The impacts of high-speed rail on health utilization among the elderly in China

Xiaojun Lin, Miao Cai

April 21, 2019

1 Introduction

The availability of transportation is an important enabling factor for the utilization of healthcare services. In addition, the distance patients are required to travel can be a barrier to the utilization of healthcare services.

The HSR is a type of bullet train capable of sustaining speeds upwards of 300 km/h. In the past decade, high-speed rail (HSR) developed rapidly in China. [*Insert the introduction of HSR in China*]. Since 2008, many cities in China have ended the history of no high-speed rail. As the development of the modern transportation mode, high-speed rail has become an important choice for people's travel tools and even a priority at certain distances. To most most population destinations in China, the HSR reduced rail travel time by about half. The accessibility of high-speed railways will shorten the time distance between cities, thus shortening psychological distance, social distance and economic distance.

In China, patients are permitted to go any hospital to seek care. [*Insert the brief introduction of the Chinese healthcare delivery system*]

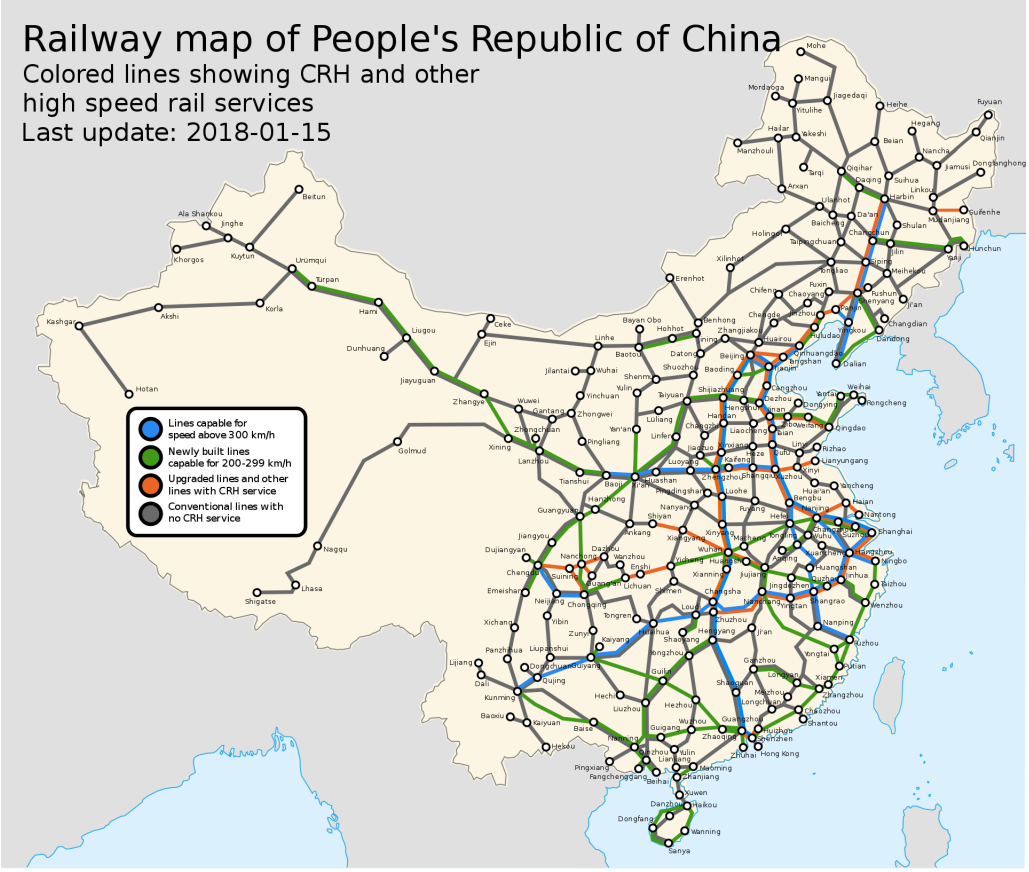
Most of the tertiary hospitals in China possess high-tech equipment and top-level health professionals, but they are located in the capital of a province. Some metropolis like Beijing, Shanghai and Guangzhou, had much more high-quality medical resources than other capital cities. Therefore, to access advanced medical technology and skills, patients with diagnosed with severe diseases have to travel a long distance. Patients with severe diseases may migrate from the rural districts to large metropolitan areas in order to utilize the high-quality health-care services available in those regions.

Although many studies worldwide have identified factors associated with the utilization of health-care service, few have investigated the effects of the development of public transportation on the utilization of health-care service. Recently, Jung-Kyu *et al.* (2019) explored the association between the Korean Train Express and the health utilization for patients diagnosed with cancer, and found that the introduction of the Korean Train Express facilitate the utilization of outpatient services rather than inpatient services, especially for high income patients.

The features of the HSR in China. First, the HSR network was constructed in a short term. Second, China has the longest HSR in the world. The rapid development of HSR provides a good natural experiment condition for our study.

This study aims to identify the effects of HSR on the utilization of available health-care services for residents with heart attack, using the data from China Health and Retirement Longitudinal Study (CHARLS). We hypothesized that the development of HSR would facilitate healthcare utilization.


2016年开通高铁线路示意图
2003年通车线路
2008年通车线路
2009年通车线路
2010年通车线路
2011年通车线路
2012年通车线路
2013年通车线路
2014年通车线路
2015年通车线路
2016年通车线路
注1：部分线路仅为示意，最终线路以国家有关部门发布为准。
注2：请勿用作商业用途，仅供铁路爱好者交流学习之用。
注3：非严格定义高铁线路，包含部分城际铁路和客货混用铁路，以200公里/小时为标准。
地图看世界
http://blog.sina.com.cn/xiaqirui
http://weibo.com/xiaqirui



2 High-speed rail in China

3 Material and methods

3.1 Data and population

The present study uses data from the survey of the China Health and Retirement Longitudinal Study (CHARLS Pilot) collected from 2011 to 2015. The detail of the pilot data is described in Yaohui *et al.* (2014). Generally speaking, CHARLS was designed in the similar way to the US Health and Retirement Study as a broad-

purposed social science and health survey of older people in China. The survey instruments is divided into seven sections: (a) Demographic Background, (b) Family, (c) Health Status and Functioning, (d) Health Care and Insurance, (e) Work, Retirement and Pension, (f and g) Income, Expenditure and Assets, (h) Interviewer Observation. The CHARLS pilot sample is representative of people aged 45 or older, and their spouses, living in households in the 28 provinces of China.

Here are the brief information for CHARLS:

1. Population: Chinese residents ages 45 and older. It includes about 10,000 households and 17,500 individuals in 150 counties/districts and 450 villages/resident committees.

2. Sampling procedure: The CHARLES sample was drawn in four stages. Within each province, 16 county level units (rural counties or urban districts) were randomly selected by Probability Proportional to Size (PPS), stratified by regions and urban/rural. Within each county, CHARLS randomly selected 3 village level units (administrative villages in rural areas or resident communities in urban areas) as Primary Sampling Units (PSUs); within each PSU, CHARLS then randomly selected 25e36 from a complete list of dwelling units generated from a map for each province.

3. Baseline and follow-up: Baseline data in 2011; Follow-up data in 2013 and 2015. The individuals will be followed up every two years. Survey data of 2017 is expected to be released in June, 2019.

4. Ethical approval: Ethical approval for this study was not required because it was based exclusively on the publicly available data, CHARLS, and the study subjects were not directly approached.

To investigate the association between HSR and health utilization, we have to know the address of each respondent. We can use the *CommunityID* in CHARLS datasets to identify the city where the respondents live, while the specific county or village information was unavailable due to privacy protection.

Finally, a total of 126 cities in 28 provinces were included in the CHARLS sample. According to the data of HSR in website, we identified the exact date of opening HSR for each city. (see *Data/PSU_HSR.xlsx*)

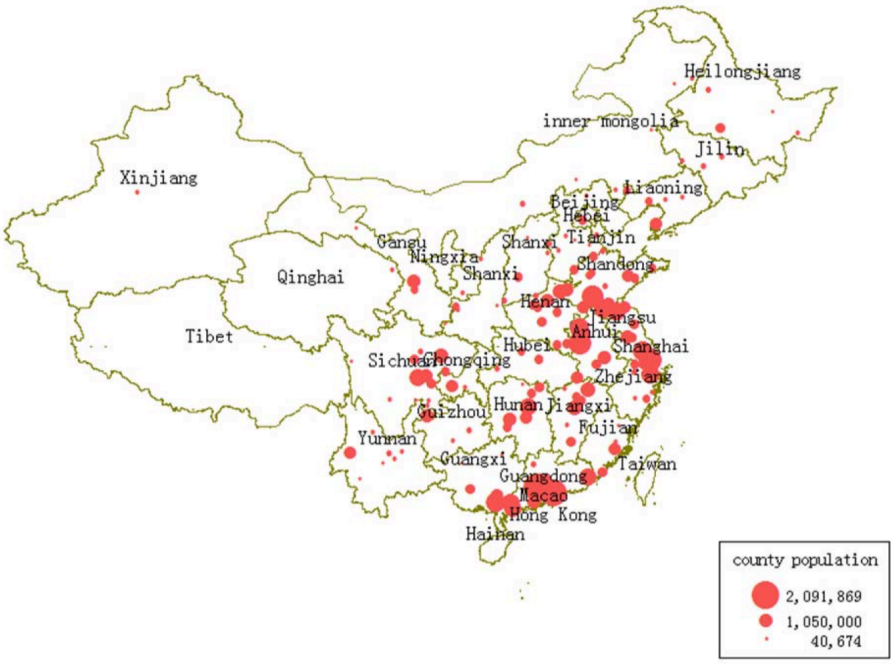


图 1 CHARLS 抽中县、区在全国的分布情况

3.2 Study variables

3.2.1 Dependent variables

The dependent variables in our analysis reflect the intensity and out-of-pocket (OOP) expenditure of different health care utilization. We consider the following

measures of health service utilization: (a) the probability of outpatient visits during the one month or being hospitalized during the year preceding the survey date; (b) the number of outpatient visits during the one month or being hospitalized during the year preceding the survey date; (c) individual OOP expenditure for the most recent doctor visit or hospitalization.

3.2.2 Independent varibles

According to the Anderson models (Andersen & Newman, 1973; Andersen, 1995), there are four types of variables that can explain health outcomes, including predisposing, enabling, need, and life-style variables.

Predisposing variables. The predisposing component centers on the idea that some individuals have a propensity to use services more than other individuals, and this tendency can be predicted from individual characteristics prior to an illness episode. In this study, we included age, gender, education, marital status, and family size. Respondents' education levels are grouped into four categories: illiterate, did not complete primary school but capable of reading, completed elementary school, and junior high and above. The lowest education level e illiterate is treated as the reference.

Enabling variables. The enabling variables represent differences in the supply of health care services, such as differences across geographical locations. In the present paper, we use city fixed effect dummies to reflect and discriminate regional variation within each province, while others represent differences in demand, such as income, health insurance. The income variable used in this study represents per capita household annual income. The health insurance variables are the key variables of interest in our paper, which are represented by three dummy variables capturing UEBMI, URBMI, and NCMS, leaving no insurance as the reference cat-

egory.

Need variables. The need variables capture the need for health care and represent the most immediate cause of health service use. In this study, we use self-reported health status, the number of respondent's chronic conditions and a set of dummy variables to indicate whether the respondent has specific chronic diseases (including hypertension, diabetes, dyslipidemia, cancer, chronic lung diseases, liver disease , heart attack, stroke, kidney disease, arthritis or rheumatism, asthma) as the proxies of need variables. Self-reported health status is captured by a dummy variable for poor reported health status, leaving fair, good, very good or excellent state of health as the reference category.

Life-style variables. Smoking behavior is the factor included in the model to measure health behavior. A dummy variable is used to identify whether the respondent ever smokes; non-smoker is the reference category.

The variable of our interest was whether the high-speed rail is access to residents preceding the survey date. We also included other characteristics of cities where the respondents live in, including the number of hospitals and clinics, hospital beds, medical personnel per 1,000 capita, hospital competition (measured by Herfindahl-Hirschman index) and population.

3.3 Statistical analysis

Following Xin & Wei (2013), we used the two-part model (TPM) model the analyze the impact of HSR on the utilization of outpatient and inpatient services.

Specifically, for the number of outpatient visits, the first part of the model is a logistic that predicts the probability of any use of outpatient service. The intensity of health care utilization is then analyzed by employing a zero-truncated Poisson model since the focus of analysis is the number of outpatient visits conditional on at

least one visit to a service provider. For the number of hospitalizations, we applied the similar estimation strategy.

For the OOP spending, the first part is the same as that for the outpatient visits; for the second part, a generalized linear model (GLM) is employed since the GLM has recently received considerable attention in modeling the health expenditure. In the present study, the GLM with a gamma distribution and a log link is used to estimate the OOP spending only among observations with positive spending.

Since CHARLS is a survey data, we consider the design characteristics in our analysis. We state the strata (county level units) and PSU (community) and use the individual sample weights provided by CHARLS to produce population representative estimates.

The first part of TPM was as follows:

$$I_{ict} = \beta_0 + \beta_1 HSR_{ct} + \beta_2 X_{ict} + \beta_4 Z_{ct} + \lambda_c + \delta_t + \varepsilon_{ict}$$

I_{ict} was a binary variable indicating the health utilization for individual i who lived in city c in year t. Health utilization was defined that the respondent has used outpatient services in last month or inpatient services in last year. HSR_{ct} was a dummy variable that equalled one if city c open the HSR station in year t. X_{ict} were the predisposing, enabling, need, and life-style variables that may affecting individual's health utilization, such as age, gender, education, marriage, and insurance type, family income level. Z_{ct} was the other city-level characteristics. We included the number of hospitals and clinics, hospital beds, medical personnel per 1,000 capita, hospital competition (measured by Herfindahl-Hirschman index) and population in this study. λ_c and $\delta_t$were city and year fixed effect. ε_{ict} was the error term.

The second part of TPM was as follows:

$$(Y_{ict}|I_{ict}=1)=\beta_0+\beta_1 HSR_{ct}+\beta_2 X_{ict}+\beta_4 Z_{ct}+\lambda_c+\delta_t+\upsilon_{ict}$$

where Y_{ict} represents the number of outpatient visits or hospitalizations or the OOP spending, and υ_{ict} is the random error term.. The standard errors were clustered at the city level.

To better understand the impact of HSR on health utilization, we also investigate whether there are differences in the effects of HSR on health utilization across geographical location (rural area vs. urban area) , age groups (45 64 years vs 64+ years) and family income level (low-income vs. high-income).

We conduct all of our analyses with STATA 11 (Stata Corporation, College Station, TX, USA), using survey commands to control for the complex design used in CHARLS.

3.3.1 Robustness check

To test the sensitivity of our estimation, should we also estimate models with city-specific linear time trend and with quadratic city-specific trends? (Hung-Hao & Chad, 2019)

References

Andersen, R. 1995. Revisiting the behavioral model and access to medical care: does it matter? *Journal of Health and Social Behavior*, **36**, 1–10.

Andersen, R., & Newman, J. F. 1973. Societal and individual determinants of medical care utilization in the United States. *Milbank Memorial Fund Quarterly Health and Society*, **51**, 95–124.

Chunyan, Li, & Weiyan, Jian. 2017. Economic burden among cardiovascular patients of middle-aged and elderly: An empirical analysis based on China Health and Retirement Longitudinal Survey data. *Chinese Journal of Health Policy*, **10**, 75–80.

Hung-Hao, Chang, & Chad, D. Meyerhoefer. 2019. Inter-brand competition in the convenience store industry, store density and healthcare utilization. *Journal of Health Economics*, **65**, 117–132.

Jung-Kyu, Choi, Se-Hyung, Kim, & Myung-Bae, Parkc. 2019. Effects of high-speed rail on health-care service utilization. *Journal of Transport & Health*, **12**, 13–20.

Xin, Li, & Wei, Zhang. 2013. The impacts of health insurance on health care utilization among the older people in China. *Social Science & Medicine*, **85**, 59–65.

Yaohui, Zhao, Yisong, Hu, James, P Smith, John, Strauss, & Gonghuan, Yang. 2014. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). *International Journal of Epidemiology*, **43**, 61–68.

CHARLS Questionnaire

Table 1: Variables and Question Index

Variable	Section	Index	Question
Gender	B	BA000_w2_3	Interviewer record R's gender.
Age		BA001_W2_1	We record your birth time is [Birth Time], is it right?
		BA002	When were you born?
Hukou		BC001	What is your current HuKou status?
Education		BD001	What is the highest level of education you have attained? (not including adult education)
Marriage		BE001	RMaritalStatus: What is your marital status?
Family size	C	CM001_W2_1	I will go through a list of people living in this household that our system recorded. Please let me know whether this list is correct.
Self-report health	D	DA001	Next, I have some questions about your health. Would you say your health is excellent, very good, good, fair, or poor?
		DA002	Next, I have some questions about your health. Would you say your health is very good, good, fair, poor or very poor?
		DA079	How would you rate your health status? Would you say your health is very good, good, fair, poor or very poor?

Table 1: Variables and Question Index

Variable Section	Index	Question
	DA080	Next I have some questions about your health. Would you say your health is excellent, very good, good, fair, or poor?
Chronic diseases	DA007	Have you been diagnosed with [conditions listed below, read one by one] by a doctor?
Smoke	DA059	Have you ever chewed tobacco, smoked a pipe, smoked self-rolled cigarettes, or smoked cigarettes/cigars?
	DA061	Do you still have the habit or have you totally quit?
Heart attack	DA007_W2_5	[Since R's LAST IW MONTH, YEAR/In the last two years], have you had a heart attack?
InsuranceE	EA001	Are you the policy holder/primary beneficiary of any of the types of health insurance listed below? (circle all that apply)
	EA002	Do you have supplemental insurance to this plan?
Outpatient visit	ED001	In the last month have you visited a public hospital, private hospital, public health center, clinic, or health worker's or doctor's practice, or been visited by a health worker or doctor for outpatient care?

Table 1: Variables and Question Index

Variable Section	Index	Question
	ED005	How many times did you visit/been visited by […] during the last month?
	ED006	How much did all the visits to [ED004 answer] cost during the last month?
	ED007	Self-paid part (Yuan)
	ED018	Could you tell me the disease name?
Inpatient service	EE003	Have you received inpatient care in the past year?
	EE004	How many times have you received inpatient care during the past year?
	EE005	What was the medical cost for all the hospitalizations you received during the past year? (Only include fees paid to the hospital, including ward fees but excluding wages paid to a hired nurse, transportation costs, and accommodation costs for yourself or family members.)
	EE006	Self-paid part (Yuan)
	EE020	Could you tell me the name of the disease?

Table 1: Variables and Question Index

Variable	Section	Index	Question
Pension	F	FN001_w2	Are you currently participating in or receiving at least one kind of pension as followings? Pension here refers to income from such pension programs as supplemental pension insurance of the firms, residents' pension insurance, rural pension insurance, Urban residents'pension and commercial pension insurance, and pension subsidy for the oldest old and so on? (choose all that apply)
Income		"INCOME_PC"	household per capita income; "exp_income_wealth.dta"

CHARLS Time Table

Table 2: Time table for CHARLS

Time	Wave
2011.5-2012.3	Wave 1 (2011)
2013.7-8	Wave 2 (2013)
2015.7-8	Wave 3 (2015)

Source: http://charls.pku.edu.cn/zh-CN/page/about/milestones

Data Exploration Note

We focus on the respondents with heart attack in CHARLS database. [Why not cancer? Because there are no more than 200 respondents who were identified with cancer in each wave]

According to Chunyan & Weiyan (2017), there are three criteria to identify the respondents with heart attack: (1) the respondents who were diagnosed with heart attack by a doctor (DA007); (2) the respondents had a heart attack in the past two years (DA007_W2_5); (3) the respondents whose primary disease in the most recent outpatient visit or hospitalization is heart attack (ED018 and EE020).

Table 3: Sample size by year

Wave	Year	N
Wave 1	2011	2127
Wave 2	2013	2576
Wave 3	2015	2929
Total		7632

Note: ED018 and EE020 are missing in the dataset of wave 3 (2015).

In Stata, you can use the command "replace householdID = householdID + "0"; replace ID = householdID + substr(ID,-2,2)" to adjust the IDs in the baseline sample (Wave 1).