China Economic Review 80 (2023) 102015

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

Health-seeking behavior and patient welfare: Evidence from China

Yang Li [a,*], Zhuo Chen [b,c]

[a] School of Economics, Faculty of Humanities and Social Sciences, The University of Nottingham Ningbo China, Room 310-2 IEB, 199 Taikang East Road, 315100 Ningbo, China
[b] University of Georgia, Department of Health Policy and Management, Athens, GA, USA
[c] School of Economics, University of Nottingham Ningbo China, Portland Building, 199 Taikang East Road, 315100 Ningbo, China

ARTICLE INFO

JEL:
I11
I13
I15
I31

Keywords:
Health-seeking behavior
Three-tier healthcare system
Welfare analysis
Random coefficient logit model
Hierarchical medical system

ABSTRACT

The inefficient use of healthcare resources is a persisting challenge to almost all healthcare systems, making it imperative to understand the underlying factors of healthcare demand. This paper investigates patients' health-seeking behavior in rural China using a random coefficients logit model. We further perform a counterfactual simulation and welfare analyses to evaluate the inefficiencies in health services utilization. The counterfactual simulation reallocates patients to more efficient health providers following the principles of the hierarchical medical system. Our analysis suggests that out-of-pocket expenses and distance to providers discourage patients from utilizing healthcare, while quality of care has a positive effect on patients' hospital choices. However, significant heterogeneity exists in patient preferences over quality of care, out-of-pocket expenses, and distance to providers. The simulation results show that the overall welfare change may have masked variations related to the hypothetical change, with societal welfare loss from switching to higher-tier providers. Our analysis provides support for policies to improve hospitals and assist rural patients in financing healthcare in China.

1. Introduction

The inefficient use of healthcare resources is a long-standing challenge to healthcare systems globally and particularly in China, making it imperative to understand patients' health-seeking behavior. The unbalanced allocation of medical resources and the inefficient use of primary care facilities remain major concerns in China (Qian, Pong, Yin, Nagarajan, & Meng, 2009; Yip et al., 2019; Zhou et al., 2021), especially in rural areas. As patients may lack trust in the quality of care provided at primary care facilities, they are increasingly bypassing primary care facilities and seeking care at higher-tier hospitals in China, regardless of the severity of their conditions (Li & Fu, 2017; Liu, Kong, Yuan, & van de Klundert, 2018; Yip et al., 2012). This mismatch leads to the underuse of primary care and the overburdening of tertiary hospitals in China. On the other hand, unlikely urban residents, patients in rural China with limited resources, even those with severe symptoms, may seek treatment from lower-tier providers because of the low out-of-pocket costs in lower-tier hospitals or the proximity of these facilities. In poor and developing areas, household income and medical expenditure have stronger effects on patients' hospital choice than the severity of their illness or hospital quality (Morey, Sharma, & Mills, 2003; Qian et al., 2009). While we use data from China for this research, the implications of health-seeking behavior and welfare analysis of healthcare utilization efficiency apply to healthcare systems elsewhere, including those that need to strengthen primary care as "gatekeepers" and those serving many residents living in rural areas.

* Corresponding author.
E-mail address: yang.li@nottingham.edu.cn (Y. Li).

https://doi.org/10.1016/j.chieco.2023.102015
Received 25 July 2022; Received in revised form 22 May 2023; Accepted 22 June 2023
Available online 24 June 2023
1043-951X/

The primary objective of this study is to investigate the determinants of healthcare demand in rural China. Using data from the China Health and Nutrition Survey (CHNS), we adopt a choice model to explore how rural residents' healthcare demands are affected by out-of-pocket (OOP) expenditures, distance, and healthcare quality. OOP expenditure is an important factor in the demand for healthcare (Ho & Pakes, 2014; Qian et al., 2009; Sahn, Younger, & Genicot, 2003). Second, it is well documented in the literature that travel distance would reduce the probability that the hospital is to be visited by patients (Ho & Pakes, 2014; Pope, 2009; Sivey, 2012). Finally, quality of care is a critical factor determining the demand for health (Grossman, 1972), with evidence of a positive effect on the demand for hospital services (Gaynor, Propper, & Seiler, 2016; Gutacker, Siciliani, Moscelli, & Gravelle, 2016; Hanson, Yip, & Hsiao, 2004).

Besides exploring the determinants of healthcare demand, this study evaluates the efficiency of healthcare utilization in rural China. China has a three-tier healthcare delivery system, consisting of primary care facilities, secondary providers, and tertiary providers. Typically, the scope of care is the narrowest and prices the lowest at primary care facilities and highest at tertiary providers. To inform the current policy debate on the efficient allocation of patients to providers (Liu, Liu, Twum, & Li, 2016; Yip et al., 2019), we perform a counterfactual analysis that is in line with the principles of the hierarchical medical system (HMS)[1]: patients are matched with providers (hospitals) based on the severity of their illness (low, moderate, or high) and type of visits they had (outpatient vs. inpatient).[2] Specifically, patients with minor illnesses or requiring outpatient visits are assigned to primary care facilities, whereas the most severely ill patients requiring inpatient treatment are assigned to tertiary providers. This counterfactual analysis helps to show how observed healthcare choices in China affect patient welfare. By exploring the healthcare utilization of residents in rural China, this study identifies potential healthcare reform measures that might improve well-being among rural residents.

The analysis demonstrates that OOP costs and distance have negative marginal effects on healthcare decisions. There is significant heterogeneity in preferences across patients in rural China. An analysis that accounts for some interactions with demographics shows that disease severity is the primary driver of heterogeneity. Beneficiaries with more severe conditions value quality of care more and are less sensitive to the OOP costs. The counterfactual results suggest that consumer surplus decreases by an average of 307 RMB (about US$48) if patients are assigned to a health provider that is most appropriate for the severity of their condition.[3] However, variations exist across groups: patients with less severe illnesses or requiring only outpatient care may benefit from a switch from city hospitals to primary healthcare providers as gatekeepers, whereas patients with severe symptoms suffer welfare losses if required to visit higher-level hospitals because the improved quality comes at a higher price. To make this policy welfare-neutral, this group would need to be compensated by a reduction in OOP costs of 988 RMB (US$153) to incentivize them to visit better hospitals.

The research contributes to several strands of literature. First, it adds to the literature on healthcare demand in China (Audibert, He, & Mathonnat, 2013; Brown & Theoharides, 2009; Qian et al., 2009; Yip, 1998). Few studies in the context of China have accounted for the role of healthcare provider quality. Hospitals providing a better quality of care tend to face increased demand and are more likely to charge higher prices. To address the potential endogeneity, quality of care is taken into account when performing this analysis. Typically, studies of hospital demand have chosen mortality, readmission rate, or hospital ranking to measure quality of care (Gaynor et al., 2016; Gutacker et al., 2016; Hanson et al., 2004; McConnell, Lindrooth, Wholey, Maddox, & Bloom, 2016). However, healthcare providers in poor areas (e.g., rural China) tend to have an inadequate supply of medical professionals; therefore, alternative measures of healthcare quality are more appropriate. For example, Sahn et al. (2003) measured quality of care in rural Tanzania based on the clinic environment and availability of medical staff. Following the quality evaluation system recommended by National Health Commission, we chose doctor-population and doctor-bed ratios to measure healthcare service quality.

This paper also contributes to the literature on the efficiency of healthcare utilization in China. Most of the existing literature focuses on the underutilization of primary care facilities (Liu et al., 2018; Yip et al., 2012; Zhou et al., 2021), whereas few studies have discussed the overutilization of primary care in rural China, particularly among low-income residents. This study attempts to fill the gap in the literature by exploring the healthcare utilization of rural residents in both directions. To our best knowledge, it is the first paper that uses counterfactual analysis to assess the implications of observed healthcare choices in China on welfare analysis. Additionally, this paper expands the empirical methodology for investigating health-seeking behavior in rural China. We adopt a random coefficient logit (RCL) model to allow for the unobserved heterogeneity in patients' preferences about healthcare costs and quality, which has rarely been discussed in the literature on hospital choices in China. The model allows a better evaluation of patient sensitivity to price or quality differences than the traditional logit model. However, concerns remain that important dimensions of heterogeneity are unobserved, including variations in patient knowledge about hospital quality based on personal experiences and unobserved differences in access to transportation.

Finally, this work adds to the literature on evaluating the New Cooperative Medical Scheme (NCMS), a heavily subsidized public insurance program for residents in rural China. Empirical studies have investigated the impacts of NCMS coverage on hospital choice and healthcare utilization but without consensus. Some research (Brown & Theoharides, 2009; Qian et al., 2009; Wagstaff, Lindelow, Jun, Ling, & Juncheng, 2009) found that the NCMS increases the probability of seeking care from better healthcare providers, whereas others suggested that it has little impact on the demand for healthcare providers (Audibert et al., 2013) and utilization of medical

[1] Hierarchical Medical System (HMS) was lauched in 2015. This paper uses the guidance of HMS policy to examine the effectiveness of health allocation.

[2] Our measure of severity is based on patient's self-reported severity, and we provide more details about the rationale of using the measure in Appendix A3.

[3] The exchange rate in 2011 between the US$ and Chinese RMB in this paper is set to US$100 = 646 RMB. Source: National Bureau of Statistics of China, http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A060J&sj=2017

services (Lei & Lin, 2009). Consistent with the latter finding, this study suggests that the NCMS has an insignificant overall effect on the choice of provider type. Besides analyzing the impact of NCMS on patients' health utilization, this study also contributes to the literature by providing insights on how to better design insurance policies to incentivize poor rural patients with severe conditions to seek healthcare from better providers. The counterfactual analysis recommends the design of an income- and diagnosis-specific benefit package in the NCMS, with the insurance coverage increasing from 10 to 70% for low-income rural residents.

The remainder of this paper is organized as follows. Section 2 illustrates the institutional background, with a focus on the three-tier healthcare system in rural China. Section 3 describes the data and descriptive statistics. Section 4 presents the model and the econometric framework. Section 5 reports the results and implications. Section 6 conducts a counterfactual simulation. Section 7 reports the results of robustness tests, and section 8 concludes.

2. Institutional background

2.1. Healthcare providers in rural China

Rural China uses a three-tier healthcare system to deliver care. The tiers, classified from lowest to highest based on the breadth of services they provide, are village clinics and township health centers (primary care facilities), county hospitals (secondary providers), and city hospitals (tertiary providers) (Wang, Wang, Wang, Zheng, & Xiao, 2014). Generally, primary care facilities in rural China provide fewer specialized services than secondary and tertiary hospitals but charge the lowest price. City hospitals are equipped with better facilities, nurses, and doctors than those in rural areas. Typically, they offer more extensive but also more expensive services. Although residents in rural China have experienced rapid increases in both income and insurance coverage, their preference for health providers tends to remain almost constant over time (Appendix A1), a pattern requiring further exploration.

In China, the central government has overall responsibility for health legislation, policy, and price schedule guidelines. Regulated by local health authorities and the Bureau of Commodity Prices, fee schedules for healthcare are set based on the medical services and provider types. Similar health services can be priced differently across different types of health facilities and usually charge higher prices for health providers with better quality (high-level providers).

To improve efficiency in healthcare delivery, the Chinese government launched the hierarchical medical system (HMS), in which different levels of hospitals are responsible for undertaking different health services.[4] Our research adopts the guidance of HMS policy to conduct the counterfactual analysis to assess the implications of observed healthcare choices on welfare in rural China.

2.2. New cooperative medical scheme (NCMS)

China's Rural Cooperative Medical System (RCMS) collapsed alongside the diminishing of communal farming in the early 1980s, leaving most rural residents without insurance coverage (Liu, 2004). As a result, 96% of rural households lacked health insurance by 2002 (Hsiao, 2005).

To improve access to healthcare and reduce the risk of illness-led poverty in rural areas, the Chinese government implemented a public insurance program for rural residents - the New Cooperative Medical Scheme (NCMS), in 2003. The NCMS enrollment rate reached 97.5% (832 million enrollees) in 2011, making it the primary insurance in rural China and the largest healthcare insurance program globally. The program is co-financed by individuals, local governments, and the central government, with county as the risk-pooling unit. Each county is allowed to design and implement its benefit packages. Hence, there is considerable variation in the cost-sharing arrangements across counties. Even within a county, variations exist in cost-sharing between outpatients and inpatients among different healthcare facility types (some evidence on cost-sharing policy is available in Appendix A2).

3. Data

This study uses high-quality panel data from the CHNS. As shown in Fig. 1, the survey provides data collected from nine provinces, which vary in economic development, public policy, and healthcare indicators. Four counties are randomly selected from each province based on income level to represent China's rural areas. The sample contains 55,245 records (person-year) from 1989 to 2011, covering approximately 6200 rural residents.[5]

Besides providing demographic information, the data provide details on all reported healthcare use during the previous month, including the patient-selected healthcare facility, its distance from the patient, its business hours, OOP expenditures, and insurance coverage. The question most relevant to this analysis relates to medical service utilization in the past four weeks. In the CHNS survey, individuals were asked whether they had been sick in the past four weeks and, if they had been, whether they had sought any healthcare services before taking the survey. The distance was reported by community healthcare workers and represented the average kilometers from the centroid of the patient's residence area to the relevant healthcare facility. Information on hospital attributes, such as the number of doctors and beds, is collected as quality measures.

Our measures of hospital quality are doctors per capita for primary care facilities and doctors per bed for high-tier hospitals (more

[4] Chinese government issued the “Guiding Opinions of the General Office of the State Council on Pushing Forward the Developing of the Hierarchical Medical System (HMS)” in September 2015.

[5] The CHNS data include nine waves, including year 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009, and 2011, respectively.

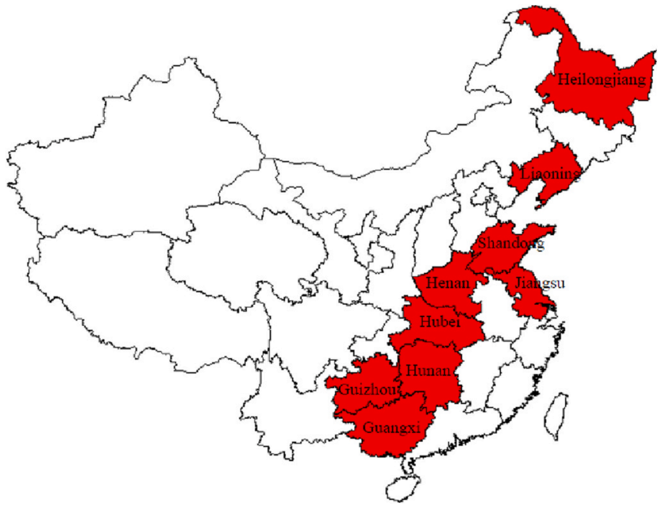

Fig. 1. Survey regions of the China Health and Nutrition Survey (CHNS). (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

details of our quality measures are available in Appendix A4). For primary care facilities (village clinics and township health centers), the doctor–population ratio is more appropriate for measuring quality than doctors per bed. These primary care facilities generally have few beds; therefore, the high doctor–bed ratio cannot be interpreted as better quality of care. In contrast, the doctor–population ratio is inappropriate for high-tier hospitals (county and city hospitals) because they serve many villages. The doctor–bed ratio is a better measure of hospital quality for these high-tier providers. To compare quality across different providers, we use quartiles to represent quality level, with higher quartiles representing better care.

Table 1 reports the summary statistics by healthcare facility type. Older patients tend to be sicker and are more likely to choose county and city hospitals because they require more sophisticated treatment from major providers (a similar pattern is observed for severity). Better-educated and high-income patients tend to seek healthcare at county or city hospitals with better quality of care. County and city hospitals often have longer business hours and more doctors and beds than village clinics but tend to be farther away from respondents' residences.

The preferences of rural patients may be heterogeneous due to their demographic characteristics, health conditions, and NCMS coverage. In addition, unobserved heterogeneity in the preferences of rural residents can potentially exist in healthcare provider choice. First, there is variation in the patients' knowledge and perception of the quality of different healthcare providers due to their experience. Second, patients' social networks vary. Third, access to specific types of transportation differs across patients. If public transport is less frequent, patients will be more likely to choose nearby hospitals for treatment. Fourth, there is subjectivity in the self-evaluation of health conditions. Patients vary in how they perceive and interpret symptoms even when they suffer from the same illness of the same severity, which can affect their healthcare provider choice. The CHNS did not include questions that could have been used as proxies for such variation. Therefore, we attempt to account for this in our modeling.

4. A model of hospital choice

4.1. Random utility framework

We model the demand for the four providers (village clinics, township health centers, county hospitals, and city hospitals) as a discrete choice derived from a random utility model. Patient i will choose the alternative that provides the highest utility when facing a choice among $J+1$ alternatives ($J=4$ alternatives plus the option of not choosing any of them). The utility can be specified as the sum of both observed and unobserved factors:

$$u_{ijt} = X_{jt}\beta - \alpha p_{ijt} + \gamma Z_{ijt} + \varepsilon_{ijt} \tag{1}$$

In the random utility model, u_{ijt} is the utility for individual i choosing alternative j at time t. p_{ijt} is the OOP cost if individual i chooses hospital type j. X_{jt} is a vector of provider attributes, including quality of care and business hours. Z_{ijt} is the average travel distance to alternative j. ε_{ijt} represents a preference shock that is unobservable to both consumers and econometricians.

As shown in Table 1, the preferences of rural patients are heterogeneous according to the observed demographic characteristics, health conditions, and NCMS coverage. However, other unobserved factors could affect the choices observed in the data (knowledge and perception of quality, for example). Allowing for such unobserved heterogeneity in the patients' preferences is important and can be done using specific functional form assumptions on the distribution of ε and the random component of preferences. The RCL demand model can approximate any random utility model (McFadden & Train, 2000) and provide more informative substitution

Table 1
Summary statistics by health provider type.

	Village clinic	Township health center	County hospital	City hospital
Demographics				
Sex (% women)	0.544	0.509	0.533	0.478
	(0.498)	(0.500)	(0.499)	(0.500)
Age (years)	49.05	48.80	52.18	53.20
	(15.22)	(15.37)	(16.27)	(16.17)
Income (RMB)	6296.9	6051.9	8735.8	10,560.7
	(11,419.6)	(10,895.2)	(11,309.4)	(18,446.4)
Education (level)				
1: Illiterate	0.394	0.379	0.361	0.338
	(0.489)	(0.486)	(0.481)	(0.474)
2: Primary school	0.263	0.270	0.239	0.216
	(0.440)	(0.444)	(0.427)	(0.412)
3: Middle school	0.253	0.244	0.248	0.270
	(0.435)	(0.430)	(0.432)	(0.444)
4: High school	0.0724	0.0891	0.0860	0.124
	(0.259)	(0.285)	(0.281)	(0.330)
5: College or above	0.0185	0.0168	0.0661	0.0520
	(0.135)	(0.128)	(0.249)	(0.222)
Severity level				
1: Low level	0.451	0.347	0.272	0.226
	(0.498)	(0.476)	(0.445)	(0.419)
2: Moderate level	0.463	0.519	0.510	0.489
	(0.499)	(0.500)	(0.500)	(0.500)
3: High level	0.0860	0.134	0.219	0.285
	(0.280)	(0.341)	(0.414)	(0.452)
NCMS (% coverage)	0.500	0.514	0.328	0.312
	(0.500)	(0.500)	(0.470)	(0.464)
Inpatient (%)	0.009	0.068	0.151	0.212
	(0.094)	(0.251)	(0.359)	(0.409)
Hospitalized days	0.101	0.549	1.724	2.601
	(1.305)	(2.399)	(4.988)	(6.141)

	Village clinic	Township health center	County hospital	City hospital
Hospital characteristics				
Distance (km)	0.543	2.101	8.707	10.10
	(3.507)	(9.478)	(11.30)	(20.64)
Business hours (per week)	95.50	129.7	138.8	134.3
	(43.77)	(50.17)	(44.34)	(42.76)
Number of doctors	2.755	29.30	195.1	391.7
	(4.594)	(30.45)	(188.1)	(366.6)
Number of beds	3.274	33.39	286.6	510.1
	(7.179)	(35.22)	(239.2)	(379.8)

Notes: Standard errors are in parentheses.
Source: China Health and Nutrition Survey (1989–2011).

patterns. The RCL model has been widely adopted in studies of healthcare markets (Bundorf, Levin, & Mahoney, 2012; Dubois & Lasio, 2018; Dunn, 2012; Einav, Finkelstein, & Cullen, 2010).

Because healthcare costs can account for a significant proportion of income for rural residents, it is unreasonable to derive indirect utility from a quasilinear utility function free of wealth effects. To account for such effects, we use the logarithm of the difference between income and healthcare OOP costs. The RCL model is:

$$u_{ijt} = X_{jt}\beta_i + \alpha_i ln\left(y_{it} - p_{ijt}\right) + \gamma_i Z_{ijt} + \varepsilon_{ijt} \tag{2}$$

where y_{it} represents the income of individual i, and variables X_{jt}, p_{ijt}, and Z_{ijt} are the same as in Eq. (1). The model allows for preferences over the provider characteristics and the price to vary in the population, as captured by the random coefficients (α_i, β_i, and γ_i can vary across individuals).

We model the random coefficients as (more details about the matrix form in Appendix B):

$$\alpha_i = \overline{\alpha} + \sum_{r=1}^{d} \pi_{1r} D_{ir} + \sigma_1 v_{i1} \tag{3}$$

$$\beta_i = \overline{\beta} + \sum_{r=1}^{d} \pi_{2r} D_{ir} + \sigma_2 v_{i2} \tag{4}$$

$$\gamma_i = \overline{\gamma} + \sum_{r=1}^{d} \pi_{3r} D_{ir} + \sigma_3 \upsilon_{i3} \tag{5}$$

We model the random coefficients as in a matrix form:

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{pmatrix} = \begin{pmatrix} \overline{\alpha} \\ \overline{\beta} \\ \overline{\gamma} \end{pmatrix} + \Pi D_i + \sigma v_i \tag{6}$$

where $\Pi = \begin{pmatrix} \pi_{11} & \cdots & \pi_{1d} \\ \vdots & \ddots & \vdots \\ \pi_{31} & \cdots & \pi_{3d} \end{pmatrix}$ is a 3 $\times d$ matrix of parameters for observed demographic variables, $D_i = (D_{i1}, \ldots, D_{id})^{'}$ is a $d \times$ 1vector of observed demographic variables, including age, education, and insurance coverage. $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ is a row of $3 + 1$ parameters for unobserved consumer attributes, and $v_i = (v_{i1}, \upsilon_{i2}, v_{i3})^{'}$ is a vector of $3 + 1$ unobserved consumer attributes, which are assumed to follow the normal distribution.

If we assume that preferences are instead homogeneous in the population ($\alpha_i = \alpha$, $\beta_i = \beta$ and $\gamma_i = \gamma$ for all i), then the choice model in (2) is estimated using a simple logit specification. Despite its simplicity, the IID assumption of the logit model leads to restricted substitution patterns. However, it provides a good comparison for illustrating the importance of unobserved preference heterogeneity across rural residents. We estimate the model parameters using maximum simulated likelihood estimation (more details in Appendix B).

4.2. Estimation

Estimating the model requires knowledge of all the OOP costs charged by all providers. In contrast, the survey data only record the costs of the providers that patients visited. Hence, the prices of alternative providers that patients did not visit need to be predicted. We use the log-linear regression specification to impute the relevant prices for the non-selected choices of users.

Previous studies dealing with a similar issue (Hanson et al., 2004; Qian et al., 2009) also adopt this imputation method:

1. for each hospital type j, we categorize patients into two groups: group 1 are those who actually chose hospital j while group 2 are respondents who used health care services other than option j
2. for group 1, we regress the observed log OOP costs on a vector of the observed variables x_{1jt} that determine the OOP cost of rural residents,

$$ln(p_{1jt}) = \alpha\, x_{1jt} + \varepsilon$$

3. the estimated parameters $\widehat{\alpha}$ are used to predict the corresponding missing logarithm of costs for group 2,

$$\widehat{ln(p_{2jt})} = \widehat{\alpha}\, x_{2jt}$$

The fees for health care in China are set based on the types of both medical services and health providers. The imputation thus assumes that the OOP cost of health services in rural China is determined by patient demographics, insurance status, illness diagnosis, severity level, days in the hospital, and type of providers. Patients with the same demographics and medical service utilization should pay the same price in the same hospital type. Thus, the imputation analyses would avoid bias only if the model provides a good prediction of the OOP cost. This paper employs two indicators, the adjusted-R squared and kernel density for the log-OOP costs, to address the validity of the imputation.

5. Results on hospital choice

5.1. Price prediction

The first stage of the analysis consists of predicting the out-of-pocket costs for alternatives that were not chosen. Theoretically, the variation in copayment rate should also be included to predict the OOP. However, the survey does not provide the county name. It is thus impossible to match the cost-sharing information and CHNS data at the county level. Since copayment rates vary across hospital types among different counties and over time, we use the three-way interaction of the hospital type, county, and year to capture the variation in copayment rates.

Table 2 reports the results of the OOP prediction based on different specifications. The first two columns in Table 2 suggest that elderly, more educated individuals, as well as high-income patients, tend to have a higher OOP cost: this can be explained by the fact that they are more likely to seek health care at better providers with higher costs (county or city hospitals). The third column includes fixed effects of hospital type, time, and county. By including fixed effects, the coefficients of education and income have the opposite sign, indicating that more educated or more affluent patients, on average, pay less than their counterparts within the same health

Table 2
Predicted out-of-pocket costs for alternatives that were not chosen.

	(1)	(2)	(3)	(4)
Sex (Woman)	0.147*	0.089	−0.041	−0.135
	(0.084)	(0.083)	(0.080)	(0.098)
Education	0.025	0.010	−0.098**	−0.098*
	(0.040)	(0.039)	(0.041)	(0.052)
Age	0.077***	0.078***	0.080***	0.077***
	(0.017)	(0.017)	(0.016)	(0.022)
Age squared	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Income	0.035	0.008	−0.236***	−0.248**
	(0.062)	(0.061)	(0.068)	(0.112)
Income squared	−0.004	−0.002	0.014**	0.003
	(0.007)	(0.007)	(0.007)	(0.015)
NCMS	0.225***	0.278***	0.121	0.497**
	(0.082)	(0.081)	(0.116)	(0.205)
Severity level (*not severe as reference*)				
Somehow severe	0.592***	0.514***	0.437***	0.395***
	(0.085)	(0.084)	(0.080)	(0.101)
Quite severe	1.547***	1.387***	1.168***	0.984***
	(0.130)	(0.128)	(0.124)	(0.168)
Inpatient	1.615***	1.398***	1.055***	0.429
	(0.275)	(0.274)	(0.263)	(0.442)
Hospitalized days	0.068***	0.074***	0.057***	0.050*
	(0.021)	(0.021)	(0.020)	(0.030)
Constant	1.656***	1.667***	−0.677	3.454
	(0.449)	(0.454)	(0.705)	(3.916)
Symptom FE		Yes	Yes	Yes
Hospital FE			Yes	Yes
Year FE			Yes	Yes
County FE			Yes	Yes
Coverage rate (triple interaction)				Yes
Observations	2045	2045	2045	2045
R-squared	0.210	0.259	0.436	0.764
Adjusted R-squared	0.205	0.247	0.378	0.495

Note: Sex is a binary variable which equals to 1 if the person is a woman, and 0 otherwise. Education takes value from 0 to 6, representing from "none" to "master or higher", respectively. Inpatient is a binary variable which equals to 1 for inpatient service. NCMS is a binary variable which equals to 1 if individual is covered by NCMS. Standard errors are in parentheses under each coefficient, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

provider.

Unsurprisingly, the coefficients on severity, whether the patient is hospitalized (inpatient), and the length of stay (hospitalized days) are statistically significant and positive in all specifications. Patients generally need to pay more out-of-pocket if they have more severe conditions, require inpatient care, or stay longer in the hospital. More importantly, the adjusted-R^2 increases from 0.2 to 0.5, suggesting the variation in copayment rates (the three-way interaction) included in column 4 increases the explanatory power of the OOP costs prediction.

Fig. 2 plots the kernel density for the log OOP costs from the data (solid black line) and from the predictions (dashed lines, columns 1–4). Predictions from column 4 (dark blue dashed line) give the best fit among all specifications as the kernel density distribution of the observed OOP cost is most close to the density estimated from regression of column 4 (more discussion about the good fit of the price prediction in Appendix C).

5.2. Hospital choice

Table 3 shows the estimated choice of a healthcare provider based on two demand specifications—the logit model and the RCL model. The first column in Table 3 displays the estimated coefficients. In both models, the coefficients of the business hours and distance are of the expected signs and are statistically significant. The number of business hours has a positive marginal utility for the average beneficiary, whereas distance has a negative marginal utility. The primary difference between the two models is the magnitude of the OOP costs and quality of care; both are small and insignificant in the logit model. The lack of statistical significance may be caused by unobserved heterogeneities in hospital quality, as described in Section 3.

In the RCL model, the coefficients of net income (after deducting the OOP costs) and quality are positive and significant, suggesting that low OOP costs and high quality of care yield positive marginal utility for rural residents.

Choosing a provider based on higher quality and lower expenditures is well documented. However, there is a dearth of discussions on the heterogeneity effects across groups. The second column in Table 3 provides the estimated standard deviations for the preference distribution of hospital attributes from the RCL model, illustrating the variance in those preferences in the population. The standard

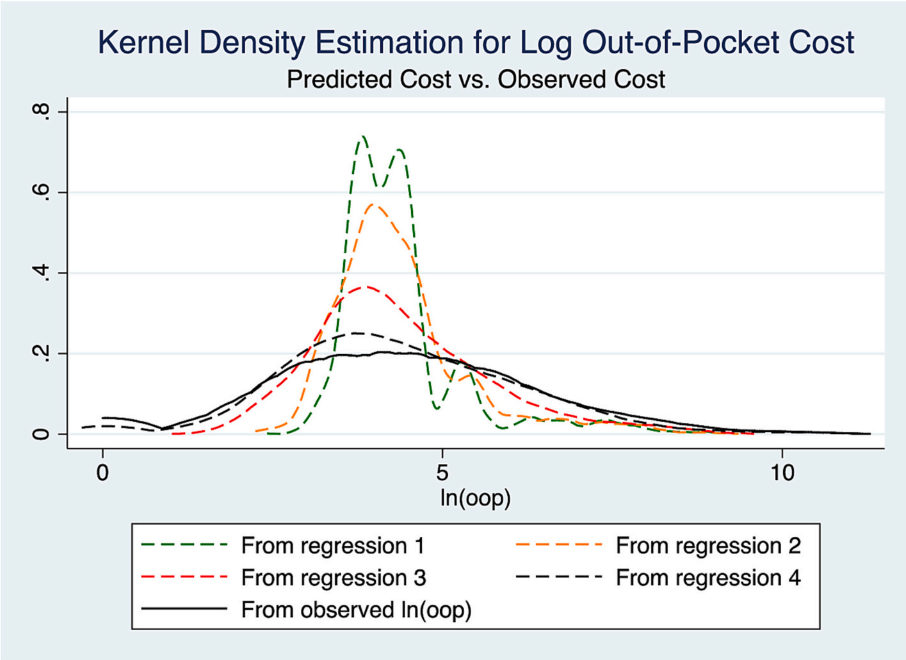

Fig. 2. Kernel density estimation for log out-of-pocket cost.
Note: Fig. 2 shows the kernel density for the log out-of-pocket costs from the data (solid black line) and from the predictions (dashed lines, columns 1–4). (for interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 3
Estimated demand from logit and random coefficient logit model.

Parameters	Variables	Logit demand	RCL demand
		Coefficient Estimation	
Mean ($\overline{\beta}$)	Quality proxy	0.021	0.129***
		(0.056)	(0.040)
Mean ($\overline{\alpha}$)	$ln(y-p)$	0.016	0.117***
		(0.020)	(0.040)
Mean ($\overline{\gamma}$)	Distance	−0.019***	−0.123***
		(0.007)	(0.020)
	Business hours	0.002**	0.012***
		(0.001)	(0.001)
		Standard Deviations	
(σ_β)	Quality proxy	–	0.410***
			(0.070)
(σ_α)	$ln(y-p)$	–	0.171***
			(0.060)
(σ_γ)	Distance	–	0.111***
			(0.020)
MRS_{qd}	MRS distance for quality	1.073	1.048***
		(2.981)	(0.370)
MRS_{pd}	MRS distance for price	−1.967	−2.210***
	(1000 RMB)	(2.411)	(0.842)
Observations	N	10,225	10,225

Notes: Cluster-robust standard errors are in parentheses under each coefficient and are clustered at the community level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Both models include time fixed effects and community fixed effects. Quality proxy refers to doctors per capita for primary care facilities and doctors per bed for high-tier hospitals (more details are available in Appendix A4). The panel of Coefficient Estimation reports the coefficients mean from logit model and RCL model, respectively. The panel of Standard Deviations reports the estimates of the standard errors of random coefficients. MRS_{qd} is the marginal rate of substitution between provider quality proxy and the distance. MRS_{pd} is the marginal rate of substitution between price and the distance.

deviations of the distribution of marginal utilities for distance and quality are significant, both economically and statistically. Thus, significant variation exists in patient preferences for distance and hospital quality. There is also significant heterogeneity in patients' preferences from price.

We also report the marginal rate of substitution between practice quality and the distance. MRS_{qd} is the additional distance in kilometers that a patient in rural China would be willing to travel to provider j if its quality increased by 1 unit.

$$MRS_{qd} = \frac{\partial Z_{ijt}}{\partial X_{jt}} = -\frac{\partial u_{ijt}/\partial X_{jt}}{\partial u_{ijt}/\partial Z_{ijt}} = -\beta_i \Big/ \gamma_i \tag{7}$$

MRS_{qd} is the coefficient on quality divided by the coefficient on distance. Following Hole (2007), we estimate standard errors using the delta method. Since it is the ratio of marginal utilities, it is invariant with respect to the scale of utility. Thus, a comparison of the MRS_{qd} for different samples of patients conveys information about differences in preferences. We calculate the marginal rate of substitution between OOP cost and the distance MRS_{pd} using the same method.

$$MRS_{pd} = \frac{\partial Z_{ijt}}{\partial p_{ijt}} = \frac{\frac{\partial u_{ijt}}{\partial p_{ijt}}}{\frac{\partial u_{ijt}}{\partial Z_{ijt}}} = \frac{\alpha_i}{\gamma_i} * \left(1/\left(\overline{y_{it} - p_{ijt}}\right)\right) \tag{8}$$

The results from RCL model suggest that, on average, a patient in rural China would be willing to travel 1.048 additional kilometers to provider j if its quality increased by 1 unit. This is consistent with the finding in Santos et al. (2017).[6] In addition, the patient in rural China would be willing to travel 2.21 additional kilometers to provider j if the OOP costs decreased by 1000 RMB (or net income increased by 1000 RMB).

Table 4 includes interactions with demographic factors to check which characteristics affected the hospital preferences of rural residents. The first and second columns present the means and standard deviations of the preference parameters. The coefficients of the means all have the expected sign, and most of them are statistically significant.

Patients generally prefer hospitals located a short distance from home and have extended business hours, high-quality services, and low prices. The significance of the estimated standard deviations suggests substantial heterogeneity in the patients' preferences. The next four columns display the parameters that measure heterogeneity in the population—the interactions with age, education level, NCMS, and severity level. The heterogeneous preference primarily comes from the severity level. Beneficiaries with more severe conditions value the quality of care more and are less sensitive to net income (or OOP costs). This conclusion that heterogeneous preferences for hospital quality exist across patients is in line with the literature (Santos et al., 2017). Patients with severe illness have been shown to prefer hospitals with superior care and treatment options (Chen et al., 2016; Yu et al., 2017) while patients have experienced relatively minor symptoms did not require first-class medical technology and high quality of care (Yu et al., 2017).

Table 4 also suggests that NCMS has a small overall impact on patient choice. Because the NCMS coverage is primarily for catastrophic illnesses, and only 13% of the patients reported the most severe symptoms in the survey, the effects of NCMS insurance might have washed out in the overall average. Therefore, we examine the choice and evolution of the market shares among the patients with the most severe illness. The proportion of those choosing city hospitals has increased from 12% to 33% since 2000, whereas the proportion choosing village clinics and township health centers has decreased by roughly 10%. Hence, insurance coverage may play a role in hospital demand among the more severely ill patients, but a more detailed analysis is necessary.

6. Counterfactual simulation

To understand the impact of hospital choices on welfare, we use our structural approach to run counterfactual simulations. Inefficient utilization of primary care is a major challenge in China. On the one hand, primary care is underused, and many patients with less severe symptoms prefer city hospitals because they distrust the quality of the service delivered by primary healthcare providers (Li & Fu, 2017; Liu et al., 2018; Yip et al., 2012). On the other hand, low-income residents in rural areas of China are more likely to visit village clinics than high-level healthcare facilities (Qian et al., 2009). Many do not visit county or city hospitals due to the high OOP costs and/or the need to travel a long distance, even when they have moderate to severe illnesses.

Table 5 summarizes the hospital choices based on the illness severity level and type of patient care. The distribution of hospital choices illustrates two types of inefficient allocation in rural China. Approximately 26% of patients with moderate or severe illness (22.21% with moderate and 3.6% with severe illness) visit village clinics for treatment, although higher-tier facilities would be more appropriate alternatives. There is also inefficiency when rural patients with less severe symptoms seek healthcare at county (6.35%) or city (2.25%) hospitals.

To assess the implications of observed healthcare choices in rural China on welfare, we conduct a counterfactual simulation by assuming that the healthcare choices are reallocated to "optimal" choices following the hierarchical medical system (HMS) policy. Specifically, we reassign rural patients to their "optimal" choices according to the requirement of HMS policy, where "optimal" is defined as the most appropriate choice based on the severity of their illness and the care type required. Table 6 illustrates the counterfactual scenario. For all patients reporting outpatient care, we assume that the choice of healthcare provider would shift from village clinics to county hospitals with increasing severity of illness, and all patients reporting inpatient care would choose township

[6] They find that MRS_{qd} is 0.00296, in which the mean of their quality measure is 955.5 while the mean of our quality is 2.2.

Table 4
Results from the full-choice model.

Variable	Mean	Standard deviations	Interactions with demographic variables			
	($\overline{\beta}$ and $\overline{\alpha}$)	(σ_β and σ_α)	Age	Education	NCMS	Severity
Quality proxy	0.1187	0.3983***	−0.0032	−0.0476	0.0104	0.1332**
	(0.2129)	(0.1042)	(0.0032)	(0.0367)	(0.0976)	(0.0568)
$ln(y-p)$	0.3767**	0.2066**	−0.0021	−0.0095	0.1784**	−0.0928**
	(0.1622)	(0.1049)	(0.0018)	(0.0249)	(0.0869)	(0.0417)
Distance	−0.2286***	0.1072***	−0.0002	0.0121*	0.0141	0.0528***
	(0.0688)	(0.0341)	(0.0005)	(0.0070)	(0.0163)	(0.0172)
Business hours	0.0107***		0.0001*	0.0009	−0.0025	−0.0018*
	(0.0036)		(0.0000)	(0.0006)	(0.0015)	(0.0010)
Observations	10,225	10,225				

Note: Cluster-robust standard errors are in parentheses under each coefficient and are clustered at the community level, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The model includes time fixed effects and community fixed effects. Quality proxy refers to doctors per capita for primary care facilities and doctors per bed for high-tier hospitals (more details are available in Appendix A4). The mean column reports the means of random coefficients and the estimates of fixed coefficients. The standard deviations column reports the estimates of the standard deviation of random coefficients. The four right-hand columns report the interactions with demographic variables, including age, education, NCMS insurance, and severity level. NCMS is a binary variable that equals 1 if an individual is covered by the NCMS.

Table 5
Summary of hospital choices based on severity level and care type.

	Low severity	Moderate severity	High severity	Percentage
Village clinic				
Outpatient	461	443	70	48.72%
Inpatient	1	1	2	0.20%
Percentage	23.11%	22.21%	3.60%	48.92%
Township health center				
Outpatient	138	191	43	18.61%
Inpatient	4	17	6	1.35%
Percentage	7.10%	10.41%	2.45%	19.96%
County hospital				
Outpatient	114	194	58	18.31%
Inpatient	13	28	23	3.20%
Percentage	6.35%	11.11%	4.05%	21.51%
City hospital				
Outpatient	39	77	32	7.40%
Inpatient	6	17	21	2.20%
Percentage	2.25%	4.70%	2.65%	9.60%

Note: Table 5 reports the summary of hospice choice based on patients' severity level and care type. Cells are numbers of patients or corresponding percentages indicatingthe proportion of patients in each severity category and care type.

Table 6
The counterfactual scenario based on optimal choice.

	Low severity	Moderate severity	High severity
Outpatient	Village clinic	Township Health Center	County hospital
Inpatient	Township Health Center	County hospital	City hospital

Note: The optimal choice is categorized based on whether the patient needs inpatient treatment and the illness severity level following the hierarchical medical system (HMS) policy.

health centers, county, or city hospitals based on illness severity. To explore the effects of quality improvement on patient welfare, our counterfactual simulation also increased the quality level for primary care facilities.[7]

We use compensating variation to measure differences in consumer welfare supplied by the counterfactual healthcare providers. Compensating variation is the monetary amount a consumer would need to receive to be indifferent to the observed choice and the counterfactual choice. Therefore, a patient would need to be compensated at the new equilibrium between optimal hospital choices to achieve the equivalent utility from the observed choice.

[7] Specifically, we increased the quality level of primary care facilities by one unit if the quality level was the lowest. The increase in quality for primary care facilities with a corresponding increase in the prices to be paid for more qualified and high-quality practitioners.

We compare the welfare under the observed and the counterfactual scenarios, given the patients' preferences. As derived by Small & Rosen (1981), consumer surplus for individual *i*:

$$E[CS_i] = \frac{1}{\alpha_i} E_\varepsilon\left[\max_j u_{ijt}\right]$$

$$= \frac{1}{\alpha_i} E_\varepsilon\left[\max_j X_{jt}\beta_i + \alpha_i ln\left(y_{it} - p_{ijt}\right) + \gamma_i Z_{ijt} + \varepsilon_{ijt}\right]$$

$$= \frac{1}{\alpha_i} ln\sum_j exp\left[X_{jt}\beta_i + \alpha_i ln\left(y_{it} - p_{ijt}\right) + \gamma_i Z_{ijt}\right] + C \tag{9}$$

With a constant marginal utility of income, the difference in expected consumer surplus (*CS*) for individual *i* is:

$$\Delta E(CS) = \int \frac{1}{\alpha_i}\left(ln\sum_j V_{ij}^{post} - ln\sum_j V_{ij}^{pre}\right) dP_D(D)dP_v(v) \tag{10}$$

where $V_{ij}^{post} = exp\left[\beta_i + \alpha_i p_{jt}^{post} + \zeta_{jt}\right]$ and $V_{ij}^{pre} = exp\left[x_j\beta_i + \alpha_i p_{jt}^{pre} + \zeta_{jt}\right]$.

Table 7 shows the results of the welfare analysis. The first row displays the average difference in welfare. On average, consumer surplus decreases by 307 RMB (US$48), equivalent to a 3.6% decrease in the average annual income of rural residents.

To explore the variation in the differences in welfare, we further categorize the patients into groups according to the switching level, where switching is defined as the difference between the counterfactual and actual healthcare providers. For instance, when a patient switches from a village clinic (1) to a city hospital (4), the healthcare choice is leveled up by three categories.[8]

There is significant heterogeneity in welfare across patients. As illustrated in the second column, consumer surplus increases when patients switch to lower-tier hospital types (down-leveled), therefore, strengthening quality for primary care facilities leads to an increase in welfare improvement for patients who are reallocated to lower-tier providers (down-leveled). As they have less severe conditions, primary care facility with better quality can meet their medical needs with lower costs compared to high-tier providers. The policy implication is that by strengthening the quality of services provided by primary care facilities, we can incentivize patients with less severe conditions to visit, which helps to alleviate the overcrowding issue in high-tier city hospitals.

If patients switch to higher-tier hospital types (up-leveled), consumer surplus decreases. Patients face losses in welfare if they are required to visit better hospitals to treat severe symptoms. The losses from high prices (or long distances) may outweigh the gains produced by better care for these groups. As shown in the second column in Table 7, patients in this scenario in the up-leveled groups typically have low incomes. For these groups of patients, we further calculate the coverage level that would leave their welfare unchanged as the counterfactual coverage rate in Table 7. When low-income patients have severe symptoms and need inpatient treatment, the coverage rate should increase to approximately 70 to 90% to incentivize them to seek healthcare from better healthcare providers. To make this policy welfare neutral, this group would need to be compensated by a decrease in OOP costs of 988 RMB (US $153), equivalent to 11% of rural residents' average annual income.

Our counterfactual simulation is most closely related to Gaynor et al. (2016), who studied the effect of removing restrictions on hospital choice in UK patients. They showed that removing constraints on a patient's choice leads to increased welfare. The difference between their findings and this study may be explained by the context. Gaynor et al. (2016) focused on the English National Health Service (NHS), and there is no price mechanism in that market. By contrast, the welfare loss in rural China is partly driven by the higher OOP costs associated with using better healthcare providers.

7. Sensitivity analysis

Due to the survey design, the key variable is defined by whether the respondents used any medical services within four weeks of the interview date of the CHNS. To check if the hospital choices reported by the respondents who were sick within four weeks are assumed to be representative of the distribution of hospital choices among rural residents, we conduct sensitivity analyses to exclude the potential sample selection issue.

To better reflect patients' preferences for healthcare, we employ the CHNS household data for sensitivity analysis, which includes additional information regarding provider characteristics for the years 1989 to 2006. The dependent variable is the health providers that individuals visit most frequently when they need medical treatment, instead of hospital choice over the last four weeks adopted in the main dataset.[9]

Table 8 presents the summary statistics of the provider attributes. Typically, county or city hospitals offer a better quality of services than primary care facilities (village clinics and township health centers). For example, there is a greater probability that a

[8] The hospital providers are assigned values from 1 to 4 for village clinics to city hospitals, repectively. If patients are up-leveled by three categories, we denote switching as "+3" in Table 7, and same rules apply to other categories.

[9] The dependent variable is based on the question "*If household members are sick or want to see a doctor, which clinics and hospitals do they most frequently visit?*".

Table 7
Welfare analysis by counterfactual simulation.

	ΔE[CS] (RMB)	Average income	Original coverage rate	Counterfactual coverage rate
Full sample	−307	8620		
Switching				
−3	14,577	14,189	–	–
−2	8560	10,377	–	–
−1	2680	9490	–	–
No switch	117	7158	–	–
+1	−5148	7402	8.6%	71%
+2	−16,420	7267	9.4%	91%
+3	−17,847	599	10%	83%

Note: The counterfactual coverage rate presents the coverage level that would leave the surplus unchanged. The exchange rate in 2011 between US$ and Chinese RMB in this paper is set at US$100 = 646 RMB. Switching is defined as the difference between counterfactual and actual healthcare providers. The switching is "3" when the healthcare choice is leveled up by three categories, while the switching is "-3" when the healthcare choice is leveled down by three categories.

Table 8
Summary statistics of provider attributes[a].

	Village clinic	Township health center	County hospital	City hospital
Travel time (mins)	8.333	17.15	32.87	71.62
	(6.068)	(13.25)	(32.32)	(240.1)
Travel cost (RMB)	0.0515	0.313	1.603	4.859
	(0.215)	(1.266)	(2.081)	(10.74)
Waiting time (mins)	6.025	12.19	22.41	31.09
	(6.620)	(12.38)	(24.10)	(45.29)
Medicine availability	0.917	0.976	0.986	0.990
	(0.177)	(0.100)	(0.0808)	(0.0721)
Doctor specialties				
Traditional Chinese medicines	0.387	0.101	0.0688	0.0934
	(0.487)	(0.301)	(0.253)	(0.291)
Western medicines	0.509	0.748	0.822	0.802
	(0.500)	(0.434)	(0.383)	(0.399)
Both Western and Chinese medicine	0.104	0.151	0.109	0.105
	(0.305)	(0.358)	(0.312)	(0.306)

Note: Numbers in parentheses are standard errors. Medicine availability is a binary variable that equals 1 if the medicine is available.
[a] The travel cost is low because roughly 70% people in the survey reported zero cost and they walked to the health providers.

given medicine will be available, as well as doctors trained in different (both Western and traditional Chinese medicines) specialties. However, these providers also cost more in terms of travel and waiting times.

Table 9 shows the healthcare choice results, estimated using more detailed provider attributes. The first and second columns of the table provide the estimates of the means and standard deviations of the preference distribution of hospital attributes, respectively. In the first column, the coefficient of $ln(y-p)$ is significantly positive under different specifications, indicating that patients with a higher net income in rural areas are generally more likely to visit better hospitals.

Compared with the first column in Table 3, column 1 in Table 9 shows the estimated results with additional attributes, including travel time and cost. The results are similar to those in Table 3. The means ($\overline{\beta}$) of the business hours and hospital quality are positive and statistically significant. Travel time has a statistically significant and large negative estimated impact. On average, beneficiaries prefer better-quality hospitals, longer business hours, and shorter travel times. The standard deviations of the quality in all specifications are insignificant, implying no variation in the preference for hospital quality after accounting for more detailed information (e.g., travel costs and waiting times). The standard deviations of the travel time, on the other hand, are always significant, which suggests an unobserved variation in the patients' preference for travel time. Finally, the standard deviations of net income are always significant, except for the last regression with medication availability.

8. Conclusions

The suboptimal and inefficient utilization of healthcare, particularly primary care, is a long-standing challenge in China. To illustrate its consequences, we attempt to develop a clear understanding of patients' healthcare-seeking behavior in rural China to inform healthcare policies that could improve their well-being. We first explore how rural residents decide what types of healthcare services to use. The RCL model suggests that OOP costs for treatment and distance to the healthcare provider have negative impacts on the decision to use a specific type of healthcare. In contrast, hospital quality has a significant positive effect.

Our structural model, particularly the counterfactual simulation, has important policy implications. First, government agencies should consider policies that enhance the quality of primary healthcare and strengthen the referral system to incentivize patients with

Table 9
Estimated demand from the random coefficient logit model.

Demand Side parameters		(1)	(2)	(3)	(4)
Mean ($\overline{\beta}$)	Quality proxy	0.123*	0.123*	0.112	0.095
		(0.065)	(0.065)	(0.068)	(0.067)
	Business hours	0.007***	0.007***	0.005***	0.003*
		(0.001)	(0.001)	(0.001)	(0.001)
	Travel time	−0.015***	−0.015**	−0.028***	−0.029***
		(0.005)	(0.006)	(0.008)	(0.009)
	Travel cost	−0.035	−0.035	−0.038	−0.029
		(0.031)	(0.031)	(0.034)	(0.033)
	Waiting time		−0.001	−0.005	−0.004
			(0.007)	(0.007)	(0.007)
	Doctor specialties			0.428***	0.142
				(0.086)	(0.101)
	Medicine availability				1.883***
					(0.414)
Mean ($\overline{\alpha}$))	$ln(y-p)$	2.048**	2.048**	2.071**	1.587*
		(0.955)	(0.957)	(0.991)	(0.882)
Std. Deviations (σ_β)	Travel time	−0.013*	−0.013*	−0.026**	−0.028**
		(0.008)	(0.008)	(0.010)	(0.011)
	Quality	0.171	0.171	0.244	0.146
		(0.214)	(0.214)	(0.216)	(0.315)
Std. Deviations (σ_α)	$ln(y-p)$	2.360*	2.360*	2.433*	1.939
		(1.268)	(1.271)	(1.332)	(1.232)
Observations	N	2617	2617	2617	2617

Note: Standard errors are in parentheses under each coefficient, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Quality proxy refers to doctors per capita for primary care facilities and doctors per bed for high-tier hospitals (more details are available in Appendix A4). The mean column reports the means of random coefficients and the estimates of fixed coefficients. The standard deviations column reports the estimates of the standard deviation of random coefficients.

less severe illnesses to visit primary healthcare providers. This approach will not only benefit patients but also alleviate overcrowding in city hospitals. Second, an income- and diagnose-specific reimbursement rate should be incorporated into the NCMS. If low-income rural residents incur lower OOP costs at city hospitals for catastrophic illnesses, they can afford care at better hospitals when they are gravely ill and require inpatient treatments.

The study has some limitations. First, the OOP costs may be endogenous to the unobserved demand factors in the utility function. In principle, the endogeneity can be addressed by using instrumental variables or by controlling for unobserved heterogeneity using fixed effects. This analysis included selected hospital quality measures but may not have accounted for the full endogeneity because the survey provided only limited information on provider quality. Because the assumption of exogenous costs is reasonable for estimating a micro-level demand model (Goldberg, 2015), we use fixed effects to account for the variation in unobserved quality.[10] Second, the counterfactual analysis assumes that patients seek healthcare from one of four types of healthcare facilities. In practice, some cost-sensitive beneficiaries potentially give up medical care if forced to go to a hospital that is too expensive. Therefore, the up-leveled group may be smaller than predicted, and average welfare gains may have been overestimated. Finally, our analysis involves partial equilibrium and does not consider any supply-side response. In the short term, it is reasonable to assume that the number of hospitals remained constant and that it is difficult for a hospital to adjust prices or quality.

Our results provide evidence and support for implementing a tiered healthcare system, with particular stress on the use of primary care physicians as “gatekeepers” to direct patients to appropriate providers. Our analysis also supports policies strengthening the quality of primary care and providing subsidized healthcare to rural patients.

Funding

Dr. Yang Li and Professor Zhuo Chen acknowledge support from the General Program of the National Natural Science Foundation of China (Grant No. 72174098).

Ethics approval

Approval from Research Ethics Office (Institutional Review Board) of McGill University and Research Ethics of University of Nottingham Ningbo China.

[10] Estimating hospital utilization at the individual level can circumvent the issue of price endogeneity by assuming that a single household has no impact on hospital prices and characteristics.

Declaration of Competing Interest

Affirm compliance with the Pre-Publication Policy: We affirm that the material in the manuscript has not been published, is not being published or considered for publication elsewhere, and will not be submitted for publication elsewhere unless rejected by the journal editor or withdrawn by the author(s).

Appendix A. Institutional background

A.1. Patient health provider choices

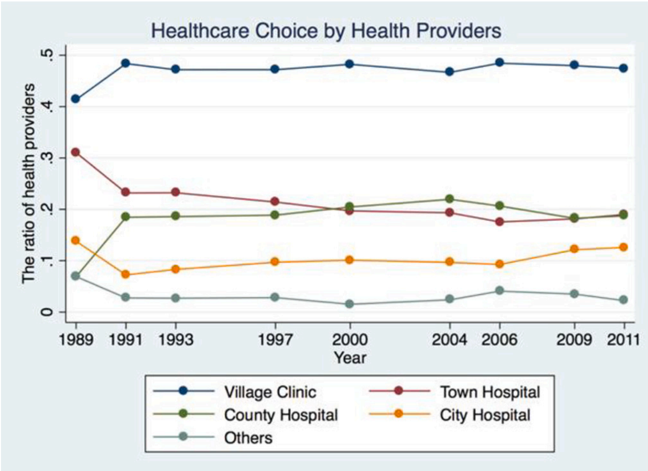

Fig. 1. Patients' healthcare choice by health providers in rural China.
Note: Fig. 1 shows patients' preference for health providers from 1989 to 2011. "Others" refer to purchasing medicines on their own, which is the outside option in our model.

Fig. 1 shows that patients' preference for health providers has remained almost unchanged over the past two decades in rural China, with almost 50% of patients visiting village clinics for treatment. The proportions of patients who sought healthcare at town hospitals or county hospitals were both at 20%. Roughly 10% of patients went to city hospitals for medical treatment.

A.2. Cost-sharing policy in NCMS

Table 1 provides an example of the cost-sharing arrangement of NCMS across health providers. In Henan province, NCMS provides more generous coverage at lower-tier hospitals than higher-tier hospitals for inpatient services. The heterogeneous insurance coverage at different health providers would affect the OOP cost, and consequently, it may affect patients' hospital choices.

Table 1
Reimbursement policy of Henan Province in 2011.

	Deductibles (RMB)	Reimbursement rate
Township health center	16	[16–79]: 50%
		[80–238]: 75%
		[239 +]: 80%
County hospital	64	[64–159]: 50%
		[160–794]: 65%
		[795 +]: 70%
City hospital	159	[159–1587]: 55%
		[1588 - 3175]: 65%
		[3176 +]: 70%

Note: Table 1 shows the reimbursement policy of Henan Provinces in 2011. Both deductibles and coverage amount are listed in Chinese Yuan. The average annual income of rural residents in Henan is approximately 6600 yuan (US $1022.. The exchange rate in 2011 between the US$ and Chinese RMB in this paper is set to US$100 = 646 RMB).

A.3. The self-reported severity

To justify whether the self-reported severity is objective and a good measure for pre-treatment severity, we examine the correlation between self-reported severity and days of daily-life difficulty, which is based on the survey question "*For how many days were you unable to carry out normal activities due to this illness?*".

As shown in Table 2, rural residents with moderate to high severity have longer days of difficulty in carrying out normal activities than those with low severity. Specifically, compared with patients with low severity, rural patients with moderate severity suffer 2 more days of difficulty, while those with high severity suffer 9 additional days of difficulty (the average difficulty time is around 3.67 days). The results are consistent even after we control for the provider fixed effects, that is, within the same hospital choice, patients who report higher severity are usually associated with more "days of difficulty" due to the symptoms. Therefore, rural residents may report their severity level based on the length of days in which they were unable to carry out normal activities due to this illness.

Table 2
Association between severity level and days of difficulty.

	(1)	(2)	(3)
VARIABLES	Days of difficulty	Days of difficulty	Days of difficulty
	Low severity (*reference group*)		
Moderate severity	2.125***	2.271***	2.000***
	(0.376)	(0.387)	(0.386)
High severity	10.093***	9.865***	9.155***
	(0.661)	(0.675)	(0.745)
Constant	6.734	7.758	7.035
	(4.729)	(4.752)	(4.777)
Mean (Days of difficulty)	3.67	3.67	3.67
Observations	8970	8970	8970
R-squared	0.198	0.235	0.251
Year FE	YES	YES	YES
County FE	–	YES	YES
Provider FE	–	–	YES

Cluster-robust standard errors in parentheses*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.4. Quality measures

We argue that our quality indicators are proper measures of hospital quality from the following perspectives. First, as a higher supply of doctors where utilization is higher, our measures of quality adopt the population size and number of beds to control for the utilization or demand effects. Specifically, we use the ratio indicators (the doctor-population ratio and doctor-bed ratio), which measure the number of doctors relative to the total number of population and beds. Second, according to the Donabedian Model, healthcare quality can be evaluated using the quality of care model and the triad of structure, process, and outcome (Donabedian, 1966).

1. "Structure," which describes the physical and organizational setting in which healthcare is delivered, includes healthcare facilities, personnel, and equipment. Our quality indicators belong to this group. It is highlighted in government guidance on quality in China. For example, released by National Health Commission (NHC)[11],[12] the evaluation standards of medical quality for both primary care and higher-tier health providers in China include the doctor per bed and doctor per capita to measure the capacity and quality of health care services. It is also documented in the literature focusing on less developed areas. For example, Sahn et al. (2003) measured the quality of care in rural Tanzania based on the availability of medical staff and clinic environment.
2. "Process," which relies on "structure" to provide the mechanism for patient care, describes the actions that allow for the adequate delivery of healthcare. These typically include diagnosis, treatment, and patient education. "Process" is often measured by patient experience and satisfaction with healthcare delivery, and studies have shown that a higher staff-to-patient ratio was associated with higher patient satisfaction rates in Western countries (Aiken et al., 2018; Kraska, Weigand, & Geraedts, 2017) and in China (Hu et al., 2020), so our quality measures are positively correlated with "process" of quality.

[11] The Evaluation Standards for Tertiary Hospitals [2020], released by National Health Commission (In Chinese) Notice of the National Health Commission on Printing and Distributing the Evaluation Standards for Tertiary Hospitals (2020 Edition)_Health_Chinese Government Website (www.gov.cn).

[12] Basic Standards of Community Hospitals and the Key Points of the Core System of Medical Quality and Safety [2019], released by National Health Commission (In Chinese) Notice of the General Office of the National Health Commission on Printing and Distributing the Basic Standards of Community Hospitals and the Key Points of the Core System of Medical Quality and Safety (Trial)_Medical Management_Chinese Government Website (www.gov.cn).

3. "Outcome" describes the effect of healthcare on patient and population health status (Donabedian, 1988; McDonald et al., 2007). Our data do not include such information, such as mortality rate or readmission rate at the hospital level, to measure "outcome." However, literature has shown that our measures of quality are associated with the quality from "outcome". For example, Lee, Kim, Cho, et al. (2017) found that the number of doctors per bed was significantly associated with pneumonia-specific readmissions and all-cause readmissions, and patients with pneumonia treated in hospitals with a higher number of doctors per bed were less likely to be readmitted.

Appendix B. Parameter estimation

Assuming that each individual's choice is independent of that of other individuals, the maximum likelihood function is

$$L(\alpha,\beta,\xi) = \Pi_i \Pi_j P_{ijt}^{d_{ijt}} \tag{12}$$

Where $d_{ijt} = 1$ if i chooses j at time t and 0 otherwise.
The log-likelihood function of the logit model is then

$$\max_{\alpha,\beta,\xi} LL(\alpha,\beta,\xi) = \max_{\alpha,\beta,\xi} \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ijt} \, lnP_{ijt}(\alpha,\beta,\xi) \tag{13}$$

At the maximum of the likelihood function, its derivative with respect to each of the parameters is zero:

$$\frac{dLL(\beta)}{d\beta} = 0 \tag{14}$$

To make the notation and discussion more succinct, let utility for the representative consumer be linear in parameters: $V_{ijt} = \beta' x_{ijt}$. The first-order condition becomes

$$\sum_i \sum_j \left(d_{ijt} - P_{ijt}\right) x_{ijt} = 0 \tag{15}$$

Rearranging and dividing both sides by N, we have

$$\frac{1}{N} \sum_i \sum_j d_{jit} \, x_{ijt} = \frac{1}{N} \sum_i \sum_j P_{jit} \, x_{ijt} \tag{16}$$

If x_{ijt} is an alternative indicator which equals to 1 if individual i chooses hospital j, and zero otherwise. In this way, the left side of Eq. (16) represents the share of people in the sample who chose alternative j, while the right side is the predicted share for alternative j. For the RCL, the integral is approximated through simulation, where we use Halton draws to increase accuracy.

Appendix C. Evidence on good fit of the price prediction

We provide additional evidence to show the good fit of price prediction from specification 4. Fig. 2 illustrates the linear prediction across different health providers. More specifically, the x-axis shows the predicted log OOP costs, while y-axis shows the observed log OOP costs. Most of the observations are around the 45-degree line, which demonstrates the robustness of our price prediction. We find the same conclusion for the price prediction over the years, which is reported in Fig. 3.

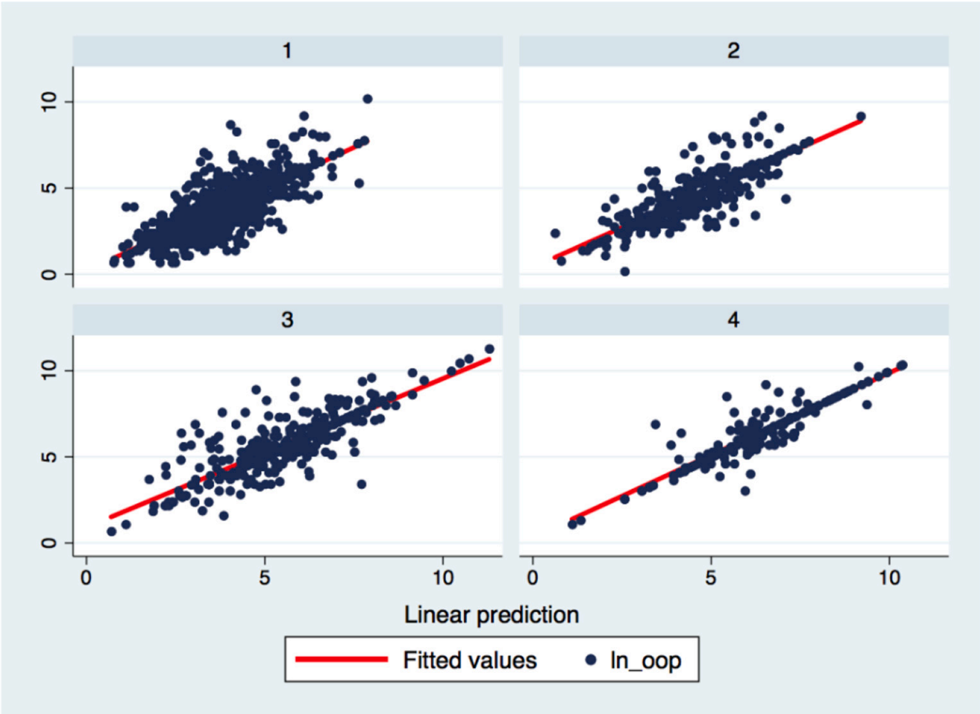

Fig. 2. Linear prediction by health providers.
Note: Fig. 2 shows the linear prediction across different health providers from the price prediction of specification 2. Numbers above each subfigure represent the health providers, with 1 for village clinics and 4 for city hospitals, respectively.

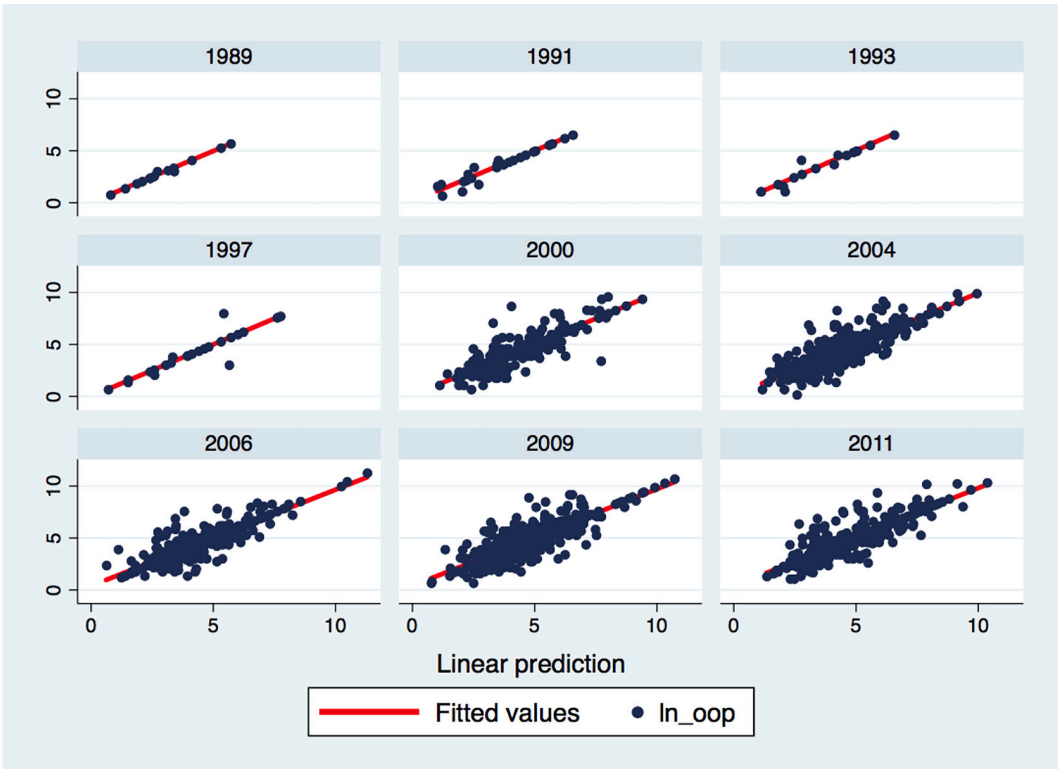

Fig. 3. Linear prediction by years.
Note: Fig. 3 shows the linear prediction over different years from the price prediction of specification 4. Numbers above each subfigure represent the survey years.

References

Aiken, L. H., Sloane, D. M., Ball, J., Bruyneel, L., Rafferty, A. M., & Griffiths, P. (2018). Patient satisfaction with hospital care and nurses in England: An observational study. *BMJ Open, 8*(1), Article e019189.

Audibert, M., He, Y., & Mathonnat, J. (2013). Multinomial and mixed logit modeling in the presence of heterogeneity: A two-period comparison of healthcare provider choice in rural China. In *No 201314, Working Paper from Centre d'Études et de Recherches sur le Développement International*.

Brown, P. H., & Theoharides, C. (2009). Health-seeking behavior and hospital choice in China's new cooperative medical system. *Health Economics, 18*(S2), S47–S64. https://doi.org/10.1002/hec.1508

Bundorf, M. K., Levin, J., & Mahoney, N. (2012). Pricing and welfare in health plan choice. *American Economic Review, 102*(7), 3214–3248. https://doi.org/10.1257/aer.102.7.3214

Chen, R., Du, X. D., Yang, Z., & Ning-Xiu, L. I. (2016). Analysis of choice of healthcare services and the influencing factors among outpatients in Chengdu city. *Modern Preventive Medicine, 43*(22), 4217–4220.

Donabedian, A. (1966). Evaluating the quality of medical care. *Milbank Mem Fund Q, 44*(3). Suppl:166–206. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/5338568. Suppl:166–206. Retrieved from.

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA, 260*(12), 1743–1748. https://doi.org/10.1001/jama.260.12.1743

Dubois, P., & Lasio, L. (2018). Identifying industry margins with price constraints: Structural estimation on pharmaceuticals. *American Economic Review, 108*(12), 3685–3724. https://doi.org/10.1257/aer.20140202

Dunn, A. (2012). Drug innovations and welfare measures computed from market demand: The case of anti-cholesterol drugs. *American Economic Journal: Applied Economics, 4*(3), 167–189. https://doi.org/10.1257/app.4.3.167

Einav, L., Finkelstein, A., & Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics, 125*(3), 877–921. https://doi.org/10.1162/qjec.2010.125.3.877

Gaynor, M., Propper, C., & Seiler, S. (2016). Free to choose? Reform, choice, and consideration sets in the english national health service. *American Economic Review, 106*(11), 3521–3557. https://doi.org/10.1257/aer.20121532

Goldberg, P. K. (2015). Product differentiation and oligopoly in international markets: The case of the U. S. Automobile Industry. *Econometrica, 63*(4), 891–951. Retrieved from http://www.jstor.com/stable/2171803. Retrieved from.

Grossman, M. (1972). On the concept of health capital and the demand for health. *J Pol Econ, 80*, 223–255. https://doi.org/10.1086/259880

Gutacker, N., Siciliani, L., Moscelli, G., & Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of Health Economics, 50*, 230–246. https://doi.org/10.1016/j.jhealeco.2016.08.001

Hanson, K., Yip, W. C., & Hsiao, W. (2004). The impact of quality on the demand for outpatient services in Cyprus. *Health Economics, 13*(12), 1167–1180. https://doi.org/10.1002/hec.898

Ho, K., & Pakes, A. (2014). Hospital choices, hospital prices, and financial incentives to physicians? *American Economic Review, 104*(12), 3814–3840. https://doi.org/10.1257/aer.104.12.3841

Hole, A. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics, 16*(8), 827–840.

Hsiao, W. C. (2005). *Plenary session. Chinese Economists Society Annual Conference, Chongqing, June 24, 2005*.

Hu, L., Ding, H., Liu, S., Wang, Z., Hu, G., & Liu, Y. (2020). Influence of patient and hospital characteristics on inpatient satisfaction in China's tertiary hospitals: A cross-sectional study. *Health Expectations, 23*(1), 115–124. https://doi.org/10.1111/hex.12974

Kraska, R. A., Weigand, M., & Geraedts, M. (2017). Associations between hospital characteristics and patient satisfaction in Germany. *Health Expectations, 20*(4), 593–600.

Lee, J. E., Kim, T. H., Cho, K. H., et al. (2017). The association between number of doctors per bed and readmission of elderly patients with pneumonia in South Korea. *BMC Health Services Research, 17*, 393. https://doi.org/10.1186/s12913-017-2352-7

Lei, X., & Lin, W. (2009). The new cooperative medical scheme in rural China: Does more coverage mean more service and better health? *Health Economics, 18*(S2), S25–S46. https://doi.org/10.1002/hec.1501

Li, L., & Fu, H. (2017). China's health care system reform: Progress and prospects. *International Journal of Health Planning and Management, 32*(3), 240–253. https://doi.org/10.1002/hpm.2424

Liu, W., Liu, Y., Twum, P., & Li, S. (2016). National equity of health resource allocation in China: Data from 2009 to 2013. *International Journal for Equity in Health, 15*(1), 1–8. https://doi.org/10.1186/s12939-016-0357-1

Liu, Y. (2004). Development of the rural health insurance system in China. *Health Policy and Planning, 19*(3), 159–165.

Liu, Y., Kong, Q., Yuan, S., & van de Klundert, J. (2018). Factors influencing choice of health system access level in China: A systematic review. *PLoS One, 13*(8), Article e0201887. https://doi.org/10.1371/journal.pone.0201887

McConnell, K. J., Lindrooth, R. C., Wholey, D. R., Maddox, T. M., & Bloom, N. (2016). Modern management practices and hospital admissions. *Health Economics, 25*(4), 470–485. https://doi.org/10.1002/hec.3171

McDonald, K. M., Sundaram, V., Bravata, D. M., Lewis, R., Lin, N., Kraft, S. A., … Owens, D. K. (2007). *Closing the quality gap: A critical analysis of quality improvement strategies* (Vol. 7). Rockville, MD: Agency for Healthcare Research and Quality (Care Coordination).

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470. https://doi.org/10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.3.co;2-t

Morey, E. R., Sharma, V. R., & Mills, A. (2003). Willingness to pay and determinants of choice for improved malaria treatment in rural Nepal. *Social Science and Medicine, 57*(1), 155–165. https://doi.org/10.1016/S0277-9536(02)00338-6

Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals.". *Journal of Health Economics, 28*(6), 1154–1165. https://doi.org/10.1016/j.jhealeco.2009.08.006

Qian, D., Pong, R. W., Yin, A., Nagarajan, K. V., & Meng, Q. (2009). Determinants of health care demand in poor, rural China: The case of Gansu Province. *Health Policy and Planning, 24*(5), 324–334. https://doi.org/10.1093/heapol/czp016

Sahn, D. E., Younger, S. D., & Genicot, G. (2003). The demand for health care services in rural Tanzania. *Oxford Bulletin of Economics and Statistics, 65*(2), 241–260. https://doi.org/10.1111/1468-0084.t01-2-00046

Santos, R., Gravelle, H., & Propper, C. (2017). Does quality affect patients' choice of doctor? Evidence from England. *The Economic Journal, 127*(600), 445–494.

Sivey, P. (2012). The effect of waiting time and distance on hospital choice for English cataract patients. *Health Economics, 21*(4), 444–456. https://doi.org/10.1002/hec.1720

Small, K. A., & Rosen, H. S. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica, 49*(1), 105–130. https://doi.org/10.2307/1911129

Wagstaff, A., Lindelow, M., Jun, G., Ling, X., & Juncheng, Q. (2009). Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics, 28*(1), 1–19. https://doi.org/10.1016/j.jhealeco.2008.10.007

Wang, J., Wang, P., Wang, X., Zheng, Y., & Xiao, Y. (2014). Use and prescription of antibiotics in primary health care settings in China. *JAMA Internal Medicine, 174*(12), 1914–1920.

Yip, W. (1998). Determinants of patient choice of medical provider: A case study in rural China. *Health Policy and Planning, 13*(3), 311–322. https://doi.org/10.1093/heapol/13.3.311

Yip, W., Fu, H., Chen, A. T., Zhai, T., Jian, W., Xu, R., … Chen, W. (2019). 10 years of health-care reform in China: Progress and gaps in universal health coverage. *The Lancet, 394*(10204), 1192–1204. https://doi.org/10.1016/S0140-6736(19)32136-1

Yip, W., Hsiao, W. C., Chen, W., Hu, S., Ma, J., & Maynard, A. (2012). Early appraisal of China's huge and complex health-care reforms. *The Lancet, 379*(9818), 833–842. https://doi.org/10.1016/S0140-6736(11)61880-1

Yu, W., Li, M., Ye, F., Xue, C., & Zhang, L. (2017). Patient preference and choice of healthcare providers in Shanghai, China: a cross-sectional study. *Bmj Open, 7*(10), Article e016418.

Zhou, Z., Zhao, Y., Shen, C., Lai, S., Nawaz, R., & Gao, J. (2021). Evaluating the effect of hierarchical medical system on health seeking behavior: A difference-in-differences analysis in China. *Social Science & Medicine, 268*, Article 113372.