Do Rural-Urban Identities Affect Individuals' Health? Evidence From China

Feng-Gang Lu, PhD[1] iD **, Yin-Ping Chen, PhD[2]** iD **,
and Fan-Jie Yang, MA[1]**

Abstract

We study the relationship between household registration status (Hukou) and the state of individuals' health to find out whether inequality in health between urban and rural population exists in China. We have used the probit model to regress the state of health on household registration using the individual-level data of the 2018 CFPS survey. We find that inequality in health between urban and rural population does exist in China. Individuals with rural Hukou have a higher probability by 1.4% to be admitted to hospital than individuals with urban Hukou. While, individuals with rural Hukou tend to over-estimate the state of their health as the probability for them to assess themselves healthy is higher by 1.7% than individuals with urban Hukou. The findings suggest that policy makers should recognize the issue of rural-urban health inequalities and take measures, such as controlling pollution in rural areas and providing high quality routine health checks for rural population to deal with the problem.

What do we already know about this topic?
Empirical studies show that disparities in health outcomes between rural and urban population are prevalent though the degree varies.

How does your research contribute to the field?
We find a contradiction in China that individuals with rural Hukou have a higher probability by 1.4% to be admitted to hospital than individuals with urban Hukou, meanwhile, individuals with rural Hukou tend to over-estimate the state of their health.

What are your research's implications toward theory, practice, or policy?
The findings suggest that policy makers should recognize the issue of rural-urban health inequalities and take measures, such as controlling pollution in rural areas and providing high quality routine health checks for rural population to deal with the problem.

Introduction

As a typical well-known "dual economy," together with a household registration system (Hukou; We will discuss China's Hukou system in details in the section of "Methods.") which separates population into rural and urban categories, we want to find out whether China features "dual" health inequalities between rural and urban population. Health inequality is a particular type of disparities in health revealing the social hierarchies of different social groups and they put the disadvantaged groups in society at further disadvantage with respect to their health, which in turn is essential for escaping from social disadvantage.[1]

Empirical studies show that disparities in health outcomes between rural and urban population are prevalent though the degree varies. The United States residents living in rural counties are more likely to have poorer health outcomes than

[1]School of Public Administration, Shandong Technology and Business University, Yantai, Shandong, China
[2]School of Marxism, Shandong Technology and Business University, Yantai, Shandong, China

Corresponding Author:
Yin-Ping Chen, East District of Shandong Technology and Business University, Binhai Middle Road 191#, Laishan District, Yantai, Shandong, 264005, China.
Email: 106976744@qq.com

their urban counterparts[2-4]; counties with a higher rural population have a higher rate of mortality of COVID-19.[5] Britain shows the same trend of health inequalities, and health inequalities are growing wider.[6] The risk for depression and/or anxiety is higher for the rural perinatal women than the urban counterparts in the UK.[7] In China, the management and control of hypertension and diabetes were worse in rural than urban areas; the coverage for breast cancer and cervical cancer screening was worse among those from rural areas and for the poorest.[8] Evidence suggests that health differences between urban and rural population may be attributed to access to health care services, health-related behaviors, and air and water quality in respective areas. Access to and the utilization of health services influence the performance of rural health care.[9-11] Compared to urban residents, rural residents had lower access to health information from sources including primary care providers, specialist doctors, blogs, and magazines, and less use of search engines.[12,13] Differences in physical aspects of a place, such as air or water quality may correlate with rural-urban health differences.[14] Particulate matter and ozone are 2 well-characterized air pollutants that can cause bad health outcomes, and contaminants in water are associated with a range of acute and chronic adverse effects.[15,16] Health-related behaviors, such as sufficient sleep, nondrinking, nonsmoking, maintaining normal body weight, meeting aerobic leisure time physical activity recommendations, can benefit individuals' health.[17] It is difficult for rural areas to deliver services and provide needed health communications about the benefits of adopting these behaviors.[18,19]

The study aims to examine the impact of household registration status on rural-urban health inequalities in China and adds to literature of this field in the following ways. First, we use rich individual-level data from the CFPS survey to identify rural-urban health inequalities across China and therefore highlights the need to address shortfalls in rural healthcare. Second, to overcome the impacts of measurement errors and the issue of outliers on estimation results, we transform the covariates into binary variables, in which approach we combine qualitative information and quantitative information into our analysis. Another advantage of this approach is that the mean of each variable also tells the probability of the event, which makes it convenient to proceed with statistical analysis. Third, as a further addition to the existing body of evidence for rural-urban health inequalities, we provide insights on the current situation of China and differentiate between the effects of rurality on individuals' self-assessed health and their medical treatment.

Methods

Data

We obtain individual-level data of 2018 released in August 2020 by the China Family Panel Studies (CFPS), which consists of 5 datasets, Person, Person-proxy, Child-proxy, Family, and Family-connection. The CFPS survey is a nationally representative, annual longitudinal survey of Chinese communities, families, and individuals launched in 2010 by the Institute of Social Science Survey (ISSS) of Peking University, China.

The 2018 CFPS data are used for our research based on 3 considerations. First, the large sample size is crucial to our analysis. The 2018 CFPS survey interviewed 14 241 families and 32 669 individuals within these families in 31 provincial districts of China. Second, the sampling methods make the overall CFPS sample representative of the country. The sample for the CFPS baseline survey through a multi-stage probability is drawn with implicit stratification. Each subsample in the CFPS study is drawn through 3 stages: county (or equivalent), then village (or equivalent), then household. Third, the scope of the survey fit our research. The 2018 CFPS survey collected individual-level and family-level data on the economic, as well as the non-economic, wellbeing of the Chinese population, with a wealth of information covering such topics as health, household registration status, economic activities, education outcomes, family dynamics and relationships.

We focus on studying the state of health of individuals over the age 16 to 65. After drawing data of individuals over 16 to 65 and deleting observations with unidentified values, we obtain a sample of 25 303 individuals, of which 18 900 individuals have urban Hukou and 6403 individuals have rural Hukou.

Model Construction

We first present our baseline regression in equation (1), which consists of regressing the state of health on registration status, controlling for 4 groups of 23 variables.

When it comes to a qualitative regress and, there are usually 3 estimators to get the estimates, the linear probability model (LPM), the logit Model, and the probit Model. Before the availability of the logit and probit models, the LPM was used quite extensively because of its simplicity. But the LPM has 2 vital drawbacks that the fitted probabilities can be less than zero or greater than one and the partial effect of any explanatory variable is constant. To explain the behavior of a binary explained variable we will have to use a suitably chosen cumulative distribution function (CDF). The logit model uses the cumulative logistic function, and the probit model uses the normal CDF, which in principle one could substitute in place of the logistic CDF. That is to say, there is no compelling reason to choose one over the other. But we choose the probit model, equation (2), as our main estimation method based on McFadden's view that the probit model is more suitable for explanations from rational choice perspective on behavior.[20] Because the LPM is simple to estimate and interpret we still use it, equation (3), as a benchmark for comparison.

Table 1. Summary Statistics.

Variable	Mean			Std. dev.	Min.	Max.	Expected sign
	Total	Rural	Urban				
Inpatient care	0.894	0.890	0.905	0.308	0	1	
Self-assessed health	0.740	0.732	0.764	0.439	0	1	
Registration	0.747	0.747	0	0.435	0	1	(-)
Per capita family income	0.292	0.205	0.556	0.456	0	1	(+)
Family deposit	0.209	0.161	0.352	0.407	0	1	(+)
Family debt	0.225	0.252	0.145	0.418	0	1	(-)
Medical insurance	0.086	0.080	0.105	0.281	0	1	(+)
Employment	0.766	0.794	0.684	0.424	0	1	(+)
Work intensity	0.410	0.494	0.322	0.492	0	1	(-)
Housing	0.875	0.880	0.858	0.331	0	1	(+)
Ventilation	0.046	0.029	0.096	0.209	0	1	(+)
Kitchen water	0.756	0.695	0.936	0.430	0	1	(+)
Food expenditure	0.269	0.240	0.353	0.443	0	1	(+)
Healthcare	0.122	0.090	0.214	0.327	0	1	(+)
Age	0.803	0.810	0.783	0.398	0	1	(+)
Workout	0.374	0.335	0.489	0.484	0	1	(+)
Smoking	0.140	0.146	0.122	0.347	0	1	(-)
Drinking	0.146	0.149	0.137	0.353	0	1	(-)
Gender	0.495	0.493	0.500	0.500	0	1	(+)/(-)
Marriage	0.796	0.798	0.789	0.403	0	1	(+)
Family bond	0.468	0.445	0.533	0.499	0	1	(+)
Internet	0.990	0.992	0.985	0.099	0	1	(+)/(-)
Worship	0.740	0.759	0.684	0.438	0	1	(+)
Trust	0.557	0.543	0.598	0.497	0	1	(+)
Helping behavior	0.697	0.694	0.705	0.460	0	1	(+)
Charity	0.250	0.222	0.334	0.433	0	1	(+)

Note. Number of total observations: 25 303. Number of urban observations: 18 900. Number of urban observations: 6403.

$$y = \alpha_0 + \alpha_1 h_i + \beta X_i + u_i \qquad (1)$$

$$P(y = 1) = G(\alpha_0 + \alpha_1 h_i + \beta X_i) \qquad (2)$$

$$P(y = 1) = \alpha_0 + \alpha_1 h_i + \beta X_i \qquad (3)$$

where the subscript i denotes individual i, y is the state of health, h is the main explanatory variable, registration status (Hukou), u is the error term, and X is a vector of 23 control variables, discussed as follows.

Explained Variable

We extract from the Person dataset our 2 main dependent variables, inpatient care and self-assessed health, to measure individuals' health state. Both variables are defined to be dummy variables, the value 1 indicating "healthy" and 0 indicating "unhealthy." Inpatient care takes on the value 1 if the respondent hasn't been admitted into hospital during the recent 12 months, and 0 if otherwise. Self-assessed health is defined to be 1 if the respondent rate his health as "excellent," "very good," or "good," and 0 if the answer is "fair" or "poor."

We are interested whether there are differences in health state between urban and rural population. As shown in Table 1, urban population are on average slightly "healthier" than rural population. To be specific, 90.5% of urban population and 89% of rural population haven't received inpatient health care services, and 76.4% of urban population and 73.2% of rural population believe that they are healthy.

Main Explanatory Variable

China has a household registration system (Hukou system) in which each person has a Hukou (registration status), classified as "rural" or "urban" status, in a specific administrative unit.[21-23] The Hukou system was used for a long history since the founding of P. R. C. as an instrument for local government to take control of population migration, because the Hukou status was related to eligibilities for social welfare benefits. People with rural Hukou were not eligible for regular urban welfare benefits (access to local schools, urban pension plans, public housing, etc.) and other rights that are

available to those with urban Hukou. So it used to be very difficult for population with rural Hukou to convert to urban Hukou. Due to seeable reasons, the government of China took steps to abolish the dualistic Hukou system since 2014, instead, established a new unified Hukou system in which each person has a residence Hukou, designating the registration status, "rural" or "urban," based on his residence and tried to provide unified social security benefits based on their contributions for all the citizens, irrespective of where they live. Today obviously, individuals have quite a freedom to move or migrate from one area to another due to the household registration reform. But rural populations still do not have perfect access to many urban public services, such as education and healthcare, which leaves a large number of rural children and elderly people behind in their hometowns.[24] Furthermore, under the current household registration system, once individuals with rural Hukou decide to transform to rural Hukou, they have to give up the ownership of their rural homes, lands, and other rights attached. With China's rapid urbanization, the market value of rural land has increased significantly, and rural Hukou, which is directly related to land rights, has acquired a higher monetary value, thus making rural individuals more inclined to retain their original Hukou.[25] These factors have led to a large group of rural individuals into cities in the form of "floating population" who work in urban areas but maintain rural Hukou and return to where they come from when they are done with their work. So Hukou is adopted here as a variable to designate urban and rural individuals.

The data on respondents' household registration status are drawn from the Person dataset. Registration status is defined to be 1 if the respondent has rural Hukou, and 0 if the respondent has urban Hukou. Table 1 shows, 74.7% of the respondents of the CFPS survey are rural and 25.3% of them are urban.

Control Variables

We control for 4 groups of variables, economic factors, standard of living, demographic statistics and habits, and social relationship. In detail, we take into consideration the effects of 23 factors, including income, wealth, debt, medical insurance, employment, work intensity, housing, ventilation, kitchen water, food expenditure, healthcare, age, gender, workout, habits of smoking and drinking, marriage, family bond, internet, faith, trust, charity and helping behavior.

There are 6 variables controlled for in the group of economic factors. Data on per capita family income, family deposit and family debt, are drawn from the Family dataset. We extract these data from the Family dataset to match to each respondent in the Person dataset based on family codes. Per capita family income is defined to be 1 if the original value is greater than mean, and 0 if otherwise. Family deposit is used as a proxy variable to measure wealth, and take on the value 1 if the original value is greater than mean, and 0 if

otherwise. Family debt is defined to be 1 if the family owe debts, and 0 if otherwise. Data on medical insurance, employment and work intensity, are drawn from the Person dataset. Medical insurance is defined to be 1 if the respondent has medical insurance, and 0 if otherwise. Employment takes on the value 1 if the respondent is employed or at school, and 0 if otherwise. Work intensity is defined to be 1 if the weekly working hour is greater than mean, and 0 if otherwise. Table 1 reports the means of these 6 variables. Compared to rural population, urban population has higher per capita family income and family deposit, less family debt and lower work intensity. To be specific, 55.6% of urban population has higher per capita family income than average, but the proportion for rural population is as low as 20.5%. 35.2% of urban population and 16.1% of rural population has higher family deposit than average. 14.5% of urban population and 25.2% of rural population owes debts. 68.4% of urban population and 79.4% of rural population is working or studying. 32.2% of urban population and 49.4% of rural population has longer working hours than average. The statistics tell us that urban population on average has a better economic background than rural population.

There are 5 variables controlled for in the group of standard of living. Data on these 5 variables, housing, ventilation, kitchen water, food expenditure, and healthcare, are drawn from the Family dataset and matched with the Person dataset. Housing is defined to be 1 if the family own the house they are living in. Ventilation takes on the value 1 if a central ventilation system or an air cleaner is used. Kitchen water takes on the value 1 if it is processed. Food expenditure takes on the value 1 if the ratio of restaurant dining expenses to family income are greater than mean, and 0 otherwise. Healthcare takes on the value 1 if the ratio of healthcare expenses to family income are greater than mean, and 0 otherwise. As shown in Table 1, 85.8% of urban population and 88% of rural population owns housing. 9.6% of urban population and 2.9% of rural population uses a new ventilation system or an air cleaner. 93.6% of urban population and 69.5% of rural population uses processed kitchen water. 35.3% of urban population and 24% of rural population has higher restaurant dining expenses than average. 21.4% of urban population and only 9% of rural population has higher healthcare expenses-income ratio than average. We find that urban population has better housing conditions, spends more on healthcare and dining in restaurants than rural population. Therefore, on the whole, urban population has higher quality of living standard than rural population.

There are 5 variables controlled for in the group of demographic statistics and habits. Data on these 5 variables, age, gender, workout, smoking, and drinking, are drawn from the Person dataset. Age is defined to be 1 if the respondent is over the age 16 to 55, and 0 if otherwise. Gender takes on the value 1 if the respondent is male, and 0 for female. Workout is 1 if workout time is longer than mean, and 0 if otherwise. Smoking is defined to be 1 if the respondent

smokes more cigarettes than mean, and 0 if not. Drinking is 1 if the respondent drinks alcohol more than 3 times a week, and 0 if otherwise. Compared to urban population, rural population has a slightly higher proportion of heavy smokers, alcohol drinkers and young people, and much less workout time. Table 1 shows, 14.6% of rural population and 12.2% of urban population smokes more than average which is 15 cigarettes a day. 14.9% of rural population and 13.7% of urban population drinks alcohol more than 3 times a week. 81% of rural population and 78.3% of urban population is over the age 16 to 55. 33.5% of rural population and 48.9% of urban population has longer workout time than average. 49.5% of rural population and 49.3% of urban population is male.

There are 7 variables controlled for in the group of social relationship. Data on these 7 variables, marriage, family bond, internet, faith, trust, charity, and helping behavior, are drawn from the Person dataset. Marriage takes on the value 1 if the respondent is in marriage or lives with a partner, and 0 if the respondent is unmarried, divorced or lost his partner. Family bond takes on the value 1 if the respondent contacts his parents or offspring more than twice a week, and 0 if otherwise. Internet takes on the value 1 if the respondent is used to using internet or Wi-Fi, and 0 if not. Faith is defined to be 1 if the respondent believes in the Buddha or the Bodhisattva, the Taoist Gods, Allah, Jesus, ancestors, ghosts, or Feng Shui (Feng Shui is also called Chinese Geomancy. The Chinese words "feng" and "shui" translate to mean "wind" and "water," respectively. The philosophy of Feng Shui is a practice of arranging the pieces in living spaces in order to create balance with the natural world. The goal is to harness energy forces and establish harmony between an individual and their environment.), and 0 if otherwise. Trust is defined to be 1 if the respondent believes that most people can be trusted, and 0 if the respondent believes that one should be careful about others. Charity is defined to be 1 if the respondent made donations in the past 12 months, and 0 if otherwise. The attitude to helping behavior is defined to be 1 if the respondent believes that most people are willing to help others, and 0 if the respondent believes that people by nature are basically selfish. Table 1 tells that rural population has better companionship and more likely has faith, but urban population is more positive to trust, helping behaviors and charity. 79.8% of rural population and 78.9% of urban population lives with partners. 99.2% of rural population and 98.5% of urban population uses internet in daily life. 53.3% of urban population and 44.5% of rural population contacts their parents or children more than twice a week. 59.8% of urban population and 54.3% of rural population believes that most people can be trusted. 70.5% of urban population and 69.4% of rural population believes most people are willing to help others. 33.4% of urban population and 22.2% of rural population made donations in the past 12 months.

The summary statistics are reported in Table 1.

Results

The Influence of Household Registration Status on Inpatient Care

In Table 2, we report the findings about the relationship between inpatient care and household registration status after controlling for further covariates. We find evidence— robust to econometric techniques of LPM and probit—that inequality in health between rural and urban population exists in China. The concern of interest here is the response probability, so we choose the results of probit for our analysis.

Table 2 shows that the effect of household registration on individuals' health is both statistically and economically significant. The probability to be admitted to hospital is about 1.4% higher for people with rural Hukou than people with urban Hukou. Because we have controlled for most economic and demographic variables, the 1.4% differential cannot be explained by the differences in the above factors between rural and urban population. We can conclude that the differential is due to household registration status or factors associated with household registration status that we have not controlled for in the above model. In this study we couldn't take into account all possible factors that determine individuals' health, such as access, quality, and utilization of clinical care, eating habits and options, air pollution, etc. These factors are related to rural and urban areas and contributing to differences in health.[11,26,27]

Controls. As Table 2 shows, regardless of different estimation methods, the findings for the control variables are quite similar. First, higher income, having medical insurance, having a job or schooling, younger age, drinking alcohol, males, close family bond, a positive attitude to trust and helping behavior are positively related to individuals' health, which means individuals with either of the above features have a higher probability to be healthy. Second, owing debts, and in-marriage or living with a partner are robustly negatively associated with individuals' health. People with either of the 2 features have a lower probability to be healthy. Third, higher family deposit, longer working time, the ownership of houses, a central ventilation system or an air cleaner, processed kitchen water, higher restaurant dining expenses, higher healthcare expenses, longer workout time, heavy smoking, use of internet, having faith, and a positive attitude to charity have no statistically significant effects on individuals' health.

Among the above factors, we need to discuss further the following ones, smoking, drinking, marriage, gender, workout, and faith. First, as shown by Table 2, smoking yields an expected negative value (-0.004) but it is not robustly associated with individuals' health indicated by the high *P*-value (.504), which runs counter to our expectation that smoking is bad to people's health. In order to test the robustness of this finding, we generate a new variable, Smoking2, which

Table 2. Inpatient Care and Household Registration.

Variable	(1)	(2)	(3)	(4)
	Inpatient care			
Registration	–0.012** (0.005)	–0.012*** (0.005)	–0.013** (0.005)	–0.014** (0.005)
Per capita family income	0.012*** (0.005)	0.015*** (0.004)	0.013*** (0.005)	0.016*** (0.005)
Family debt	–0.037*** (0.005)	–0.037*** (0.005)	–0.035*** (0.004)	–0.035*** (0.004)
Medical insurance	0.037*** (0.006)	0.038*** (0.006)	0.046*** (0.008)	0.046*** (0.008)
Employment	0.034*** (0.006)	0.035*** (0.005)	0.032*** (0.005)	0.034*** (0.005)
Age	0.085*** (0.006)	0.086*** (0.006)	0.071*** (0.005)	0.072*** (0.005)
Drinking	0.022*** (0.005)	0.022*** (0.006)	0.02*** (0.006)	0.02*** (0.006)
Gender	0.009** (0.004)	0.01** (0.004)	0.011** (0.004)	0.011** (0.004)
Marriage	–0.027*** (0.005)	–0.026*** (0.005)	–0.03*** (0.006)	–0.03*** (0.006)
Family bond	0.015*** (0.004)	0.015*** (0.004)	0.016*** (0.004)	0.016*** (0.004)
Trust	0.008** (0.004)	0.009** (0.004)	0.009** (0.004)	0.009** (0.004)
Helping behavior	0.009** (0.004)	0.009* (0.005)	0.009** (0.004)	0.009** (0.004)
Charity	–0.007* (0.004)	–0.007 (0.004)	–0.008* (0.005)	–0.008 (0.006)
Healthcare	–0.01* (0.006)	–0.009 (0.006)	–0.01* (0.006)	–0.009 (0.006)
Family deposit	0.006 (0.005)		0.006 (0.005)	
Work intensity	0.006 (0.004)		0.006 (0.004)	
Housing	–0.005 (0.006)		–0.007 (0.006)	
Ventilation	0.007 (0.009)		0.007 (0.01)	
Kitchen water	0.001 (0.005)		0.001 (0.005)	
Food expenditure	0.004 (0.004)		0.005 (0.005)	
Workout	–0.008 (0.006)		–0.07 (0.006)	
Smoking	–0.001 (0.006)		–0.004 (0.006)	
Internet	–0.012 (0.017)		–0.012 (0.021)	
Worship	–0.004 (0.006)		–0.003 (0.005)	
Estimation technique	LPM	LPM	Probit	Probit
F/Wald chi^2	27.78	40.87	715.03	704.52
Pro>F/Pro>chi^2	0.0000	0.0000	0.0000	0.0000
Observations	25 303	25 303	25 303	25 303

Note. Average marginal effects and Delta-method standard errors in parentheses for probit estimation; Coefficients and Robust standard errors in parentheses for LPM estimation. Columns (2) and (4) report the results of regression excluding insignificant variables in (1) and (3).
*$P<.1$. **$P<.05$. ***$P<.01$.

is defined to be 1 if the respondent smokes, and 0 if otherwise, to redo the regression. The new regression results show that the average marginal effect of smoking2 is 0.004 and the P-value is .482. There is still no clear-cut evidence that smoking leads to bad health. This result may arise from the measurement that we quantify health state by whether the respondent was admitted to hospital. The influence of smoking on individuals' health may not come so far to cause smokers being admitted even if smoking indeed influences individuals' health in some way. We will discuss this in the next section concerning the effects of smoking on self-assessed health. Second, contrary to popular belief and our expectation, Table 2 indicates that drinking alcohol more than 3 times a week raises the probability to be healthy by 2%. One explanation for this is that people in China usually drink together with a group of relatives or friends, from which the gossips, family love, and friendship may benefit people's health. Third, the negative value on marriage (−0.03) tells that being married or living with a partner harms

individuals' health, which is quite contrary to what we expected. We need to find what happens with our marriage if this finding stands. One hypothetical cause of this may lie in the differences of age between married people and unmarried people. Unmarried people are on average younger than married people because the Marriage Law of China have requirements of age for people to be married, which is above 22 for males and 20 for females. To test the hypothesis, we drop observations of males under 22 and females under 20, and redo the regression. The new regression gives the average marginal effect of marriage (−0.015) and the P-value (.016). The above finding about marriage and health still stands. Fourth, we also need to mention the finding that males have a higher probability to be healthy by 1.1% than females. The causes behind the differential need to be looked into by researchers and policy makers. Fifth, the negative relationships between workout, faith, and health are not as what we expected though explicable, which indicates that individuals suffering from bad health are more motivated to work out

Table 3. Self-assessed Health and Household Registration.

Variable	(1)	(2)	(3)	(4)
	Self-assessed health			
Registration	0.013* (0.007)	0.013** (0.007)	0.016** (0.007)	0.017** (0.007)
Per capita family income	0.047*** (0.006)	0.046*** (0.006)	0.051*** (0.007)	0.05*** (0.007)
Family deposit	0.023*** (0.007)	0.023*** (0.007)	0.023*** (0.007)	0.023*** (0.007)
Family debt	−0.034*** (0.007)	−0.034*** (0.007)	−0.033*** (0.006)	−0.034*** (0.006)
Medical insurance	0.016* (0.009)	0.016* (0.009)	0.019** (0.01)	0.019** (0.01)
Employment	0.029*** (0.007)	0.032*** (0.007)	0.029*** (0.007)	0.031*** (0.007)
Ventilation	0.047*** (0.012)	0.047*** (0.012)	0.051*** (0.014)	0.05*** (0.014)
Kitchen water	0.023*** (0.007)	0.023*** (0.007)	0.021*** (0.006)	0.021*** (0.006)
Food expenditure	0.043*** (0.006)	0.043*** (0.006)	0.045*** (0.006)	0.044*** (0.006)
Age	0.174*** (0.008)	0.175*** (0.008)	0.151*** (0.006)	0.152*** (0.006)
Workout	0.017*** (0.006)	0.016*** (0.006)	0.018*** (0.006)	0.017*** (0.006)
Gender	0.057*** (0.006)	0.056*** (0.006)	0.058*** (0.006)	0.058*** (0.005)
Marriage	−0.09*** (0.007)	−0.09*** (0.007)	−0.105*** (0.007)	−0.105*** (0.008)
Family bond	0.06*** (0.006)	0.06*** (0.005)	0.061*** (0.005)	0.061*** (0.005)
Worship	−0.031 (0.026)		−0.039 (0.030)	
Trust	0.039*** (0.006)	0.039*** (0.006)	0.039*** (0.006)	0.039*** (0.006)
Helping behavior	0.065*** (0.007)	0.065*** 0.007)	0.061*** (0.006)	0.062*** (0.006)
Charity	0.019*** (0.006)	0.019*** (0.006)	0.019*** (0.006)	0.019*** (0.006)
Healthcare	−0.014 (0.009)		−0.014 (0.009)	
Drinking	0.016 (0.011)		0.012 (0.008)	
Work intensity	0.01 (0.006)		0.007 (0.006)	
Housing	0.008 (0.008)		0.004 (0.008)	
Smoking	−0.009 (0.009)		−0.014 (0.009)	
Internet	−0.025 (0.024)		−0.023 (0.029)	
Estimation technique	LPM	LPM	Probit	Probit
F/Wald chi^2	100.3	119.9	1193.15	1984.33
Pro>*F*/Pro>chi^2	0.0000	0.0000	0.0000	0.0000
Observations	25 303	25 303	25 303	25 303

Note. Average marginal effects and Delta-method standard errors in parentheses for probit estimation; Coefficients and Robust standard errors in parentheses for LPM estimation. Columns (2) and (4) report the results of regression excluding insignificant variables in (1) and (3).
*$P<.1$. **$P<.05$. ***$P<.01$.

and worship. To account for the potential endogeneity arising from reverse causality between workout, faith and health, we use the lagged data of workout and faith from the 2016 CFPS survey matching with the 2018 database based on person IDs to do the regression. The average marginal effects and the p-values for workout and worship are −0.008 and 0.018, and −0.004 and 0.429, which means the relationships between workout, worship and health are still negative and insignificant.

The Influence of Registration Status on Self-Assessed Health

Table 3 displays the estimation results of equation (2) using probit, and equation (3) using LPM, with self-assessed health as the dependent variable. Due to the same reason as section 4.1, we use the results of probit for our analysis.

As Table 3 shows, the positive and significant average marginal effect of household registration on individuals'

self-assessed health indicates that individuals with rural Hukou have a higher probability by 1.7% to assess themselves healthy than individuals with urban Hukou. Considering that we have controlled for a number of economic and demographic variables, the differential of 1.7% in probability may reveal the differences in attitudes of rural and urban population toward health. For one thing, urban population may have better knowledge of their health than rural population due to more frequent health checks and professional diagnoses. For another thing, urban population may have higher standards than rural population to claim themselves healthy. The CFPS survey shows that 62% of urban population goes to general hospitals or specialty hospitals, but the proportion of rural population is as low as 33%, and 67% of rural population chooses to go to small hospitals, such as township hospitals or village clinics.

We can never neglect the 1.7% differential in self-assessed health between rural and urban population because maybe it is the over-optimism of their health that partly

leads to the 1.4% higher probability discussed in the above section for rural individuals to be admitted to hospital than their urban counterparts. Therefore we can conclude that the over-optimism of health of rural population also reflects deep concerns in health inequality between rural and urban population.

Controls. First, higher income, higher deposit, having medical insurance, having a job or schooling, using a central ventilation system or an air cleaner, using processed kitchen water, higher restaurant dining expenses, younger age, workout, males, close family bond, a positive attitude to trust, helping behavior, and charity are significantly and positively related to individuals' self-assessed health. Second, owing debts, and in marriage or living with a partner are significantly and negatively related to individuals' self-assessed health. Third, the ownership of housing, longer working time, higher healthcare expenses, drinking alcohol, heavy smoking, having faith, and use of internet have no statistically significant effects on individuals' self-assessed health.

Compared to the effects of covariates on individuals' health, on one hand, drinking alcohol now becomes statistically insignificant. On the other hand, family deposit (0.023), ventilation (0.05), kitchen water (0.021), and food expenditure (0.044) now become statistically significant, which means that individuals with higher family deposit, higher restaurant dining expenses, using a central ventilation system or an air cleaner, and using processed kitchen water, have a higher probability by 2.3%, 5%, 2.1%, and 4.4% respectively, to believe themselves healthy even though the 4 factors have no statistically significant effects on their inpatient care.

Seemingly similar to the effects on health, smoking still yields the expected negative value (-0.014) but is still not robustly related to self-assessed health. However, the p-value reported for smoking here is 0.106, which means that at the 11% statistical significance level we find evidence that smoking is negatively related to self-assessed health. That is, smoking more than 15 cigarettes a day will lower down the probability by 1.4% for individuals to assess themselves healthy. Combining the finding here and the finding about the effect of smoking on health in section 4.1, we can conclude that heavy smoking does harm individuals' health though the influence is not reflected on the probability for smokers to be admitted to hospital.

Discussion

Endogeneity Discussion

The problem of endogeneity arises mainly from 2 causes, omitted variables and simultaneity. As shown in Table 1, we have already controlled for as many variables as possible to avoid the endogeneity problem of omitted variables. We tried to take into our regression the factor, birthweight, but there

are only 4371 respondents in the CFPS survey know/remember their birthweights, which means we will lose the valuable information of the other 20 932 observations. So we chose to exclude the variable, birthweight, in our regression process above to make full use of and extract valuable information from the data. For comparison, we have used the 4371 observations to regress the state of health on covariates including birthweight. Table 4 displays the regression results of eq. (2). Columns (1) and (4) replicate the results shown in column (3) of Tables 2 and 3. Columns (2) and (5) show the estimation results taking birthweight into account. The results in columns (2) and (5) are on average similar to those in columns (1) and (3). And birthweight itself is insignificantly associated with individuals' inpatient care, but significantly and positively related to individuals' self-assessed health.

Simultaneity arises when explanatory variables are jointly determined with the dependent variable. First, based on what we introduced about the Hukou system in China, the main explanatory variable, household registration (Hukou) status, is basically predetermined and exogenous. Therefore, the problem of biased and inconsistent estimators will not arise from endogeneity of registration status. Second, simultaneity may arise between workout, worship and health. To deal with the potential endogeneity, we use the lagged data of workout and worship from the 2016 CFPS survey to replace the 2018 data. After matching with the 2018 database based on person IDs, we redo the regression using the 21 169 observations left. The estimation gives similar results on worship and workout which are presented in section 4.1.

Other Robustness Checks

We have carried out several measures and other robustness checks which we now describe. First, with the large sample size of 25 303 observations, we report the heteroscedasticity-robust standard errors and the heteroskedasticity-robust *F*/Wald statistics in Tables 2 and 3 to address the issue of heteroscedasticity. Second, we use 2 estimation methods, LPM and probit to estimate model equation (1), and we use the regression results of probit for our analysis considering the properties of binary explained variables. As shown by Tables 2 and 3, the findings for covariates are quite similar irrelevant of the different methods we use. Third, we have tried to minimize the impact of measurement errors and the issue of outliers by transforming the variables in equation (1) into binary variables and get the results of Tables 2 and 3. For comparison, we have also done the regression of equation (2) using the original values of covariates after we delete outliers and observations with errors detected by the econometric software package. Among these covariates, household registration, family debt, medical insurance, employment, housing, ventilation, kitchen water, drinking, gender, marriage, family bond, faith, trust, helping behavior, charity, and internet are the same binary variables as used above, and per capital family income, family deposit, work intensity (measured

Table 4. Endogeneity and Other Robustness Checks.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
		Inpatient care			Self-assessed health	
Birthweight		0.007 (0.004)			0.021*** (0.006)	
Registration	−0.013** (0.005)	−0.016*** (0.002)	−0.019*** (0.006)	0.016** (0.007)	0.039** (0.016)	0.017** (0.008)
Per capita family income	0.013*** (0.005)	0.015*** (0.002)	1.94e−08** (8.38e−09)	0.051*** (0.007)	0.040** (0.016)	5.31e−07*** (1.70e−07)
Family debt	−0.035*** (0.004)	−0.033*** (0.011)	−0.036*** (0.005)	−0.033*** (0.006)	−0.018*** (0.006)	−0.045*** (0.007)
Medical insurance	0.046*** (0.008)	0.045** (0.020)	0.053*** (0.010)	0.019* (0.011)	0.025** (0.011)	0.024* (0.012)
Employment	0.032*** (0.005)	0.030*** (0.013)	0.050*** (0.009)	0.029*** (0.007)	0.051*** (0.017)	0.062*** (0.013)
Age	0.071*** (0.005)	0.068*** (0.000)	−0.003*** (0.000)	0.151*** (0.006)	0.104*** (0.015)	−0.008*** (0.000)
Drinking	0.020*** (0.006)	0.016** (0.004)	0.023*** (0.007)	0.012 (0.008)	0.009 (0.019)	0.022 (0.029)
Gender	0.011** (0.004)	0.006*** (0.001)	0.014*** (0.005)	0.058*** (0.006)	0.057*** (0.015)	0.056*** (0.008)
Marriage	−0.030*** (0.006)	−0.031** (0.014)	−0.013** (0.006)	−0.105*** (0.007)	−0.082*** (0.019)	−0.016*** (0.006)
Family bond	0.016*** (0.004)	0.022** (0.010)	0.015*** (0.004)	0.061** (0.005)	0.067*** (0.013)	0.016*** (0.006)
Trust	0.009** (0.004)	0.012 (0.101)	0.006* (0.003)	0.039*** (0.006)	0.034*** (0.014)	0.035** (0.007)
Helping behavior	0.009** (0.004)	0.009 (0.010)	0.006* (0.003)	0.061** (0.006)	0.092*** (0.014)	0.053*** (0.007)
Charity	−0.008* (0.005)	−0.001 (0.011)	−0.014** (0.005)	0.019*** (0.006)	0.048*** (0.016)	0.018*** (0.006)
Healthcare	−0.010* (0.006)	−0.008 (0.015)	−0.165 (0.107)	−0.014 (0.009)	−0.040 (0.102)	−0.004 (0.126)
Family deposit	0.006 (0.005)	0.032** (0.013)	2.75e−08 (3.37e−08)	0.023*** (0.007)	0.019*** (0.003)	2.35e−08** (1.14e−08)
Work intensity	0.006 (0.004)	0.003 (0.010)	0.000 (0.000)	0.007 (0.006)	0.002 (0.014)	0.000 (0.000)
Housing	−0.007 (0.006)	0.025 (0.016)	−0.007 (0.007)	0.004 (0.008)	0.015 (0.021)	0.015 (0.009)
Ventilation	0.007 (0.010)	0.019 (0.026)	0.009 (0.012)	0.051*** (0.014)	0.053*** (0.016)	0.037** (0.017)
Kitchen water	0.001 (0.005)	0.011 (0.010)	0.002 (0.005)	0.021*** (0.006)	0.016*** (0.001)	0.023*** (0.007)
Food expenditure	0.005 (0.005)	0.006 (0.011)	0.006 (0.016)	0.045*** (0.006)	0.048*** (0.015)	0.109*** (0.032)
Workout	−0.007 (0.006)	−0.015 (0.021)	−0.002 (0.002)	0.018*** (0.006)	0.033** (0.013)	0.004*** (0.001)
Smoking	−0.004 (0.000)	−0.002 (0.014)	−0.000 (0.000)	−0.014 (0.009)	−0.014 (0.019)	−0.000 (0.000)
Internet	−0.012 (0.021)	−0.009 (0.057)	−0.004 (0.024)	−0.023 (0.029)	−0.098 (0.090)	−0.036 (0.036)
Worship	−0.003 (0.005)	−0.019 (0.016)	−0.021 (0.020)	−0.039 (0.030)	−0.035 (0.026)	−0.014 (0.032)
Estimation technique	Probit	Probit	Probit	Probit	Probit	Probit
Wald chi^2	715.03	137.65	524.79	1193.15	318.63	1574.98
Pro>chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	25303	4371	18091	25303	4371	18091

Note. Average marginal effects and Delta-method standard errors in parentheses.
*$P < .1$. **$p < .05$. ***$p < .01$.

by weekly working hours), food expenditure (measured by the ratio of restaurant dining expenses to family income), healthcare (measured by the ratio of healthcare expenses to family income), age, workout (measured by weekly workout times), and smoking (measured by the number of cigarettes smoked 1 day), take on the original values of the CFPS survey. Columns (3) and (6) of Table 4 display the probit estimation results using the original values of covariates. Because of the change in units of measurement of the covariates, there is no point to compare the absolute values of average marginal effects for covariates. But the same signs and similar statistical significance of the covariates show the robustness of our findings in Tables 2 and 3. Although the signs of age are opposite, they show the same effect of age on individuals' health, which is that the younger the individuals, the healthier they are.

Limitations

We now note some limitations of our analysis. The data used to measure self-assessed health are self-reported by respondents and could be subject to errors and misinterpretations in reporting. As explained in the Methods section, the measurement of self-assessed health is based on how the respondent answer question QP201 in the questionnaire of CFPS. Question QP201 states, "How would you rate your health status," below which, there is a note for interviewers, "Do not explain the concept of health. Record according to the respondent's own opinion." We can deduce that respondents rate their health based on their own perception of "health." Besides, there are still missing factors in our framework due to the availability of data even though we have controlled for 23 variables. Factors, such as pollution control and local medical facilities, haven't been introduced into our model, which may leads to a less sound conclusion about the relationship between health and household registration. Therefore, future studies may wish to deal with these issues by using different statistical techniques and databases.

Conclusion

As Marmot et al argued,[28] the health of the population is not just a matter of how well the health service is funded and functions, but also the conditions in which people are born, live, work, and age, and inequities in resources. Health is a good measure of social and economic progress. When a society has large social and economic inequalities, it also has large inequalities in health.

The objective of this paper is to find out whether inequality in health exists between urban and rural population in China. We find evidence that the probability to be admitted to hospital for individuals with rural Hukou is higher by 1.4% than urban Hukou, whereas there is a tendency of overestimating their health for individuals with rural Hukou. The findings raise the alarm for policy makers to take serious the inequality in health between rural and urban population and take measures to deal with the problem.

Author Contributions

LFG designed the structure of the paper, and did the regression using Stata. CYP processed the original data and formatted them into binary variables. Material Preparation and data collection were performed by YFJ.

Data Availability

The datasets generated and/or analyzed during the current study are available in the China Family Panel Studies (CFPS) repository, http://www.isss.pku.edu.cn/cfps/en/index.htm?CSRFT=9LRP-W6RR-KHJJ-JOWT-RSZR-09D3-2JLB-SJ1W.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Feng-Gang Lu https://orcid.org/0000-0001-5131-5641
Yin-Ping Chen https://orcid.org/0000-0002-7752-8124

Ethical Approval

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Academic and Ethics Committee of SDTBU (AEC2022-PA-02).

Informed Consent

Consent was obtained from participants prior to the interviews.

References

1. Braveman P. Health disparities and health equity: concepts and measurement. *Annu Rev Public Health*. 2006;27(1):167-194.
2. Singh GK, Daus GP, Allender M, et al. Social determinants of health in the United States: addressing major health inequality trends for the nation, 1935-2016. *Int J MCH AIDS*. 2017;6(2):139-164.
3. Anderson TJ, Saman DM, Lipsky MS, Lutfiyya MN. A cross-sectional study on health differences between rural and non-rural U.S. counties using the county health rankings. *BMC Health Serv Res*. 2015;15(441):441-448.
4. Chaudhary MA, Shah AA, Zogg CK, et al. Differences in rural and urban outcomes: a national inspection of emergency general surgery patients. *J Surg Res*. 2017;218:277-284.

5. Abedi V, Olulana O, Avula V, et al. Racial, economic, and health inequality and COVID-19 infection in the United States. *J Racial Ethn Health Disparities*. 2021;8:732-742.

6. Marmot M. Health equity in England: the Marmot review 10 years on. *BMJ*. 2020;368:m693.

7. Ginja S, Jackson K, Newham JJ, Henderson EJ, Smart D, Lingam R. Rural-urban differences in the mental health of perinatal women: a UK-based cross-sectional study. *BMC Pregnancy Childbirth*. 2020;20(1):464-514.

8. Qin VM, McPake B, Raban MZ, et al. Rural and urban differences in health system performance among older Chinese adults: cross-sectional analysis of a national sample. *BMC Health Serv Res*. 2020;20(1):372-414.

9. Aday A, Andersen R. A framework for the study of access to medical care. *Health Serv Res*. 1974;9(3):208-220.

10. Vedom J, Cao H. Health care access and regional disparities in China. *Espace Popul Soc*. 2011;1:63-78.

11. Wang Z, Chen Y, Pan T, Liu X, Hu H. The comparison of healthcare utilization inequity between URRBMI and NCMS in rural China. *Int J Equity Health*. 2019;18:90.

12. Graves JM, Mackelprang JL, Moore M, et al. Rural-urban disparities in health care costs and health service utilization following pediatric mild traumatic brain injury. *Health Serv Res*. 2019;54(2):337-345.

13. Chen X, Orom H, Hay JL, et al. Differences in rural and urban health information access and use. *J Rural Health*. 2019;35(3):405-417.

14. Qasim M, Pervaiz Z, Chaudhary AR. Do poverty and income inequality mediate the association between agricultural land Inequality and human development? *Soc Indic Res*. 2020; 151(1):115-134.

15. EPA. *Final Report: Integrated Science Assessment for Particulate Matter*. US EPA; 2009.

16. EPA. *Final Report: Integrated Science Assessment of Ozone and Related Photochemical Oxidants*. US EPA; 2013.

17. Wingard DL, Berkman LF, Brand RJ. A multivariate analysis of health-related practices: A nine-year mortality follow-up of the Alameda county study. *Am J Epidemiol*. 1982;116(5): 765-775.

18. Matthews KA, Croft JB, Liu Y, et al. Health-related behaviors by urban-rural county classification—United States, 2013. *Morb Mortal Wkly Rep*. 2017;66(5):1-8.

19. Liu Y, Croft JB, Wheaton AG, et al. Clustering of five health-related behaviors for chronic disease prevention among adults, United States, 2013. *Prev Chronic Dis*. 2016;13:E70.

20. McFadden D. *Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka P. Frontiers in Econometrics*. Academic Press; 1973.

21. Cheng T, Selden M. The origins and social consequences of China's *Hukou* system. *China Q*. 1994;139:644-668.

22. Chan KW, Zhang L. The Hukou system and rural-urban migration in China: processes and changes. *China Q*. 1999;160: 818-855.

23. Chan KW. The household registration system and migrant labor in China: notes on a debate. *Popul Dev Rev*. 2010;36(2): 357-364.

24. Afridi F, Li SX, Ren Y. Social identity and inequality: the impact of China's Hukou system. *J Public Econ*. 2015;123: 17-29.

25. Chen CB, Fan C. China's Hukou puzzle: why don't rural migrants want urban Hukou? *China Review*. 2016;16(3): 9-39.

26. Ricketts TC. The changing nature of rural health care. *Annu Rev Public Health*. 2000;21(1):639-657.

27. Liese AD, Weis KE, Pluto D, Smith E, Lawson A. Food store types, availability, and cost of foods in a rural environment. *J Am Diet Assoc*. 2007;107(11):1916-1923.

28. Marmot M, Atkinson T, Bell J. *Fair Society, Healthy Lives: The Marmot Review*. Department for International Development; 2010. Accessed April 21, 2021. https://www.gov.uk/research-for-development-outputs/fair-society-healthy-lives-the-marmot-review-strategic-review-of-health-inequalities-in-england-post-2010