Physician beneficence and profit-taking among private for-profit clinics in China: A field study using a mystery shopper audit

Roland Cheo
Centre for Behavioural and Implementation Science Interventions, Yong Loo Lin School of Medicine, National University of Singapore

Ge Ge
Department of Health Management and Health Economics, University of Oslo

Rugang Liu
School of Public Health and Policy, Nanjing Medical University

Jian Wang
Dong Fureng Institute of Economic and Social Development and Center for Health Economics and Management, Wuhan University

Qiqi Wang
School of Economics, Xian School of Economics and Finance

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK
Working paper 2023:6

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-280-3

Physician beneficence and profit-taking among private for-profit clinics in China: A field study using a mystery shopper audit

Roland Cheo[*] Ge Ge[‡] Rugang Liu[§] Jian Wang[¶] Qiqi Wang[‖]

Abstract: This study employs a mystery shopper audit on a random sample of 96 for-profit private clinics in Jinan, China. We investigate two instruments which reflect beneficence among for-profit clinicians in private practice. The first is whether physicians returned a lost wallet "accidentally" left next to the physician's table; and the second, whether physicians prescribed antibiotics to pseudo-patients who displayed no symptoms of any illness but had complained of fever the night before. These measures quantify beneficence under two different valence framing: returning a wallet represents clinicians who "do good" at personal cost to themselves, while not prescribing antibiotics represents a choice "not to do harm" to patients. We look at the correlation between these beneficence measures and the physicians' prescription behaviours, and their revenues from the consultation. We find that whether doctors return a lost wallet or not, and prescribe antibiotics or not, such physicians are still as likely to prescribe medications which increase their incomes.

Keywords: mystery shopper audit, antibiotic prescription, lost wallet, beneficence, for-profit clinics, physicians

JEL Classification Codes: C93, D64, D91, I18

[*] *Corresponding author*, Centre for Behavioural and Implementation Science Interventions, Yong Loo Lin School of Medicine, National University of Singapore, email: blangah@live.cn

[†]We acknowledge the support received from the Research Council of Norway (IRECOHEX, Project-No. 231776) and the financial contribution from the Research Council of Norway through the Norwegian Center for Health Service Research (NORCHER project number 296114)

[‡]Department of Health Management and Health Economics, University of Oslo

[§]School of Public Health and Policy, Nanjing Medical University

[¶]Dong Fureng Institute of Economic and Social Development and Center for Health Economics and Management, Wuhan University

[‖]School of Economics, Xian School of Economics and Finance

1 Introduction

In the market for credence goods such as medical services (Balafoutas and Kerschbamer, 2020; Dulleck and Kerschbamer, 2006; Dulleck et al., 2011; Kerschbamer and Sutter, 2017), a disparity in knowledge exists between patients and healthcare professionals. Patients often lack information about available treatments and their effectiveness in promoting recovery. It is expected that doctors who practice beneficence, whether driven by genuine concern for their patients or merely adhering to professional norms, will not exploit this information gap for personal gain but still recommend treatments which maximise their patients' welfare (Balafoutas and Kerschbamer, 2020). However both theoretical and experimental literature on physician beneficence suggests that doctors may sometimes prioritize their income over recommending the most suitable treatments for their patients. Our study, like Brock et al. (2016), aims to investigate the external validity of these findings in real-world healthcare settings.

This study aims to address a potential knowledge gap by exploring two tools which might be used to assess physician beneficence in a real-world context. We label our approach as an "experiment" because we directly manipulate the clinical environment, departing from the conventional use of the term to signify randomized treatments. Our unique instruments gauge two facets of beneficence: the avoidance of harm, represented by the judicious use of antibiotics, and the promotion of good, symbolized by the return of lost items, such as wallets found beneath their consultation tables. To achieve this, we conduct a field investigation employing a mystery shopper audit at 96 clinics adopting similar methodology used in Currie et al. (2011, 2014). Our primary findings from this field study encompass the total cost of medication prescribed by attending physicians, their utilization of diagnostic procedures and the total time required for diagnoses. These observations enable us to assess whether doctors tend to encourage greater medication purchases or reduce consultation times to increase their profit per minute of consultation. As far as we can tell, this paper is among the first to examine whether these two novel instruments can predict the overall profit-oriented behavior of physicians.

Our study is also related to the literature on credence goods (Balafoutas and Kerschbamer, 2020; Dulleck and Kerschbamer, 2006; Dulleck et al., 2011; Gottschalk et al., 2020; Kerschbamer and Sutter, 2017). In particular, our experimental design is very similar in spirit to Gottschalk et al. (2020) which also incorporates a mystery shopper audit in their setup. In the Gottschalk et al. (2020) study, a single patient visits 180 dental clinics using different scripts suggesting higher or lower socio-economic characteristics and whether the patient intends to seek a second opinion.They find that overcharging occurs in 50 out of 180 dental clinics. They also find that dentists with lower utilization rates are more likely to propose unnecessary treatments. Unlike studies like Lu (2014), Gottschalk et al. (2020) and our study can rule out third party payments through insurance or company subsidies as a reason for the prevalence of overcharging.

Prior theoretical research on induced demand such as McGuire and Pauly (1991) describe a

physician's practice in a utility maximizing framework where doctors care about their incomes and their leisure while they are negatively affected by inducement or trying to get their patients to overbuy medication or services they do not actually need. In such a model, McGuire and Pauly (1991) show that if there is an expectation of a fall in incomes, doctors are likely to engage in inducement in order to maintain their incomes. The empirical literature affirms such findings suggesting that physicians respond to income pressures on their practice by striving to increase their incomes (Iversen, 2004; Gruber and Owings, 1994; Plotzke and Courtemanche, 2011; Quast et al., 2008; Rizzo and Zeckhauser, 2003, 2007) at some cost to patients' welfare. The literature shows that this phenomena is present both in developing and developed countries.

Another strand of literature shows how institutional reforms favoring payment systems which reward doctors for the number of services they prescribe, result in costly treatments and maximum treatment recommendations over flat fee payment systems which lead to under provided services (Brosig-Koch et al., 2016; McGuire, 2000; Park et al., 2007; Quinn et al., 2020) though Li et al. (2014) finds contrary evidence for this when exploiting a policy change in Canada. Other studies show how the risk attitudes (Martinsson and Persson, 2019) and altruistic tendencies of physicians (Brosig-Koch et al., 2017; Godager and Wiesen, 2013; Martinsson and Persson, 2019) might mitigate this profit-taking motive or under-provision of medical care due to a lack of monetary incentives. Our paper adds to the growing literature measuring physician altruism (Attema et al., 2023; Byambadalai et al., 2023; Ge et al., 2022; Ge and Godager, 2021; Li et al., 2022; Li, 2018; Li et al., 2017) and determining whether such measures promote patient welfare at the cost of the physician's profit[1]. Previous research on hospital care have shown that prosocial doctors are less likely to be concerned with profit and more for the welfare of the patient. Hence, they are willing to sacrifice personal welfare for the sake of others (Hellerstein, 1998; Allaby, 2003; Hennig-Schmidt et al., 2011). We also note the distinction of the literature examining physician beneficence and physician altruism (Harris, 2018; Galizzi et al., 2023; Glannon and Ross, 2002). While altruism is often reflected in the weight of doctor's utility function which includes the patient's well-being, beneficence might reflect professional duty towards the patient without regards to other regarding preferences. Hence, beneficence may work through several channels such as self-image concerns, or the adhering to professional norms by the doctor, or altruism.

Currently, only a handful of studies try to quantify either physician altruism or physician beneficence (Brosig-Koch et al., 2016, 2017; Hennig-Schmidt et al., 2011; Godager and Wiesen, 2013; Martinsson and Persson, 2019). Two ways are currently being employed. The first involves laboratory experiments where subjects take on the role of physicians in a controlled scenario with hypothetical patients who are identical in all areas excepting the efficacy of how the treatments affect their health. In such a setting, it is clear what the correct diagnosis and optimal treatment is and therefore variation in altruism of subjects can be measured by how far

[1]The proposed instrument is targeting a more general notion of altruism (if at all) as it does not focus on a physician altruism in the sense of Arrow (1963) as the recipient of the returned wallet is not necessarily the simulated patient and hence it is more appropriate for us to consider it as beneficence.

they deviate from this standard. A second way researchers have measured altruism is by using the giving amounts elicited from the standard dictator game. In this line of literature, subjects are randomly paired with a partner and decide how much to allocate to their partner from a fixed pie. Due to the lack of reputation building, social image pressures and other possible strategic concerns, the amount given to a random, anonymous partner in the dictator game can be interpreted as a measure for altruism. Two main criticisms might be levied at such papers - (1) the use of students in a laboratory setting and hence the lack of external validity in the field, especially since the scenarios presented in the lab are described as hypothetical; and (2) the appropriateness of the measurement of altruism itself. To elucidate the problems of the measure of altruism elicited from these two forms of experiments, we highlight that the choice to give in both types of experiments involves house money. To this end, any giving amount is pareto improving from the lens of a Kaldor Hicks criterion of social welfare perspective since it improves both players' initial conditions. In the Hennig-Schmidt design, both the subject and the charity which receives the money; and in the dictator game, both the sender and receiver. This is not to say that they do not reflect elements of altruistic preferences however these measures are imprecise and may not necessarily correlate with how physicians treat patients in the field. There have also been criticism levied on the dictator game in the literature pointing to evidence that giving behavior is inconsistent with the altruism interpretation as the decision to give can be manipulated through the framing of decisions, i.e. giving or taking (Bardsley, 2008; List, 2007) or that alternative interpretations exist, such as behavior based on the possibility of opting out of the game by burning money (Dana et al., 2007).

The complication arises when we attempt to measure beneficence involving personal cost to the clinician, without the involvement of external resources. Additionally, drawing insights from the literature on valence framing, it is evident that there are substantial behavioral disparities in strategic interactions when employing "give" or "take" framing (Goerg et al., 2019; Grossman and Eckel, 2015; Khadjavi and Lange, 2015; Korenok et al., 2014; Tappin and Capraro, 2018). In our specific context, physician beneficence encompasses not only the avoidance of harm but also actively promoting the good of patients, extending the scope beyond a simple allocation of resources as reflected in the dictator game. It is clear that these two different dimensions of beneficence may lead to very different types of recommendations for treatment[2].

Our results show that our use of these two beneficence measures do not predict profit-oriented behaviors. Both types of doctors - those who do not harm by not prescribing antibiotics to their patients, and those who return a lost wallet, are still as likely to engage in profit-taking behaviors as other doctors who prescribe other types of medications.

[2]These two dimensions may invoke different concerns for the doctor such as a concern for personal reputation and social image should his errant diagnosis come to public scrutiny, while doing good to a patient might reflect a desire for warm glow. This may explain why papers like Brosig-Koch et al. (2016) and Godager and Wiesen (2013) find that physician beneficence varies depending on who the doctor treats. In particular, doctors may experience greater levels of beneficence for patients whose medical conditions are more severe than others. In this way, different motivations accompany the prescribing decision when such doctors treat different patients.

This paper makes several contributions to the literature using behavioral economics to improve health service delivery. Firstly, we show that beneficence is a multidimensional concept where doctors may choose to do good in one context, such as returning a lost wallet, but fail to prevent harm by prescribing antibiotics to obviously healthy patients. Secondly, we show that regardless which beneficence measure we use, either doctors who return a wallet or doctors who do not prescribe antibiotics but prescribe other medications, all earn statistically similar revenues and profits as doctors who did not return a wallet or those who prescribed antibiotics.

The remainder of the paper is organized as follows. Section 2 presents the literature regarding these two instruments. Section 3 elucidates the context in which the field experiment was conducted. Section 4 explains the experimental design and Section 5 presents the results.

2 Lost wallets and inappropriate prescriptions of antibiotics as candidate measures for physician beneficence

This paper explores the connection between the return of a lost wallet as well as the over prescription of antibiotic drugs by physicians on profiteering through the over-prescription of drugs to obviously healthy patients in registered single physician for-profit clinics. In particular, we examine in our paper whether there is inducement through a deliberate prescription of more expensive drugs with minimal consultation time.

The use of these two instruments provide two dimensions of how we might understand beneficence under two possible types of valence framing. The first, returning a lost wallet, might be considered as an active decision to "do good" to a patient, while the second, the decision not to prescribe antibiotics, can be conceived as an active decision "not to do harm". The literature for valence framing shows that the choice to do good and not to do harm affects behaviors in both the laboratory - eliciting more cooperative behaviors in public goods games (Böhm and Theelen, 2016), and invoking higher levels of beneficient giving in dictator games (Cappelen et al., 2013), as well as in the field, improving stakeholders' willingness to cooperate with firms (Crilly et al., 2016). However it is unclear whether this applies to the medical field, especially in the context of patient and physician interactions, though Holmes Jr et al. (2020) discusses how in light of the Covid 19 pandemic, the choice of doing good and doing no harm has huge implications for who receives critical treatment at the appropriate time given the current strain on hospital resources.

2.1 Lost wallet as a measure of beneficence

The return of a lost wallet requires deliberate action not to profit over someone else's misfortune while at the same time implies expending costly action to return the wallet to the rightful owner. We postulate that physicians who first call the number contained in the wallet and thereafter willingly courier the wallet to the student using a pickup service are exhibiting a type of

beneficence consistent with prosocial preferences that benefit the patient at added personal cost to the physician which is independent of the patient-physician relationship. In some sense, this notion of beneficence might be associated with warm glow preferences or pure altruism since there is no way for the rightful owner to suspect that the wallet was dropped in the physician's clinic, therefore the physician is free from any negative sentiments or social image concerns.

Ours is not the first study to consider a real effort task in the form of the return of a lost wallet (Beres, 2013; Cohn et al., 2019; Knack and Keefer, 1997) as a predictor of prosocial preferences. Data from experiments conducted by Readers Digest (Beres, 2013; Knack and Keefer, 1997) found that the number of intact wallets returned[3] correlated with the *trust* measure of the World Values Survey at 0.67. Previous research has shown that returning a lost wallet correlates with certain prosocial measures of the World Values Survey (Beres, 2013; Knack and Keefer, 1997) such as trust and civic honesty. Recently, Cohn et al. (2019) carried out a separate lost wallet field experiment in 40 countries and placed China at the bottom of the list. Altogether 17,303 wallets were dropped at various locations in 355 cities. Their study found that China ranked lowest in terms of the return of these lost wallets at around 12%. Various reasons have been used to explain the low return rate in China. The most controversial reason is that the Chinese have lower prosocial preferences than the peoples of other countries, however other possible reasons exist. For one, that study considered a particular group of people: receptionists, who may not be representative of the general population. Another possible reason is the manner of which the study recorded a successful return. An email address was provided in the wallet and those who sent an email to that address provided the main study's data for 'civic honesty'. Critics of this design have pointed out that most Chinese do not have an email address and instead use social networking apps like WeChat and QQ to communicate (Huynh et al., 2020). Therefore the low return rate may be a result of technological rather than prosocial preferences. Our paper incorporates these criticisms to the Cohn et al. (2019) experimental design. Our subjects are primarily considered a more prosocial group - physicians, and in order to increase the return rate, includes a mobile number instead of an email. Our results support the validity of the results in the Cohn et al. (2019) study in that even among physicians[4], the return rate of the wallets stood at around 21.3%.

In the economics literature, the return of a lost wallet comes from an economic device referred to as the 'lost wallet' protocol in a series of papers using laboratory (Dufwenberg and Gneezy, 2000; Cox et al., 2010) or internet (Charness et al., 2007) experiments. In Dufwenberg and Gneezy (2000), X finds a wallet belonging to Y. The value of the wallet to X is much lower than the value to Y. X has to decide whether to return the wallet, and gets a possible reward from Y (which will be less than the value of keeping the wallet) or keep the wallet. Based on the

[3]The study dropped twenty wallets in twenty cities in Western European countries, as well as ten wallets in twelve US cities.

[4]One of the criticisms levied on the Cohn et al. (2019) study is that the wallets were entrusted to receptionists who may not be representative of the general population. Our study shows that even if the target group were doctors the return rate in China is still low.

reward structure, since the reward is more likely to be lower than x, X should keep the wallet. Hence the equilibrium prediction is that X will always keep the wallet. The open puzzle of this literature is that both the reward size of the wallet owner, nor the size of the outside option of the wallet picker influences whether the wallet is returned. This suggests that this action is more related to some constant prosocial preferences which do not vary with the economic environment surrounding the wallet. Instead what has been shown in the literature has been that certain beliefs have been known to be correlated with the return of the wallet, such as honesty and other regarding preferences.

The use of the lost wallet protocol is also closely related to other variants, with a difference in one aspect: there is no chance for a reward from Y if X chooses to return. These studies use different devices such as money errantly mailed to people via a misdirected letter (Andreoni et al., 2021; Franzen and Pointner, 2013; Stoop, 2014) or short solicitation messages to mobile phones after errant deposits of money are made into that mobile account(Alem et al., 2018).

2.2 Over-prescription of antibiotics as a measure of beneficence

While it seems trivial to say that physicians should not prescribe antibiotics to healthy patients, the literature on antibiotic abuse in China, like other developing countries, show that Chinese doctors still prescribe antibiotics when they should not (Reynolds and McKee, 2009; Currie et al., 2011, 2014). About 70% of out-patients and 80% of in-patients are treated with antibiotics (Li, 2014). Other papers also draw the same conclusion for doctors in other countries, even those in the developed world (for French doctors, see Delattre and Dormont 2003; for Dutch doctors, see Dijk et al. 2013; for Taiwanese doctors, see Xirasagar and Lin 2006). Given the prevalence of this type of misuse in both the developed and developing world, and despite doctors knowing the consequences of such actions[5], the choice to prescribe antibiotics to healthy pseudo-patients requires deliberate action to harm patients for profit. We therefore postulate that physicians who choose not to prescribe antibiotics to healthy patients are making a deliberate decision not to harm their patients. This is also consistent with beneficence under a take (i.e. taking away health capital) framing. This second instrument for measuring beneficence cannot be separated from a physicians' level of risk taking preferences, social image or reputational concerns.

While some critics might argue that doctors in these countries are only responding to patient demand and are independent of the physicians' preferences, however the evidence provided by our study and others is that antibiotic abuse by Chinese doctors is largely driven by supply side concerns rather than patient demand (Currie et al., 2011, 2014).

[5]Such misuse has resulted in resistant strands of deadly bacteria and this public health crisis is an endemic problem all over the developing world (Reardon, 2014; Chukwuani et al., 2002; Arya, 2004; Gani et al., 1991).

3 The Chinese context

To comprehend the context of our research, it is important to be familiar with the pathway to obtaining a General Practioner (GP) licence in China, especially for the majority of the primary care physicians in our sample. Prior to 2012, to qualify as a GP, individuals needed to have completed a three year medical programme from a junior college, which encompassed secondary technical schools and junior medical colleges. In 2012, an initiative aimed at enhancing the quality of care delivered by primary practitioners was introduced, known as the "3+2" program. This program required individuals to not only complete the 3-year junior college medical education, but also undergo an additional 2-year general practitioner training. This may include apprenticeship in a clinic under the guidance of an experienced GP. A distinctive feature of junior college medical education is its has lower entry requirements and shorter training duration compared to other medical education programs, such as the five-year medical university programme. The 3+2 program is the primary route where Chinese primary care physicians are trained and plays a significant role in China's primary medical services(Lian et al., 2019; Liu et al., 2022). Conversely, individuals pursuing a five-year medical university program for their first degree are unlikely to become primary care physicians in single doctor clinics, preferring to specialize in particular fields of medicine. Figure 1 and Figure 2 depict the typical setup of a single physician practice, comprising of one treatment room and a dispensary.

Figure 1: Typical private clinic entrance

Figure 2: Typical clinic's treatment room

It is therefore important to distinguish that those who choose to practise privately as a GP tend to take the 3+2 route rather than an actual five-year comprehensive university medical degree because it provides the shortest route to be certified. It is also important to note that doctors in private practice will not be licensed to practice in public hospitals. It is from this kind of primary care facilities that we conducted a field experiment on physicians in Jinan, China which is the capital city in Shandong province. As the capital of a province and also a first-tier city, Jinan can be representative of most urban centers in China. The majority of the physicians in our sample are owners or co-owners of the clinics. The profit from drug sales is often their main source of income, since no consultation fees are charged.

4 Experimental design and data

Our field experiment's aims were as follows: (1) using a mystery shopper audit, we examine whether the doctors who recommended antibiotic drugs compared to those who did not mattered in the treatment decisions of otherwise healthy simulated patients in 96 single doctor primary healthcare clinics in Jinan, China. (2) In the process 96 wallets are "lost" on (or under) the doctor's desk and we examine whether the doctors who returned the wallet differed in their earlier prescription behavior.

4.1 Sample Selection

Based on all registered one doctor clinics in official Chinese registers, over 136 primary care clinics in Jinan, characterized by having a single responsible doctor, or having an hour schedule, were sampled to be included in the study. Our sample focused only on small walk-in private clinics where the general admission practice does not include a formal procedure for registration of patient IDs and where no formal patient records are kept. For such clinicians, they do not have access to any form of digital health records as such a system is not practised in these single physicians private clinics which is very much unlike the practice in public hospitals. This allowed us to ensure the anonymity of our pseudo-patients and we could randomly assign them to clinic visits[6] without the possibility that clinicians would be aware that the same patient had visited other clinics.

From official Chinese registers in the Health and Family Planning Commission of Jinan Municipality, out of the 136 initial sample, we chose 118 primary care clinics in Jinan based on four substantive criteria:(1) the clinic is for-profit with only one practicing physician, (2) it is located within the five districts of Jinan city,(3) the clinic has a valid license on the date of the experiment, and (4) the clinic provides general medicine as opposed to traditional chinese medicine (TCM). From the list of suitable clinics, we then randomly randomly chose 96 clinics, and kept the remaining 22 clinics as backups in case the clinic was shut on the day of the field study. It was expected that some clinics were permanently closed, so in the randomization process, backup clinics were assigned to either intervention or control group, so we could replace closed clinics in our initial visit.

4.2 Mystery Shopper Audit experiment

Following Moriarty et al. (2003); Bisgaier and Rhodes (2011), we carried out a mystery shopper audit with 96 clinics in November and December 2015.

The mystery shopper audit is considered a scientifically sound experimental method that straddles ethical concerns over healthcare practitioners and the patients they serve. The use of deception is controversial in science, and there is no unanimous classification across disciplines.

[6]It might be more challenging to conduct a similar field experiment in a system where durable physician-patient relations, often formalized as patient list systems, are the norm.

The main ethical dilemma in our study is that the healthy pseudo-patients provide incorrect information to the physician when describing their state of health. However, following the ethical analysis of Rhodes and Miller (2012), it can be ethically justified as long as confidentiality of research subjects is ensured, that risks to the research subjects are minimal and that the research is potentially valuable to human knowledge. This project was also subject to ethical assessment and received IRB approval before we carried on with the research[7].

We established protocols to ensure the anonymity of the physicians in our study. We generated a unique series of ID numbers for each clinic. After the visit, pseudo-patients destroyed the sheet of paper linking ID numbers with the actual clinic addresses - removing the possibility of tracking particular clinics or physicians. In order to track wallets, each wallet had a ID number to match with the clinic ID number. In addition, the field experiment also contributed positively to the revenues of the clinics in the study, because the physicians profited from the sale of the prescribed medications.

To ensure the safety of the pseudo-patients, a team of two students always traveled together. Furthermore, the pseudo-patients, 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 offered by the physician excepting temperature measuring or the visual inspection of the throat.

We recruited 13 students from School of Public Health, Shandong University to take the role as simulated patients - 4 males and 9 females. One female patient was given the role of backup patient, so that in total 12 patients were assigned to visit clinics. We applied a script detailing cold symptoms similar to that of Currie et al. (2011, 2014). Each patient underwent 10 hours of training on how to use the scripted dialogue presenting cold symptoms to the doctor. We used an experienced actor who instructed the pseudo-patients on how to speak in a clear and unthreatening way, and guided them on not including any information other than what was in the script. The actor also showed the pseudo-patients how to place the wallet under their backpack either on the table or the floor and to leave the clinic only with the backpack[8]. The experimenters then assigned each team to visit a clinic that was not in the designated list of 136 clinics in the study in order to practice. In this practice session, the companion student recorded the practice consultation between the pseudo-patient and the physician and afterwards the team filled out the baseline survey assessing the doctor after they had left the clinic. Finally, a separate training was arranged where the experimenters and the actor went through all recorded dialogues (12 consultations in total) in order to ensure that the script was adhered to and the tone of the conversation was convincing and then, the experimenters went through the survey to ensure that patient assessed subjective evaluations of the doctor were consistent. This involved all teams listening to each conversation and grading each attribute

[7]See the Acknowledgement section

[8]In this way, the physician would not see a student purposely leaving her wallet should there be a CCTV camera in the clinic.

of the doctor simultaneously and finally mutually agreeing on what grade to be given. This was in preparation for the actual field study a few days later in order to limit the amount of variation due to subjective interpretation by the simulated patients.

During the actual field study, to ensure the safety of simulated patients, patients were always accompanied by a fellow student, so a team of two students always travelled together. Furthermore, the patients, being students of the School of Public Health, had at least one semester of basic medical training. Our simulated patients met with the attending physicians and noted treatment advice. These were later reported in a survey after the patient left the practice.

The teams visited clinics and followed a script describing cold symptoms. We minimized selection biases on the part of of the pseudo-patients by giving them a predetermined list of 8 randomly selected clinics from the sample of 118 clinics and instructed them to follow the list strictly. They hailed official taxis to their destinations and claimed reimbursement from the experimenters. Upon arrival and departure to their destinations, teams sent messages confirming their locations. Students took note of whether doctors followed appropriate diagnostic procedures, and noted their service attitudes. They waited for medical advice to be given and recorded all given advice. They also recorded observable information describing the doctor, the consultation, the waiting time etc in a structured data collection sheet. Students were instructed to refuse any treatment and/or diagnostic test except for the measurement of temperature and a visual inspection of the throat. Since doctors also operated their own pharmacies, if students were recommended drugs to purchase, they were instructed to take note of the medication and then ask about the purchase price and buy them if possible. Students had to take note of both prices and drug information such as medication name as well as the pharmaceutical company which manufactured them in order to later be able to determine the retail and wholesale prices of those drugs[9].

4.3 The Lost Wallet experiment

During each visit, the accompanying student with the simulated patient would accidentally forget her wallet (left on the physician's table or dropped under his table) at the clinic. Cohn et al. (2019) found that wallets with the following items were more likely to be returned: wallets with more money, personal items with value only to the owner of the wallet like keys and contact information present in a business card. Based on these findings, we included in each wallet the following contents: A train ticket to be used a month later valued at 130 yuan, a key, a female student ID with a phone number and student dormitory address, and 20 Yuan RMB[10]. Furthermore, inside each wallet, we had carefully marked patient and sequence numbers

[9]Only 8% of our observations had gaps of information such as a lack of the recommended price, or missing information such as the manufacturer's name.

[10]While the cash value of the wallet was relatively small, however the train ticket of 130 yuan was dated to be used during the Chinese New Year holiday and would represent a great inconvenience for any student hoping to go home for the holidays. It was also important to show any physician that the wallet belonged to a student

(however indecipherable to anyone except the research team and later destroyed so that clinics remained anonymous even to the researchers). For physicians who called the number provided in the wallets, we had a female student answer phone calls from physicians with a set script. She was instructed to say thank you, and state that she was out of town on a school field trip, and state a request to forward the wallet by courier service using the 20 yuan RMB in the wallet[11]. Hence, upon the return of a missing wallet, information about the return could be linked to returned data while maintaining anonymity of the physicians and his practice. At the end of the study, none of the data could be individually linked to any specific doctor or the clinic's address.

4.4 Data

The main dependent variables that we are interested in are the continuous variables - *total price, and total price per minute* and the dummy variable, *total price >20 yuan.*

In our field study, the doctors in our sample only profit from the sale of medications and do not charge a consultation fee. Hence profit motivated doctors will attempt to sell more costly drugs to these healthy patients or try to minimize the time spent with each patient but at the same time sell costlier drugs to increase their patient turnover. A limitation with our experimental design is that we didn't ask students to purchase the recommended drugs but instead take note of the treatment cost, hence some of our simulated patients failed to take note of the total price of drugs being quoted or if doctors quoted prices for each individual drugs, the simulated patient failed to remember all of them. This resulted in a loss of 8% of our sample data. To account for this, we also include another outcome measure *total price >20 yuan* to account for 6 out of 9 left out physicians in our data. These three variables are our chief measures of profit maximizing behavior of physicians.

Our paper primarily tests for the relationship between doctors who returned the lost wallet and those who did not prescribe antibiotics in the cost of treatment levied on simulated patients. Another measure of profit maximizing behavior is if doctors tried to maximize the total price per minute of consultation he earned from simulated patients. Such doctors may try to maximize revenues by seeing as many patients as possible and hence may sell drugs at a lower price but reduce consultation times so that they can see more patients at any fixed amount of time.

5 Results

5.1 Summary statistics

We present the summary statistics of the treatment variables of doctors in Table 1 (overall) and Table 2 (segmented by our beneficence measures). Our data comes from the Mystery

and not a working adult and hence increase the opportunity cost of losing the wallet.

[11]Courier service within the city generally costs around 10 yuan RMB.

Shopper Audit Survey and whether the physician returned the lost wallet. We also consider variables denoting whether certain diagnostic procedures were employed as well as the subjective evaluations made by the pseudo-patients.

From Table 1, we can see that 63.5% of all doctors in our sample prescribed antibiotics while only 19 out of the 96 doctors returned the wallet. Table 2 looks at whether there are differences in profit-oriented behaviours of doctors who returned the wallet compared to those who did not; and doctors who prescribed antibiotics compared to those who did not. We can see from our sample that there is little evidence for gender bias in the choice to return the wallet (53% male) compared to doctors who did not (48% male) however there is some concern that gender bias may be present in the choice to prescribe antibiotics - of those who prescribed antibiotics, 56% were male while for doctors who did not, 37% were male. We can see from the Mann Whitney tests that there is no difference in the prescribing behaviors of doctors who returned the wallet versus those who did not, however we do find a significant difference for doctors who prescribed antibiotics compared to those who did not (MW p-value=0.00) for both total price and total price per minute of consultation and for doctors who recommended drugs at a price above 20 yuan (MW p-value=0.00). Since doctors charged patients mainly through the drugs they sold, the insignificance of our three prescribing behavior outcome variables when using the lost wallet as a measure of beneficence, informs us that the choice to do good to patients does not over-ride physicians' profit-taking behaviors. However for doctors who had a preference for not doing harm to patients - not prescribing antibiotics, there is a clear correlation with lower profit-taking behaviors.

Under diagnostic procedures in Table 2, we can see that there are no significant differences in physicians taking temperature, the checkup of the throat, using a stethoscope, asking whether there are medications at home and providing other medical information between doctors who prescribe antibiotics and those who do not, suggesting that these doctors are comparable. Yet those who prescribe antibiotics do ask about *cough* and *allergy* and give information about *how to use* versus those who did not prescribe. These three procedures are consistent with the recommendation of antibiotics and may suggest confirmatory bias on the part of the physicians

Table 1: Description of Key Variables

Variable	Description	Mean	Std Dev	Min	Max	Obs
Beneficence Instruments						
Wallet	= 1 if wallet was returned	0.198	0.401	0	1	96
Antibiotics	= 1 if antibiotics was prescribed	0.635	0.484	0	1	96
Doctor's characteristics						
Male	= 1 if the doctor is male	0.49	0.503	0	1	96
Age	Age(1=$\leq$ 30,2=[31,40], 3=[41,50],4=$\geq$ 51)	2.521	0.767	1	4	96
Prescription behavior						
Total price	Total price of the prescribed drugs	17.262	9.812	0	46	87
Total price per minute	Ratio of total price and consultation time, doctor's per minute revenue	1.36	0.876	0	4.117	87
Total price >20 yuan	=1 if doctor charged more than 20 yuan for drugs	0.398	0.492	0	1	93
Diagnostic procedure						
Cough	= 1 if the doctor asked (wet)coughing	0.438	0.499	0	1	96
Temperature	= 1 if the doctor checked temperature	0.667	0.474	0	1	96
Stethoscope	= 1 if the doctor used a stethoscope	0.083	0.277	0	1	96
Throat	= 1 if the doctor checked throat	0.823	0.384	0	1	96
Allergy	= 1 if the doctor asked allergic reaction	0.375	0.487	0	1	96
Medication at home	= 1 if the doctor asked if the patient has medication at home	0.240	0.429	0	1	96
How to use	= 1 if the doctor informed how to use prescribed medication	0.625	0.487	0	1	96
Other information	= 1 if the doctor informed other information, e.g. possible side effects	0.625	0.487	0	1	96
Polite	= 1 if the doctor was polite	0.896	0.307	0	1	96
Assessed subjective evaluations (results were given on a scale of 1 to 10)						
Concern	To what extent was the doctor concerned of you?	5.063	1.514	2	9	96
Professional skills	How do you evaluate the doctor's professional skills?	4.927	1.517	2	8	96
Respect	To what extent did the doctor respect your opinions?	5.260	1.438	2	8	96
Treatment information	To what extent did the doctor explain about the illness and treatment?	4.823	1.496	2	8	96
Overall	How do you evaluate overall treatment experience?	4.948	1.496	2	8	96
Recommend	To what extent will you recommend this doctor to your parents?	4.198	1.798	1	9	96

Table 2: Summary Statistics of Doctors' Characteristics, Prescription Behavior, Diagnostic Procedures and Assessed Subjective Evaluations (By Beneficence Measure)

	Beneficence Measure											
	Returned (Did not return) Lost Wallet						Prescribed (Did not prescribe)Antibiotics					
Variable	Mean	Min.	Max.	Sd.	*N*	MW p-value	Mean	Min.	Max.	Sd.	*N*	MW p-value
Doctor's characteristics												
Male	0.53(0.48)	0	1	0.51(0.50)	19(77)	0.72	0.56(0.37)	0	1	0.50(0.49)	61(35)	0.08
Age(1=$\leq$ 30,2=[31,40], 3=[41,50],4=$\geq$ 51)	2.26(2.58)	1	4	0.65(0.78)	19(77)	0.09	2.57(2.43)	1	4	0.76(0.78)	61(35)	0.34
Prescription behavior												
Total price	16.82(17.39)	0	25(46)	8.82(10.13)	19(68)	0.76	20.69(11.65)	3(0)	46(42.5)	10.18(7.87)	54(33)	0.00
Total price per minute	1.32(1.37)	0	2.28(4.12)	0.77(0.91)	19(68)	0.79	1.64(0.91)	0.20(0)	4.12(3.76)	0.77(0.87)	54(33)	0.00
Total price >20 yuan	0.42(0.42)	0	1	0.51(0.50)	19(77)	0.97	0.52(0.20)	0(0)	1(1)	0.50(0.41)	58(35)	0.00
Diagnostic procedures												
Cough	0.37(0.44)	0	1	0.50(0.50)	19(77)	0.57	0.54(0.26)	0	1	0.50(0.44)	61(35)	0.00
Temperature	0.47(0.71)	0	1	0.51(0.45)	19(77)	0.05	0.72(0.57)	0	1	0.45(0.50)	61(35)	0.14
Use stethoscope	0(0.10)	0	1	0(0.31)	19(77)	0.14	0.07(0.11)	0	1	0.25(0.32)	61(35)	0.40
Throat	0.84(0.82)	0	1	0.38(0.39)	19(77)	0.81	0.85(0.77)	0	1	0.36(0.426)	61(35)	0.32
Allergy	0.26(0.40)	0	1	0.45(0.49)	19(77)	0.26	0.56(0.06)	0	1	0.50(0.24)	61(35)	0.00
Medication at home	0.37(0.21)	0	1	0.50(0.41)	19(77)	0.14	0.25(0.23)	0	1	0.43(0.43)	61(35)	0.85
How to use	0.63(0.62)	0	1	0.50(0.49)	19(77)	0.95	0.74(0.43)	0	1	0.44(0.50)	61(35)	0.00
Other information	0.42(0.66)	0	1	0.51(0.48)	19(77)	0.13	0.67(0.51)	0	1	0.47(0.51)	61(35)	0.21
Polite	0.95(0.88)	0	1	0.23(0.32)	19(77)	0.41	0.92(0.86)	0	1	0.28(0.36)	61(35)	0.35
Assessed subjective evaluations (results were given on a scale of 1 to 10)												
Concern	4.84(5.12)	3(2)	7(9)	1.46(1.53)	19(77)	0.49	5.15(4.91)	2(3)	9(8)	1.61(1.34)	61(35)	0.58
Professional skills	4.79(4.96)	3(2)	7(8)	1.40(1.55)	19(77)	0.65	4.87(5.02)	2(2)	8(8)	1.55(1.47)	61(35)	0.45
Respect	5.42(5.22)	3(2)	7(8)	1.35(1.47)	19(77)	0.548	5.34(5.11)	2(2)	8(8)	1.45(1.43)	61(35)	0.54
Treatment information	4.68(4.86)	2	7(8)	1.57(1.56)	19(77)	0.701	4.93(4.63)	2(2)	8(7)	1.55(1.57)	61(35)	0.42
Overall	4.89(4.96)	3(2)	8(8)	1.41(1.53)	19(77)	0.81	4.98(4.89)	2(2)	8(8)	1.57(1.39)	61(35)	0.87
Recommend	4.11(4.22)	1(1)	7(9)	1.63(1.85)	19(77)	0.95	4.16(4.26)	1(1)	9(7)	1.88(1.67)	61(35)	0.61

who are predisposed to recommending antibiotics. As further evidence that doctors, regardless whether they returned or did not return a wallet, prescribe or did not prescribe antibiotics, are comparable, it is observed that the Mann Whitney p-values for all *assessed subjective evaluations* show no differences between the groups.

5.2 Non-Parametric Results

Table 3 compares the outcome variables over two dimensions. For physicians who(do not) prescribe antibiotics, we compare whether prescription behavior is different for doctors who return the wallet versus those who did not and present the Mann Whitney p-values in the parenthesis next to their mean values vertically. For physicians who(do not) return the wallet, we compare whether prescription behaviors are different for doctors who prescribe antibiotics versus those who did not and present the Mann Whitney p-values in the last column. From the table, we can see that when we look at the prescribing behavior of physicians who both returned the wallet and did not prescribe antibiotics, their prescribing behavior is not significantly different from physicians who did not return the wallet but did not prescribe antibiotics. The only significant results we can observe is between doctors who prescribe antibiotics versus those who did not. These preliminary comparisons show that on average, doctors who prescribe antibiotics charge about double for their treatment drugs than doctors who do not. This is true even for doctors who returned the wallet which suggests that while their actions are clearly altruistic and costly, however they still operate as revenue maximising entities when running their clinics.

Of the 96 physicans in our sample, there were 52 physicians who did not return the wallet and prescribed antibiotics, 25 physicians who did not return the wallet but did not prescribe antibiotics, 9 physicians who returned the wallet but prescribed antibiotics and 10 physicians who both returned the wallet and did not prescribe antibiotics. We report to readers an oversight in our design: our pseudo patients were instructed to remember the prices of drugs being recommended by physicians, however as some doctors chose to relay the information of each individual drug they recommended, there was a recall error committed by our patients. As a result, 9 of the 96 observations in our dataset do not have the total cost of the treatment charge. As a result we only record 87 observations for the results shown in Table 3. We rectify this problem in the next section where we carry out logit analysis looking at physicians who charged above 20 yuan (thereby including 6 out of 9 observations that we omitted here).

5.2.1 The multidimensional relationship between these two measures of beneficence

Are doctors who return the wallet the same doctors who choose not to prescribe antibiotics? This question is important because it answers the question whether doctors who choose to do good are the same doctors who choose to do no harm. A correlation test suggests a very weak relationship between the two measures (ρ= 0.1669) - only 10.4% of the 96 physicians do both,

Table 3: Summary statistics of the prescription behavior of physicians using a 2x2 matrix: Returning or not returning wallet x prescribing or not prescribing antibiotics

		Prescribed Antibiotics									
		1					0				
		Mean	Std dev	Min	Max	N	Mean	Std dev	Min	Max	N	MW p-value
	Prescipt. Behav.											
	Total price	20.78	5.17	8	25	9	13.25	10.10	0	25	10	0.06
1	Total price/min	3.73	2.13	0.8	7.67	9	3.40	3.56	0	11.5	10	0.51
	Total price >20	0.56	0.53	0	1	9	0.3	0.48	0	1	10	0.27
Returned		MW					MW					
		p-value					p-value					
wallet												
	Total price	20.68 (0.78)	8.36	3	46	45	10.95 (0.29)	10.36	0	42.5	23	0.00
0	Total price/min	4.02 (0.41)	4.17	0.2	20	45	1.78 (0.29)	2.37	0	10.63	23	0.00
	Total price >20	0.54 (0.92)	0.50	0	1	52	0.16 (0.36)	0.37	0	1	25	0.00

64.6% of physicians do neither, and 25% of physicians do one of the two.

5.3 Parametric Results

In order to account for other omitted variables as well as possible interaction effects, we present both the logit and OLS analyses in Table 4. We include age, gender and their interaction variable as controls.

Table 4: Logit and OLS analyses of gender and altruistic preferences on physicans' prescription behavior

Model	Logit	OLS		Logit	OLS	
	(3) Total price >20yuan	(4) Total price	(5) Total price per minute	(6) Total price >20yuan	(7) Total price	(8) Total price per minute
Wallet	-0.16	-1.03	-0.11			
	(0.47)	(2.81)	(0.22)			
Antibiotics				**1.46*****	**8.55*****	**0.67*****
				(0.46)	(1.84)	(0.18)
Male	1.89	-4.55	-0.09	1.67	-5.76	-0.19
	(1.67)	(8.88)	(0.79)	(1.88)	(7.28)	(0.68)
Age	-0.60	**-5.91****	**-0.47****	-0.69	**-5.57****	**-0.45***
	(0.42)	(2.38)	(0.22)	(0.65)	(2.48)	(0.24)
Male x Age	-0.25	3.96	0.24	-0.23	3.46	0.21
	(0.62)	(3.18)	(0.28)	(0.73)	(2.83)	(0.27)
Constant	0.44	**29.03*****	**2.27*****	-0.25	**23.96*****	**1.87*****
	(1.08)	(6.36)	(0.54)	(1.53)	(7.05)	(0.66)
Log likelihood	-57.04			-52.90	-52.92	
Pseudo R^2	0.09			0.15		
R^2		0.12	0.13		0.29	0.25
N	93	87	87	93	87	87

Standard errors clustered by patient ID in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 shows us that returning a wallet does not significantly predict profit-taking behaviors, however prescribing antibiotics does. We can see from both the logistic and OLS regressions that physicians who prescribe antibiotics are more likely to offer drugs at a cost higher than 20

yuan (odds ratio of 4.31) and offer drugs at more than 10 yuan higher and earn 0.67 yuan per minute more than doctors who do not prescribe antibiotics. We also see some evidence that older doctors are less likely to overcharge patients[12].

We add the logit regressions to include 6 of the 9 physicians left out in our OLS regressions in our sample for which we have some of the treatment costs but not all. We also note that younger doctors earn significantly more both in terms of higher revenues and total price per minute of consultation. This suggests that younger doctors are running their clinics more like business entities by prescribing costlier medications and lowering consultation times.

5.4 Sensitivity Test 1: Could higher revenues earned by doctors reflect better quality drugs rather than profit-taking behavior?

We obtain the wholesale prices of 86.2% of all prescribed medications prescribed by our sample of doctors online and derive estimates of total profit and total profit earned per minute of consultation by taking total price of treatment minusing the wholesale prices of those drugs available through online sources. We then rerun the OLS regression analyses in Table 4 to see if the results are consistent with a motive of profit-taking.

Table 5: Do physicians who prescribe antibiotics (return wallet) earn greater profit than those who do not?

Model	(9)	(10)	(11)	(12)
Dep Var	Total profit	Total profit/min	Total profit	Total profit/min
Wallet	-0.63	-0.03		
	(1.93)	(0.14)		
Antibiotics			**3.91****	**0.34****
			(1.71)	(0.16)
Male	-5.12	-0.53	-5.51	-0.55
	(7.68)	(0.63)	(6.91)	(0.59)
Age	-3.34	0.32	-3.14	-0.31
	(2.26)	(0.20)	(2.40)	(0.22)
Male x Age	3.04	0.28	2.69	0.25
	(3.03)	(0.25)	(2.91)	(0.25)
Constant	**16.84****	**1.47****	**14.50****	**1.28****
	(5.64)	(0.51)	(6.78)	(0.62)
R^2	0.07	0.07	0.15	0.14
N	79	79	79	79

Note: Standard errors clustered by patient ID in parentheses * Significant at the 10% level. ** Significant at the 5% level.

We can see from Table 5 that physicians who prescribed antibiotics profited around 4 yuan more than their counterparts who did not. *Profit per minute* is also higher by 0.34 yuan for doctors prescribing antibiotics than doctors who did not. Hence we can dismiss the notion

[12]The negative and significant coefficient for age may reflect a wealth or income effect in that older doctors may be richer and have a weaker incentive to over-charge or profit from patients than younger doctors. Alternatively, it may also indicate that in such a market, younger and more profit-oriented physicians may choose to move out of the profession in favour of more lucrative opportunities, and hence those who remain who are more beneficent.

that the increased earnings of physicians who prescribe antibiotics solely result from elevated drug costs or superior medication quality. We rule out the possibility that antibiotic drugs are more costly than other drugs, as the profit margin remains greater for doctors prescribing antibiotics. We also see that the age of doctors no longer explains this profit-taking phenomena. This implies that younger doctors may be prescribing more expensive medications, yet their profits are no par with those who prescribe cheaper alternatives.

5.5 Sensitivity Test 2: Does this result hold if we remove from the sample physicians who did not prescribe any medication?

Our earlier OLS regression results include the 19 physicians who do not prescribe any medication at all. This inclusion has the effect of weighting the average treatment cost downwards for physicians not prescribing antibiotics, and may be the reason for our positive result. To address this bias, we remove physicians who do not prescribe drugs and directly compare the profit made by physicians who prescribe antibiotics and physicians who prescribe all other types of medications except antibiotics. This allows us to compare doctors who practise some form of beneficence (choosing not to do harm by prescribing antibiotics) with those who do not. Table

Table 6: Do physicians who prescribe antibiotics earn greater profit than those who prescribe other medications?

Model	(13)	(14)	(15)	(16)	(17)	(18)
Dep Var	Total profit			Total profit/min		
Total price of drugs	>0	>12	>15	>0	>12	>15
Antibiotics	1.37	0.59	-0.17	0.14	0.08	0.09
	(1.34)	(1.53)	(2.16)	(0.14)	(0.18)	(0.26)
Male	-5.99	-3.79	-6.63	-0.63	-0.56	-0.72
	(7.81)	(10.02)	(10.71)	(0.66)	(0.84)	(0.90)
Age	-2.98	-3.12	-3.27	-0.32	-0.37	-0.38
	(2.59)	(3.39)	(3.40)	(0.24)	(0.31)	(0.32)
Male x Age	2.46	1.81	3.18	0.25	0.25	0.33
	(3.21)	(4.13)	(4.31)	(0.27)	(0.35)	(0.37)
Constant	**17.37****	**19.62****	**20.79****	**1.55****	**1.84****	**1.87***
	(6.94)	(8.58)	(9.24)	(0.66)	(0.81)	(0.87)
R^2	0.58	0.09	0.07	0.08	0.12	0.10
N	70	56	47	70	56	47

Note: Standard errors clustered by patient ID in parentheses * Significant at the 10% level. ** Significant at the 5% level.
We restrict our full sample in Table 6 to only the physicians who actually prescribe medications (above 0 yuan, above 12 yuan, and above 15 yuan).

6 further clarifies our earlier results. When we narrow down our sample solely to doctors who prescribe medications, we discover that both total profit and the total profit per minute of consultation, regardless of the type of medication (including antibiotics or others), exhibit no significant differences. This suggests that our earlier results were indeed picking up a positive effect only because of the inclusion of doctors who were earning zero revenues from the patient. Consequently, we conclude that doctors, as long as they were willing to prescribe medication to generally healthy patients, were focused on maximising their revenues and profits. Thus, the

choice to "do no harm" by prescribing less harmful medications did not diminish their profit-oriented behaviors. These over-charging practices align with the findings in the literature of credence goods, especially in situations where liability rules are in place[13] (Balafoutas et al., 2013; Dulleck et al., 2011).

One particular reason might explain why we find a weak link between physician beneficence and profit-oriented behaviors. Since our study looks at primary care physicians who own their own clinics and whose sole income comes from the sale of medications, our findings may be due primarily to sample selection bias and may not reflect the behaviors of physicians in public hospitals who have other sources of income on top of the incentives linked to the treatments they prescribe.

6 Limitations

We thank the anonymous referees for pointing to us the possible weaknesses in our paper. We acknowledge the existence of some possible omitted variable bias as we are unable to provide any neighborhood, clinic or physician socio-demographic covariates to control for prescribing behavior. We had agreed to an IRB protocol whereby no physician's data or clinic's data could potentially be linked to one another, hence this oversight is attributed to an initial design which did not include these variables.

We are also unable to quantify the effect of the individual 12 simulated patients' characteristics on physicians' behavior in our study, owing to the fact that we have decoupled all information linking pseudo-patient and clinic, clinic and doctor. However we do have information on which patient visited which 8 clinics and can add patient dummy variables as control variables. Given the concern that overcharging might be related to patient characteristics, we have run separate regressions similar to Table 6, looking at whether profit or profit per minute was influenced by both physician characteristics as well as controlling for patient dummy variables. Our results show that these patient dummies were not significant[14]. We also thank an anonymous referee for the suggestion of incorporating a measure which sheds light on whether the clinician believed the patient to be in pain or in a state of discomfort. This we reserve for future research.

7 Conclusion

Employing a mystery shopper audit field experiment in a provincial capital city in China, our initial findings reveal a significant statistical correlation between doctors who prescribe antibiotics and higher total revenues compared to those who abstained from doing so. Additionally, when we factor in the wholesale prices of the recommended drugs obtained by

[13]Such liability rules exist especially for medical practitioners who under-provide medical care to patients and risk losing their license or are censured by a professional body.

[14]We acknowledge that using patient dummies does not rule out that individual patient characteristics could have influenced the physician's prescription behavior.

online suppliers, we find that these doctors also earn greater profits. In contrast, we do not observe similar disparities in earnings and profits between doctors who return a lost wallet and those who do not. When we unpack this result further by limiting our sample solely to doctors who prescribe medications, these differences in revenues and profits become insignificant. This highlights the reality that doctors who choose to prescribe medications, whether choosing to include antibiotics or not, exhibit profit-driven motivations in our study.

All in all, our investigation to understand how two novel measures of beneficence might affect profit taking among physicians in private for-profit clinics in a first tier city in China has found that physician beneficence seems to be an independent decision from the decision to profit from the doctor-patient relationship. This suggests that while those who return the wallet have an beneficent preference to "do good", they are still willing to profit on otherwise healthy patients, providing them with medications that they do not need. In the same vein, physicians who choose to "do no harm" to patients by not prescribing antibiotics earn the same as their counterparts who do prescribe antibiotics. Hence we find that in general, doctors who prescribe any medications are still willing to engage in inducement and earn profits from their otherwise healthy patients. This is in line with the adage that "business is business" and primary care private clinics in China operate as purely profit-maximising entities.

Acknowledgements

IRB Approval

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), which serves as the institutional review board for the University of Oslo.

Conflicts of interest/Competing interests

The authors have no conflicts of interest in this research.

References

Alem, Y., Eggert, H., Kocher, M. G., and Ruhinduka, R. D. (2018). Why (field) experiments on unethical behavior are important: Comparing stated and revealed behavior. *Journal of Economic Behavior & Organization*, 156:71–85.

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

Andreoni, J., Nikiforakis, N., and Stoop, J. (2021). Higher socioeconomic status does not predict decreased prosocial behavior in a field experiment. *Nature Communications*, 12(1):4266.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–973.

Arya, S. C. (2004). Antibiotics prescription in hospitalized patients at a chinese university hospital. *Journal of Infection*, 48:117 – 118.

Attema, A. E., Galizzi, M. M., Groß, M., Hennig-Schmidt, H., Karay, Y., L'haridon, O., and Wiesen, D. (2023). The formation of physician altruism. *Journal of Health Economics*, 87:102716.

Balafoutas, L., Beck, A., Kerschbamer, R., and Sutter, M. (2013). What drives taxi drivers? a field experiment on fraud in a market for credence goods. *Review of Economic Studies*, 80(3):876–891.

Balafoutas, L. and Kerschbamer, R. (2020). Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions. *Journal of Behavioral and Experimental Finance*, 26:100285.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11:122–133.

Beres, D. (2013). Most honest cities. *The Reader's Digest "lost wallet" test. In Reader's Digest.*, 24:2017.

Bisgaier, J. and Rhodes, K. V. (2011). Auditing access to specialty care for children with public insurance. *New England Journal of Medicine*, 364(24):2324–2333.

Böhm, R. and Theelen, M. M. (2016). Outcome valence and externality valence framing in public good dilemmas. *Journal of Economic Psychology*, 54:151–163.

Brock, J. M., Lange, A., and Leonard, K. L. (2016). Generosity and prosocial behavior in healthcare provision: Evidence from the laboratory and field. *Journal of Human Resources*, 51(1):133–162.

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

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

Byambadalai, U., Ma, C.-t. A., and Wiesen, D. (2023). Changing preferences: An experiment and estimation of market-incentive effects on altruism. *Journal of Health Economics*, 92:102808.

Cappelen, A. W., Nielsen, U. H., Sørensen, E. Ø., Tungodden, B., and Tyran, J.-R. (2013). Give and take in dictator games. *Economics Letters*, 118(2):280–283.

Charness, G., Haruvy, E., and Sonsino, D. (2007). Social distance and reciprocity: An internet experiment. *Journal of Economic Behavior & Organization*, 63(1):88–103.

Chukwuani, C., Onifade, M., and Sumonu, K. (2002). Survey of drug use practices and antibiotic prescribing pattern at a general hospital in nigeria. *Pharmacy World and Science*, 24(5):188–195.

Cohn, A., Maréchal, M. A., Tannenbaum, D., and Zünd, C. L. (2019). Civic honesty around the globe. *Science*, 365(6448):70–73.

Cox, J. C., Servátka, M., and Vadovič, R. (2010). Saliency of outside options in the lost wallet game. *Experimental Economics*, 13(1):66–74.

Crilly, D., Ni, N., and Jiang, Y. (2016). Do-no-harm versus do-good social responsibility: Attributional thinking and the liability of foreignness. *Strategic Management Journal*, 37(7):1316–1329.

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.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33:67–80.

Delattre, E. and Dormont, B. (2003). Fixed fees and physician-induced demand: A panel data study on french physicians. *Health Economics*, 12(9):741–754.

Dijk, C. E., Berg, B., Verheij, R. A., Spreeuwenberg, P., Groenewegen, P. P., and Bakker, D. H. (2013). Moral hazard and supplier-induced demand: Empirical evidence in general practice. *Health Economics*, 22(3):340–352.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30(2):163–182.

Dulleck, U. and Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*, 44(1):5–42.

Dulleck, U., Kerschbamer, R., and Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*, 101(2):526–555.

Franzen, A. and Pointner, S. (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*, 16:155–169.

Galizzi, M., Godager, G., Li, J., Linnosmaa, I., Tammi, T., and Wiesen, D. (2023). Economics of healthcare provider altruism. In Zimmermann, K. F., editor, *Handbook of Labor, Human Resources and Population Economics*. Springer.

Gani, L., Arif, H., Widjaja, S. K., Adi, R., Prasadja, H., Tampubolon, L. H., Lukito, E., and Jauri, R. (1991). Physicians' prescribing practice for treatment of acute diarrhoea in young children in jakarta. *Journal of Diarrhoeal Diseases Research*, 9(3):194–199.

Ge, G. and Godager, G. (2021). Predicting strategic medical choices: An application of a quantal response equilibrium choice model. *Journal of Choice Modelling*, 39:100282.

Ge, G., Godager, G., and Wang, J. (2022). Exploring physician agency under demand-side cost sharing—an experimental approach. *Health Economics*, 31(6):1202–1227.

Glannon, W. and Ross, L. F. (2002). Are doctors altruistic? *Journal of Medical Ethics*, 28(2):68–69.

Godager, G. and Wiesen, D. (2013). Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32(6):1105–1116.

Goerg, S. J., Rand, D., and Walkowitz, G. (2019). Framing effects in the prisoner's dilemma but not in the dictator game. *Journal of the Economic Science Association*, pages 1–12.

Gottschalk, F., Mimra, W., and Waibel, C. (2020). Health services as credence goods: A field experiment. *The Economic Journal*, 130(629):1346–1383.

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

Gruber, J. and Owings, M. (1994). Physician financial incentives and cesarean section delivery. Technical report, National Bureau of Economic Research.

Harris, J. (2018). Altruism: Should it be included as an attribute of medical professionalism? *Health Professions Education*, 4(1):3–8.

Hellerstein, J. K. (1998). The importance of the physician in the generic versus trade-name prescription decision. *The Rand Journal of Economics*, 29(1):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.

Holmes Jr, D. R., Stefanini, G. G., Durand-Zaleski, I., Ntsekhe, M., and Wijns, W. (2020). From 'do no harm'to 'do maximal good': an evolving concept. *European Heart Journal*.

Huynh, T. L. D., Wang, M., and Rieger, M. O. (2020). How does the email matter to the civic honesty? a comment on cohn et al.(2019). *Business and Society Review*, 125(4):387–391.

Iversen, T. (2004). The effects of a patient shortage on general practitioners' future income and list of patients. *Journal of Health Economics*, 23(4):673 – 694.

Kerschbamer, R. and Sutter, M. (2017). The economics of credence goods–a survey of recent lab and field experiments. *CESifo Economic Studies*, 63(1):1–23.

Khadjavi, M. and Lange, A. (2015). Doing good or doing harm: Experimental evidence on giving and taking in public good games. *Experimental Economics*, 18(3):432–441.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *Quarterly Journal of Economics*, 112(4):1251–1288.

Korenok, O., Millner, E. L., and Razzolini, L. (2014). Taking, giving, and impure altruism in dictator games. *Experimental Economics*, 17(3):488–500.

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

Li, J., Casalino, L. P., Fisman, R., Kariv, S., and Markovits, D. (2022). Experimental evidence of physician social preferences. *Proceedings of the National Academy of Sciences*, 119(28):e2112726119.

Li, J., Dow, W. H., and Kariv, S. (2017). Social preferences of future physicians. *Proceedings of the National Academy of Sciences*, 114(48):E10291–E10300.

Li, J., Hurley, J., DeCicca, P., and Buckley, G. (2014). Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, 23(8):962–978.

Li, Y. (2014). China's misuse of antibiotics should be curbed. *British Medical Journal*, 348:g1083.

Lian, S., Chen, Q., Yao, M., Chi, C., and Fetters, M. (2019). Training pathways to working as a general practitioner in china. *Family Medicine*, 51(3):262–270.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.

Liu, X., Feng, J., Liu, C., Chu, R., Lv, M., Zhong, N., Tang, Y., Li, L., and Song, K. (2022). Medical education systems in china: development, status, and evaluation. *Academic Medicine*, 98(1):43–49.

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.

Martinsson, P. and Persson, E. (2019). Physician behavior and conditional altruism: the effects of payment system and uncertain health benefit. *Theory and Decision*, 87(3):365–387.

McGuire, T. G. (2000). Physician agency. In *Handbook of Health Economics*, volume 1, pages 461–536. Elsevier.

McGuire, T. G. and Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics*, 10(4):385–410.

Moriarty, H., McLeod, D., and Dowell, A. (2003). Mystery shopping in health service evaluation. *British Journal of General Practice*, 53(497):942–946.

Park, M., Braun, T., Carrin, G., Evans, D. B., Organization, W. H., et al. (2007). Provider payments and cost-containment: lessons from oecd countries. Technical report, World Health Organization.

Plotzke, M. R. and Courtemanche, C. (2011). Does procedure profitability impact whether an outpatient surgery is performed at an ambulatory surgery center or hospital? *Health Economics*, 20(7):817–830.

Quast, T., Sappington, D. E., and Shenkman, E. (2008). Does the quality of care in medicaid mcos vary with the form of physician compensation? *Health Economics*, 17(4):545–550.

Quinn, A. E., Trachtenberg, A. J., McBrien, K. A., Ogundeji, Y., Souri, S., Manns, L., Rennert-May, E., Ronksley, P., Au, F., Arora, N., et al. (2020). Impact of payment model on the behaviour of specialist physicians: A systematic review. *Health Policy*, 124(4):345–358.

Reardon, S. (2014). Antibiotic resistance sweeping developing world. *Nature*, 509(7499):141.

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.

Rizzo, J. A. and Zeckhauser, R. J. (2003). Reference incomes, loss aversion, and physician behavior. *Review of Economics and Statistics*, 85(4):909–922.

Rizzo, J. A. and Zeckhauser, R. J. (2007). Pushing incomes to reference points: Why do male doctors earn more? *Journal of Economic Behavior & Organization*, 63(3):514–536.

Stoop, J. (2014). From the lab to the field: envelopes, dictators and manners. *Experimental Economics*, 17:304–313.

Tappin, B. M. and Capraro, V. (2018). Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. *Journal of Experimental Social Psychology*, 79:64–70.

Xirasagar, S. and Lin, H.-C. (2006). Physician supply, supplier-induced demand and competition: empirical evidence from a single-payer system. *The International Journal of Health Planning and Management*, 21(2):117–131.

Appendices

A Scripts of pseudo-patient used in 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 degrees celsius.

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 is it?
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?
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.

B Experimental protocol for the pseudo-patient 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.

C Data collection sheet

Survey of doctor:

1. Gender of doctor.
2. Patient's subjective assessment of doctor age, in the following categories:
   ≤ 30,[31-40],[41,50],≥ 50.
3. (a) Is more than 1 doctor sharing office?
   (b) If yes, how many addititional doctors?
4. How many other patients are present at the office during wait
5. How many minutes did you wait before seeing the doctor?
6. How many minutes did the doctor spend on the patient treated before you?
7. How many minutes did the doc spend on you?
8. When you entered doctors office, how many additional patients were present inside the office?

Diagnostic procedures:

9. Did the doctor ask if you had symptoms of coughing?
10. Did the doctor ask if you have wet cough?
11. Did the doctor measure your temperature?
12. (a) Did the doctor check your throat?
    (b) If yes did the doctor tell you the result?
13. (a) Did the doctor use a stethoscope?
    (b) If yes did the doctor tell you the results
14. Did the doctor ask if you have any allergies?

15. Did the doctor ask you if you have medicine for cold at home?

16. (a) Did the doctor prescribe you any medicine(s)? If yes write down name of medicine(s), name of pharmaceutical company, price of medicine(s). Total price for whole recommended bundle.

    (b) If the doctor prescribed any medicine to you, did the doctor describe how to use the drugs? (for example before or after meal, how many times per day, how big dose to apply each time).

    (c) If the doctor has prescribed medicine to you, did the doctor describe any side effects of the medicine?

17. Did the doctor tell you other relevant information you should pay attention to? (for example drink more water, rest more, put on more clothes, eat more fruit, avoid excessive and intensive exercise etc.)

Departure:

18. After you said thank you to the doctor, did the doctor reply you back? (for example you are welcome)

Assessment (scale 1-10):

19. During the visit, to what extent was the doctor concerned about you?

20. Subjective impression of professional skills.

21. During the process of diagnosis, to what extent do you think the doctor respected your own opinion?

22. To what extent did the doctor inform you about your illness and treatment?

23. What is your overall assessment of the consultation.

24. To what extent would recommend your parents to visit this doctor if they experienced the similar illness?