The effect of a mystery shopper scheme on prescriptions in primary care

Roland Cheo
Center for Economic Research, Shandong University, China

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

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

Rugang Liu
Center for Health Economic Experiments and Public Policy, School of Public Health, Shandong University, China

Qiqi Wang
School of Economics, Shandong University, China

Jian Wang
Department of Health Management and Health Economics, University of Oslo, Norway

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK

Working paper 2018: 1

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-262-9

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

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

Abstract

Health care systems in many countries are still characterized by limited availability of provider performance data which 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. Using a randomized treatment-control design, we conduct a field experiment in primary care clinics in a Chinese city. We investigate whether informing clinics in the treatment group of a forthcoming mystery shopper audit influences the physicians' prescribing behavior. As expected, we find that antibiotic medications are prescribed to patients in the majority of cases, even though such prescribing is not in accordance with current recommendations or guidelines. While the intervention did not cause significant reduction in antibiotic prescriptions, our results show that a mystery shopper scheme reduces overall unnecessary prescribing.

Keywords: Field Experiment, Analysis of Health Care Markets, Government Policy, Information and Product Quality, Social Responsibility.

JEL-Classification: C93, I11, I18, L15, M14.

☆Financial support by the Research Council of Norway (IRECOHEX, Project-No. 231776) is gratefully acknowledged.

*Corresponding author. Email: gege@medisin.uio.no

[1]Center for Economic Research, Shandong University, China

[2]Institute of Health and Society, Department of Health Management and Health Economics, University of Oslo, Norway.

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

[4]Center for Health Economic Experiments and Public Policy, School of Public Health, Shandong University, China

[5]School of Economics, Shandong University, China

1. Introduction

Asymmetric information about product quality is a fundamental characteristic of the market for medical care, as noted by Arrow (1963). The seller of the product is an expert who typically holds information that is superior to that of the buyer. When provider quality assurance is limited by the presence of asymmetric information, it affects the provider's 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 advancement in information and communication technology in the development of policy measures, for example by designing mechanisms for provider payment that are based on routinely collected data on provider activity and performance. The Quality and Outcomes Framework (QOF) in United Kingdom, is an example of an extensive pay-for-performance program which relies on advanced infrastructure in the form of health registers and patient lists when measuring provider performance.

Countries differ in the feasibility of these type of policy measures. Many health care systems are still characterized by limited availability of provider performance data and patient registers which can be used to implement pay-for-performance schemes. In the presence of asymmetric information on product quality, the degree of asymmetry can be influenced by introducing simple auditing schemes. Such performance auditing can be implemented without necessarily linking financial incentives to performance. As described by Dranove (2011), health plans and hospitals frequently contribute actively to quality assurance mechanisms, by collecting and voluntary disclosing quality information. While knowledge of hospital performance is a necessity in modern hospital management, auditing of privately practicing physicians will more likely require an external initiative. As described in the review by Jamtvedt et al. (2006) most intervention studies on auditing focus on the effect of auditing when it is combined with other measures such as *reminders* (Baker et al., 1997; Eccles et al., 2001), *feedback* (White et al., 1995; O'connell et al., 1999; Wells et al., 2000; Eccles et al., 2001; Kiefe et al., 2001; Baker et al., 2003; Godager et al., 2016) or *educational interventions* (Feder et al., 1995; Kerse et al., 1999). In a recent study by Östervall (2017), however, the effect of auditing primary care physicians' practice in Sweden is separated from the effect of reminding physicians and patients about inappropriate use of antibiotics. The reminders are found to have a substantial effect on prescribing, whereas the introduction of auditing does not significantly influence physician prescribing behavior. Our study relates to the study by Östervall (2017) in that we aim at quantifying the effect of 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. The method of mystery shopping is frequently used for performance measurement to reduce asymmetry of information in industries organized as chains. Mystery shopper schemes enables decision makers to acquire performance information on subdivisions of an organization, information that can be used for pure monitoring purposes as well as performance based payment (Wilson, 1998). Mystery shopper schemes can be customized to suit different purposes, and the use of 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 or when audits will occur. Decades ago, the mystery shopping approach was adopted in the health domain to study provider behavior and it is proved valuable to the society (Madden et al., 1997). In a health context, mystery shoppers are commonly refereed to as *pseudo patients*, *simulated patients*, *standardized patients* or *surrogate patients*. The use of pseudo patients involves an element of deception, which generally involves careful ethical considerations, especially in the health and research domains. The application of this method can be ethically justified, however, as long as individual's confidentiality is protected, the risks to the research subjects are minimal and the research is potentially valuable in furthering our knowledge on the subject (Rhodes and Miller, 2012)[1].

The quality measure applied in our study is the physician's prescribing behavior when the patient presents a specific set of symptoms. The specific symptoms presented by the pseudo patients in this study are symptoms of the common cold. Medical guidelines recommend that no medication is prescribed for common cold symptoms. Hence, whether or not medication is prescribed is an observable and convenient quality measure. Prescribing behavior in primary care is also a highly relevant quality aspect, as inappropriate prescribing of medication is a global public health challenge. According to World Health Organization (2012), more than one half of medical prescriptions worldwide are inappropriate, causing adverse health outcomes and raising health expenditures. Over-prescribing of antibiotics is common in many countries, leading to a widespread resistance against medication for treatable bacterial infections (Gani et al., 1991; Chukwuani et al., 2002; Arya, 2004; Reardon, 2014). Governments are increasingly implementing guidelines and regulations to curb such misuse of medication. The literature reveals, however, that antibiotics are prescribed

[1]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).

too often, even in the presence of guidelines and gate-keeping (Reynolds and McKee, 2009; Currie et al., 2011, 2014).

We conduct a field experiment where we randomize clinics to either a treatment or control group. We apply similar audit methodology as Currie et al. (2011, 2014) and announce a forthcoming mystery shopper audit only to the clinics in the treatment group. We find that a mystery shopper scheme reduces overall unnecessary prescribing. For the case of non-antibiotic medication, the mystery shopper scheme reduces the frequency of inappropriate prescribing by 15.46%. For the case of antibiotic medication, the intervention does not cause significant reduction in inappropriate prescribing. This paper contributes to the literature using field experiments to acquire knowledge on key mechanisms in health service delivery. It provides new evidence suggesting that auditing primary care providers can have a direct effect on physician behavior even when it is not combined with pay-for-performance, or other measures such as reminders, feedback or educational interventions. The remainder of the paper is organized as follows. We present our model and hypotheses in Section 2. The experimental design and procedure is described in Section 3. The descriptive statistics and results are presented in Section 4. In section 5 we conclude and discuss.

2. Model and hypotheses

The patient-physician relation is commonly described as a case of (imperfect) agency (McGuire, 2000). The patient (principal) consults the physician (agent), who is an expert with superior information regarding health and expected treatment effects. Under perfect physician-agency, the treatment alternative which is optimal for the patient will coincide with the treatment alternative which is optimal for the physician. In our study setting, income from selling prescribed medication comprises a substantial share of physician income. Financial incentives to prescribe drugs result in conflicting objectives between patient and physician, as it becomes costly to always behave as a perfect agent on behalf of the patient.

We study the case of a patient with the common cold, where prescribed medication is not expected to contribute towards 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 patient and physician were equally well informed. Upon seeing a patient with the common cold, the physician decides whether to prescribe, or not to prescribe, medication. We assume that the patient passively accept the physician's treatment recommendation and indicate the prescribing choice by a, where $a = 0$ if the alternative *not prescribe* is chosen, whereas $a = 1$ if physician chooses *prescribe*. We assume that the physician's net profit π from prescribing is positive. The physician's choice affects patient's net benefit (health benefit measured in money minus cost of medication). We let $V(a)$ denote net patient-benefit. In the case of a patient with the common cold we assume that prescribing reduces patient's net benefit, $V(1) - V(0) < 0$. We assume further that physicians are partly altruistic, and similar to Farley (1986) we include the physician's concern for the patient's overall well-being when specifying the physician's objective. Letting $\alpha > 0$ denote a physician preference parameter indicating the weight attached to patient's net benefit, the objective for a physician who is only concerned about profit and patient's net benefit can be expressed as:

$$U(a) = \pi a + \alpha V(a) \tag{1}$$

where $U(1) = \pi + \alpha V(1)$ when *prescribe* is chosen and $U(0) = \alpha V(0)$ when *not prescribe* is chosen. A rational physician would prefer to prescribe if $U(1) - U(0) > 0$, to not prescribe if $U(1) - U(0) < 0$ and be indifferent if $U(1) - U(0) = 0$. Under the assumption that the physician maximize (1), physicians with low altruism, $\alpha < \frac{\pi}{V(0)-V(1)}$, will prescribe, those with a high altruism, $\alpha > \frac{\pi}{V(0)-V(1)}$, will not prescribe, while physicians with $\alpha = \frac{\pi}{V(0)-V(1)}$ will be indifferent to prescribing alternatives. In the case of preference heterogeneity in the population of physicians, preference variation will cause practice variation in terms of heterogeneous prescribing choice for a given patient. Under the assumption that the physician maximize (1), a mystery shopper scheme will not influence prescribing behavior. Hence, we may state the refutable hypothesis:

H_0**: Physician prescribing behavior will not be affected by introducing a mystery shopper scheme.**

We propose however that the choice alternative *not prescribe*, being medically appropriate while yielding low physician profit, becomes more rewarding when the physician is informed of a forthcoming mystery shopper scheme, since it implies that his service quality and professionalism can be acknowledged by a relevant institution. 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. If one assumes that physician objective includes valuation of other elements such as "recognition by others" or "social stigma", introducing a mystery shopper scheme can influence behavior. See for example Bénabou and Tirole (2006) for an elaborate description. We indicate the existence of a mystery shopper scheme by T, where $T = 0$ if a mystery shopper scheme does not exist, whereas $T = 1$ if a mystery shopper scheme does exist. 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 absence of a mystery shopper scheme, $T = 0$, stigma does not affect provider objective, $S(1;0) - S(0;0) = 0$. In the case of mystery shopping, $T = 1$, however, prescribing unnecessary medication brings about a negative stigma effect: $S(1;1) - S(0;1) < 0$. The objective for a physician who cares about social stigma in addition to profit and patient's net benefit can be expressed as:

$$\tilde{U}(a;T) = \pi a + \alpha V(a) + \beta S(a;T) \tag{2}$$

where the preference parameter $\beta > 0$ indicates the preference weight of social stigma in the physicians objective function.

In the absence of a mystery shopper scheme, $T = 0$, a physician would prescribe if $\tilde{U}(1;0) - \tilde{U}(0;0) > 0$, where $\tilde{U}(1;0) = \pi + \alpha V(1) + \beta S(1;0)$ and $\tilde{U}(0;0) = \alpha V(0) + \beta S(0;0)$. In this case, the social stigma effect is absent since, by assumption, $S(1;0) = S(0;0)$. The choice situation with $T = 0$ is identical to the scenario discussed above where physicians who are less altruistic towards the patient's overall well-being, $\alpha < \frac{\pi}{V(0)-V(1)}$, will prescribe.

In the presence of mystery shopper scheme, $T = 1$, a physician's decision depends on the sign of $\tilde{U}(1;1) - \tilde{U}(0;1)$, where $\tilde{U}(1;1) = \pi + \alpha V(1) + \beta S(1;1)$ and $\tilde{U}(0;1) = \alpha V(0) + \beta S(0;1)$. It can be shown that in a population of physicians that maximize (2) with varying α, introducing a mystery shopping scheme will result in 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 in the absence of mystery shopping, the physician with preference weight α such that $\alpha = \frac{\pi}{V(0)-V(1)}$. This physician will now strictly prefer *not prescribe*, since $\tilde{U}(1;1) - \tilde{U}(0;1) = \beta(S(1;1) - S(0;1)) < 0$. We may specify an alternative hypothesis:

H_A: **Physician prescribing will be reduced by introducing a mystery shopper scheme.**

3. Experimental design and procedure

The literature reveals that Chinese physicians prescribe medication, especially antibiotics, when they should not (Reynolds and McKee, 2009; Currie et al., 2011, 2014). An important cause of medication over-prescribing in China is the financial incentive. Revenues from selling medication have become more important to hospitals since the early 1980s when the government began to cut down financial support to hospitals (Yip and

Hsiao, 2008). For physicians in private clinics, profit from medication sales is often the main source of income, as they most often do not charge consultation fees. To reduce the physicians' financial incentives of antibiotic over-prescription in China, multiple reforms have been implemented by the Chinese government since 2009. However these reforms have not proven effective (Yip et al., 2012).

Our field experiment was carried out in Jinan, the capital city of Shandong province in China. It would be difficult to conduct a similar field experiment in the U.S. or in an European country, where durable physician-patient relations, often formalized as patient list systems, are common. By carrying out the experiment in China, we are able to randomly assign pseudo patients to clinic visits. We chose Jinan as the study location for two main reasons. First of all, because of its status as a first tier city in China and its position as a provincial capital, Jinan has a vibrant economy which mirrors other provincial capitals in China. Secondly, the support we received from School of Public Health, Shandong University and Qilu Health Service Center, affiliated to the largest public hospital in Jinan, added substantial credibility to the mystery shopper intervention.

From official Chinese registers, Health and Family Planning Commission of Jinan Municipality, we identified 118 primary care clinics in Jinan based on the following criteria: the clinic is for-profit[2] with only one practicing physician; must be located in the 5 districts of Jinan city[3]; has a valid license at the date of experiment; and provides general medicine[4]. From the list of suitable clinics we then randomly assigned 48 clinics to the control group, 48 to the treatment group and the rest 22 clinics as backups. In case any visited clinic is 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 antibiotics to be prescribed in a majority of consultations. Our aim is to assess whether the intervention could bring about a substantial reduction in inappropriate prescriptions. 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 percent when significance level is set at the conventional level of 5 percent.

[2]Non-profit clinics do not pursue economic profits by definition, so we assume that the problem of medication abuse is much less prevalent and severe in non-profit clinics.

[3]Other districts or counties are too far to reach.

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

Mystery shopper audit

Following Moriarty et al. (2003) and Bisgaier and Rhodes (2011), 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 as Table 1. Through the first audit we collected baseline data on characteristics of the clinics and practicing physicians and their prescribing behavior. Based on the second audit we compare differences in prescribing behavior between the treatment and control group. In both audits, pseudo patients present symptoms of the common cold to the physician based on a script (see Appendix A) and a protocol (see Appendix B). They describe their symptoms as "feel fatigued...have a low grade fever, slight dizziness, a sore throat and a poor appetite", and they tell the physician that the body temperature is 37 °C in the morning. The pseudo patients are clearly instructed not to say to the physician that they have a cold. They then let the physician measure the temperature and/or visually inspect the throat. The pseudo patients are instructed to refuse any other treatment or diagnostic test by the physician. If the physician prescribe any medication and the total price is lower than 20 Yuan, the pseudo patient buys the medication. If it is more than 20 Yuan, the patient is instructed to memorize the name(s) and the pharmaceutical company of the medication prescribed, and not to buy it. A pseudo patient is always accompanied by a fellow student on their visits. The fellow students have the tasks to observe 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 to help the pseudo patient memorize the medication names. The pseudo patient and the accompanying student fill out a data collection sheet together after they leave the clinic.

	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

Table 1: Timeline of the field experiment

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 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 Figure C.1 for the original project description letter in Chinese and Figure C.2 for an English translation in Appendix C). The project is about quality evaluation of primary care services in Jinan, particularly on 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 hence evaluate the quality of care. To enhance the credibility of the research project, we offered the clinics three ways to receive feedback of the quality assessment: publicly available feedback (results will be published on Shandong University website), feedback in private (result will only be received by the clinic) or no feedback. The representative read the project description together with the physician and ensured that the physician understood the project. In addition, Qilu Health Science Center, affiliated to Shandong University and one of the largest public hospitals (Qilu Hospital) in Jinan, provided an endorsement letter to support the project (see Figure C.3 for the original endorsement letter in Chinese and Figure C.4 for an English translation in Appendix C). The representative presented the endorsement letter to the physician and left both the stamped project description and endorsement letter at the clinic.

Training of the pseudo patients

The audits were carried out by 12 healthy pseudo patients, each accompanied by a fellow student, recruited from School of Public Health, Shandong University[5]. Each pair of students (a pseudo patient and an accompanying student) underwent 10 hours of training in total on 10th and 11th October 2015. The purpose was to ensure adherence to the script and the protocol. On the first day, they went through a review on types of antibiotics in the market place. 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 identified 118 clinics. To further ensure that the script was adhered to, the data collection sheets as well as physician-patient dialogs from the practice visits were discussed. We expect that the extensive training of pseudo patients and accompanying students reduces data variation due to subjective interpretations by the pseudo patients. The teams of pseudo patients were randomly assigned to clinics, and they were not informed about whether clinics were in the treatment or control group. We further ensure that none of the teams visited the same clinic on both the first and second audits.

[5]In total 13 pseudo patients were recruited (4 males and 9 females), allowing for one female student as a backup; 13 accompanying students were recruited (2 males and 11 females), allowing for one female student as a backup.

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 which provides unique and valuable knowledge to the society (Madden et al., 1997; Rhodes and Miller, 2012). The use of deception is controversial within social science research. However, following Rhodes and Miller (2012)'s ethical analysis, it can be ethically justified as long as confidentiality of research data is protected, the risks to the research subjects are minimal and the research is potentially valuable to the human knowledge.

During our actual field study, to ensure the safety of pseudo patients, they were always accompanied by a fellow student, so a team of two students always traveled together. Further, the patients, being students of School of Public Health, had at least one semester of basic medical training and they were especially instructed to refuse any treatment and/or diagnostic test by the physician except for temperature measuring and visual inspection of 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 were destroyed after the visits, so that data from the clinics could not be traced to a particular clinic or physician, even by the researchers. The field experiment also contributed positively to the revenues of clinics in the study sample.

4. Descriptive statistics and results

The 96 clinics were randomized into the treatment and control group. The map (see Figure 1) indicating the locations of the clinics in the treatment and control group provides a rough impression that the treatment and control clinics are randomly scattered in Jinan city. Table 2 reports the inclusion of treatment and control clinics over the 5 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).

	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

Table 2: Table of locations of sampled 96 clinics

During the experiment, it was discovered that many of the clinics registered as single-physician units had more than one physician employed. Due to design and confidentiality of individual physicians we cannot ensure a one-to-one link of physicians in the first and

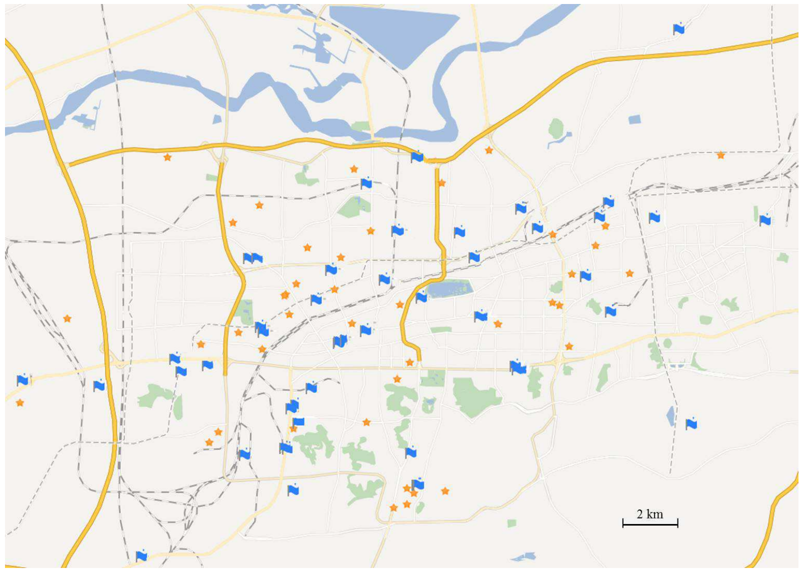


Figure 1: Map of locations of sampled 96 clinics
Stars: the control group; Flags: the treatment group.

second audit. However, based on the data from the first audit we may describe the baseline balance between the treatment and control group at both physician and clinic level. We collected data on gender and observed age of practicing physicians and the location and size (number of additional patients in the waiting room, number of additional physicians and patients in physician's office) of the clinics. Table 3 and Table 4 show that based on the results from χ^2 and Mann-Whitney tests, there is no significantly systematic differences between the treatment and the control group at both physician and clinic level.

		Control		**Treatment**	
Variables		Frequency	N	Frequency	N
Gender	Male	24	48	23	48
	Female	24	48	25	48
χ^2 test p-value: 0.838					
Age	≤ 30	2	48	2	48
	[31,40]	24	48	26	48
	[41,50]	12	48	18	48
	≥ 51	10	48	2	48
χ^2 test p-value: 0.085					

Table 3: Check of randomization at physician level

We evaluate physicians' prescribing behavior using the rate of overall prescribing indicating if a physician has provided prescription of any medication. In addition, prescribing behavior in antibiotics and non-antibiotics are investigated separately as well. Table 5 summarizes the baseline information of physicians' prescribing behavior in the first audit. The large majority of the physicians wrote prescriptions to the patients in both control

Variables	Control Mean	Sd.	N	Treatment Mean	Sd.	N
Number of additional physicians in the office Mann-Whitney test p-value: 0.792	0.333	0.808	48	0.354	0.758	48
Number of additional patients in the waiting room Mann-Whitney test p-valuee: 0.182	0.250	0.636	48	0.375	0.672	48
Number of additional patients in physician's office Mann-Whitney test p-value: 0.865	0.979	1.436	48	0.938	1.359	48

Table 4: Check of randomization at clinic level

(93.8%) and treatment (87.5%) group. Around two thirds of the physicians over-prescribed antibiotics (62.5% in the control group and 66.7% in the treatment group) and even more of them provided non-antibiotic prescriptions to the patients, namely 87.5% and 85.4% in the control and treatment group respectively. The observations from the first audit clearly confirm the prevalence of over-prescribing of medication in China in the case of common cold where no medication, especially antibiotics, is recommended to be prescribed. We use χ^2 test of independence to test the null hypothesis that the assignment of intervention and the prescribing behavior are independent. The first audit data shows the independence and presented a well balanced baseline on prescribing in any medication ($p = 0.294$), antibiotics ($p = 0.670$) and non-antibiotics ($p = 0.765$).

	Control Mean	Sd.	Freq.	N	Treatment Mean	Sd.	Freq.	N	χ^2 test p-value
Overall	93.8%	0.245	45	48	87.5%	0.334	42	48	0.294
Antibiotics	62.5%	0.489	30	48	66.7%	0.476	32	48	0.670
Non-antibiotics	87.5%	0.334	42	48	85.4%	0.357	41	48	0.765

Table 5: Prescribing behavior in the first audit

Results

We present the effect of the mystery shopper scheme on physicians' prescribing behavior in the second audit[6] in Table 6 by comparing the control and the treatment group[7]. Similar as the findings from the first audit, the rate of overall prescribing is very high

[6]In the second audit, one clinic in the control group turned into a drug store, and one clinic in the treatment group was closed. We removed these two clinics from our sample and hence data from 94 clinics was used in our study of the second audit.

[7]We use a χ^2 to evaluate the differences. Since we have a random assignment of the treatment, the χ^2 test is both an efficient and robust test of the treatment effect. Although the difference-in-differences estimator provides robust estimators when treatment is potentially non-random assigned, the use of difference-in-differences estimation is not warranted in our situation, and would be inefficient (Bertrand et al., 2004).

in both the control and the treatment group. Whereas all the physicians in the control group provided some prescription to the pseudo patients, significantly fewer physicians in the treatment group, namely 89.4%, did that. The prescribing behavior in antibiotics and non-antibiotics are then investigated respectively. The two groups are not statistically significantly different in antibiotic prescribing behavior ($p = 0.286$), which means the mystery shopping scheme has little effect on mitigating antibiotic over-prescribing. However due to the intervention of mystery shopping scheme, the rate of prescribing in non-antibiotics significantly ($p = 0.025$) decreases by 14.8 percentage points in the treatment group, which is equivalent to a 15.46% reduction compared to the control group. In other words, those physicians who have the knowledge of a future mystery shopping audit are less likely to prescribe any medication, in particular non-antibiotics to the patients, and the odds of prescribing non-antibiotics is reduced by 81.2% due to the announcement of the mystery shopping audit.

	Control				**Treatment**				χ^2 test	Odds
	Mean	Sd.	Freq.	N	Mean	Sd.	Freq.	N	p-value	Ratio
Overall	100.0%	0.000	47	47	89.4%	0.312	42	47	0.022	–
Antibiotics	57.4%	0.500	27	47	68.1%	0.471	32	47	0.286	1.580
Non-antibiotics	95.7%	0.204	45	47	80.9%	0.398	38	47	0.025	0.188

Notes: Odds ratio for overall prescribing is not defined due to a 100% prescribing rate in the control group.

Table 6: Prescribing behavior in the second audit

Robustness check

To model the intervention effect on prescribing behavior and to check the robustness of the non-parametric test results, we fit logistic regressions on treatment for prescribing behavior in both antibiotics and non-antibiotics[8]. Table 7 presents the estimated odds ratios with standard errors in parentheses. Model 1 and 2 are simple logistic regressions. By definition, the estimated odds ratios are the same as the calculated ones from the cross tabulation presented in Table 6. While the different distribution assumptions result in different standard errors, the significance from logistic regression is consistent compared to the non-parametric tests. The observed characteristics of physicians, for example age and gender, might affect their prescribing behavior systematically. To control for this, model 3 and 4 fit the (conditional) fix-effects logistic regressions[9] for matched treatment-control

[8] The results of overall prescribing rate is omitted here due to a perfect prediction of prescribing behavior in the control group.

[9] For our study, the conditional fixed effects Logit model is preferred to unconditional fixed effects logit model. This is because in the unconditional model the structural estimator has a larger amount of bias than in the conditional one and the bias amplify as the number of observations in each group is getting smaller (Neyman and Scott, 1948; Katz, 2001).

groups (Chamberlain, 1980). The groups are matched according to physicians' gender (male or female) and age (young or old), and a summary of the matched groups is presented in Table 8. In additional to the fixed effects, we allow for the heterogeneous treatment effects across the groups. Therefore, the standard errors are adjusted for clustering on group level in model 3 and 4 (Abadie et al., 2017). Conditional on physicians' gender and age, the average treatment effect on antibiotic prescribing is not significantly different from zero, but the intervention on average significantly reduces the odds of prescribing non-antibiotic by 82.8% compared to the control group. The estimates are consistent across models in terms of mean, standard errors and the significance.

	Logistic regression		**Fixed-effects logistic regression**	
	Antibiotics	**Non-antibiotics**	**Antibiotics**	**Non-antibiotics**
	Model 1	Model 2	Model 3	Model 4
Intervention	1.580	0.188^{**}	1.927	0.172^{**}
	(0.272)	(4.328)	(0.224)	(4.415)
Number of observations	94	94	94	94
Log-Likelihood	-61.49	-31.22	-52.85	-25.37
Pseudo R^2	0.00919	0.0797	0.0198	0.105

Estimated odds ratios are presented with standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Robustness check

	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 group grouped as "Old", and those younger than 40 are grouped as "Young".

Table 8: Summary of matched groups

5. Conclusion and discussion

We conduct a randomized field experiment to assess the impact of a mystery shopper scheme on prescribing behavior in primary care in China. We find that a mystery shopper scheme reduces overall unnecessary prescribing. For the case of non-antibiotic medication, the mystery shopper scheme reduces the frequency of inappropriate prescribing by 15.46%. For the case of antibiotic medication, the intervention does not cause a significant reduction in inappropriate prescribing. This paper provides new evidence suggesting that auditing primary care providers can have a direct effect on physician behavior even when it is not combined with pay-for-performance, or other measures such as reminders, feedback or educational interventions.

There are several underlying mechanisms which might potentially explain our result of no significant intervention effect on antibiotic prescribing behavior. First, the intervention message did not provide any tangible assessment criteria on quality of primary care. Hence, the safety aspect of antibiotic prescribing might not be the top priority for small private clinics in China. On the contrary, providing a good service by prescribing medications to satisfy the patients' expectations might be one of the quality aspects that is considered more important to clinics for attracting more patients in the long run. In health care systems where information asymmetry is an issue, patients' low awareness of antibiotic resistance and lack of knowledge on antibiotics misuse leads to high prevalence of self-medication (Grigoryan et al., 2007; Togoobaatar et al., 2010; Pan et al., 2012; Yu et al., 2014) and thus high expectation of receiving antibiotic prescription from the physicians (Reynolds and McKee, 2009; Jin et al., 2011). Second, our intervention only considered whether the knowledge of being audited by a relevant institution would change clinics' antibiotic prescribing behavior, but did not provide any direct incentives in the form of pay-for-behavior, or negative consequence or punishment for inappropriate prescribing. We postulate that the behavioral response to the intervention might have been stronger if the clinics considered the evaluation results of behavior to be influential for reputation among patients and hence demand of patients. Third, the education and thus knowledge of appropriate use of antibiotics might still be suboptimal even among primary care providers (Huang et al., 2013; Bai et al., 2016). If that is the case, the intervention might not be strong enough for some physicians to improve behavior.

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 treatment and control group. 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 base, we cannot rule out the possibility of information spillover about the intervention among individual physicians from different groups. Due to the characteristics of the intervention, however, we expect information spillover to have minor impact, if present at all. If information about the intervention reaches clinics in the control group, they would know that the mystery shopper audit is followed by an announcement. Hence, a reasonable strategy for a clinic in the control group is to not change behavior following a spillover of information from a clinic in the treatment group.

Our study investigated the intervention effect three weeks after the intervention. More studies are warranted if one is to access any long-term effect of the mystery shopper scheme.

Acknowledgments

We are grateful to Yiu-Shing Lau, Eline Aas, Tron Anders Moger, Daniel Wiesen and Matteo Galizzi for their valuable advice. We also 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. Financial support by the Research Council of Norway (IRECOHEX, Project-No. 231776) is gratefully acknowledged. Jian Wang acknowledges the National Natural Science Foundation of China (grant no. 71373146).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–969.

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

Bai, Y., Wang, S., Yin, X., Bai, J., Gong, Y., and Lu, Z. (2016). Factors associated with doctors' knowledge on antibiotic use in china. *Scientific reports*, 6.

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. *Quality in Health Care*, 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. *Br J Gen Pract*, 53(489):284–291.

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

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, pages 249–275.

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.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238.

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.

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.

Dranove, D. (2011). Health care markets, regulators, and certifiers. In Pauly, M. V., Mcguire, T. G., and Barros, P. P., editors, *Handbook of Health Economics*, volume 2 of *Handbook of Health Economics*, pages 639 – 690. Elsevier.

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.

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.

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*, pages 194–199.

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, Part B:36 – 46.

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.

Huang, Y., Gu, J., Zhang, M., Ren, Z., Yang, W., Chen, Y., Fu, Y., Chen, X., Cals, J. W., and Zhang, F. (2013). Knowledge, attitude and practice of antibiotics: a questionnaire study among 2500 chinese students. *BMC medical education*, 13(1):163.

Jamtvedt, G., Young, J., Kristoffersen, D., O'Brien, M., and Oxman, A. (2006). Audit and feedback: effects on professional practice and health care outcomes. cochrane database of systematic reviews 2006, issue 2. *Cochrane Database of Systematic Reviews*, (2).

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.

Katz, E. (2001). Bias in conditional and unconditional fixed effects logit estimation. *Political Analysis*, 9(4):379–384.

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.

Kiefe, C., Allison, J., Williams, O., Person SD, M. W., and Weissman, N. (2001). Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA*, 285:2871–9.

Madden, J. M., Quick, J. D., Ross-Degnan, D., and Kafle, K. K. (1997). Undercover careseekers: simulated clients in the study of health provider behavior in developing countries. *Social Science & Medicine*, 45(10):1465–1482.

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).

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

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

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.

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

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.

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. *Milbank Quarterly*, 90(4):706–724.

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.

Wells, K. B., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unützer, J., Miranda, J., Carney, M. F., and Rubenstein, L. V. (2000). Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *Jama*, 283(2):212–220.

White, P., Atherton, A., Hewett, G., and Howells, K. (1995). Using information from asthma patients: a trial of information feedback in primary care. *BMJ*, 311(7012):1065–1069.

Wilson, A. M. (1998). The role of mystery shopping in the measurement of service performance. *Managing Service Quality: An International Journal*, 8(6):414–420.

World Health Organization (2012). The pursuit of responsible use of medicines: sharing and learning from country experiences.

Yip, W. and Hsiao, W. C. (2008). The chinese health system at a crossroads. *Health Affairs*, 27(2):460–468.

Yip, W. C.-M., Hsiao, W. C., Chen, W., Hu, S., Ma, J., and Maynard, A. (2012). Early appraisal of china's huge and complex health-care reforms. *The Lancet*, 379(9818):833–842.

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.

Appendix A. Script of pseudo patient 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.
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.

Appendix 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

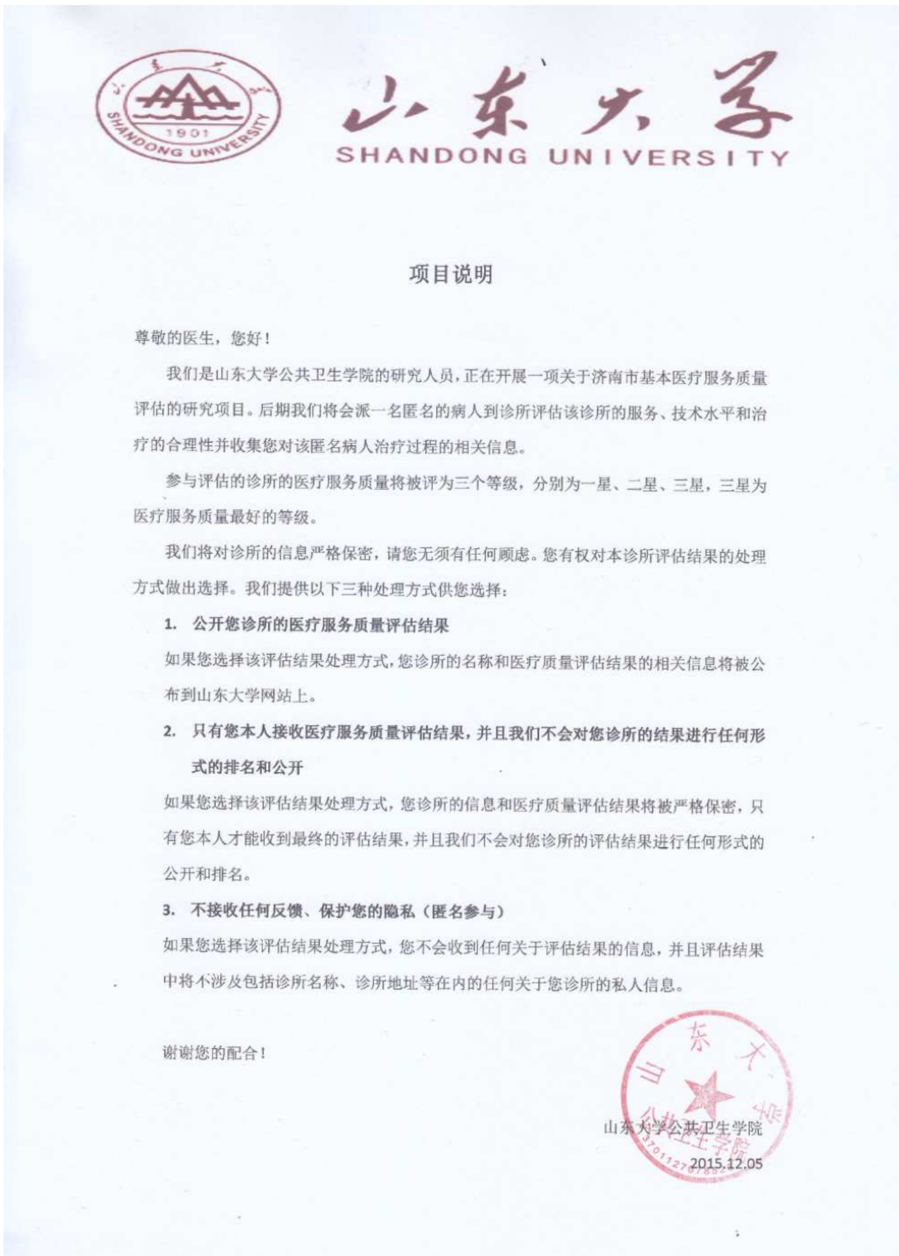
1. Observe the number of additional patients in the waiting room.
2. Observe the number of additional physicians and patients in the office, the gender and age of the practicing physician.
3. Memorize the name(s) of the medication and the pharmaceutical company.

Out of the Clinic

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

Appendix C. Letters used in intervention

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



山东大学
SHANDONG UNIVERSITY

项目说明

尊敬的医生，您好！

我们是山东大学公共卫生学院的研究人员，正在开展一项关于济南市基本医疗服务质量评估的研究项目。后期我们将会派一名匿名的病人到诊所评估该诊所的服务、技术水平和治疗的合理性并收集您对该匿名病人治疗过程的相关信息。

参与评估的诊所的医疗服务质量将被评为三个等级，分别为一星、二星、三星，三星为医疗服务质量最好的等级。

我们将对诊所的信息严格保密，请您无须有任何顾虑。您有权对本诊所评估结果的处理方式做出选择。我们提供以下三种处理方式供您选择：

1. **公开您诊所的医疗服务质量评估结果**

如果您选择该评估结果处理方式，您诊所的名称和医疗质量评估结果的相关信息将被公布到山东大学网站上。

2. **只有您本人接收医疗服务质量评估结果，并且我们不会对您诊所的结果进行任何形式的排名和公开**

如果您选择该评估结果处理方式，您诊所的信息和医疗质量评估结果将被严格保密，只有您本人才能收到最终的评估结果，并且我们不会对您诊所的评估结果进行任何形式的公开和排名。

3. **不接收任何反馈、保护您的隐私（匿名参与）**

如果您选择该评估结果处理方式，您不会收到任何关于评估结果的信息，并且评估结果中将不涉及包括诊所名称、诊所地址等在内的任何关于您诊所的私人信息。

谢谢您的配合！

山东大学公共卫生学院

2015.12.05

Figure C.1: Original project description issued by School of Public Health, Shandong University (in Chinese)

山东大学
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)

Figure C.2: English translation of the project description issued by School of Public Health, Shandong University

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

山东大学
SHANDONG UNIVERSITY

介绍信

山东大学齐鲁医学部支持山东大学公共卫生学院开展“济南市基本医疗服务评估”研究项目。

山东大学齐鲁医学部
2015. 12. 4

Figure C.3: Original endorsement letter issued by Qilu Health Service Center, Shandong University (in Chinese)

山东大学
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)

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