effects (γ_t in Model 2) are reported in panel C. For both models, the mean marginal utility of prescribing without the intervention (panel A) differs substantially over the four drug types. The mean marginal utilities are positive for Antibiotics and OTC, and negative for Other Rx and Alternatives. Negative mean marginal utilities are expected for Other Rx and Alternatives, as only a minority of physicians included these types of drugs when treating a pseudopatient. In panel B for Model 1, we see that the estimated average intervention effect is negative and statistically significant. The interpretation is that the mystery shopper intervention caused a reduction in the mean marginal utility, and thus reduced the probability of prescribing drugs to the pseudopatient.[8]

An important aspect of homogeneous effect models like Model 1 is that they may conceal systematic intervention effects in cases where the intervention increases prescribing of some drugs and reduces that of other drugs, implying that the intervention causes behavioral changes. The fact that substitution is a rational response by an economic agent is in general an important issue to consider when conducting experiments in the field [43]. In Model 2, we account for the possibility of substitution by allowing for between-drug-variation in the intervention effect. The heterogeneous intervention effects are presented in panel C. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) do not indicate substantial differences in fit when comparing the two models. However, the hypothesis that the intervention effect, γ_t, is independent of drug type can be rejected ($p - value = 0.0029$, Wald test). The interpretation of the heterogeneous intervention effects is that the announcement of a mystery shopper audit led to a reduction in prescribing of Other Rx and OTC.[9]

[8] Compared to the control group, the mystery shopper intervention reduced the odds of prescribing by 19.2% as measured by the average intervention effect.

[9] Compared to the control group, the mystery shopper intervention reduced the odds of prescribing Other Rx by 61.0%, and reduced the odds of prescribing OTC by 42.7%.

While the conditional logit model provides consistent estimates of the mean marginal utilities and intervention effects, heterogeneity in these parameters are not modeled explicitly. To provide inference on differences in means, while allowing for the possibility of heterogeneous effects, we apply cluster-robust standard errors [44].[10] We estimate the cluster-robust standard errors by grouping physicians according to their gender and age,[11] and a summary of the clustered groups is presented in Table 7. We describe the robustness to alternative criteria for clustering physicians in Appendix B.

To enhance the credibility of the research project in the intervention, we offered the clinics in the treatment group options for receiving feedback of the quality assessment. The three options were: publicly available feedback (results will be published on the Shandong University website), feedback in private (results will only be received by the clinic) or no feedback. Table 8 summarizes physician prescribing behavior by their feedback choices.

Among all 47 physicians, 33 chose to receive no feedback, 11 opted into receive private feedback, while only 3 were willing to publish their evaluation results on the University website. It is worth mentioning that providing physicians with feedback options was not designed to reveal any causal relation between feedback choice and prescribing behavior. The reason is that feedback choices made by the physicians are endogenously decided, not exogenously assigned to clinics. Nevertheless, we report the prescribing behavior of four types of drugs in three feedback groups below and encourage future study designs on the relationships between feedback choices and prescribing behavior.

Discussion

Overprescribing of medications contributes to rising health expenditures and possibly adverse health outcomes. Unlike many previous studies, which have focused only on the overprescribing of antibiotics, we investigated the intervention effects on four types of drugs, including antibiotics, other prescription drugs, over-the-counter drugs, and alternative and nonpharmacological treatments. We quantified the change in composition of prescriptions caused by the intervention. Our results provide evidence that there is substantial variation in prescribing in the case of a mild common cold. Moreover, we found that the average intervention effect is mostly driven by reductions in Other Rx and OTC medications.

The finding that an announcement of a mystery shopper audit does not have significant effect on antibiotic prescribing might have several explanations: The intervention message did not provide any specific assessment criteria on the quality of primary care, and thus physicians' response to the intervention might reflect their prioritization of good quality. Furthermore, prescribing medications that satisfy the patients' expectations might be one of the quality aspects that is considered important to clinics for attracting patients. Due to the limited awareness of antibiotic resistance and lack of knowledge on antibiotic misuse in the population, patients demand antibiotics for self-medication [45–48], and expect primary care physicians to provide antibiotics [25, 49]. Antibiotics are often prescribed due to diagnostic uncertainty as it is difficult to distinguish whether an infection is viral or bacterial, especially at the early stage [50–52].

The credibility of the pseudopatients is a key issue, and, in particular, it is important that the physicians were not able to identify them. It is important to note that the script for the symptom presentation from Currie et al. [26, 27] is deliberately developed so that the physicians cannot observe from an examination whether or not the pseudopatient's presentation is true. The pseudopatient's presentation cannot be proven false objectively. While the announcement of audits might make physicians alert for pseudopatients, vague symptoms of the common cold are so prevalent among patients in general that it is hardly feasible for physicians to dismiss this type of patient. There are obviously other patients who have symptoms that can easily be verified, hence, physicians might feel confident that those patients are not the mystery shoppers. Therefore, our effect estimates should be interpreted in the context of the common cold where the issue of overprescribing is highly relevant.

One might be concerned about information spillover among individual physicians from different groups. Since the intervention was randomly assigned to the clinics, we could not control for the distance between clinics in the treatment and control groups. Even though we were informed that there was no association or organized union of primary care clinics in Jinan where physicians could exchange information on a regular basis, we cannot rule out the possibility of information spillover about the intervention among individual physicians from different groups. Given our experimental design, however, we expect information spillover to have a minor impact, if present at all. If information about the intervention reached the clinics in the control group, they would most likely expect a mystery shopper audit to be preceded by an announcement. Hence, one reasonable strategy for a clinic in the control group is to not change behavior. In the case

[10] One might argue that applying a mixed logit model instead would be preferable, since it would account for the heterogeneity of preferences. The estimation of a mixed logit model with random coefficients would double the number of unknown parameters to estimate, and hence require either a larger number of decision-makers or a larger number of choice occasions for each decision-maker to provide sufficient statistical power.

[11] We group physicians who are older than 40 years old as "Old", and those who are younger than 40 as "Young". This method of grouping reflects the reality in China that physicians in general start their careers in their early 20s and retire around age 60. Moreover, this method provides relatively balanced group sizes. See Appendix B for a detailed description of other clustering criteria and the robustness check of the average intervention effect.

Table 7 Summary of groups

	Young female	Young male	Old female	Old male	Total
Control	13	5	15	14	47
Treatment	13	15	9	10	47
Total	26	20	24	24	94

Notes: Physicians older than 40 years old are grouped as "Old", and those younger than 40 are grouped as "Young"

where a clinic in the control group does change behavior and reduce prescribing, it would result in a smaller intervention effect compared to a situation where information spillover is absent.

Field experiments cannot facilitate a perfectly controlled environment. The behavior of individuals in the treatment group might affect that in the control group in some indirect way which is unobservable to the researchers. While the pseudopatients' behaviors in our experiment were predetermined and therefore unaffected by physician behavior, one can never completely rule out the possibility of behavioral spillovers when conducting experiments in the field.

Our study investigated the intervention effect three weeks after the intervention. More research is needed in order to provide knowledge on the long-term effects of a mystery shopper scheme.

Conclusion

In health care systems where provider performance data and patient registers are not available, interventions that can be implemented to influence asymmetric information and thus improve health care quality are of great interest to policy makers. This study provides new evidence suggesting that announced performance auditing of primary care providers could directly affect physician behavior, even when it is not combined with pay-for-performance or measures such as reminders, feedback or educational interventions. In our study, we conducted a field experiment to assess the impact of a preannounced mystery shopper audit on prescribing behavior in primary care in China. We find that the mystery shopper intervention reduces the probability of prescribing. Moreover, we find that the intervention effects are heterogeneous and differ across types of medicine. We present robust evidence suggesting that a simple announcement of a mystery shopper scheme influences medical treatment decisions. Hence, our results suggest that, upon making medical decisions, physicians have a rich set of motives that do not only include profit and health benefits. More knowledge regarding these motives is needed to develop policies that improve welfare.

Appendix

A First audit

In this section, we show the balance of the randomization by analyzing the "intervention effect" on prescribing behavior in the first audit. The first audit was conducted one week before the intervention, and 96 clinics were randomly grouped into control and treatment. We expect that the assignment of groups does not affect physicians' prescribing behavior. The analytical models used here are identical to those for the second audit analyses. Table 9 below reports the results in terms of marginal utilities. Not surprisingly, no intervention effect was detected in the first audit. In addition to the balance of randomization at clinic level which we demonstrated in "Data" section, the results here reinforce the balance at the individual level, providing evidence that physicians in the treatment and control group did not behave significantly differently prior to the intervention. The standard errors are adjusted for clustering on matched groups of physicians by gender and age. Table 10 summarizes the matched groups.

B Robustness of average intervention effect

Now we check the robustness of the average intervention effect to different criteria of clustering levels on which the standard errors are adjusted for. The physicians were grouped according to their gender (male or female) and age (young or old). In "Results" section, We grouped physicians who were older than 40 years old as "Old", and those who were younger than 40 as "Young" (referred to as

Table 8 Prescribing behavior and feedback choices

	No feedback				Private feedback				Public feedback			
	Mean	Sd.	Freq.	*N*	Mean	Sd.	Freq.	*N*	Mean	Sd.	Freq.	*N*
Antibiotics	72.7%	0.452	24	33	45.5%	0.522	5	11	100%	0	3	3
Other Rx	12.1%	0.331	4	33	0	0	0	11	33.3%	0.577	1	3
OTC	78.8%	0.415	26	33	72.7%	0.467	8	11	66.7%	0.577	2	3
Alternatives	0	0	0	33	18.2%	0.405	2	11	33.3%	0.577	1	3

Table 9 Intervention effects on physician prescribing, audit 1

	Model 1	Model 2
Panel A: Prescribing pattern		
Antibiotics	0.580*	0.511
	(0.306)	(0.264)
Other Rx	-1.967***	-1.946***
	(0.450)	(0.356)
OTC	1.589***	1.768***
	(0.304)	(0.499)
Alternatives	-3.455***	-3.850***
	(0.376)	(0.914)
Panel B: Average intervention effect		
	0.041	
	(0.101)	
Panel C: Heterogeneous intervention effects		
Antibiotics		0.182
		(0.336)
Other Rx		0.000
		(0.267)
OTC		-0.301
		(0.407)
Alternatives		0.715
		(1.480)
Number of observations	768	768
Log-Likelihood	-155.2	-154.8
Pseudo R^2	0.417	0.419
AIC	320.3	325.5
BIC	343.5	362.7

Other Rx: Other prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments. Marginal utilities are presented with standard errors in parentheses. In both models, the standard errors are adjusted for clustering on groups of physicians by gender and age
* ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$)

clustering 2 in Table 11 below). This method of grouping reflects the reality in China that physicians in general start their careers in their early 20s and retire around age 60. Moreover, this method provides relatively balanced group sizes.

Table 12 and Table 13 show the summaries of clustering 1 and clustering 3 where we define "old" and "young" using threshold age 30 and age 50, respectively. Table 14, clustering 4, presents the summary of clustering by all the combinations of age and gender. In all tables, we report the number of physicians in each group.

Applying the same analytical model presented in "Results" section, we tested the robustness of the average intervention effect to each clustering criteria. As it shows in Table 15, no significant intervention effect was detected in the first audit, while in Table 16, the intervention resulted in a significant reduction of mean marginal utility of prescribing. The estimates of the average intervention effects and their significance are consistent across four clustering strategies.

C Scripts of pseudopatient used in first and second audit

Step one: Statement of the Chief Complaint

Patient: Hello, doctor. For the last two days, I've been feeling fatigued. I have been having a low grade fever, slight dizziness, a sore throat, and a poor appetite. This morning, the symptoms worsened so I took my body temperature. It was 37 °C.

If pseudo patients are asked questions about symptoms mentioned in the chief complaint, they are supposed to answer appropriately. If the doctor asks about other symptoms not in the chief complaint, then they should say that there are no such symptoms. Answer NO if asked the following questions:

Do you feel nauseous?
Do you have any phlegm?
Do you have any muscle soreness?
Have you eaten anything bad or unclean recently?
Are you currently taking any medications?
Do you have medication at home?

Step two: Physical Examination

Physician: I'll give you a physical examination/I will now conduct a physical exam.
Physical Examination.

Step three: Physician's Diagnoses and Explanation of Findings

Physician: I'll prescribe [...] for you.
If the doctor wants to give you medication, ask what medication it is.

Table 10 Summary of matched groups, audit 1

	Young female	Young male	Old female	Old male	Total
Control	16	10	8	14	48
Treatment	16	12	9	11	48
Total	32	22	17	25	96

Notes: Physicians older than 40 years old are grouped as "Old", and those younger than 40 are grouped as "Young"

Table 11 Summary of groups, clustering 2

	Young female	Young male	Old female	Old male	Total
Panel A: Audit 1					
Control	16	10	8	14	48
Treatment	16	12	9	11	48
Total	32	22	17	25	96
Panel A: Audit 2					
Control	13	5	15	14	47
Treatment	13	15	9	10	47
Total	26	20	24	24	94

Notes: Physicians older than 40 years old are grouped as "Old", and those younger than 40 are grouped as "Young"

Table 12 Summary of groups, clustering 1

	Young female	Young male	Old female	Old male	Total
Panel A: Audit 1					
Control	0	2	24	22	48
Treatment	2	0	23	23	48
Total	2	2	47	45	96
Panel A: Audit 2					
Control	2	0	26	19	47
Treatment	0	2	22	23	47
Total	2	2	48	42	94

Notes: Physicians older than 30 years old are grouped as "Old", and those younger than 30 are grouped as "Young"

Table 13 Summary of groups, clustering 3

	Young female	Young male	Old female	Old male	Total
Panel A: Audit 1					
Control	20	18	4	6	48
Treatment	25	21	0	2	48
Total	45	39	4	8	96
Panel A: Audit 2					
Control	25	14	3	5	47
Treatment	21	19	1	6	47
Total	46	33	4	11	94

Notes: Physicians older than 50 years old are grouped as "Old", and those younger than 50 are grouped as "Young"

Table 14 Summary of groups, clustering 4

	Female	Male	Female	Male	Female	Male	Female	Male	Total
	$\leq$30	$\leq$30	[31,40]	[31,40]	[41,50]	[41,50]	$\geq$51	$\geq$51	Total
Panel A: Audit 1									
Control	0	2	16	8	4	8	4	6	48
Treatment	2	0	14	12	9	9	0	2	48
Total	2	2	30	20	13	17	4	8	96
Panel A: Audit 2									
Control	2	0	11	5	12	9	3	5	47
Treatment	0	2	13	13	8	4	1	6	47
Total	2	2	24	18	20	13	4	11	94

Notes: Age is categorized into four levels: younger than or equal to 30 years old; between 31 to 40 years old; between 41 to 50 years old; older than or equal to 51 years old

Table 15 Robustness of average intervention effects on physician prescribing, audit 1

	clustering 1	clustering 2	clustering 3	clustering 4
Panel A: Prescribing pattern				
Antibiotics	0.580*	0.580*	0.580***	0.580*
	(0.313)	(0.306)	(0.216)	(0.309)
Other Rx	-1.967***	-1.967***	-1.967***	-1.967***
	(0.485)	(0.450)	(0.596)	(0.446)
OTC	1.589***	1.589***	1.589***	1.589***
	(0.440)	(0.304)	(0.341)	(0.292)
Alternatives	-3.455***	-3.455***	-3.455***	-3.455***
	(0.302)	(0.376)	(0.378)	(0.441)
Panel B: Average intervention effects				
Average intervention effect	0.041	1.073	0.976	0.939
	(0.070)	(0.101)	(0.232)	(0.231)
Number of observations	768	768	768	768
Log-Likelihood	-155.2	-155.2	-155.2	-155.2
Pseudo R^2	0.417	0.417	0.417	0.417

Other Rx: Other prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments. Estimated odds ratios are presented with standard errors in parentheses. The standard errors are adjusted for clustering on groups following four clustering criteria.
* ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$)

Patient: what kind of medication it is? Patient takes a look at the medication and memorizes the name and the pharmaceutical company of the medication.

Ask the physician for information regarding side effects of the medication after 3-4 seconds if the physician does not voluntarily inform you of the side effects.
Patient: Ok. [...] (pause for 3-4 seconds) [...] Does it have any side effects?
If the total is under 20 yuan, buy the medication.
Patient: How much is each medication?

Table 16 Robustness of average intervention effects on physician prescribing, audit 2

	clustering 1	clustering 2	clustering 3	clustering 4
Panel A: Prescribing pattern				
Antibiotics	0.630***	0.630***	0.630***	0.630***
	(0.069)	(0.197)	(0.151)	(0.180)
Other Rx	-1.481***	-1.481***	-1.481***	-1.481***
	(0.117)	(0.294)	(0.132)	(0.277)
OTC	1.551***	1.551***	1.551***	1.551***
	(0.080)	(0.269)	(0.117)	(0.295)
Alternatives	-2.418***	-2.418***	-2.418***	-2.418***
	(0.353)	(0.531)	(0.350)	(0.466)
Panel B: Average intervention effects				
Average intervention effect	-0.214**	-0.214***	-0.214**	-0.214**
	(0.086)	(0.077)	(0.091)	(0.085)
Number of observations	752	752	752	752
Log-Likelihood	-175.4	-175.4	-175.4	-175.4
Pseudo R^2	0.327	0.327	0.327	0.327

Other Rx: Other prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments. Estimated odds ratios are presented with standard errors in parentheses. The standard errors are adjusted for clustering on groups following four clustering criteria. Clustering 2 is applied in the models in the main paper
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

SHANDONG UNIVERSITY

Project Description

Dear Doctor,

I represent the School of Public Health at Shandong University. We have been given the task of evaluating the quality of primary care services in Jinan as part of a research project. We will be coming to clinics and evaluating each clinic on its service, professionalism, and adequacy of treatment. An anonymous patient will come into your clinic and collect information about your treatment decision. You will be evaluated based on the service, the professionalism and the adequacy of treatment experienced by that patient.

Service quality will be indicated by one, two or three stars, and clinics with the highest quality will receive three stars. The collected data on individual clinics will be kept securely, and the clinics can decide if they want the information on quality assessment and clinic name to become publicly available, or if they want to receive the quality assessment of their clinic in a private correspondence, or if they are not interested in receiving any information about the quality assessment.

Publicly available quality assessment

I am here to ask you if you would like to let us present your clinic's quality assessment on a public website? If you choose this alternative, your clinic's name and quality assessment will be on our list of clinics to include in public ranking, and it will be published on the Shandong University website.

Private quality assessment

If you do not choose to participate in public ranking, we can also offer you the opportunity to receive your quality assessment only in private without any public ranking.

Protection of privacy. No participation in feedback of quality assessment.

It is also an alternative to not receive information about the quality assessment at all. In this case the quality assessment of your clinic will not be linked to the name or address of your clinic.

Thank you!

School of Public Health, Shandong University
05.12.2015
(Shandong University School of Public Health stamp)

Fig. 3 English translation of the project description issued by School of Public Health, Shandong University

SHANDONG UNIVERSITY

Endorsement Letter

Shandong University Qilu Health Service Center supports research project "Quality evaluation of primary care services in Jinan" conducted by School of Public Health, Shandong University.

Qilu Health Service Center, Shandong University
04.12.2015
(Shandong University Qilu Health Service Center stamp)

Fig. 4 English translation of the endorsement letter issued by Qilu Health Service Center, Shandong University

If it is over 20 yuan, say,
Patient: Doctor, I do not have enough money with me today, I can come back later to buy.

Step four: Departure

Patient: Thank you!
Physician: You are welcome.

D Experimental protocol for the pseudopatient and accompanying student

Pseudo patient Before entering the clinic

1 Ensure that you have the questionnaire and IDs are correct.
2 Notify in the chat group that you have arrived at the clinic: WRITE Group XXX arrive at Clinic YYYY.

In the clinic

1 DO NOT say to the doctor that you have a cold.
2 MUST say that you had a slight fever.

Out of the Clinic

1 The two of you fill out the data collection sheet.

Accompanying student In the clinic

1 Observe the number of additional patients in the waiting room.
2 Observe the number of additional physicians and patients in the office, the gender and age of the practicing physician.
3 Memorize the name(s) of the medication and the pharmaceutical company.

Out of the Clinic

1 The two of you fill out data collection sheet.

E Letters used in the intervention

The project description letter was issued by School of Public Health, Shandong University.

The endorsement letter was issued by Qilu Health Service Center, Shandong University.

Abbreviations

QOF: Quality and outcomes framework; Rx: Prescription drugs; OTC: Over-the-counter drugs; Alternatives: Alternative and nonpharmacological treatments; MWW: Mann-whitney-wilcoxon; AIC: Akaike information criterion; BIC: Bayesian information criterion

Acknowledgments

We are grateful to Eline Aas, Jeannette Brosig-Koch, Matteo Galizzi Heike Hennig-Schmidt, Tor Iversen, Yiu-Shing Lau, Tron Anders Moger, Julie Riise and Daniel Wiesen for their valuable comments and advice. We thank Dr. Quan Li from Cheeloo College of Medicine, Shandong University, for providing the endorsement letter. We would also like to thank participants at the 37th Nordic Health Economists' Study Group meeting at University of Southern Denmark (2016), the 4th Workshop on Behavioral and Experimental Health Economics at University of Cologne (2016) and iHEA Biennial World Congress at Boston University (2017) for comments and suggestions. Two anonymous reviewers have contributed substantially to improving the article. This project and some of the authors have received funding from the Research Council of Norway (IRECOHEX, No. 231776 and NORCHER, No. 296114) and Natural Science Foundation of China (No. 71373146 and No. 71703081).

Authors' contributions

RC, GGo, JW, and QW designed the experiment. GGe and RL prepared the data. GGe conducted the analysis. RC, GGe, GGo, and JW drafted and revised the manuscript. All author(s) conducted the experiment, interpreted the results, and read and approved the final manuscript.

Funding

Research Council of Norway (IRECOHEX, No. 231776), Research Council of Norway (NORCHER, No. 296114); Natural Science Foundation of China (No. 71373146); Natural Science Foundation of China (No. 71703081).

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

This project was subject to ethical assessment and was approved by the Data Protection Official for Privacy in Research, Norwegian Social Science Data Services (case number: 44243).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Center for Economic Research, Shandong University, 27 Shanda Nanlu, 250100 Jinan, Shandong, P.R. China. [2]Department of Health Management and Health Economics, University of Oslo, P.O. Box 1089 Blindern, 0317 Oslo Norway. [3]Health Services Research Unit, Akershus University Hospital, Sykehusveien 25, 1478 Nordbyhagen Norway. [4]School of Health Policy & Management, Nanjing Medical University, 101 Longmian Avenue, Jiangning District, Nanjing P.R. 211166 China. [5]Center for Global Health, Nanjing Medical University, 101 Longmian Avenue, Jiangning District, Nanjing P.R. 211166 China'. [6]Dong Fureng Institute of Economic and Social Development, Wuhan University, 54 Lishi Hutong, Dongcheng District, Beijing 100010 China. [7]Center for Health Economics and Management in School of Economics and Management, Wuhan University, 299 Bayi Road Wuchang District, Wuhan 430072 China. [8]School of Economics, Xi'an University of Finance and Economics, 360 Changning Avenue, Chang'an District, Xi'an Shanxi 710100 China.

Received: 5 March 2020 Accepted: 10 September 2020
Published online: 24 September 2020

References

1. Arrow KJ. Uncertainty and the welfare economics of medical care. Am Econ Rev. 1963;53:941–69.
2. Dranove D Pauly MV, Mcguire TG, Barros PP, (eds). Health care markets, regulators, and certifiers. Elsevier; 2011. p. 639–90. https://doi.org/10.1016/B978-0-444-53592-4.00010-4.
3. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard - Jensen J, French SD, O'Brien M, Johansen M, Oxman A. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2012;6:. https://doi.org//10.1002/14651858.CD000259.pub3.
4. Baker R, Farooqi A, Tait C, Walsh S. Randomised controlled trial of reminders to enhance the impact of audit in general practice on management of patients who use benzodiazepines. Qual Health Care. 1997;6(1):14–18.
5. Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, Soutter J, Wilsdon J, Matowe L, Needham G, Gilbert F, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. Lancet. 2001;357(9266):1406–9.

6. White P, Atherton A, Hewett G, Howells K. Using information from asthma patients: a trial of information feedback in primary care. BMJ. 1995;311(7012):1065–9.
7. O'connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in australia. Bmj. 1999;318(7182):507–11.
8. Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unützer J, Miranda J, Carney MF, Rubenstein LV. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. Jama. 2000;283(2):212–20.
9. Kiefe C, Allison J, Williams O, Person SD MW, Weissman N. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA. 2001;285:2871–9.
10. Baker R, Fraser RC, Stone M, Lambert P, Stevenson K, Shiels C. Randomised controlled trial of the impact of guidelines, prioritized review criteria and feedback on implementation of recommendations for angina and asthma. Br J Gen Pract. 2003;53(489):284–91.
11. Godager G, Hennig-Schmidt H, Iversen T. Does performance disclosure influence physicians' medical decisions? an experimental study. J Econ Behav Organ. 2016;131 Part B:36–46. https://doi.org/10.1016/j.jebo.2015.10.005.
12. Feder G, Griffiths C, Highton C, Eldridge S, Spence M, Southgate L. Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? a randomised controlled trial in general practices in east london. BMJ. 1995;311(7018):1473–8.
13. Kerse NM, Flicker L, Jolley D, Arroll B, Young D. Improving the health behaviours of elderly people: randomised controlled trial of a general practice education programme. BMJ. 1999;319(7211):683–7.
14. Östervall LW. Nudging to prudence? the effect of reminders on antibiotics prescriptions. J Econ Behav Organ. 2017;135:39–52.
15. Wilson AM. The role of mystery shopping in the measurement of service performance. Manag Serv Qual Int J. 1998;8(6):414–20.
16. Madden JM, Quick JD, Ross-Degnan D, Kafle KK. Undercover careseekers: simulated clients in the study of health provider behavior in developing countries. Soc Sci Med. 1997;45(10):1465–82.
17. Rhodes KV, Miller FG. Simulated patient studies: An ethical analysis. Milbank Q. 2012;90(4):706–24.
18. Simasek M, Blandino DA. Treatment of the common cold. Am Fam Physician. 2007;75(4):515–20.
19. Allan GM, Arroll B. Prevention and treatment of the common cold: making sense of the evidence. Can Med Assoc J. 2014;186(3):190–9.
20. World Health Organization. The pursuit of responsible use of medicines: sharing and learning from country experiences. 2012;WHO/EMP/MAR/2012.3:.
21. Gani L, Arif H, Widjaja SK, Adi R, Prasadja H, Tampubolon LH, Lukito E, Jauri R. Physicians' prescribing practice for treatment of acute diarrhoea in young children in jakarta. J Diarrhoeal Dis Res. 1991;9(3):194–9.
22. Chukwuani C, Onifade M, Sumonu K. Survey of drug use practices and antibiotic prescribing pattern at a general hospital in nigeria. Pharm World Sci. 2002;24(5):188–95.
23. Arya SC. Antibiotics prescription in hospitalized patients at a chinese university hospital. J Infect. 2004;48:117–8.
24. Reardon S. Antibiotic resistance sweeping developing world. Nature. 2014;509(7499):141.
25. Reynolds L, McKee M. Factors influencing antibiotic prescribing in china: An exploratory analysis. Health Policy. 2009;90(1):32–36.
26. Currie J, Lin W, Zhang W. Patient knowledge and antibiotic abuse: Evidence from an audit study in china. J Health Econ. 2011;30(5):933–49.
27. Currie J, Lin W, Meng J. Addressing antibiotic abuse in china: An experimental audit study. J Dev Econ. 2014;110:39–51.
28. McGuire TGCuyler, Newhouse, (eds). Physician agency. North-Holland, Amsterdam (The Netherlands): Elsevier; 2000. p. 461–536.
29. Farley PJ. Theories of the price and quantity of physician services: A synthesis and critique. J Health Econ. 1986;5:315–33.
30. Bénabou R, Tirole J. Incentives and prosocial behavior. Am Econ Rev. 2006;96(5):1652–78.
31. Yip W, Hsiao WC. The chinese health system at a crossroads. Health Aff. 2008;27(2):460–8. https://doi.org/10.1377/hlthaff.27.2.460 http://content.healthaffairs.org/content/27/2/460.full.pdf+html.
32. Hvistendahl M. China takes aim at rampant antibiotic resistance. Am Assoc Adv Sci. 2012;336(6083):795.
33. Xiao Y, Zhang J, Zheng B, Zhao L, Li S, Li L. Changes in chinese policies to promote the rational use of antibiotics. PLoS Med. 2013;10(11):1001556.
34. Yip WC-M, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of china's huge and complex health-care reforms. Lancet. 2012;379(9818):833–42.
35. Moriarty H, McLeod D, Dowell A. Mystery shopping in health service evaluation. Br J Gen Pract. 2003;53(497):942–6.
36. Bisgaier J, Rhodes KV. Auditing access to specialty care for children with public insurance. N Engl J Med. 2011;364(24):2324–33.
37. Fahey T, Stocks N, Thomas T. Systematic review of the treatment of upper respiratory tract infection. Arch Dis Child. 1998;79(3):225–30.
38. Kenealy T, Arroll B. Antibiotics for the common cold and acute purulent rhinitis. Cochrane Database Syst Rev. 2013;6:. https://doi.org//10.1002/14651858.CD000247.pub3.
39. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka PE, editor. Frontiers in Econometrics. New York: Academic Press; 1974. p. 105–42.
40. McFadden D. Economic choices. Am Econ Rev. 2001;91(3):351–78.
41. Chamberlain G. Analysis of covariance with qualitative data. Rev Econ Stud. 1980;47:225–38.
42. Hole AR. Fitting mixed logit models using maximum simulated likelihood. Stata J. 2007;7:388–401.
43. Harrison GW, List JA. Field experiments. J Econ Lit. 2004;42(4):1009–55.
44. Abadie A, Athey S, Imbens GW, Wooldridge J. When should you adjust standard errors for clustering?. 2017.
45. Grigoryan L, Burgerhof JG, Degener JE, Deschepper R, Lundborg CS, Monnet DL, Scicluna EA, Birkin J, Haaijer-Ruskamp FM. Attitudes, beliefs and knowledge concerning antibiotic use and self-medication: a comparative european study. Pharmacoepidemiol Drug Saf. 2007;16(11):1234–43.
46. Togoobaatar G, Ikeda N, Ali M, Sonomjamts M, Dashdemberel S, Mori R, Shibuya K. Survey of non-prescribed use of antibiotics for children in an urban community in mongolia. Bull World Health Organ. 2010;88(12):930–6.
47. Pan H, Cui B, Zhang D, Farrar J, Law F, Ba-Thein W. Prior knowledge, older age, and higher allowance are risk factors for self-medication with antibiotics among university students in southern china. PloS ONE. 2012;7(7):41314.
48. Yu M, Zhao G, Lundborg CS, Zhu Y, Zhao Q, Xu B. Knowledge, attitudes, and practices of parents in rural china on the use of antibiotics in children: a cross-sectional study. BMC Infect Dis. 2014;14(1):112.
49. Jin C, Ely A, Fang L, Liang X. Framing a global health risk from the bottom-up: user perceptions and practices around antibiotics in four villages in china. Health Risk Soc. 2011;13(5):433–49.
50. Arnold SR, To T, McIsaac WJ, Wang EE. Antibiotic prescribing for upper respiratory tract infection: the importance of diagnostic uncertainty. J Pediatr. 2005;146(2):222–6.
51. Kotwani A, Wattal C, Katewa S, Joshi P, Holloway K. Factors influencing primary care physicians to prescribe antibiotics in delhi india. Fam Pract. 2010;27(6):684–90.
52. Xue H, Shi Y, Huang L, Yi H, Zhou H, Zhou C, Kotb S, Tucker JD, Sylvia SY. Diagnostic ability and inappropriate antibiotic prescriptions: a quasi-experimental study of primary care providers in rural china. J Antimicrob Chemother. 2019;74(1):256–63.
53. Xu T, de Almeida N, Abilio C, Moles RJ. A systematic review of simulated-patient methods used in community pharmacy to assess the provision of non-prescription medicines. Inter J Pharma Pract. 2012;20(5):307–319.
54. Adams SA, Paul KT, Ketelaars C, Robben P. The use of mystery guests by the Dutch Health Inspectorate: results of a pilot study in longterm intramural elderly care. Health Policy. 2015;119(6):821–830.
55. Cheraghi-Sohi S, Hole AR, Mead N, McDonald R, Whalley D, Bower P, Roland M. What patients want from primary care consultations: a discrete choice experiment to identify patientsŠ priorities. Annals Fam Med. 2008;6(2):107–115.
56. Hole AR. Modelling Heterogeneity in Patients' Preferences for the Attributes of a General Practitioner Appointment. J Health Econ. 2008;27(4):1078–1094.
57. Hole AR. Discrete Choice Methods in Health Economics', Health Econometrics (Contributions to Economic Analysis, Volume 294): Emerald Publishing Limited; 2018.

58. Wang J, Iversen T, Hennig-Schmidt H, Godager G. Are patient-regarding preferences stable? Evidence from a laboratory experiment with physicians and medical students from different countries. Euro Econ Rev. 2020103411.

Publisher's Note

II | Paper II

II

III | Paper III

Predicting strategic medical choices:
An application of a quantal response equilibrium choice model

Ge Ge*[1], Geir Godager[1,2]

Abstract

We revisit the question of how market competition affects pro-social behavior. We apply a quantal response equilibrium choice (QREC) model to data from an incentivized laboratory experiment, where the participants make decisions on medical treatment for abstract patients in *monopoly*, *duopoly* and *quadropoly* games. Our results demonstrate that competition can cause substantial behavioral responses without any changes in pro-social preferences if one allows for the scale parameter to depend on the market setting. We find that a QREC model with fixed preference parameters provides precise out-of-sample predictions of behavior in games with vector payoffs. A Monte Carlo study is performed to show that the two-step estimator is accurate.

Declarations of interest: none.

Keywords:
Behavioral game theory, bounded rationality, prediction, quantal response equilibrium, discrete choice modelling,

JEL-Classification: C25, C57, C70 C92, D43, I11

*Corresponding author. Email: gege@medisin.uio.no
[1]Department of Health Management and Health Economics, University of Oslo, Norway.
[2]Health Services Research Unit, Akershus University Hospital, Norway

Preprint submitted to Journal of Choice Modelling *November 10, 2020*

1. Introduction

In their seminal contribution, McKelvey and Palfrey (1995) proved that a quantal response equilibrium (QRE) always exists for finite games. A QRE can be described as a statistical version of a Nash equilibrium (Camerer, 2011), where sub-optimal choice alternatives have non-zero choice probabilities. By allowing for sub-optimal strategic choices, the QRE introduces a weaker assumption on human rationality to game theory. Further, McKelvey and Palfrey (1995) bridge the gap between Behavioral Game Theory and Choice Modelling by showing that behavior by individuals who maximize a linear combination of expected utility and noise is a game theoretic equilibrium. Assuming the noise to be type I extreme value distributed implies a logistic quantal response equilibrium (LQRE). The objective in the LQRE model of McKelvey and Palfrey (1995) can be written:

$$U_j = \lambda V_j + \varepsilon_j \quad , \tag{1}$$

where the λ is the scale parameter of a multinomial logit (MNL) model (σ in Fiebig et al., 2010), and is typically referred to as the "rationality parameter" in the behavioral game theory literature inspired by McKelvey and Palfrey (1995).

Twenty-five years after McKelvey and Palfrey (1995), most applications are still focusing on strategic scenarios where the choice alternatives are characterized by a single attribute. Generalizations to study strategic agents caring about multiple attributes is straight forward, and familiar to choice modellers. Yet, there are few choice modellers applying quantal response equilibrium choice (QREC) models, i.e. QRE models where decision-makers have preferences for multiple attributes. One of the aims of this paper is to inspire the use of QREC models in applied research. We show how ready-made software can be applied to obtain a two-step QREC estimator, and conduct Monte Carlo simulations to provide documentation on its precision. Not only is this two-step QREC estimator a convenient approach, the results from our Monte Carlo simulations show that it is also accurate even with a moderate sample size.

We conduct an incentivized lab experiment to study strategic behavior under various levels of competition. The experiment is framed in a medical setting where the individuals play the role of physician. In the game, individuals make treatment choices that generate own profits as well as positive health benefits for the patients. Three levels of competition are considered: monopoly, duopoly, and quadropoly. In the competitive scenarios, subjects' joint choices of treatment alternatives determine the subjects' profit as well as the health benefit of patients in the market.

The results show that the substantial difference in observed behavior under different competition intensities can be attributed to a change in the scale parameter. In particular, the choices become less random as the competition intensifies. This result provides useful nuances to the recent literature on whether markets erode social responsibility (Falk and Szech, 2013; Bartling et al., 2015; Kirchler et al., 2015; Bartling et al., 2020), and contributes to the small empirical literature on physician competition (Gaynor and Vogt, 2000; Gaynor and Town, 2011).

The paper proceeds as follows: In section 2, we give a brief overview of recent literature on the scale in choice models and behavioral game theory. In Section 3, we apply a QREC model to a physician oligopoly and describe the two-step estimation procedure. In Section 4, we give a brief description of the experimental protocol. Descriptive statistics and results from estimation and Monte Carlo simulations are presented in Section 5. Finally, we discuss and conclude in Section 6.

2. Related literature

2.1. Stochastic choice

As described by McFadden et al. (1999), random utility models allow for weaker forms of human rationality than the strong rationality assumption typically employed in textbook economics, and theoretical contributions have showed that logit specifications are highly non-restrictive.[3] The MNL model in (1) is highly versatile and can be motivated and deduced in many different ways:

[3] See Dagsvik (2016) and the references therein.

- According to Thurstone (1927a; 1927b), the source of randomness in behavior is that individuals' utilities vary from moment to moment in a stochastic manner.

- A different way of motivating (1) is to regard individuals as incapable of making perfectly rational choices. In the model deduction by Luce (1959), Tversky (1972), and McKelvey and Palfrey (1995), the individual's (expected) utility is assumed deterministic, while randomness stems from randomness in implementation of choice. Luce describes a kind of perception error (McFadden et al., 1999) in motivating randomness. He proposes that individuals are unable to discriminate perfectly between utility levels of available alternatives.[4]

- Swait and Marley (2013) and Wallin et al. (2018) show that (1) follows by implication from the optimizing behavior of a decision-maker balancing the competing goals of achieving high utility and product variation.

- Erlander (1998) shows that (1) can be motivated by the implication from "the efficiency assumption" that samples with higher total observable utility are more probable.

- The MNL model has often been presented as a practical econometric specification where the noise term is introduced to account for variables that are unobservable to the analyst (McFadden, 1974).

The fact that there are many ways to motivate and deduce (1) does not mean that a given way of motivating the MNL model suits every purpose. Introducing the noise term as *unobservables* appear less convincing when analyzing data collected in a controlled lab experiment. Using the motivation by Thurstone (1927a; 1927b) or Swait and Marley (2013) appear less plausible for models of (errors in) medical decision making (Mackowiak, 2020), a pilot's choice of applying flaps for takeoff (Loukopoulos et al., 2009) or losses by Grandmasters in chess.

The bounded rationality perspective is the motivation applied by McKelvey and Palfrey

[4]The difference in motivation by Luce and Thurstone becomes superficial in practical applications of standard choice models. McFadden (1981) shows that the two types of probabilistic choice models are equivalent in many cases.

(1995), and they interpreted the scale parameter as individuals' *degree of rationality*. This interpretation can be criticized, as one might argue that *rationality* would have to be very narrowly defined to be represented by one single parameter. Further, for investment decisions and many other examples, randomizing can be a rational strategic choice. Regardless of motivation, the scale parameter is a measure of the *degree of randomness in behavior*, and we take this broad perspective in this paper.

2.2. Scale in choice models

Controlled lab experiments provide favorable conditions for identifying the scale parameter in choice models. Experiments performed in a controlled environment facilitate implementation of ceteris paribus changes in the choice context, and enable the researcher to confront each decision-maker with exactly the same set of choice scenarios. A context-dependent scale parameter is identified under the assumption that preference parameter are fixed and independent of context.

The scale in choice models has generated much confusion (Swait and Louviere, 1993; Louviere and Eagle, 2006; Hess and Rose, 2012; Hess and Train, 2017). Empirical contributions have often ignored or avoided addressing questions about how the scale might differ systematically between individuals or between choice occasions for the same individual. An important practical reason is probably that many available data sets do not contain sufficient information to address questions regarding, for example, scale differences caused by systematical contextual differences between choice occasions. While data collected in the field may record context differences, such as whether commuters on public transport were interviewed in rush hour or not, it would rarely be the case that decision-makers are randomly assigned to a choice context. As a result of unobserved selection mechanisms, it is highly plausible that individuals sampled at different choice contexts differ on both preference parameters and scale.

Louviere and Eagle (2006) argue that scale is highly unlikely to be constant, as the impact of noise on choices can vary over conditions, contexts circumstances or situations, as well as between decision-makers. Further, as illustrated with examples by Louviere and Eagle (2006), if scales vary across decision-makers, decision-makers will seem heterogeneous in preferences even if they differ only in scale. For the same reason, differences in scale across choice occasions, can make individuals' preferences appear context-dependent even when preferences are stable.

As discussed by Louviere et al. (1999), whether preferences depend on the context, or whether such findings are a result of ignoring between-context differences in scale, is an important question for choice modellers: Evidence of stable preferences provides support for the validity of results from stated preference experiments when aiming to predict market behavior.

Many studies provides conclusions on preferences without modelling preferences or scale. An example is Falk and Szech (2013), who draw the conclusion that market competition causes experimental subjects to value the life of a mouse less, based on the observation that competition in the laboratory had unfortunate consequences for mice.

In our experiment, individuals' choices result in more benefit for patients when there is more competition. If one assumes the scale to be fixed, and allow preferences to be context-dependent, the results would suggest that competition causes preferences to change, and that individuals become more altruistic in market settings with more competition. As competition is beneficial to the third party, one could say that such results provide a conclusion that is opposite of that reported by (Falk and Szech, 2013).

2.3. Empirical game theory

The Nash equilibrium (NE) (1950) is useful in providing theoretical predictions in single-criterion games, and has undoubtedly been pivotal to the development of game theory and empirical analysis of strategic behavior. The fact that the Nash equilibrium fits poorly with observed behavior in many cases is much discussed in the literature on empirical game theory, and an example is the enlightening contribution by Goeree and Holt (2001). For decades, the poor fit of the NE assumption inspired explorations of new equilibrium concepts that could explain the observed behavior (Harsanyi, 1973; Harsanyi et al., 1988; Ma and Manove, 1993).

Strategic economic agents often care about more than a single criterion upon making economic choices: Firms are concerned about both long run and short run profits, politicians deciding on foreign trade policies might be concerned about several aspects of policy impacts, such as effects on the environment, unemployment, budget deficits as well as the likelihood of re-election. Medical doctors who compete for patients, as in the framing of the experiment studied in this paper, are likely concerned about profits as well as health effects for patients when choosing medical treatments. A game with vector payoff is often referred to as a multi-criteria game, or a multi-payoff game (Shapley and Rigby, 1959; Zeleny, 1975; Voorneveld et al., 2000), and would perhaps be translated to a "multiple attribute game" in choice modelling terminology. If players' valuation of payoff elements is known to the researcher, the multi-criteria game can be scalarized into a standard single-criterion game, which can be solved for NE using standard approaches.

Despite the obvious relevance of multi-criteria games for analyzing strategic behavior when the payoff is a vector, its literature is small compared to that of the single-criterion game theory. A plausible explanation is the lack of useful equilibrium concepts for multi-criteria games. Although equilibrium concepts, such as, the *Pareto equilibrium* (Shapley and Rigby, 1959; Voorneveld et al., 1999) and *ideal equilibrium* (Voorneveld et al., 2000) have been developed for solving such games, there are clear limitations when individuals' preferences are latent. Shapley and Rigby (1959) show that narrowing down the set of plausible actions in multi-criteria games is difficult when individuals' valuation of the elements in the vector payoff is unknown. Therefore, the *Pareto equilibrium* discussed by Shapley and Rigby (1959) is often not useful, as many plausible actions might remain after removing dominated strategies. Further, the number of pure strategy NE, and whether or not a given set of actions by the players constitute pure strategy NE, can depend on the unknown preference weights.

In their seminal contribution, McKelvey and Palfrey (1995) proved the existence of a Quantal Response Equilibrium. As described by Jessie and Saari (2016), one of the contributions of McKelvey and Palfrey (1995) is to link behavioral game theory to the choice modelling paradigm. Choices in QRE models are assumed to be the result of individuals maximizing a linear combination of expected utility and noise as in (1).[5] The unknown value of λ has motivated much critique of the QRE assumption. While the QRE assumption is shown to fit empirical data in many applications, it has been criticized for being difficult to falsify. As noted by Haile et al. (2008), an estimated scale parameter in a particular game might not be useful in predicting behavior in other games, or by other players. For this reason, some authors, such as McCubbins et al. (2013), even argue that QRE *"...is of limited use in explaining human behavior across even a small range of similar decisions"*. However, others, such as Anderson et al. (2001); Goeree et al. (2005); Goeree and Holt (2005); Goeree et al. (2010); Matějka and McKay (2015) and Wright and Leyton-Brown (2017), are more optimistic in their view on the role of QRE in future empirical research on strategic decision-making. Wright and Leyton-Brown (2017) suggest an agenda for acquiring more knowledge on the usefulness of the QRE assumption when aiming to provide ex-ante predictions of strategic behavior.

3. A quantal response equilibrium choice model

3.1. The basic 2×2 game for two players

For clarity and ease of exposition, we start with the most basic example of a strategic scenario: A symmetric two-player game where two identical agents have an identical binary choice set. Our game differs from more common games where players' payoff is a scalar. In our game, is a vector and the values of its elements are determined by players' joint choices. The two players comprise a duopoly of physicians, who can choose between either Low or $High$ treatment intensity for their patients. The two physicians' joint choices of treatment plan determine the vector of profit and patient health benefit that each of the two players receives. We assume that physicians value both profit, Π, and patient's benefit, B.

The game is described in a normal form representation with a row player and a column player in Table 1. $\Pi_{r|c}$ and $B_{r|c}$ denote the payoff to the row player when he chooses row r and the column player chooses column c, $r, c = L, H$. We study symmetric games in this paper, so

[5] With a structure similar to Swait and Marley (2013), one could regard a QRE as an equilibrium of a multi-criteria game where the players' weighting of the two criteria, a scalar payoff and noise, is unknown. In this text, however, the term multi-criteria refers to games where payoffs are vectors of observable attributes.

Table 1: The matrix of payoff vectors to the row player in a symmetric duopoly game with two pure strategies.

		The opponent L	H
Row	L	$\Pi_{L\|L}, B_{L\|L}$	$\Pi_{L\|H}, B_{L\|H}$
player	H	$\Pi_{H\|L}, B_{H\|L}$	$\Pi_{H\|H}, B_{H\|H}$

reporting the payoffs to the row player is sufficient for describing the game: In a symmetric two-player game, the payoff to the row player when he chooses treatment L and the column player chooses treatment H is identical to the payoff received by the column player in the mirrored situation where the column player chooses treatment L and the row player chooses treatment H. We assign the first-person perspective to the row player and let all payoff vectors represent payoffs to the row player. In this way, we economize on notation and we highlight the common ground of strategic decision-making and choices under uncertainty.

The utility for the row player of choosing, for example, L is uncertain, as it depends on whether the opponent chooses L or H. Conditional on the column player choosing c, the row player's utility from choosing row r, modeled as a simple linear-in-parameter preference function, can be expressed as:

$$v_{r|c} = \beta_{\Pi}\Pi_{r|c} + \beta_B B_{r|c} \quad . \tag{2}$$

Assuming observed choices to be the result of individuals maximizing a linear combination of *(objective) expected utility* and noise involves only a small and natural augmentation of the basic choice model in (1).[6] Let P_L and P_H denote the unobserved probabilities that the opponent chooses alternative L or H, respectively. Given the observed payoff matrix in Table 1, the row

[6]There is a growing literature on *Random Expected Utility Models*. See for example Gul and Pesendorfer (2006), Blavatskyy (2007), Dagsvik (2008; 2015) and Ke (2018).

player's expected utility from choosing row r can be expressed as:

$$V_r(\mathbf{P}) = P_L v_{r|L} + P_H v_{r|H}. \tag{3}$$

Inserting (2) into (3), and rearranging terms, we get:

$$\begin{aligned} V_r(\mathbf{P}) &= P_L(\beta_\Pi \Pi_{r|L} + \beta_B B_{r|L}) + P_H(\beta_\Pi \Pi_{r|H} + \beta_B B_{r|H}) && \text{(4a)} \\ &= \beta_\Pi(P_L \Pi_{r|L} + P_H \Pi_{r|H}) + \beta_B(P_L B_{r|L} + P_H B_{r|H}) && \text{(4b)} \\ &= \beta_\Pi E\Pi_r + \beta_B EB_r && \text{(4c)} \end{aligned}$$

In (4c), we have introduced notation for the expected attribute vector of r: $E\Pi_r = P_L\Pi_{r|L} + P_H\Pi_{r|H}$ and $EB_r = P_L B_{r|L} + P_H B_{r|H}$. Equation (4) highlights that expected utility is a function of the unobserved multinomial probability vector $\mathbf{P} = (P_L, P_H)$.

Inserting $V_r(\mathbf{P})$ for V_r in (1) and assuming ε_r to be type 1 extreme value distributed, the probability that the row player chooses L becomes:

$$P_L = \frac{e^{\lambda V_L(\mathbf{P})}}{e^{\lambda V_L(\mathbf{P})} + e^{\lambda V_H(\mathbf{P})}} \quad , \tag{5}$$

where λ is the scale parameter. As $\lambda \longrightarrow 0$, the behavior becomes purely random, and the player plays each pure strategy with equal probability. As $\lambda \longrightarrow \infty$, the players make no errors, and behavior becomes deterministic. In the limit, the players are perfectly responsive to the differences in expected utility across alternatives, and the behavior converges to a NE.

Data generated from the game in Table 1 differ from typical choice data in that the expected attribute vector in (4c) is unobservable. This is a key feature that distinguishes (5) from the standard scaled logit model. An intuitive two-step estimation procedure can be applied to estimate (β, λ): The first step is to replace unobserved probabilities $\mathbf{P}$ on the right-hand side of (4c) with observed relative choice frequencies, $\mathbf{f}$, to obtain an estimate of the expected attribute vector. The second step is to estimate (β, λ) by means of maximum likelihood, treating the estimated attribute vector as if it was an observed attribute vector. In the following we refer to this as a two-step QREC estimator. An important note is that a choice modeller tasked with analyzing the data from this game might likely proceed with the two-step procedure without being knowledgeable of McKelvey and Palfrey's contribution of QRE.[7]

A key feature of (5) from a behavioral game theory perspective is that **P** is a function of itself, as the left-hand side of (5) is one of the elements of **P** on the right-hand side. McKelvey and Palfrey (1995) proved that there always exists a QRE for a finite game. While solving the fixed point of (5) or having knowledge of the existence of QRE is not necessary for obtaining the two-step estimator, it should be reassuring for applied choice modellers that a QRE always exists in a game with well defined choice sets.

In the following subsection, we introduce functional form and generalize the QREC model to more than two players and J pure strategies.

3.2. Generalized model

There are three different market settings denoted by t, t =[monopoly, duopoly, quadropoly], and eight independent games are played within each market setting. We present the payoff matrices of the 24 games in Appendix A and elaborate one example in the next Section. A complete description of the formulas for computing payoff matrices based on the experimental parameters can be found in Appendix E.[8]

Consider a physician selecting one treatment alternative j from a finite set $C = 1, 2, ..., J$. The choice set comprises J mutually exclusive alternatives within a game. To simplify notation, in this subsection, we suppress the index for each game within a market setting. We assume that physicians' preference parameters are homogeneous and are fixed over games. While assuming a representative individual model appears restrictive after decades of applications of mixed logit models, a parsimonious model specification is suitable for our purpose.[9]

[7]This was indeed the the case for the authors of this paper.

[8]Appendix E attached to this paper is submitted to and currently under review at *Data in Brief*.

[9]While generalizing the model to account for the unobserved heterogeneity in λ or preference parameters is an obvious extension, specifying a representative individual model is a convenient choice as it substantially reduces the burden of conducting the Monte Carlo simulations of repeated equilibrium game play. Our parsimonious specification can also be defended by arguing that preferences for the attributes in the experiment (money and health benefit) are likely to be less heterogeneous than preferences for, for example, household appliances.

The payoff players receive is a vector consisting of two elements: physician profit, Π, and patient benefit, B. Similar to Goeree et al. (2002), we assume that a healthy patient population in the market is a shared good, and a physician's valuation of the patient benefit in the market is independent of which physician provides treatment to which patient. In other words, B here is the total benefit of all the patients in the market. We represent preferences by a linear-in-parameter function of the two payoff elements. We consider two alternative specifications which are a quadratic function and a Cobb-Douglas function in a log-linear form. While the quadratic specification has obvious drawbacks, such as saturation, it is convenient and simple. The log-linear Cobb-Douglas arguably has a more solid axiomatic foundation following the proof by Dagsvik (2018), but offers less flexibility than the quadratic specification. The non-linear utility specification allows us to identify the curvature of the utility function (see for example Van Der Pol et al. (2010); Kolstad (2011); van der Pol et al. (2014); Holte et al. (2016); Wang et al. (2020)). By Taylors theorem, further expanding the polynomial in the two specifications would provide better approximations. However, such improvements in functional forms are costly, as more and richer data is required to quantify additional parameters. In addition, larger samples and additional parameters also raise computational costs. Hence, a quadratic form and log-linear Cobb-Douglas, rather than a more general translog, is a convenient choice.

The payoff elements of treatment alternative j in monopoly is denoted by Π_{jt} and B_{jt}. In duopoly, $\Pi_{jt|x}$ and $B_{jt|x}$ are the payoff elements of choosing j given that the opponent's choice is x. In quadropoly, $\Pi_{jt|xyz}$ and $B_{jt|xyz}$ are the payoff elements of choosing j given that the combination of choices by opponents' are xyz. The conditional utility of a physician choosing alternative j given the opponent(s)' choice(s) is denoted by v_{jt}, $v_{jt|x}$, and $v_{jt|xyz}$ in monopoly, duopoly and quadropoly, respectively. Preference parameters are denoted by the vector β. The conditional utility for the two functional forms in, for example, quadropoly, can be written:

Quadratic utility:

$$v_{jt|xyz} = \beta_{\Pi}\Pi_{jt|xyz} + \beta_B B_{jt|xyz} + \beta_{\Pi\Pi}\Pi^2_{jt|xyz} + \beta_{BB}B^2_{jt|xyz} + \beta_{\Pi B}\Pi_{jt|xyz}B_{jt|xyz} \quad , \tag{6}$$

Cobb-Douglas utility:

$$v_{jt|xyz} = \underline{U} + \beta_{\Pi} ln\Pi_{jt|xyz} + \beta_B ln B_{jt|xyz} \quad , \tag{7}$$

where $j, x, y, z \in C$ and t =[monopoly, duopoly, quadropoly], and $\underline{U}$ is a reference utility.

The opponent(s)' action(s) are unknown ex ante. The probability that the opponent chooses alternative x in duopoly is denoted by P_{xt}, and the probabilities of the three opponents choosing x, y and z in quadropoly is denoted by P_{xt}, P_{yt} and P_{zt}, respectively. When a physician chooses alternative j, his expected utility, $V_{jt}(\mathbf{P}_t)$, is given by:

$$\text{Monopoly: } V_{jt}(\mathbf{P}_t) = v_{jt} \quad ; \tag{8a}$$

$$\text{Duopoly: } V_{jt}(\mathbf{P}_t) = \sum_{x \in C} P_{xt} v_{jt|x} \quad ; \tag{8b}$$

$$\text{Quadropoly: } V_{jt}(\mathbf{P}_t) = \sum_{x \in C}\sum_{y \in C}\sum_{z \in C} P_{xt}P_{yt}P_{zt}v_{jt|xyz} \quad , \tag{8c}$$

where $\mathbf{P}_t = (P_{1t}, P_{2t}, ..., P_{Jt})$ denotes the vector of the opponent(s)' choice probabilities for all alternatives in market setting t. We generalize (1) by assuming that the physician maximizes a linear combination of the expected utility from choosing j and an error term, ε_{jt}, in each game:

$$\lambda_t V_{jt}(\mathbf{P}_t) + \varepsilon_{jt} \; , \tag{9}$$

In our symmetric games with homogeneous players, the probability of choosing j in market setting t is the same for each player and given by:

$$P_{jt} = \frac{e^{\lambda_t V_{jt}(\mathbf{P}_t)}}{\sum_{r \in C} e^{\lambda_t V_{rt}(\mathbf{P}_t)}}, \tag{10}$$

One cannot identify both the scale parameter and the preference parameters (Train, 2009; Hess and Rose, 2012). In the quadratic utility specification, we normalize $\beta_{\Pi} + \beta_{\Pi\Pi} = 1$ in order to identify λ in all three markets.[10] In the Cobb-Douglas specification, we apply the normalization $\beta_\Pi + \beta_B = 1$.[11] We handle zeros in the Cobb-Douglas specification by introducing a dummy equal to one whenever $\Pi = 0$ or $B = 0$. This dummy identifies the reference utility $\underline{U}$ in (7), and enables us to replace $ln(0)$ by 0 wherever necessary.[12]

3.3. Estimation

We now let $g = 1, 2, \ldots, 8$ index the eight games in each market setting, and $n = 1, 2, \ldots, 136$ index the individuals, the log-likelihood function given our QREC model specification can be written as:

$$LL(\lambda_t, \beta)|_{\mathbf{P}_{gt}} = \sum_n \sum_j \sum_g \sum_t y_{njgt} ln \left(\frac{e^{\lambda_t V_{jgt}(\mathbf{P}_{gt},\beta)}}{\sum_{r\in C} e^{\lambda_t V_{rgt}(\mathbf{P}_{gt},\beta)}} \right), \tag{11}$$

where $y_{njgt} = 1$ if physician n chose j in game g in market setting t, and zero otherwise. If $\mathbf{P}_{gt}$ is known, preference estimates from maximizing equation (11) can be acquired directly.

In this paper we use the two-step procedure, where $\mathbf{P}_{gt}$ is replaced by relative frequencies,$\mathbf{f}_{gt}$, and document the accuracy by means of Monte Carlo simulations.[13] With $\mathbf{P}_{gt}$ replaced by $\mathbf{f}_{gt}$, the likelihood function can be written:

$$LL(\lambda_t, \beta)|_{\mathbf{f}_{gt}} = \sum_n \sum_j \sum_g \sum_t y_{njgt} ln \left(\frac{e^{\lambda_t V_{jgt}(\mathbf{f}_{gt},\beta)}}{\sum_{r\in C} e^{\lambda_t V_{rt}(\mathbf{f}_{gt},\beta)}} \right). \tag{12}$$

Since $\mathbf{f}_{gt}$ is a consistent estimator of $\mathbf{P}_{gt}$, (12) converges in probability to equation (11) as the number of individuals increases towards infinity. It follows that the estimators $\hat{\lambda}_t$ and $\hat{\beta}$ which maximize equation (12) are consistent. They are estimated by the `gmnl` module in STATA 16 (Fiebig et al., 2010; Gu et al., 2013). Note that differently from the full information likelihood procedure described by Moffatt (2015), computing the fixed point assuming subjects play a LQRE is not necessary for using the two-step procedure.

---

[10]In this way, the estimates are measured in the marginal utility of the first dollar of profit when the health benefit is zero.

[11]See e.g. Swait and Marley (2013) for an example of constraining the sum of the coefficients.

[12]A similar procedure is described by Battese (1997).

[13]Despite that the two-step procedure, is frequently used, (See, e.g. (Bajari and Hortacsu, 2005)) we are unaware of any studies that document the accuracy of the two-step estimator.

4. Experimental data

In this section, we briefly describe the experimental design and show an example of the payoff matrices with illustration. A complete description of the experimental protocol and the formulas for computing payoff matrices based on the experimental parameters can be found in Appendix E.[14]

4.1. Experiment

The experiment has a medical framing. Participants are instructed to play the role of a physician and choose one of eleven different medical treatments for eight "patients" in each of three markets with various levels of competition; monopoly, duopoly and quadropoly (see Appendix B for instructions for participants). The treatment choices of the participants jointly determine their own profits and patients' health benefits. By our design, the payoffs reflect the intensity of competition in the market. The profit and benefit accrued in the laboratory are converted into monetary transfers to the participants and a charity dedicated to providing surgeries for ophthalmic patients. This element of our protocol, which is identical to Hennig-Schmidt et al. (2011), motivates participants' patient-regarding behavior in the laboratory. The experiment was designed to accommodate independent games of complete information with , simultaneous move. It was programmed in z-Tree (Fischbacher, 2007). Funding for the development of experimental design, the programming, as well as the payments to participants and *Christoffel Blindenmission* was provided by Research Council of Norway, through a grant to IRECOHEX, Project # 231776. The ethical review and approval of experimental procedure was given by Norwegian Social Science Data Services (reference #43709).

4.2. Payoff matrices in three markets

We now present the payoff matrix for patient game 1 in duopoly in Table 2 as an example, and a complete set of payoff matrices of the experiment is provided in Appendix A. Since the game is symmetric, we only present player n's (the row player) payoffs. Each cell represents a vector of profit and patient benefit which player n receives given the combination of treatments offered

[14] Appendix E attached to this paper is submitted to and currently under review at *Data in Brief*.

by player n and his opponents. If, for instance, player n chooses alternative 3 for patient 1 in duopoly, and the opponent chooses alternative 2, then player n will receive 664 Taler of profit, and 273 Taler is provided to the patients (100 Taler = 1 Euro). The subjects could inspect the

Table 2: Payoff matrix, patient game 1 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	500,0	270,73	120,176	50,285	20,392	10,495	0,600	0,700	0,800	0,900	0,1000
1	723,73	495,100	267,173	119,276	50,385	20,492	10,595	0,700	0,800	0,900	0,1000
2	845,176	701,173	480,200	259,273	115,376	48,485	19,592	10,695	0,800	0,900	0,1000
3	865,285	801,276	664,273	455,300	246,373	109,476	46,585	18,692	9,795	0,900	0,1000
4	823,392	798,385	739,376	613,373	420,400	227,473	101,576	42,685	17,792	8,895	0,1000
5	743,495	735,492	713,485	660,476	548,473	375,500	203,573	90,676	38,785	15,892	8,995
6	640,600	634,595	627,592	608,585	563,576	467,573	320,600	173,673	77,776	32,885	13,992
7	510,700	510,700	505,695	500,692	485,685	449,676	372,673	255,700	138,773	61,876	26,985
8	360,800	360,800	360,800	356,795	353,792	342,785	317,776	263,773	180,800	97,873	43,976
9	190,900	190,900	190,900	190,900	188,895	186,892	181,885	167,876	139,873	95,900	51,973
10	0,1000	0,1000	0,1000	0,1000	0,1000	0,995	0,992	0,985	0,976	0,973	0,1000

Payoffs are measured in Taler (100 Taler = 1 Euro).

payoff elements in each cell of the matrix by means of a "calculator" when choice combinations by players were inserted (Requate and Waichman, 2011).[15]

5. Results

5.1. Descriptive statistics

We report the observed choice frequencies in Table 3 and the relative frequencies in Figure 2.[16] We see in Table 3 that 24 out of 136 subjects chose pure strategy 0 for patient type 1 in monopoly, whereas none of the decision-makers chose this alternative for patient type 1 in duopoly and quadropoly. Health benefit to patiens is proportional to the numbering of the pure strategies, so that when more decision-makers pick alternatives in columns further to the right in the Table 3, more health benefits are transferred to the charity. Hence, from observing Table 3 and Figure 2, it is clear that, in this experiment, more competition benefits the patients in the market. This result is in contrast with Falk et al. (2013). See Section 6 for further discussion.

[15] Interested readers may also calculate the payoff matrix using the formulas and game parameters given in Appendix E.

[16] Note that Table 3 and the payoff matrices in Appendix A are sufficient for reproducing all empirical results in this paper.

Table 3: Observed frequencies of strategy choice in the 24 patient games by the 136 subjects.

Market	Patient game	Pure strategy										
		0	1	2	3	4	5	6	7	8	9	10
Monopoly	1	24	12	4	18	14	27	16	15	2	2	2
	2	21	9	9	13	11	18	24	14	10	1	6
	3	23	6	10	9	11	14	10	22	14	6	11
	4	23	5	7	13	11	19	9	29	11	3	6
	5	24	7	7	19	18	15	14	21	5	4	2
	6	23	4	8	21	14	17	11	15	13	4	6
	7	21	6	7	14	7	14	11	13	24	7	12
	8	21	4	14	12	7	16	14	8	20	7	13
Duopoly	1	0	0	1	3	3	12	24	41	26	22	4
	2	0	0	0	2	4	7	12	27	36	26	22
	3	0	0	0	3	4	3	9	18	37	34	28
	4	1	0	3	1	0	4	6	21	30	29	41
	5	1	0	0	1	4	18	22	48	32	7	3
	6	1	0	0	1	4	9	14	33	42	19	13
	7	1	1	0	1	1	1	3	15	18	28	67
	8	0	1	0	0	1	2	7	14	25	28	58
Quadropoly	1	0	0	0	2	2	4	14	31	48	30	5
	2	0	0	0	2	0	3	10	11	42	31	37
	3	1	0	0	0	0	0	3	12	26	41	53
	4	1	0	1	0	0	0	4	9	24	39	58
	5	0	0	0	0	0	10	9	37	58	17	5
	6	0	0	0	1	1	1	8	15	40	43	27
	7	0	0	0	1	0	1	3	2	16	22	91
	8	0	0	0	0	1	1	4	5	14	27	84

5.2. Estimation results

We present the estimated scale and preference parameters in Table 4. While all estimated preference parameters are reported to be significant, the reported standard errors need to be interpreted in a conditional sense, as estimates of $\mathbf{P}_{gt}$ are used for approximating the likelihood function. We include the 95% confidence intervals from the Monte Carlo simulations in the results table in addition to the estimated 95% confidence intervals. For both quadratic and Cobb-Douglas specifications, the differences between the two confidence intervals are negligible, and the interpretation is that reported the standard errors of the two-step estimator are accurate. The estimated preference parameters from the Cobb-Douglas specification is similar in magnitude to that reported by Wang et al. (2020), who fit a scaled logit model with log-linear Cobb-Douglas utility to the data from experiments based on an extended version of the design by Hennig-Schmidt et al. (2011). The confidence interval for the preference parameter in Table 4 overlaps with the confidence interval in Wang et al. (2020).

Table 4: Results from estimation and Monte Carlo simulation.

Model	Quadratic utility*					Cobb-Douglas utility				
	Estimates	95% C.I. Estimated†		95% C.I. Simulated‡		Estimates	95% C.I. Estimated†		95% C.I. Simulated‡	
Preference parameters										
β_{Π}	1.05	1.04	1.05	1.05	1.05	0.69	0.67	0.71	0.67	0.71
β_{B}	0.59	0.46	0.73	0.45	0.70	0.31	0.29	0.33	0.29	0.33
$\beta_{\Pi\Pi}$	-0.05	-0.05	-0.04	-0.05	-0.05	n.a.	n.a.	n.a.	n.a.	n.a.
β_{BB}	-0.03	-0.04	-0.02	-0.04	-0.02	n.a.	n.a.	n.a.	n.a.	n.a.
$\beta_{\Pi B}$	-0.05	-0.05	-0.04	-0.05	-0.04	n.a.	n.a.	n.a.	n.a.	n.a.
$\underline{U}$	n.a.	n.a.	n.a.	n.a.	n.a.	1.88	1.77	1.98	1.77	1.98
λ Parameters										
Monopoly	0.54	0.40	0.69	0.41	0.69	1.80	1.39	2.21	1.40	2.24
Duopoly	2.39	2.13	2.64	2.17	2.68	2.17	1.98	2.37	2.02	2.35
Quadropoly	4.13	3.76	4.50	3.76	4.55	2.51	2.30	2.73	2.31	2.73
Log-Likelihood		-6121.90					-6243.36			
# subjects:					136					
# Games:					24					

∗ Variables in the quadratic specifications were scaled in Euro instead of Taler.
† C.I based on Standard errors from Maximum Likelihood as reported by software.
‡ C.I based on the Monte Carlo simulations described in Section 5.3.
Parameter normalisations: $\beta_{\Pi} + \beta_{\Pi\Pi} = 1$ in the quadratic specification, and $\beta_{\Pi} + \beta_{B} = 1$ in the Cobb-Douglas specification.

Assuming the estimated standard errors are accurate, we apply simple Wald tests to test the null hypothesis that the three scale parameters are identical across market settings. We may reject this hypothesis for both quadratic ($p - value < 0.0001$) and Cobb-Douglas ($p - value < 0.0007$) specifications. We also perform pairwise tests. For the quadratic specification we may reject the null hypothesis that scale parameters in monopoly and duopoly are equal ($p - value < 0.0001$) and the null hypothesis that scale parameters in duopoly and quadropoly are equal ($p - value < 0.0001$). For the Cobb-Douglas specification, we cannot reject the null hypothesis that scale parameters in monopoly and duopoly are equal ($p - value = 0.0856$). However, we may reject the null hypothesis that scale parameters in duopoly and quadropoly are equal ($p - value = 0.0056$). The interpretation is that higher competition raises the scale parameter. We return to the interpretation of this result in the discussion section.

For the quadratic specification, we observe that the estimates of $\beta_{\Pi\Pi}$ and β_{BB} are negative, which implies that the marginal utilities are declining in profit and patient benefit. In addition, the estimate of $\beta_{\Pi\mathrm{B}}$ is also negative, indicating that profit and patient benefit are substitutes for this utility specification.

Conditional on parameter estimates from Table 4, and the 16 payoff matrices where competition is present (Tables A.2-A.9 and Tables A.10-A.17 in Appendix A), we can compute the matrices comprising the scalar utility payoffs in each game (see Appendix C for the computed utility payoff game assuming the quadratic preferences). We see that the numbers of dominated strategies and NE differ substantially between games. For example, the pure strategy Nash equilibrium in patient game 1 in duopoly is unique, but we find many cases where the scalarized games have multiple NE, such as game 7 in duopoly has six different NE.

5.3. Monte Carlo simulation

We conduct Monte Carlo simulations to assess the precision of the two-step QREC estimator. Let $(\hat{\beta}\ ,\ \hat{\lambda}_t)$ denote our estimated preference and scale parameters which are presented in Table 4. We apply these parameters as the fixed parameters in the Monte Carlo simulations. From the proof by McKelvey and Palfrey (1995), we know that the preference and scale parameter $(\hat{\beta},\ \hat{\lambda}_t)$ determine a LQRE given the payoff matrix, and we denote this LQRE by $\hat{\mathbf{P}}_t^*$.[17] In order to study the performance of the two-step estimator in estimating parameters of an assumed equilibrium, we start by computing the fixed point $\hat{\mathbf{P}}_t^*$. Under QRE, each player's belief of the opponent(s)'s choice probabilities is identical to the equilibrium probabilities. Hence, the fixed point is the solution to the following set of equations:

$$\hat{P}_{jt} = \frac{e^{\hat{\lambda}_t V_{jt}(\hat{\mathbf{P}}_t^*)}}{\sum_{r\in C} e^{\hat{\lambda}_t V_{rt}(\hat{\mathbf{P}}_t^*)}}, \tag{13}$$

[17] McKelvey and Palfrey (1995) claim that uniqueness of the LQRE applies for "almost all games". Dagsvik (forthcoming) establishes necessary and sufficient conditions for unique and multiple QRE, and provides useful algorithms for concluding on the number of QRE. We have applied the algorithms of Dagsvik (forthcoming) to examine the uniqueness of the equilibrium in the eight games in duopoly, and found that the equilibrium is unique for all eight games in duopoly.

where $\hat{P}_{jt}$ are the elements of $\hat{\mathbf{P}}_t^*$.[18] After having computed the fixed point, we generate 5 000 synthetic data sets for both functional forms. We then simulate choices by individuals with preference and scale parameters given by $(\hat{\beta},\ \hat{\lambda}_t)$ under the assumption that they play the LQRE in (13). When generating the synthetic choice data, we draw type I extreme value distributed error terms, and let simulated players pick the treatment alternative which maximizes:

$$\hat{\lambda}_t V_{jt}(\hat{\mathbf{P}}_t^*) + \varepsilon_{jt}. \tag{14}$$

Using the 5 000 synthetic data sets where the data generating process is known, we estimate preference and scale parameters $(\hat{\beta}_1,\ \hat{\lambda}_{t1})...(\hat{\beta}_{5000},\ \hat{\lambda}_{t5000})$ by applying the two-step procedure where we maximize (12). For each synthetic data set, we compute the observed relative frequencies of actions in a game, and apply the program `gmnl` by Gu et al. (2013) to estimate 5 000 scaled logit models for both functional forms. We describe the distribution of $\hat{\lambda}_t$ in Figure 1 and the confidence intervals of estimated scale and preference parameters that are based on the Monte Carlo simulations are reported in Table 4.

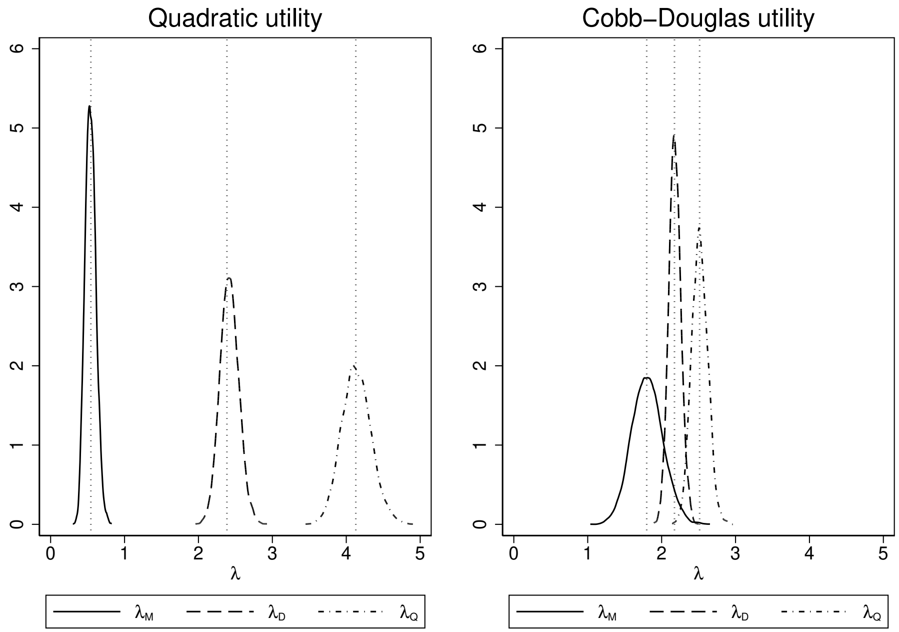
The distributions in Figure 1 are kernel density plots of estimated scale parameters in the three markets based on the synthetic data sets. We see that they do not overlap for the quadratic specification, whereas a substantial overlap is observed for the Cobb-Douglas specification. Hence, under the assumption that the quadratic model is the correct specification, it is highly unlikely that one would draw the wrong conclusion based on results from the two-step estimator, even with a small sample size. However, if the Cobb-Douglas model is the correct specification, one would need a larger sample size to provide statistical power to reject the null hypothesis that scale parameters in monopoly and duopoly are equal.

5.4. Model fit

As discussed by Wright and Leyton-Brown (2017), flexible models of human behavior are vulnerable for overfitting. Hence, the assessment of model fit should be based on out-of-sample predictions. By applying the following "jackknife-procedure", we show that estimation results

[18]See McKelvey and Palfrey, 1995, p. 11.

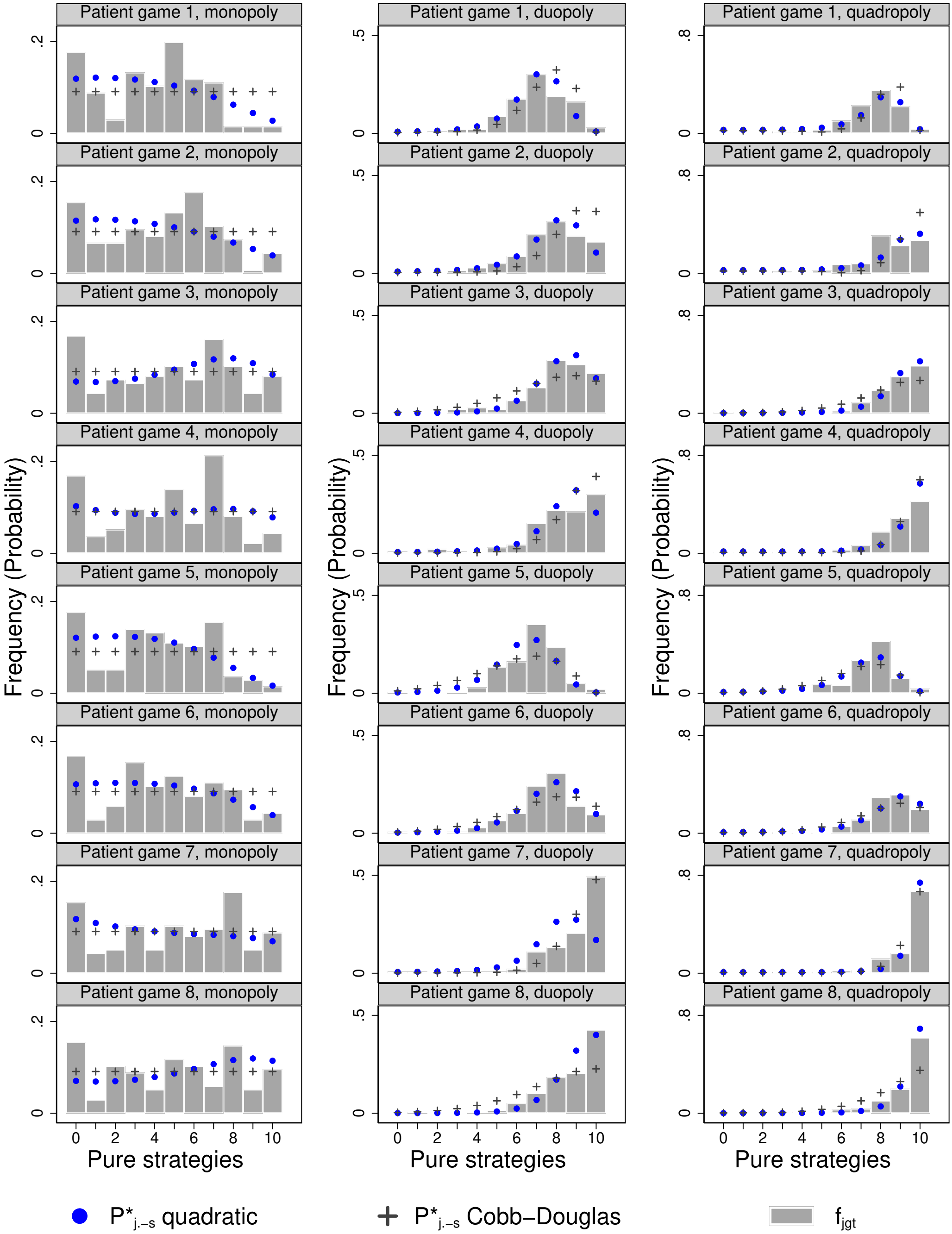
Figure 1: Distribution of λ estimates from Monte Carlo simulation. Dotted vertical lines indicate the given parameter values.



from our QREC model can be used to predict behavior ex ante. We first estimate the model 24 times, excluding one game each time. Then we compute the predicted fixed point in the game that was excluded from the estimation. We denote these "jackknife fixed points" by $\hat{\mathbf{P}}^*_{-s}$, where $-s \in 1, 2, 3 \ldots 24$ denotes the excluded patient games.

In Figure 2, we plot the observed relative frequencies for each strategy in each game, f_{jgt}, with the corresponding out-of-sample prediction, $\hat{\mathrm{P}}^*_{j,-s}$. Our QREC model captures the substantial differences in behavior across games and market settings quite well that the out-of-sample predictions are quite similar to observed behavior. We do not reject the null hypothesis that the two distributions, f_{jgt} and $\hat{\mathrm{P}}^*_{j,-s}$, are the same by means of Wilcoxon matched-pairs signed-ranks test ($p = 0.1861$ for quadratic, $p = 0.1948$ for Cobb-Douglas), and Fisher-Pitman

Figure 2: Relative frequencies and fixed points predicted out of sample for all pure strategies in the experiment. Patient benefit in a game increases from left to right.



permutation test for paired replicates ($p > 0.99$ for both quadratic and Cobb-Douglas forms). Judged by the mean squared error (MSE), the quadratic specification (MSE= 0.0021) provides a better fit than the Cobb-Douglas specification (MSE= 0.0033).

6. Discussion and conclusion

In this paper, we apply experimental data to study strategic medical choices in three market settings: monopoly, duopoly and quadropoly. Our results suggest that a substantial change in behavior can be explained by the fixed preference QREC model. The results provide useful nuances to the recent literature on whether markets erode social responsibility.

The descriptive results show that subjects provide significantly more patient benefit as markets becomes more competitive. If a theoretical model was absent, it might be tempting to conclude that *"Competition raises moral values"* or, alternatively, that *"Competition crowds in pro-social motivation"*. However, we interpret the results with reference to economic theory and the assumption of fixed preferences.

The substantial difference in behavior between monopoly, duopoly and quadropoly market settings can be attributed to changes in individuals' scale parameter while keeping preferences fixed. The scale parameter is a measure of the randomness in behavior, or equivalently, a measure of determinism in behavior. We find that the scale parameter rises as markets become more competitive, implying a higher degree of determinism in behavior. One possible intuitive explanation is that competition triggers decision-makers' attention. Another possible explanation is that competition positively affects the perceived return from cognitive effort.

An important aspect of our experiment is the large number of cells in payoff matrices; 11^2 in duopoly and 11^4 in quadropoly. As a result, many potential choice combinations will remain unrealized in the data. Therefore, one may argue that the experimental design provide unfavorable conditions for maximum likelihood estimation of the parameters of a QREC model. However, results from Monto Carlo simulations show that the two-step estimator produces accurate estimates with a moderate sample size, even under these unfavorable conditions. Further, since the two-step estimator does not rely on the assumption that an equilibrium is reached, it is robust with regard to a hypothetical scenario where the data is *not* generated by a QRE.

The seminal contribution by McKelvey and Palfrey (1995) revealed that the toolkit of choice modellers can be applied to the study of strategic behavior. Yet, applications of choice models in studying strategic decision-making in games with vector-payoff have not reached its potential. One may argue that the empirical game theory literature can benefit from knowledge translation from the advances in the vast choice modelling literature (Revelt and Train, 1998; McFadden, 2001; Hess and Daly, 2010; Jessie and Saari, 2016), where context-dependent scale parameters have been a central research topic for decades (Louviere et al., 1999, 2002; Louviere and Eagle, 2006; Louviere and Meyer, 2008; Wang et al., 2020).

The economic literature comprises many contributions that reject the assumptions of perfect rationality and purely selfish behavior. The QREC model provides a consistent economic model of strategic behavior that relies on less restrictive assumptions. Over the last decades, numerous laboratory experiments have generated a large base of data from, for example, public good games and ultimatum games. The fact that these games will constitute multi-criteria games for altruistic decision-makers are typically ignored, and NE or other equilibrium concepts for single-criterion games are applied in the analyses. Re-examination of the existing data by fitting QREC models has substantial research potential. Behaviors which have been referred to as anomalies can be consistent with a QREC model with a plausible utility function. Hence, re-examination of the data is likely to contribute to new insights, and provide explanations for phenomena that have been poorly understood. Moore methodological research on QREC models would benefit this proposed research. For example, while the two-step estimator is robust and consistent, it is a research question whether it is more or less efficient than the full information maximum likelihood estimator described by Moffatt (2015).

Acknowledgements

Comments, questions and suggestions from two anonymous reviewers have contributed to substantial improvements of this paper. We thank Daniel Wiesen for joint collaboration on experimental design in the early stages of this project, and for managing the experimental sessions in Cologne. We thank Emanuel Castillo for z-Tree programming; Anne Classen, Tobias Danzeisen, Lena Kuhne and Theresa Schwan for research assistance. We are grateful for comments and suggestions from John K. Dagsvik, Karen E. Hauge, Arne Risa Hole, Albert Ma, Knut R. Wangen, Erik Ø. Sørensen, and seminar participants at Norwegian School of Economics. We gratefully acknowledge financial support from the Research Council of Norway (IRECOHEX, No. 231776 and NORCHER, No. 296114) who provided the funds for designing experiment, development of z-Tree program, conducting experiment as well as writing and publishing of manuscript.

References

Anderson, S. P., J. K. Goeree, and C. A. Holt (2001): "Minimum-effort coordination games: Stochastic potential and logit equilibrium," *Games and Economic Behavior*, 34, 177–199.

Bajari, P. and A. Hortacsu (2005): "Are structural estimates of auction models reasonable? Evidence from experimental data," *Journal of Political Economy*, 113, 703–741.

Bartling, B., E. Fehr, and Y. Özdemir (2020): "Does market interaction erode moral values?" *Working Paper available at SSRN.*

Bartling, B., R. A. Weber, and L. Yao (2015): "Do markets erode social responsibility?" *The Quarterly Journal of Economics*, 130, 219–266.

Battese, G. E. (1997): "A note on the Estimation of Cobb-Douglas Production Functions when Some Explanatory Variables Have Zero Values." *Journal of Agricultural Economics*, 48, 250–252.

Blavatskyy, P. R. (2007): "Stochastic expected utility theory," *Journal of Risk and Uncertainty*, 34, 259–286.

Camerer, C. F. (2011): *Behavioral game theory: Experiments in strategic interaction*, Princeton University Press.

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

——— (2015): "Stochastic models for risky choices: A comparison of different axiomatizations," *Journal of Mathematical Economics*, 60, 81–88.

——— (2016): "What independent random utility representations are equivalent to the IIA assumption?" *Theory and Decision*, 80, 495–499.

——— (2018): "Invariance axioms and functional form restrictions in structural models," *Mathematical Social Sciences*, 91, 85–95.

——— (forthcoming): "On the Equilibria in Models of Social Interaction and Quantal Response Equilibrium," *HERO Working Paper Series.*

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

Falk, A., S. Meier, and C. Zehnder (2013): "Do Lab Experiments Misinterpret Social Preferences? The Case of Self-Selected Student Samples," *Journal of the European Economic Association*, 11, 839–852.

Falk, A. and N. Szech (2013): "Morals and Markets," *Science*, 340, 707–711.

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.

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

Gaynor, M. and R. J. Town (2011): "Competition in health care markets," in *Handbook of health economics*, Elsevier, vol. 2, 499–637.

Gaynor, M. and W. B. Vogt (2000): "Antitrust and competition in health care markets," *Handbook of health economics*, 1, 1405–1487.

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

——— (2005): "An experimental study of costly coordination," *Games and Economic Behavior*, 51, 349–364.

Goeree, J. K., C. A. Holt, and S. K. Laury (2002): "Private costs and public benefits: unraveling the effects of altruism and noisy behavior," *Journal of Public Economics*, 83, 255–276.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2005): "Regular quantal response equilibrium," *Experimental Economics*, 8, 347–367.

——— (2010): "Quantal response equilibria," in *Behavioural and Experimental Economics*, Springer, 234–242.

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

Gul, F. and W. Pesendorfer (2006): "Random expected utility," *Econometrica*, 74, 121–146.

Haile, P. A., A. Hortaçsu, and G. Kosenok (2008): "On the empirical content of quantal response equilibrium," *American Economic Review*, 98, 180–200.

Harsanyi, J. C. (1973): "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, 2, 1–23.

Harsanyi, J. C., R. Selten, et al. (1988): "A general theory of equilibrium selection in games," *MIT Press Books*, 1.

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.

Hess, S. and A. Daly (2010): *Choice modelling: the state-of-the-art and the state-of-practice: Proceedings from the Inaugural International Choice Modelling Conference*, Emerald Group Publishing Limited.

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

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

Holte, J. H., P. Sivey, B. Abelsen, and J. A. Olsen (2016): "Modelling nonlinearities and reference dependence in general practitioners' income preferences," *Health Economics*, 25, 1020–1038.

Jessie, D. T. and D. G. Saari (2016): "From the Luce choice axiom to the quantal response equilibrium," *Journal of Mathematical Psychology*, 75, 3–9.

Ke, S. (2018): "Rational expectation of mistakes and a measure of error-proneness," *Theoretical Economics*, 13, 527–552.

Kirchler, M., J. Huber, M. Stefan, and M. Sutter (2015): "Market design and moral behavior," *Management Science*, 62, 2615–2625.

Kolstad, J. R. (2011): "How to make rural jobs more attractive to health workers. Findings from a discrete choice experiment in Tanzania," *Health Economics*, 20, 196–211.

Loukopoulos, L. D., R. Dismukes, and I. Barshi (2009): "The perils of multitasking," *AeroSafety World*, 4, 18–23.

Louviere, J., D. Street, R. Carson, A. Ainslie, J. Deshazo, T. Cameron, D. Hensher, R. Kohn, and T. Marley (2002): "Dissecting the random component of utility," *Marketing letters*, 13, 177–193.

Louviere, J. J. and T. Eagle (2006): "Confound it! That pesky little scale constant messes up our convenient assumptions," in *Sawtooth Software Conference*, Sawtooth Software Inc.

Louviere, J. J. and R. J. Meyer (2008): "Formal choice models of informal choices," in *Review of Marketing Research*, Emerald Group Publishing Limited, 3–32.

Louviere, J. J., R. J. Meyer, D. S. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher, and J. Irwin (1999): "Combining sources of preference data for modeling complex decision processes," *Marketing Letters*, 10, 205–217.

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

Ma, C.-t. A. and M. Manove (1993): "Bargaining with deadlines and imperfect player control," *Econometrica: Journal of the Econometric Society*, 1313–1339.

Mackowiak, P. A. (2020): "George Washington's Recurrent Health Problems and Fatal Infection Reexamined," *Clinical Infectious Diseases*, ciaa1324.

Matějka, F. and A. McKay (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–98.

McCubbins, M. D., M. Turner, and N. Weller (2013): "Testing the foundations of quantal response equilibrium," in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, Springer, 144–153.

McFadden, D. (1974): "The measurement of urban travel demand," *Journal of Public Economics*, 3, 303–328.

——— (1981): "Econometric models of probabilistic choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, MIT Press, 198–272.

——— (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*, ed. by F. B. and M. C.F., Springer, 73–110.

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

Moffatt, P. G. (2015): *Experimetrics: Econometrics for experimental economics*, Macmillan International Higher Education.

Nash, J. F. (1950): "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

Requate, T. and I. Waichman (2011): ""A profit table or a profit calculator?" A note on the design of Cournot oligopoly experiments," *Experimental Economics*, 14, 36–46.

Revelt, D. and K. Train (1998): "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, 80, 647–657.

Shapley, L. S. and F. D. Rigby (1959): "Equilibrium points in games with vector payoffs," *Naval Research Logistics Quarterly*, 6, 57–61.

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

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

Thurstone, L. L. (1927a): "A law of comparative judgment." *Psychological Review*, 34, 273.

——— (1927b): "Psychophysical analysis," *The American Journal of Psychology*, 38, 368–389.

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

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

van der Pol, M., G. Currie, S. Kromm, and M. Ryan (2014): "Specification of the utility function in discrete choice experiments," *Value in Health*, 17, 297–301.

Van Der Pol, M., A. Shiell, F. Au, D. Jonhston, and S. Tough (2010): "Eliciting individual preferences for health care: a case study of perinatal care," *Health Expectations*, 13, 4–12.

Voorneveld, M., S. Grahn, and M. Dufwenberg (2000): "Ideal equilibria in noncooperative multicriteria games," *Mathematical methods of operations research*, 52, 65–77.

Voorneveld, M., D. Vermeulen, and P. Borm (1999): "Axiomatizations of Pareto equilibria in multicriteria games," *Games and economic behavior*, 28, 146–154.

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

Wang, J., T. Iversen, H. Hennig-Schmidt, and G. Godager (2020): "Are patient-regarding preferences stable? Evidence from a laboratory experiment with physicians and medical students from different countries," *European Economic Review*, 125.

Wright, J. R. and K. Leyton-Brown (2017): "Predicting human behavior in unrepeated, simultaneous-move games," *Games and Economic Behavior*, 106, 16–37.

Zeleny, M. (1975): "Games with multiple payoffs," *International Journal of game theory*, 4, 179–191.

Appendix A. Payoff matrices in three markets

We present here the 24 payoff matrices of the experiment. Table A.1 reports the payoff matrices in the eight monopoly dictator game. The payoff matrices of the eight duopoly games are reported in Tables A.2-A.9. A complete payoff matrix in quadropoly contains 11^4 cells, we therefore report an excerpt of player n's (row) payoffs when three opponents (column) act identically. The payoff matrices of the eight quadropoly games are reported in Tables A.10-A.17.

Table A.1: Payoffs in monopoly.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8
	Π, B	Π, B	Π, B	Π, B	Π, B	Π, B	Π, B	Π, B
0	1000 , 0	1000 , 0	1500 , 0	1500 , 0	1000 , 0	1000 , 0	1500 , 0	1500 , 0
1	990 , 100	993 , 100	1490 , 50	1490 , 100	990 , 50	993 , 50	1493 , 100	1493 , 50
2	960 , 200	970 , 200	1460 , 100	1460 , 200	960 , 100	970 , 100	1470 , 200	1470 , 100
3	910 , 300	932 , 300	1410 , 150	1410 , 300	910 , 150	932 , 150	1433 , 300	1433 , 150
4	840 , 400	880 , 400	1340 , 200	1340 , 400	840 , 200	880 , 200	1380 , 400	1380 , 200
5	750 , 500	813 , 500	1250 , 250	1250 , 500	750 , 250	813 , 250	1313 , 500	1313 , 250
6	640 , 600	730 , 600	1140 , 300	1140 , 600	640 , 300	730 , 300	1230 , 600	1230 , 300
7	510 , 700	632 , 700	1010 , 350	1010 , 700	510 , 350	632 , 350	1133 , 700	1133 , 350
8	360 , 800	520 , 800	860 , 400	860 , 800	360 , 400	520 , 400	1020 , 800	1020 , 400
9	190 , 900	393 , 900	690 , 450	690 , 900	190 , 450	393 , 450	893 , 900	893 , 450
10	0 , 1000	250 , 1000	500 , 500	500 , 1000	0 , 500	250 , 500	750 , 1000	750 , 500

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.2: Payoff matrix, patient game 1 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	500,0	270,73	120,176	50,285	20,392	10,495	0,600	0,700	0,800	0,900	0,1000
1	723,73	495,100	267,173	119,276	50,385	20,492	10,595	0,700	0,800	0,900	0,1000
2	845,176	701,173	480,200	259,273	115,376	48,485	19,592	10,695	0,800	0,900	0,1000
3	865,285	801,276	664,273	455,300	246,373	109,476	46,585	18,692	9,795	0,900	0,1000
4	823,392	798,385	739,376	613,373	420,400	227,473	101,576	42,685	17,792	8,895	0,1000
5	743,495	735,492	713,485	660,476	548,473	375,500	203,573	90,676	38,785	15,892	8,995
6	640,600	634,595	627,592	608,585	563,576	467,573	320,600	173,673	77,776	32,885	13,992
7	510,700	510,700	505,695	500,692	485,685	449,676	372,673	255,700	138,773	61,876	26,985
8	360,800	360,800	360,800	356,795	353,792	342,785	317,776	263,773	180,800	97,873	43,976
9	190,900	190,900	190,900	190,900	188,895	186,892	181,885	167,876	139,873	95,900	51,973
10	0,1000	0,1000	0,1000	0,1000	0,1000	0,995	0,992	0,985	0,976	0,973	0,1000

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.3: Payoff matrix, patient game 2 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	500,0	270,73	120,176	50,285	20,392	10,495	0,600	0,700	0,800	0,900	0,1000
1	725,73	496,100	268,173	119,276	50,385	20,492	10,595	0,700	0,800	0,900	0,1000
2	854,176	708,173	485,200	262,273	116,376	49,485	19,592	10,695	0,800	0,900	0,1000
3	886,285	821,276	681,273	466,300	252,373	112,476	47,585	19,692	9,795	0,900	0,1000
4	862,392	836,385	774,376	642,373	440,400	238,473	106,576	44,685	18,792	9,895	0,1000
5	804,495	796,492	772,485	715,476	593,473	406,500	219,573	98,676	41,785	16,892	8,995
6	730,600	723,595	715,592	694,585	642,576	533,573	365,600	197,673	88,776	37,885	15,992
7	633,700	633,700	626,695	620,692	601,685	557,676	462,673	316,700	171,773	76,876	32,985
8	520,800	520,800	520,800	515,795	510,792	494,785	458,776	380,773	260,800	140,873	62,976
9	393,900	393,900	393,900	393,900	389,895	385,892	373,885	345,876	287,873	196,900	106,973
10	250,1000	250,1000	250,1000	250,1000	250,1000	248,995	245,992	238,985	220,976	183,973	125,1000

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.4: Payoff matrix, patient game 3 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	750,0	570,31	405,73	270,123	180,176	120,230	75,285	45,340	30,392	15,446	15,495
1	924,31	745,50	566,81	402,123	268,173	179,226	119,280	75,335	45,390	30,442	15,496
2	1066,73	905,81	730,100	555,131	394,173	263,223	175,276	117,330	73,385	44,440	29,492
3	1156,123	1029,123	874,131	705,150	536,181	381,223	254,273	169,326	113,380	71,435	42,490
4	1179,176	1099,173	978,173	831,181	670,200	509,231	362,273	241,323	161,376	107,430	67,485
5	1150,230	1100,226	1025,223	913,223	775,231	625,250	475,281	338,323	225,373	150,426	100,480
6	1083,285	1049,280	1003,276	935,273	832,273	707,281	570,300	433,331	308,373	205,423	137,476
7	980,340	960,335	929,330	889,326	828,323	737,323	626,331	505,350	384,381	273,423	182,473
8	843,392	834,390	817,385	791,380	757,376	705,373	628,373	533,381	430,400	327,431	232,473
9	683,446	676,442	669,440	656,435	635,430	607,426	566,423	504,423	428,431	345,450	262,481
10	495,495	495,496	490,492	485,490	475,485	460,480	440,476	410,473	365,473	310,481	250,500

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.5: Payoff matrix, patient game 4 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	750,0	405,73	180,176	75,285	30,392	15,495	0,600	0,700	0,800	0,900	0,1000
1	1088,73	745,100	402,173	179,276	75,385	30,492	15,595	0,700	0,800	0,900	0,1000
2	1285,176	1066,173	730,200	394,273	175,376	73,485	29,592	15,695	0,800	0,900	0,1000
3	1340,285	1241,276	1029,273	705,300	381,373	169,476	71,585	28,692	14,795	0,900	0,1000
4	1313,392	1273,385	1179,376	978,373	670,400	362,473	161,576	67,685	27,792	13,895	0,1000
5	1238,495	1225,492	1188,485	1100,476	913,473	625,500	338,573	150,676	63,785	25,892	13,995
6	1140,600	1129,595	1117,592	1083,585	1003,576	832,573	570,600	308,673	137,776	57,885	23,992
7	1010,700	1010,700	1000,695	990,692	960,685	889,676	737,673	505,700	273,773	121,876	51,985
8	860,800	860,800	860,800	851,795	843,792	817,785	757,776	628,773	430,800	232,873	103,976
9	690,900	690,900	690,900	690,900	683,895	676,892	656,885	607,876	504,873	345,900	186,973
10	500,1000	500,1000	500,1000	500,1000	500,1000	495,995	490,992	475,985	440,976	365,973	250,1000

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.6: Payoff matrix, patient game 5 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	500,0	380,31	270,73	180,123	120,176	80,230	50,285	30,340	20,392	10,446	10,495
1	614,31	495,50	376,81	267,123	178,173	119,226	79,280	50,335	30,390	20,442	10,496
2	701,73	595,81	480,100	365,131	259,173	173,223	115,276	77,330	48,385	29,440	19,492
3	746,123	664,123	564,131	455,150	346,181	246,223	164,273	109,326	73,380	46,435	27,490
4	739,176	689,173	613,173	521,181	420,200	319,231	227,273	151,323	101,376	67,430	42,485
5	690,230	660,226	615,223	548,223	465,231	375,250	285,281	203,323	135,373	90,426	60,480
6	608,285	589,280	563,276	525,273	467,273	397,281	320,300	243,331	173,373	115,423	77,476
7	495,340	485,335	469,330	449,326	418,323	372,323	316,331	255,350	194,381	138,423	92,473
8	353,392	349,390	342,385	331,380	317,376	295,373	263,373	223,381	180,400	137,431	97,473
9	188,446	186,442	184,440	181,435	175,430	167,426	156,423	139,423	118,431	95,450	72,481
10	0,495	0,496	0,492	0,490	0,485	0,480	0,476	0,473	0,473	0,481	0,500

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.7: Payoff matrix, patient game 6 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	500,0	380,31	270,73	180,123	120,176	80,230	50,285	30,340	20,392	10,446	10,495
1	615,31	496,50	377,81	268,123	179,173	119,226	79,280	50,335	30,390	20,442	10,496
2	708,73	601,81	485,100	369,131	262,173	175,223	116,276	78,330	49,385	29,440	19,492
3	765,123	681,123	578,131	466,150	354,181	252,223	168,273	112,326	75,380	47,435	28,490
4	774,176	722,173	642,173	546,181	440,200	334,231	238,273	158,323	106,376	70,430	44,485
5	748,230	715,226	666,223	593,223	504,231	406,250	309,281	219,323	146,373	98,426	65,480
6	694,285	672,280	642,276	599,273	533,273	453,281	365,300	277,331	197,373	131,423	88,476
7	614,340	601,335	582,330	557,326	519,323	462,323	392,331	316,350	240,381	171,423	114,473
8	510,392	504,390	494,385	478,380	458,376	426,373	380,373	322,381	260,400	198,431	140,473
9	389,446	385,442	381,440	373,435	361,430	345,426	322,423	287,423	243,431	196,450	149,481
10	248,495	248,496	245,492	243,490	238,485	230,480	220,476	205,473	183,473	155,481	125,500

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.8: Payoff matrix, patient game 7 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	750,0	405,73	180,176	75,285	30,392	15,495	0,600	0,700	0,800	0,900	0,1000
1	1090,73	746,100	403,173	179,276	75,385	30,492	15,595	0,700	0,800	0,900	0,1000
2	1294,176	1073,173	735,200	397,273	176,376	74,485	29,592	15,695	0,800	0,900	0,1000
3	1361,285	1261,276	1046,273	716,300	387,373	172,476	72,585	29,692	14,795	0,900	0,1000
4	1352,392	1311,385	1214,376	1007,373	690,400	373,473	166,576	69,685	28,792	14,895	0,1000
5	1299,495	1286,492	1247,485	1155,476	958,473	656,500	354,573	158,676	66,785	26,892	13,995
6	1230,600	1218,595	1205,592	1169,585	1082,576	898,573	615,600	332,673	148,776	62,885	25,992
7	1133,700	1133,700	1121,695	1110,692	1076,685	997,676	827,673	566,700	306,773	136,876	57,985
8	1020,800	1020,800	1020,800	1010,795	1000,792	969,785	898,776	745,773	510,800	275,873	122,976
9	893,900	893,900	893,900	893,900	884,895	875,892	848,885	785,876	652,873	446,900	241,973
10	750,1000	750,1000	750,1000	750,1000	750,1000	743,995	735,992	713,985	660,976	548,973	375,1000

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.9: Payoff matrix, patient game 8 in duopoly.

	The opponent										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	750,0	570,31	405,73	270,123	180,176	120,230	75,285	45,340	30,392	15,446	15,495
1	925,31	746,50	567,81	403,123	269,173	179,226	119,280	75,335	45,390	30,442	15,496
2	1073,73	911,81	735,100	559,131	397,173	265,223	176,276	118,330	74,385	44,440	29,492
3	1175,123	1046,123	888,131	716,150	544,181	387,223	258,273	172,326	115,380	72,435	43,490
4	1214,176	1132,173	1007,173	856,181	690,200	524,231	373,273	248,323	166,376	110,430	69,485
5	1208,230	1155,226	1076,223	958,223	814,231	656,250	499,281	354,323	236,373	158,426	105,480
6	1169,285	1132,280	1082,276	1009,273	898,273	763,281	615,300	467,331	332,373	221,423	148,476
7	1099,340	1076,335	1042,330	997,326	929,323	827,323	702,331	566,350	430,381	306,423	204,473
8	1000,392	989,390	969,385	938,380	898,376	836,373	745,373	632,381	510,400	388,431	275,473
9	884,446	875,442	866,440	848,435	821,430	785,426	732,423	652,423	553,431	446,450	339,481
10	743,495	743,496	735,492	728,490	713,485	690,480	660,476	615,473	548,473	465,481	375,500

Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.10: Payoff matrix, patient game 1 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	250,0	110,89	40,191	20,295	10,397	0,500	0,599	0,699	0,799	0,899	0,999
1	475,48	247,100	109,189	40,291	20,395	10,497	0,600	0,699	0,799	0,899	0,999
2	682,142	461,148	240,200	106,289	38,391	19,495	10,597	0,700	0,799	0,899	0,999
3	792,261	646,242	437,248	228,300	100,389	36,491	18,595	9,697	0,800	0,899	0,999
4	798,379	731,361	596,342	403,348	210,400	92,489	34,591	17,695	8,797	0,900	0,999
5	735,490	713,479	653,461	533,442	360,448	188,500	83,589	30,691	15,795	8,897	0,1000
6	634,596	627,590	608,579	557,561	454,542	307,548	160,600	70,689	26,791	13,895	6,997
7	510,698	505,696	500,690	485,679	444,661	362,642	245,648	128,700	56,789	20,891	10,995
8	360,799	360,798	356,796	353,790	342,779	313,761	256,742	173,748	90,800	40,889	14,991
9	190,900	190,899	190,898	188,896	186,890	181,879	165,860	135,842	91,848	48,900	21,989
10	0,1000	0,1000	0,999	0,998	0,995	0,991	0,979	0,960	0,941	0,948	0,1000

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.11: Payoff matrix, patient game 2 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	250,0	110,89	40,191	20,295	10,397	0,500	0,599	0,699	0,799	0,899	0,999
1	476,48	248,100	109,189	40,291	20,395	10,497	0,600	0,699	0,799	0,899	0,999
2	689,142	466,148	243,200	107,289	39,391	19,495	10,597	0,700	0,799	0,899	0,999
3	811,261	662,242	448,248	233,300	103,389	37,491	19,595	9,697	0,800	0,899	0,999
4	836,379	766,361	625,342	422,348	220,400	97,489	35,591	18,695	9,797	0,900	0,999
5	796,490	772,479	707,461	577,442	390,448	203,500	89,589	33,691	16,795	8,897	0,1000
6	723,596	715,590	694,579	635,561	518,542	350,548	183,600	80,689	29,791	15,895	7,997
7	633,698	626,696	620,690	601,679	550,661	449,642	304,648	158,700	70,789	25,891	13,995
8	520,799	520,798	515,796	510,790	494,779	452,761	369,742	250,748	130,800	57,889	21,991
9	393,900	393,899	393,898	389,896	385,890	373,879	341,860	279,842	188,848	98,900	43,989
10	250,1000	250,1000	250,999	250,998	248,995	245,991	238,979	218,960	178,941	120,948	63,1000

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.12: Payoff matrix, patient game 3 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	375,0	255,42	165,89	105,140	60,191	45,243	30,295	15,347	15,397	0,448	0,500
1	522,18	373,50	253,92	164,139	104,189	60,241	45,293	30,345	15,397	15,447	0,498
2	701,48	511,68	365,100	248,141	161,189	102,239	58,291	44,343	29,395	15,447	15,497
3	846,90	677,98	494,118	353,150	240,191	155,239	99,289	56,341	42,393	28,445	14,497
4	951,142	804,140	643,148	469,168	335,200	228,241	147,289	94,339	54,391	40,443	27,495
5	1000,201	888,192	750,190	600,198	438,218	313,250	213,291	138,339	88,389	50,441	38,493
6	992,261	912,251	809,242	684,240	547,248	399,268	285,300	194,341	125,389	80,439	46,491
7	929,321	879,311	808,301	717,292	606,290	485,298	354,318	253,350	172,391	111,439	71,489
8	817,379	791,371	748,361	688,351	611,342	516,340	413,348	301,368	215,400	146,441	95,489
9	669,436	656,429	635,421	600,411	552,401	490,392	414,390	331,398	242,418	173,450	117,491
10	490,490	485,486	475,479	460,471	435,461	400,451	355,442	300,440	240,448	175,468	125,500

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.13: Payoff matrix, patient game 4 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	375,0	165,89	60,191	30,295	15,397	0,500	0,599	0,699	0,799	0,899	0,999
1	715,48	373,100	164,189	60,291	30,395	15,497	0,600	0,699	0,799	0,899	0,999
2	1037,142	701,148	365,200	161,289	58,391	29,495	15,597	0,700	0,799	0,899	0,999
3	1227,261	1001,242	677,248	353,300	155,389	56,491	28,595	14,697	0,800	0,899	0,999
4	1273,379	1166,361	951,342	643,348	335,400	147,489	54,591	27,695	13,797	0,900	0,999
5	1225,490	1188,479	1088,461	888,442	600,448	313,500	138,589	50,691	25,795	13,897	0,1000
6	1129,596	1117,590	1083,579	992,561	809,542	547,548	285,600	125,689	46,791	23,895	11,997
7	1010,698	1000,696	990,690	960,679	879,661	717,642	485,648	253,700	111,789	40,891	20,995
8	860,799	860,798	851,796	843,790	817,779	748,761	611,742	413,748	215,800	95,889	34,991
9	690,900	690,899	690,898	683,896	676,890	656,879	600,860	490,842	331,848	173,900	76,989
10	500,1000	500,1000	500,999	500,998	495,995	490,991	475,979	435,960	355,941	240,948	125,1000

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.14: Payoff matrix, patient game 5 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	250,0	170,42	110,89	70,140	40,191	30,243	20,295	10,347	10,397	0,448	0,500
1	347,18	247,50	168,92	109,139	69,189	40,241	30,293	20,345	10,397	10,447	0,498
2	461,48	336,68	240,100	163,141	106,189	67,239	38,291	29,343	19,395	10,447	10,497
3	546,90	437,98	319,118	228,150	155,191	100,239	64,289	36,341	27,393	18,445	9,497
4	596,142	504,140	403,148	294,168	210,200	143,241	92,289	59,339	34,391	25,443	17,495
5	600,201	533,192	450,190	360,198	263,218	188,250	128,291	83,339	53,389	30,441	23,493
6	557,261	512,251	454,242	384,240	307,248	224,268	160,300	109,341	70,389	45,439	26,491
7	469,321	444,311	408,301	362,292	306,290	245,298	179,318	128,350	87,391	56,439	36,489
8	342,379	331,371	313,361	288,351	256,342	216,340	173,348	126,368	90,400	61,441	40,489
9	184,436	181,429	175,421	165,411	152,401	135,392	114,390	91,398	67,418	48,450	32,491
10	0,490	0,486	0,479	0,471	0,461	0,451	0,442	0,440	0,448	0,468	0,500

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.15: Payoff matrix, patient game 6 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	250,0	170,42	110,89	70,140	40,191	30,243	20,295	10,347	10,397	0,448	0,500
1	347,18	248,50	169,92	109,139	69,189	40,241	30,293	20,345	10,397	10,447	0,498
2	466,48	340,68	243,100	165,141	107,189	68,239	39,291	29,343	19,395	10,447	10,497
3	560,90	448,98	326,118	233,150	159,191	103,239	65,289	37,341	28,393	19,445	9,497
4	625,142	528,140	422,148	308,168	220,200	150,241	97,289	62,339	35,391	26,443	18,495
5	650,201	577,192	488,190	390,198	284,218	203,250	138,291	89,339	57,389	33,441	24,493
6	635,261	584,251	518,242	438,240	350,248	256,268	183,300	124,341	80,389	51,439	29,491
7	582,321	550,311	506,301	449,292	380,290	304,298	221,318	158,350	108,391	70,439	44,489
8	494,379	478,371	452,361	416,351	369,342	312,340	250,348	182,368	130,400	88,441	57,489
9	381,436	373,429	361,421	341,411	314,401	279,392	236,390	188,398	137,418	98,450	67,491
10	245,490	243,486	238,479	230,471	218,461	200,451	178,442	150,440	120,448	88,468	63,500

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.16: Payoff matrix, patient game 7 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	375,0	165,89	60,191	30,295	15,397	0,500	0,599	0,699	0,799	0,899	0,999
1	716,48	373,100	164,189	60,291	30,395	15,497	0,600	0,699	0,799	0,899	0,999
2	1044,142	706,148	368,200	162,289	59,391	29,495	15,597	0,700	0,799	0,899	0,999
3	1246,261	1017,242	688,248	358,300	158,389	57,491	29,595	14,697	0,800	0,899	0,999
4	1311,379	1201,361	980,342	662,348	345,400	152,489	55,591	28,695	14,797	0,900	0,999
5	1286,490	1247,479	1142,461	932,442	630,448	328,500	144,589	53,691	26,795	13,897	0,1000
6	1218,596	1205,590	1169,579	1070,561	873,542	590,548	308,600	135,689	49,791	25,895	12,997
7	1133,698	1121,696	1110,690	1076,679	985,661	804,642	544,648	283,700	125,789	45,891	23,995
8	1020,799	1020,798	1010,796	1000,790	969,779	887,761	724,742	490,748	255,800	112,889	41,991
9	893,900	893,899	893,898	884,896	875,890	848,879	776,860	634,842	428,848	223,900	98,989
10	750,1000	750,1000	750,999	750,998	743,995	735,991	713,979	653,960	533,941	360,948	188,1000

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Table A.17: Payoff matrix, patient game 8 in quadropoly.

	The 3 opponents										
	0	1	2	3	4	5	6	7	8	9	10
Player n	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B	Π,B
0	375,0	255,42	165,89	105,140	60,191	45,243	30,295	15,347	15,397	0,448	0,500
1	522,18	373,50	254,92	164,139	104,189	60,241	45,293	30,345	15,397	15,447	0,498
2	706,48	515,68	368,100	250,141	162,189	103,239	59,291	44,343	29,395	15,447	15,497
3	860,90	688,98	501,118	358,150	244,191	158,239	100,289	57,341	43,393	29,445	14,497
4	980,142	828,140	662,148	483,168	345,200	235,241	152,289	97,339	55,391	41,443	28,495
5	1050,201	932,192	788,190	630,198	459,218	328,250	223,291	144,339	92,389	3,441	39,493
6	1070,261	984,251	873,242	738,240	590,248	431,268	308,300	209,341	135,389	86,439	49,491
7	1042,321	985,311	906,301	804,292	680,290	544,298	396,318	283,350	193,391	25,439	79,489
8	969,37	938,371	887,361	816,351	724,342	612,340	490,348	357,368	255,400	173,441	112,489
9	866,436	848,429	821,421	776,411	714,401	634,392	536,390	428,398	312,418	223,450	152,491
10	735,490	728,486	713,479	690,471	653,461	600,451	533,442	450,440	360,448	263,468	188,500

We present an excerpt where the three opponents (column) act identically. Payoffs are measured in Taler (100 Taler = 1 Euro).

Appendix B. Instructions for participants of the experiment

You are taking part in an economic decision-making experiment. Please carefully read the instructions. It is very important that you do not speak with other participants for the duration of the experiment. If you break these rules, you could be excluded from the experiment and not receive any payment. If you do not understand something, please take another look at the instructions. If you still have questions, please raise your hand. We will come to you at your cubicle and answer your questions in private.

You can earn money in the course of the experiment. The amount of your earnings depends on your decisions and decisions made by other participants. At no time will you be told the names of the other participants. They will also not at any time be informed about your identity.

For showing up you will receive a fee of EUR 2.50.

All monetary amounts in this experiment are expressed in Taler, whereby the following applies: Taler 100 = EUR 1.

At the end of the experiment, the amount of money you earned will be paid to you in cash. Your decisions are made on the computer screen present in your cubicle. All data and answers will be evaluated anonymously. You were asked to draw your own personal cubicle number in order to maintain anonymity.

The experiment will last around 60 minutes and consists of three parts. Before each of the three parts you will receive detailed instructions and be asked to answer control questions pertaining to these instructions. Please note: Neither your decisions in the first part nor in the second part of the experiment have an influence on the other parts of the experiment.

We will ask you to answer a few questions at the end of the experiment. You will receive an additional payment for answering this questionnaire.

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

The demand for medical care by the various patient types is determined only after you have made your decisions about the quality of care for all eight types.

[Duopoly: You are randomly matched with another participant. This participant also decides in the role of a physician. Also this physician determines the quality for the same eight types of patients. The matching with this participants remains throughout the entire second part of the experiment. You and the other physician choose the quality simultaneously and independently from each other.]

[Quadropoly: You are randomly matched with three other participants. These participants also decide in the role of a physicians. Also these physicians determine the quality for the same eight types of patients. The matching with these participants remains throughout the entire third part of the experiment. You and the other physicians choose the quality simultaneously and independently from each other.]

In total, 100 patients of each type demand medical care. It will only be determined after you have made your decisions about the quality of care for all eight types how many of the 100 patients of each type who seek treatment from you.

[Duopoly: Only after you and the other physician, you are matched with, decided upon the quality of medical treatment for the eight patients, it is determined how many of the 100 patients seek treatment from you and the other physician.]

[Quadropoly: Only after you and the others physicians, you are matched with, decided upon the quality of medical treatment for the eight patients, it is determined how many of the 100 patients seek treatment from you and the other physicians.]

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

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

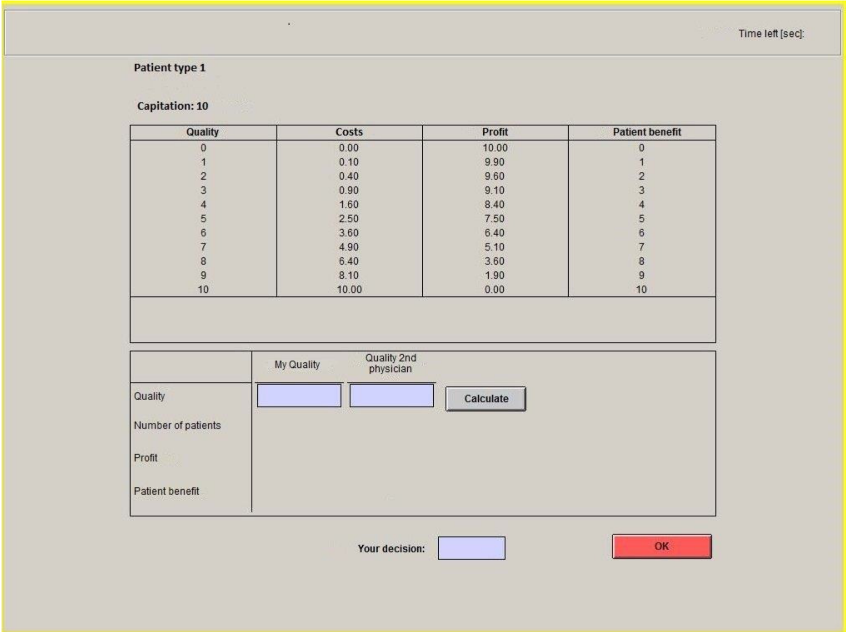
(when read: your earnings are equal to the difference between the lump sum and the costs that arise from the quality of care you have chosen, multiplied by the number of patients who seek treatment from you.)

With the quality of care you choose, you determine not only your own earnings, but also the utility enjoyed by the patient. The amount of the lump sum, your costs, your earnings, and the patient's utility will be displayed on your screen (as illustrated below) for each patient type. Before you choose the quality of care for each patient type, you have the opportunity to click on the "calculator" button and thereby calculate patients' potential demand for treatment (as illustrated below). You can enter the quality you would like to provide as many times as you want. Clicking on the "calculate" button provides you with information about the number of patients who would seek care given the quality level you entered. In addition, you receive information about the resulting earnings and patient utility. You define the quality of care that you wish to provide by entering that quality in the field "your decision" and confirming this entry with "OK."

Payment. After the conclusion of the experiment, one of the 8 decisions will be randomly chosen to function as the relevant round for determining your payment for this part of the experiment. The earnings from this randomly-chosen round will be converted into Euro at the end of the experiment and paid out to you in cash. There are no participants present in the lab who take on the role of patients. An actual patient will benefit from the patient utility resulting from the quality of care you selected in the randomly-chosen round: A monetary value equalling the patient utility derived from your decision, multiplied by the number of patients who seek treatment from you, will be transferred to Christoffel Blindenmission Deutschland e.V., 64625 Bensheim. This organisation will use the funds to enable the treatment of patients suffering from cataracts, a serious eye condition.

Control questions. Before proceeding to the decisions in the experiment, we would like to ask you to answer several control questions. These control questions should make it easier for you to become acquainted with the decision-making situation. If you have questions about

[Duopoly: 2nd example screen]



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

this, please raise your hand. The first part of the experiment will begin after all participants have correctly answered the control questions.

Payment Procedure. In order to ensure that payments to the participants and the transfer of the monetary donation to Christoffel Blindenmission Deutschland e.V. are carried out correctly, an overseer will be randomly chosen after the third part of the experiment. The overseer receives a fee of EUR 5 in addition to his or her regular payment from the experiment. The overseer will affirm that the transfer to Christoffel Blindenmission is correctly carried out by the financial administration of the University of Cologne. For the transfer to Christoffel Blindenmission, the overseer will fill out a payment order to Christoffel Blindenmission with the amount, in Euro, that corresponds to the patient utility realized in the randomly selected round. The financial administration of the University of Cologne will then execute payment of the donation to Christoffel Blindenmission using funds allocated for this experiment. The form will be placed in a stamped envelope addressed to the financial administration of the University of Cologne. The overseer and the experimenter will jointly deposit this envelope in the nearest mailbox.

The overseer will confirm by signing a form that he or she properly carried out the assigned tasks, as described above. A copy of this form, as well as a copy of the confirmation from Christoffel Blindenmission that the donation was received, can be requested by all participants from the office of the Seminar of Personnel Economics and Human Resource Management. The copies will be sent by e-mail.

Appendix C. Scalarized games

Conditional on parameter estimates from Table 4, and the 16 payoff matrices where competition is present (Tables A.2-A.9 and Tables A.10-A.17 in Appendix A), we can compute the matrices comprising the scalar utility payoffs in each game. In the following Tables (C.1-C.16), we present utility payoff matrices for duopoly and quadropoly, assuming the quadratic preferences. We indicate the symmetric pure strategy NE as well as the dominated pure strategies in each of the games. There are no dominating pure strategies in any of the scalarized games. We see that the number of dominated strategies and the number of NE differ substantially between games. For example, while the pure strategy Nash equilibrium in patient game 1 in duopoly is unique, we find many cases where the scalarized games have multiple NE. The scalarized patient game 7 in duopoly, for example, has six different NE.

Table C.1: Duopoly, quadratic utility payoff, patient game 1

	The opponent										
Player n	0§	1§	2§	3	4	5	6	7	8	9	10§
0§	4.03	2.80	2.03	1.86	1.98	2.20	2.37	2.53	2.63	2.66	2.62
1§	5.21	4.32	3.16	2.40	2.21	2.27	2.43	2.53	2.63	2.66	2.62
2§	5.63	5.31	4.51	3.42	2.69	2.48	2.49	2.59	2.63	2.66	2.62
3	5.68	5.61	5.32	4.59	3.59	2.91	2.67	2.65	2.68	2.66	2.62
4	5.60	5.59	5.52	5.25	4.57	3.67	3.04	2.80	2.73	2.71	2.62
5	5.46	5.45	5.43	5.36	5.09	4.48	3.67	3.10	2.86	2.75	2.67
6	5.23	5.22	5.22	5.19	5.11	4.85	4.29	3.59	3.10	2.85	2.70
7	4.88	4.88	4.88	4.87	4.84	4.75	4.51	4.02	3.44	3.02	2.77
8	4.37	4.37	4.37	4.37	4.36	4.33	4.25	4.04	3.66†	3.22	2.88
9	3.64	3.64	3.64	3.64	3.64	3.63	3.61	3.56	3.43	3.19	2.92
10§	2.62	2.62	2.62	2.62	2.62	2.63	2.63	2.63	2.64	2.64	2.62

† Pure strategy NE. § Dominated strategy.

Table C.2: Duopoly, quadratic utility payoff, patient game 2

	The opponent										
Player n	§0	1§	2	3	4	5	6	7	8	9	10
0§	4.03	2.80	2.03	1.86	1.98	2.20	2.37	2.53	2.63	2.66	2.62
1§	5.22	4.33	3.16	2.40	2.21	2.27	2.43	2.53	2.63	2.66	2.62
2§	5.64	5.33	4.53	3.44	2.70	2.48	2.50	2.59	2.63	2.66	2.62
3§	5.69	5.63	5.37	4.64	3.63	2.93	2.68	2.65	2.68	2.66	2.62
4	5.62	5.61	5.57	5.33	4.66	3.73	3.07	2.81	2.74	2.71	2.62
5	5.49	5.49	5.49	5.44	5.22	4.61	3.77	3.15	2.88	2.76	2.67
6	5.31	5.32	5.32	5.31	5.26	5.04	4.48	3.72	3.16	2.88	2.71
7	5.07	5.07	5.07	5.07	5.06	5.00	4.79	4.28	3.61	3.10	2.81
8	4.75	4.75	4.75	4.75	4.74	4.73	4.67	4.46	4.01	3.44	2.98
9	4.31	4.31	4.31	4.31	4.31	4.30	4.29	4.23	4.04	3.67†	3.19
10	3.72	3.72	3.72	3.72	3.72	3.73	3.72	3.71	3.66	3.53	3.25†

† Pure strategy NE. § Dominated strategy.

Table C.3: Duopoly, quadratic utility payoff, patient game 3

	The opponent										
Player n	0§	1§	2	3	4	5	6	7	8	9	10§
0§	5.15	4.50	3.72	3.00	2.52	2.24	2.08	2.02	2.07	2.11	2.24
1§	5.61	5.24	4.62	3.87	3.16	2.69	2.42	2.25	2.18	2.22	2.24
2	5.73	5.64	5.28	4.70	3.98	3.29	2.84	2.57	2.39	2.32	2.35
3	5.65	5.74	5.64	5.29	4.72	4.03	3.38	2.95	2.69	2.52	2.44
4	5.58	5.69	5.74	5.61	5.26	4.71	4.05	3.44	3.03	2.78	2.62
5	5.57	5.65	5.72	5.71	5.55	5.18	4.64	4.03	3.46	3.08	2.85
6	5.61	5.65	5.69	5.71	5.65	5.43	5.05	4.54	3.96	3.44	3.10
7	5.65	5.66	5.68	5.67	5.64	5.51	5.26	4.87	4.38	3.85	3.39
8	5.61	5.61	5.60	5.58	5.54	5.46	5.29	5.01	4.62	4.16	3.70
9	5.41	5.40	5.39	5.36	5.32	5.25	5.14	4.93	4.64	4.28†	3.89
10§	4.94	4.94	4.92	4.90	4.87	4.81	4.73	4.61	4.41	4.15	3.85

† Pure strategy NE. § Dominated strategy.

Table C.4: Duopoly, quadratic utility payoff, patient game 4

	The opponent										
Player n	0§	1	2	3	4	5	6	7	8	9	10
0§	5.15	3.72	2.52	2.08	2.07	2.24	2.37	2.53	2.63	2.66	2.62
1	5.72	5.33	4.02	2.86	2.41	2.35	2.47	2.53	2.63	2.66	2.62
2	5.34	5.72	5.42	4.23	3.13	2.66	2.57	2.63	2.63	2.66	2.62
3	4.94	5.29	5.68	5.44	4.35	3.32	2.85	2.72	2.72	2.66	2.62
4	4.75	4.93	5.25	5.61	5.39	4.39	3.43	2.97	2.80	2.74	2.62
5	4.72	4.78	4.93	5.21	5.51	5.28	4.36	3.47	3.01	2.81	2.70
6	4.71	4.77	4.82	4.94	5.16	5.37	5.11	4.25	3.44	2.99	2.75
7	4.75	4.75	4.79	4.83	4.92	5.07	5.18	4.87	4.08	3.34	2.91
8	4.72	4.72	4.72	4.75	4.78	4.83	4.92	4.92	4.57†	3.84	3.18
9	4.57	4.57	4.57	4.57	4.59	4.60	4.63	4.65	4.57	4.19†	3.54
10	4.22	4.22	4.22	4.22	4.22	4.23	4.24	4.24	4.22	4.09	3.72†

† Pure strategy NE. § Dominated strategy.

Table C.5: Duopoly, quadratic utility payoff, patient game 5

	The opponent										
Player n	0§	1§	2§	3§	4	5	6	7	8	9§	10§
0§	4.03	3.41	2.80	2.30	2.03	1.91	1.86	1.89	1.98	2.07	2.20
1§	4.70	4.17	3.57	2.98	2.49	2.22	2.09	2.04	2.06	2.14	2.20
2§	5.14	4.75	4.25	3.67	3.10	2.64	2.37	2.25	2.19	2.20	2.27
3§	5.37	5.11	4.74	4.25	3.71	3.17	2.74	2.49	2.38	2.32	2.33
4	5.42	5.28	5.02	4.66	4.20	3.69	3.20	2.81	2.58	2.48	2.43
5	5.35	5.26	5.11	4.85	4.50	4.07	3.61	3.18	2.83	2.64	2.55
6	5.17	5.10	5.00	4.85	4.59	4.26	3.87	3.47	3.10	2.82	2.67
7	4.82	4.77	4.70	4.60	4.45	4.22	3.92	3.59†	3.27	2.98	2.77
8	4.24	4.22	4.17	4.10	4.02	3.89	3.69	3.46	3.22	3.00	2.81
9§	3.38	3.36	3.34	3.31	3.26	3.20	3.11	2.99	2.86	2.74	2.65
10§	2.12	2.13	2.12	2.11	2.10	2.08	2.07	2.06	2.06	2.09	2.14

† Pure strategy NE. § Dominated strategy.

Table C.6: Duopoly, quadratic utility payoff, patient game 6

	The opponent										
Player n	0§	1§	2§	3§	4	5	6	7	8	9	10§
0§	4.03	3.41	2.80	2.30	2.03	1.91	1.86	1.89	1.98	2.07	2.20
1§	4.71	4.18	3.58	2.98	2.49	2.22	2.09	2.04	2.06	2.14	2.20
2§	5.17	4.78	4.27	3.69	3.12	2.65	2.38	2.26	2.20	2.21	2.27
3§	5.41	5.17	4.80	4.31	3.76	3.22	2.77	2.52	2.39	2.33	2.33
4	5.50	5.37	5.13	4.77	4.30	3.78	3.27	2.86	2.62	2.50	2.45
5	5.49	5.41	5.28	5.03	4.68	4.24	3.76	3.29	2.91	2.70	2.59
6	5.41	5.35	5.26	5.12	4.88	4.54	4.13	3.69	3.27	2.94	2.75
7	5.23	5.19	5.13	5.04	4.89	4.65	4.34	3.96	3.57	3.21	2.93
8	4.93	4.91	4.87	4.80	4.71	4.57	4.34	4.06	3.73†	3.41	3.12
9	4.49	4.46	4.44	4.40	4.34	4.25	4.13	3.93	3.69	3.44†	3.19
10§	3.83	3.83	3.81	3.79	3.75	3.70	3.63	3.54	3.40	3.23	3.07

† Pure strategy NE. § Dominated strategy.

Table C.7: Duopoly, quadratic utility payoff, patient game 7

	The opponent										
Player n	0§	1	2	3	4	5	6	7	8	9	10
0§	5.15	3.72	2.52	2.08	2.07	2.24	2.37	2.53	2.63	2.66	2.62
1	5.72	5.33	4.03	2.87	2.41	2.35	2.47	2.53	2.63	2.66	2.62
2	5.31	5.71	5.43	4.24	3.14	2.67	2.57	2.63	2.63	2.66	2.62
3	4.86	5.23	5.66	5.47	4.38	3.33	2.86	2.72	2.72	2.66	2.62
4	4.58	4.78	5.15	5.58	5.43	4.44	3.46	2.98	2.80	2.74	2.62
5	4.46	4.53	4.72	5.06	5.47	5.34†	4.43	3.51	3.03	2.82	2.70
6	4.36	4.44	4.50	4.67	4.97	5.33	5.19†	4.35	3.49	3.02	2.76
7	4.35	4.35	4.41	4.47	4.61	4.87	5.15	4.99†	4.21	3.41	2.94
8	4.32	4.32	4.32	4.38	4.42	4.54	4.74	4.93	4.73†	4.01	3.27
9	4.26	4.26	4.26	4.26	4.30	4.33	4.42	4.56	4.67	4.42†	3.74
10	4.12	4.12	4.12	4.12	4.12	4.15	4.17	4.23	4.32	4.34	4.05†

† Pure strategy NE. § Dominated strategy.

Table C.8: Duopoly, quadratic utility payoff, patient game 8

	The opponent										
Player n	0§	1§	2	3	4	5	6	7	8	9	10
0§	5.15	4.50	3.72	3.00	2.52	2.24	2.08	2.02	2.07	2.11	2.24
1§	5.61	5.24	4.63	3.88	3.17	2.70	2.42	2.25	2.18	2.22	2.24
2	5.72	5.64	5.30	4.71	3.99	3.31	2.85	2.57	2.40	2.32	2.35
3	5.63	5.74	5.66	5.32	4.76	4.07	3.41	2.97	2.70	2.53	2.45
4	5.51	5.65	5.74	5.65	5.32	4.77	4.11	3.48	3.06	2.80	2.63
5	5.46	5.57	5.68	5.73	5.61	5.28	4.75	4.12	3.53	3.13	2.89
6	5.46	5.54	5.62	5.69	5.70	5.55	5.20	4.69	4.10	3.54	3.17
7	5.51	5.55	5.60	5.65	5.68	5.64	5.45	5.09	4.60	4.04	3.53
8	5.56	5.58	5.60	5.62	5.64	5.62	5.53	5.30	4.94	4.46	3.95
9	5.56	5.57	5.57	5.57	5.57	5.55	5.49	5.35	5.10	4.74	4.29
10	5.46	5.46	5.45	5.45	5.43	5.41	5.36	5.27	5.09	4.83	4.48†

† Pure strategy NE. § Dominated strategy.

Table C.9: Quadropoly, quadratic utility payoff, patient game 1

	The 3 opponents										
Player n	0§	1§	2§	3	4	5	6	7	8	9	10§
0§	2.32	1.55	1.39	1.64	1.92	2.14	2.37	2.53	2.63	2.66	2.62
1§	4.06	2.74	1.99	1.80	1.99	2.21	2.37	2.53	2.63	2.66	2.62
2§	5.21	4.28	3.06	2.34	2.14	2.28	2.43	2.53	2.63	2.66	2.62
3	5.59	5.24	4.40	3.28	2.62	2.40	2.49	2.59	2.63	2.66	2.62
4	5.59	5.51	5.18	4.42	3.42	2.82	2.60	2.64	2.68	2.66	2.62
5	5.45	5.44	5.35	5.04	4.35	3.48	2.94	2.73	2.72	2.70	2.62
6	5.22	5.22	5.20	5.10	4.81	4.20	3.45	2.99	2.79	2.74	2.66
7	4.88	4.88	4.87	4.84	4.75	4.47	3.96	3.35	2.98	2.78	2.68
8	4.37	4.37	4.37	4.36	4.33	4.25	4.02	3.62	3.18	2.89	2.71
9	3.64	3.64	3.64	3.64	3.63	3.62	3.56	3.42	3.19	2.94†	2.75
10§	2.62	2.62	2.62	2.62	2.63	2.63	2.64	2.64	2.65	2.65	2.62

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.10: Quadropoly, quadratic utility payoff, patient game 2

	The 3 opponents										
Player n	0§	1§	2§	3	4	5	6	7	8	9	10
0§	2.32	1.55	1.39	1.64	1.92	2.14	2.37	2.53	2.63	2.66	2.62
1§	4.06	2.74	1.99	1.80	1.99	2.21	2.37	2.53	2.63	2.66	2.62
2§	5.23	4.30	3.08	2.35	2.14	2.28	2.43	2.53	2.63	2.66	2.62
3	5.62	5.28	4.45	3.32	2.64	2.41	2.50	2.59	2.63	2.66	2.62
4	5.62	5.56	5.26	4.51	3.49	2.85	2.61	2.65	2.68	2.66	2.62
5	5.50	5.50	5.44	5.17	4.49	3.57	2.99	2.74	2.73	2.71	2.62
6	5.32	5.32	5.32	5.26	5.01	4.40	3.59	3.06	2.81	2.75	2.67
7	5.08	5.07	5.08	5.07	5.01	4.77	4.23	3.53	3.06	2.81	2.70
8	4.75	4.75	4.75	4.75	4.74	4.67	4.46	3.99	3.40	2.99	2.74
9	4.31	4.31	4.31	4.31	4.31	4.30	4.24	4.05	3.68	3.21	2.87
10	3.72	3.72	3.73	3.73	3.72	3.73	3.72	3.68	3.55	3.28	2.95†

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.11: Quadropoly, quadratic utility payoff, patient game 3

	The 3 opponents										
Player n	0	1	2	3	4	5	6	7	8	9	10
0§	3.25	2.55	2.03	1.74	1.57	1.65	1.73	1.79	1.96	1.99	2.14
1§	4.21	3.43	2.75	2.24	1.95	1.78	1.85	1.91	1.96	2.11	2.13
2§	5.09	4.31	3.56	2.91	2.41	2.13	1.96	2.02	2.07	2.11	2.25
3§	5.55	5.11	4.37	3.65	3.03	2.56	2.29	2.12	2.17	2.21	2.24
4§	5.72	5.52	5.09	4.37	3.70	3.11	2.67	2.42	2.26	2.30	2.34
5	5.73	5.69	5.46	5.02	4.33	3.70	3.15	2.75	2.52	2.38	2.41
6	5.71	5.71	5.61	5.34	4.90	4.24	3.66	3.17	2.81	2.60	2.47
7	5.68	5.68	5.62	5.46	5.16	4.72	4.10	3.58	3.14	2.83	2.66
8	5.61	5.59	5.53	5.42	5.22	4.90	4.47	3.91	3.45	3.09	2.83
9	5.39	5.36	5.32	5.23	5.08	4.86	4.53	4.13	3.66	3.28	3.00
10	4.92	4.90	4.86	4.80	4.70	4.54	4.32	4.04	3.71	3.34	3.07†

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.12: Quadropoly, quadratic utility payoff, patient game 4

	The 3 opponents										
Player n	0§	1§	2	3	4	5	6	7	8	9	10
0§	3.25	2.03	1.57	1.73	1.96	2.14	2.37	2.53	2.63	2.66	2.62
1§	5.14	3.61	2.44	1.97	2.08	2.25	2.37	2.53	2.63	2.66	2.62
2	5.74	5.27	3.88	2.77	2.30	2.36	2.47	2.53	2.63	2.66	2.62
3	5.36	5.72	5.33	4.06	3.02	2.56	2.57	2.62	2.63	2.66	2.62
4	4.94	5.32	5.66	5.32	4.16	3.20	2.74	2.71	2.71	2.66	2.62
5	4.78	4.95	5.28	5.56	5.23	4.18	3.30	2.86	2.79	2.73	2.62
6	4.77	4.82	4.96	5.22	5.43	5.09	4.13	3.33	2.91	2.79	2.69
7	4.76	4.79	4.84	4.94	5.13	5.24	4.87	4.01	3.30	2.90	2.74
8	4.72	4.73	4.75	4.78	4.85	4.97	4.97	4.59	3.82	3.20	2.82
9	4.57	4.57	4.58	4.59	4.61	4.65	4.69	4.62	4.22	3.57	3.03
10	4.22	4.22	4.23	4.23	4.23	4.24	4.26	4.26	4.14	3.78	3.25†

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.13: Quadropoly, quadratic utility payoff, patient game 5

	The 3 opponents										
Player n	0§	1§	2§	3§	4§	5	6	7	8	9§	10§
0§	2.32	1.85	1.55	1.42	1.39	1.52	1.64	1.74	1.92	1.99	2.14
1§	3.13	2.53	2.07	1.77	1.64	1.60	1.72	1.82	1.91	2.07	2.13
2§	3.97	3.25	2.68	2.24	1.96	1.83	1.79	1.89	1.99	2.07	2.21
3§	4.55	4.00	3.32	2.78	2.38	2.11	2.00	1.95	2.05	2.13	2.20
4§	4.90	4.48	3.95	3.33	2.84	2.47	2.23	2.13	2.10	2.18	2.26
5	5.02	4.73	4.33	3.84	3.27	2.84	2.52	2.33	2.24	2.22	2.30
6	4.95	4.75	4.47	4.10	3.66	3.16	2.80	2.54	2.39	2.33	2.32
7	4.68	4.55	4.36	4.10	3.76	3.39	2.99	2.71	2.52	2.42	2.39
8	4.16	4.09	3.97	3.80	3.58	3.31	3.04	2.75	2.57†	2.46	2.42
9§	3.33	3.29	3.24	3.15	3.03	2.88	2.72	2.57	2.44	2.38	2.37
10§	2.11	2.10	2.08	2.06	2.03	2.00	1.97	1.97	1.99	2.05	2.14

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.14: Quadropoly, quadratic utility payoff, patient game 6

	The 3 opponents										
Player n	0§	1§	2§	3§	4§	5§	6	7	8	9	10§
0§	2.32	1.85	1.55	1.42	1.39	1.52	1.64	1.74	1.92	1.99	2.14
1§	3.13	2.53	2.07	1.77	1.64	1.60	1.72	1.82	1.91	2.07	2.13
2§	4.00	3.28	2.70	2.26	1.97	1.84	1.79	1.90	1.99	2.07	2.21
3§	4.61	4.06	3.37	2.83	2.41	2.13	2.01	1.96	2.05	2.14	2.20
4§	5.01	4.59	4.07	3.42	2.91	2.52	2.27	2.16	2.11	2.19	2.27
5§	5.19	4.92	4.52	4.02	3.42	2.96	2.61	2.38	2.28	2.24	2.31
6	5.22	5.04	4.77	4.39	3.92	3.38	2.97	2.66	2.46	2.38	2.35
7	5.12	5.00	4.81	4.54	4.19	3.77	3.29	2.94	2.68	2.52	2.46
8	4.86	4.79	4.67	4.49	4.24	3.92	3.55	3.15	2.87	2.67	2.56
9	4.44	4.39	4.32	4.21	4.05	3.83	3.56	3.27	2.97	2.76†	2.63
10§	3.80	3.78	3.74	3.69	3.59	3.46	3.30	3.11	2.92	2.73	2.62

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.
† Pure strategy NE. § Dominated strategy.

Table C.15: Quadropoly, quadratic utility payoff, patient game 7

	The 3 opponents										
Player n	0§	1§	2	3	4	5	6	7	8	9	10
0§	3.25	2.03	1.57	1.73	1.96	2.14	2.37	2.53	2.63	2.66	2.62
1§	5.14	3.62	2.44	1.97	2.08	2.25	2.37	2.53	2.63	2.66	2.62
2	5.74	5.29	3.90	2.78	2.30	2.36	2.47	2.53	2.63	2.66	2.62
3	5.30	5.71	5.36	4.09	3.04	2.56	2.57	2.63	2.63	2.66	2.62
4	4.80	5.23	5.65	5.37	4.21	3.23	2.75	2.72	2.71	2.66	2.62
5	4.54	4.74	5.14	5.54	5.31	4.26	3.34	2.88	2.79	2.74	2.62
6	4.43	4.50	4.69	5.05	5.41	5.19	4.23	3.39	2.94	2.81	2.69
7	4.36	4.41	4.47	4.63	4.95	5.24	5.01	4.15	3.37	2.93	2.75
8	4.33	4.33	4.37	4.43	4.56	4.81	5.02	4.78	3.99	3.29	2.85
9	4.26	4.26	4.27	4.30	4.34	4.45	4.63	4.75	4.49	3.78	3.14
10	4.12	4.12	4.12	4.13	4.15	4.18	4.26	4.38	4.42	4.14	3.51†

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.

† Pure strategy NE. § Dominated strategy.

Table C.16: Quadropoly, quadratic utility payoff, patient game 8

	The 3 opponents										
Player n	0§	1§	2§	3§	4	5	6	7	8	9	10
0§	3.25	2.55	2.03	1.74	1.57	1.65	1.73	1.79	1.96	1.99	2.14
1§	4.22	3.44	2.75	2.24	1.95	1.78	1.85	1.91	1.96	2.11	2.13
2§	5.10	4.33	3.58	2.92	2.42	2.14	1.96	2.02	2.07	2.11	2.25
3§	5.57	5.15	4.41	3.69	3.05	2.58	2.30	2.13	2.18	2.22	2.24
4	5.73	5.57	5.15	4.44	3.76	3.16	2.70	2.44	2.27	2.31	2.34
5	5.72	5.72	5.54	5.12	4.44	3.80	3.23	2.80	2.55	2.40	2.43
6	5.65	5.72	5.68	5.48	5.06	4.41	3.80	3.27	2.88	2.65	2.50
7	5.61	5.67	5.69	5.61	5.38	4.95	4.33	3.77	3.29	2.93	2.72
8	5.61	5.63	5.65	5.62	5.49	5.23	4.81	4.22	3.70	3.27	2.96
9	5.58	5.58	5.58	5.55	5.47	5.31	5.02	4.61	4.07	3.61	3.23
10	5.45	5.45	5.44	5.41	5.35	5.23	5.04	4.74	4.35	3.87	3.48†

We report an excerpt of player n's utility payoffs where the 3 opponents act identically.

† Pure strategy NE. § Dominated strategy.

Appendix D. Material available for downloading.

1. *GeGodager_prep_est_Fixedpoint.do*
   Stata commands for data preparation, estimation, and computing the fixed point.

2. *GeGodager_MonteCarlo.do*
   Stata commands for Monte Carlo simulation.

Appendix E. Data in Brief material

(This manuscript is submitted to and currently under review at *Data in Brief.*)

Article Title

Data from an incentivized laboratory experiment on strategic medical choices

Authors

Ge Ge[1], Geir Godager[1,2]

Affiliations

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.

Corresponding author

Ge Ge (gege@medisin.uio.no)

Abstract

This paper presents data of medical choices determining physicians' profit and patients' health benefit under three levels of market competition: monopoly, duopoly, and quadropoly. The data was collected from 136 German university students in an incentivised laboratory experiment. The designed experimental parameters and the formula for computing the payoff matrices of the games are described in this paper as well. This data was analyzed by Ge and Godager (2020) who employed generalized multinomial logit models to investigate the relationship between market competition and determinism in behavior under a quantal response equilibrium paradigm. This data contributes to future investigation on alternative game theoretic equilibrium concepts and the development of empirical methods for studying strategic choice behavior.

Keywords

Incentivized laboratory Experiment, Oligopoly, Competition, Behavioral Game Theory, Quantal Response Equilibrium, Physician behavior

Specifications Table

Subject	Economics and Econometrics
Specific subject area	Behavioral game theory
Type of data	Table
How data were acquired	Incentivised laboratory experiment programmed and implemented using zTree
Data format	Raw, Partially analyzed
Parameters for data collection	Strategic choices in Monopoly, Duopoly and Quadropoly market settings
Description of data collection	Data was collected in 5 experimental sessions at Cologne Laboratory for Experimental Research of the University of Cologne
Data source location	University of Cologne, Cologne, Germany
Data accessibility	With the article
Related research article	G. Ge, G. Godager, Predicting strategic medical choices: An application of a quantal response equilibrium choice model, Journal of Choice Modelling.

Value of the Data

- The data enables the analysis of strategic medical choices that determine profit and health benefit for real patients.
- This data can be beneficial for researchers interested in incentivized choice experiments and behavioral game theory.
- The data can be useful for exploring several alternative game theoretic equilibrium concepts, the development of new empirical methods and teaching.

1 Data Description

The data includes 136 subjects' medical choice decisions under three market conditions. The three market conditions, namely *monopoly*, *duopoly*, and *quadropoly* have various levels of competition where 1, 2, and 4 players make strategic choices simultaneously, respectively. Decisions are made for 8 "patients" in each market condition, hence there are 24 games in total in the experiment. In each game, the subjects choose among 11 medical treatments that determine their own profit and patients' health benefit. In duopoly and quadropoly, the subjects are randomly matched, and the joint decisions by the matched group determine their payoffs. The random match is dissolved after completion of all the decision tasks in each market setting.

The complete choice data of the 136 subjects is presented in the Excel file *GeGodager_2020_rawdata.xlsx*. Table 1 below summarizes the frequencies of each strategy being chosen in each game.

Table 1: Observed frequencies of strategy choice in the 24 patient games by the 136 subjects.

Market	**Patient game**	**Pure strategy**										
		0	1	2	3	4	5	6	7	8	9	10
Monopoly	1	24	12	4	18	14	27	16	15	2	2	2
	2	21	9	9	13	11	18	24	14	10	1	6
	3	23	6	10	9	11	14	10	22	14	6	11
	4	23	5	7	13	11	19	9	29	11	3	6
	5	24	7	7	19	18	15	14	21	5	4	2
	6	23	4	8	21	14	17	11	15	13	4	6
	7	21	6	7	14	7	14	11	13	24	7	12
	8	21	4	14	12	7	16	14	8	20	7	13
Duopoly	1	0	0	1	3	3	12	24	41	26	22	4
	2	0	0	0	2	4	7	12	27	36	26	22
	3	0	0	0	3	4	3	9	18	37	34	28
	4	1	0	3	1	0	4	6	21	30	29	41
	5	1	0	0	1	4	18	22	48	32	7	3
	6	1	0	0	1	4	9	14	33	42	19	13
	7	1	1	0	1	1	1	3	15	18	28	67
	8	0	1	0	0	1	2	7	14	25	28	58
Quadropoly	1	0	0	0	2	2	4	14	31	48	30	5
	2	0	0	0	2	0	3	10	11	42	31	37
	3	1	0	0	0	0	0	3	12	26	41	53
	4	1	0	1	0	0	0	4	9	24	39	58
	5	0	0	0	0	0	10	9	37	58	17	5
	6	0	0	0	1	1	1	8	15	40	43	27
	7	0	0	0	1	0	1	3	2	16	22	91
	8	0	0	0	0	1	1	4	5	14	27	84

2 Experimental Design, Materials and Methods

An incentivised laboratory experiment was conducted at University of Cologne. The experiment has a medical framing and the participants are instructed to play the role of a physician and choose medical treatment for eight "patients" in each of three different market settings; monopoly, duopoly and quadropoly. The treatment choices of participants determine their own profit and patients' health benefit. The profit and patient benefit accrued in the laboratory are converted into monetary transfers to the participants and a charity dedicated to providing surgeries for ophthalmic patients, respectively. This element of our protocol, which is identical to Hennig-Schmidt et al. [4], motivates participants' patient-regarding behavior in the laboratory. The sessions lasted about 90 minutes on average, and participants earned on average about 14 Euros and provided about 8 Euros patient benefits. In total, 1 102 Euros were transferred to the Christoffel Blindenmission. The experiment was programmed in zTree [1] and the subjects were recruited through ORSEE [3].

2.1 Construction of payoff matrices

Physicians receive profit by treating patients and patients gain health benefit. In each market, there is a fixed demand of 100 patients of the same type. In monopoly, the physician serves the whole market, while in duopoly and quadropoly, the demand is assumed to respond positively to the patient benefit from the treatment. In other words, in the competitive markets, the physician who provides larger health benefit is more likely to attract more patients, and his market share is therefore positively related to the health benefit he provides and negatively related to the health benefit provided by his opponent(s). On the other hand, providing larger health benefits increases costs and hence reduces physician's profit margin. In each game, the payoffs are constructed based on the specified demand function and the experimental parameters designed to characterize a "patient type". In total, there are eight patient types and 24 games. In the following, we describe in details the experimental parameters, the specifications of the market demand function and the computation of the payoff elements. Interested readers can reproduce the payoff matrices using Stata code *Data_in_Brief.do* presented in the supplementary materials.

Experimental parameters

We use three parameters, F, ϕ, and δ, to characterise different patient types. The capitation payment parameter, F, denotes the payment to the physician for each patient he treats, and takes the value of either 10 or 15. The cost parameter, ϕ, specifies the cost of the treatment, and is either 0.075 or 0.1. The patient benefit parameter, δ, denotes the benefit the patient receives, and is either 0.5 or 1. The $2 \times 2 \times 2$ combinations of parameter levels make up a total of eight unique configurations characterizing eight patient types for each market in the experiment.

Consider a choice set C of 11 treatment strategies ($C = \{0, 1, 2 \ldots 10\}$). For any strategy j from the choice set C ($j \in C$), we denote per-patient profit as π_j, and per-patient benefit as b_j, and they are given by:

$$\pi_j = F - \phi j^2, \tag{1a}$$

$$b_j = \delta j. \tag{1b}$$

Table 2 below describes variation in per-patient profit and per-patient benefit over the eight patient types.

Table 2: Per-patient profit and per-patient benefit over patient types, $j = 0, 1, 2 \ldots 10$

Patient Type	π_j	b_j
1	$10 - 0.1j^2$	j
2	$10 - 0.075j^2$	j
3	$15 - 0.1j^2$	$0.5j$
4	$15 - 0.1j^2$	j
5	$10 - 0.1j^2$	$0.5j$
6	$10 - 0.075j^2$	$0.5j$
7	$15 - 0.075j^2$	j
8	$15 - 0.075j^2$	$0.5j$

Market demand function

A physician's demand is determined by the treatment choices of all competing physicians through a logistic demand system:

$$D_j = 100 \qquad \text{Monopoly} \tag{2a}$$

$$D_{j|x} = \frac{100 \exp(b_j)}{\exp(b_j) + \exp(b_x)} \qquad \text{Duopoly} \tag{2b}$$

$$D_{j|xyz} = \frac{100 \exp(b_j)}{\exp(b_j) + \exp(b_x) + \exp(b_y) + \exp(b_z)} \qquad \text{Quadropoly} \tag{2c}$$

where $j, x, y, z \in C$. In monopoly, a physician serves the whole market, the demand of choosing j, denoted as D_j, is therefore fixed to 100 patients. In duopoly, a physician's demand of choosing j given the opponent's choice is x, denoted as $D_{j|x}$, is a function of his chosen per-patient benefit b_j and the per-patient benefit b_x chosen by the opponent. Similarly, in quadropoly, a physician's demand of choosing j given the combination of the opponents' choices is xyz, $D_{j|xyz}$, depends on the per-patient benefit chosen by all the four physicians in the game, b_j, b_x, b_y and b_z. The design of this demand response reflects a potential competitive scenario among health care providers in the market.

Payoff elements

Subjects receives a vector of payoff comprising two elements: total profit, Π, and total patient benefit, B. The total profit and health benefit of choosing strategy j, Π_j and B_j, depend on per-patient-profit, per-patient-benefit and the demand. We assume that subjects value both own profit and benefits to patients. Further, a healthy patient population in the market is assumed to be a a shared good and hence a physician's valuation of patient benefit is independent of the care provider [2]. In other words, B here is the total benefit of all the patients in the market. We let Π_j and B_j denote the payoff elements from alternative j in monopoly, $\Pi_{j|x}$ and $B_{j|x}$ denote the payoff elements from choosing j given the opponent's choice is x in duopoly, and similarly $\Pi_{j|xyz}$ and $B_{j|xyz}$ denote the payoff elements from choosing j given the combination of the opponents' choices is xyz in quadropoly. A physician's payoffs from choosing j given the opponent(s)' choice(s) can be expressed as:

$$\Pi_j = D_j \pi_j \qquad \text{Monopoly} \qquad (3a)$$

$$B_j = D_j b_j \qquad \text{Monopoly} \qquad (3b)$$

$$\Pi_{j|x} = D_{j|x} \pi_j \qquad \text{Duopoly} \qquad (3c)$$

$$B_{j|x} = D_{j|x} b_j + D_{x|j} b_x \qquad \text{Duopoly} \qquad (3d)$$

$$\Pi_{j|xyz} = D_{j|xyz} \pi_j \qquad \text{Quadropoly} \qquad (3e)$$

$$B_{j|xyz} = D_{j|xyz} b_j + D_{x|jyz} b_x + D_{y|jxz} b_y + D_{z|jxy} b_z \qquad \text{Quadropoly} \qquad (3f)$$

where $j, x, y, z \in C$.

We now illustrate with examples how to calculate physicians' payoff (Π and B) in each market.

- Example 1: Monopoly, Patient Type 3, $j = 7$.
  We see from Table 2, when the physician chooses $j = 7$, he receives $15 - 0.1 \times 7^2 = 10.1$

Taler of profit for each patient and the per-patient benefit is $0.5 \times 7 = 3.5$ Taler. Under monopoly, a physician serves the whole market with 100 patients. Therefore, the physician's payoff contains $10.1 \times 100 = 1010$ Taler profit and $3.5 \times 100 = 350$ Taler patient benefit.

- Example 2: Duopoly, Patient Type 3, $j = 7$, $x = 5$.

  Under duopoly, the physician's demand is determined by the per-patient benefits chosen by him and his opponent. Hence, the physician's demand is $\frac{100\exp(0.5\times7)}{\exp(0.5\times7)+\exp(0.5\times5)}$ and his opponent's demand is $\frac{100\exp(0.5\times5)}{\exp(0.5\times7)+\exp(0.5\times5)}$. The total patient benefit included in the physician's payoff is therefore: $3.5\times\frac{100\exp(3.5)}{\exp(3.5)+\exp(2.5)}+2.5\times\frac{100\exp(2.5)}{\exp(3.5)+\exp(2.5)} = 323$ Taler. The player's profit is $10.1 \times \frac{100\exp(3.5)}{\exp(3.5)+\exp(2.5)} = 737$ Taler.

- Example 3: Quadropoly, Patient Type 3, $j = 7$, $x = 5, y = 5, z = 5$.

  Under quadropoly, the physician's demand is determined by the per-patient benefits chosen by him and his three opponents. The physician's demand is therefore $\frac{100\exp(3.5)}{\exp(3.5)+\exp(2.5)+\exp(2.5)+\exp(2.5)}$. The total patient benefit in the physician's payoff vector is $3.5\times\frac{100\exp(3.5)}{\exp(3.5)+\exp(2.5)+\exp(2.5)+\exp(2.5)}+3\times2.5\times\frac{100\exp(2.5)}{\exp(3.5)+\exp(2.5)+\exp(2.5)+\exp(2.5)} = 298$ Taler. The total profit is $10.1 \times \frac{100\exp(3.5)}{\exp(3.5)+\exp(2.5)+\exp(2.5)+\exp(2.5)} = 485$ Taler.

2.2 Experimental procedure

Upon arrival, participants were randomly assigned to cubicles. The instructions (available in the PDF file *Instructions.pdf*) informed participants of the structure of the experiment and the payoffs. To facilitate non-cooperative decision-making, the employment of random matching and re-matching of participants was described clearly to them. The participants had available a "calculator" in the zTree program to see how a combination of treatment choices determine payoffs. In other words, the participants could inspect each cell of the payoff matrices. The sequence of the three markets to be played by the participants varied in order to mitigate the "order effects".[1] Participants were given adequate time to read the instructions and ask clarifying questions in private. For each market setting, participants answered control questions to make sure they understood how (joint) choices affect their profit and the patient benefit.

At the end of the experiment, one randomly drawn outcome from each market determined the participant's payment and the patient benefits to be transferred to the charity. To

[1] Among all 136 participants, 56 played in the order of monopoly, duopoly, then quadropoly; 28 played in the order of monopoly, quadropoly, then duopoly; 24 played in the order of duopoly, quadropoly, and monopoly; and 28 played in the order of quadropoly, duopoly, then monopoly.

ensure participants' trust in the transfer of patient benefits, we applied a procedure similar to Hennig-Schmidt et al. [4]. One participant was chosen at random to be a *monitor*. After the experiment, the monitor verified that a money order, equivalent to the total patient benefit provided by all participants, was issued by the Finance Department of the University of Cologne. The money order was payable to the *Christoffel Blindenmission*, which supports ophthalmologists performing cataract surgeries in a hospital in Masvingo, Zimbabwe. After sealing the money order in an envelope, the monitor and an experiment assistant walked together to the nearest mailbox and deposited the envelope. The monitor was paid an additional 5 Euros.

Ethics Statement

The ethical review and approval of experimental procedure was given by Norwegian Social Science Data Services (reference number 43709). Informed consent was obtained orally from all subjects.

Acknowledgments

Financial support from the Research Council of Norway (IRECOHEX, No. 231776 and NORCHER, No. 296114) is gratefully acknowledged. We thank Daniel Wiesen for joint collaboration on experimental design in the early stages of this project, and for managing the experimental sessions in Cologne. We thank Emanuel Castillo for z-Tree programming; Anne Classen, Tobias Danzeisen, Lena Kuhne and Theresa Schwan for research assistance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

References

[1] Urs Fischbacher. Z-tree: Zurich Toolboox for Readymade Economic Experiments. *Experimental Economics*, 10:171–178, 2007.

[2] Jacob K Goeree, Charles A Holt, and Susan K Laury. Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *Journal of Public Economics*, 83(2):255–276, 2002.

[3] Ben Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

[4] Heike Hennig-Schmidt, Reinhard Selten, and Daniel Wiesen. How Payment Systems Affect Physicians' Provision Behavior – An Experimental Investigation. *Journal of Health Economics*, 30:637–646, 2011.