Urbanization and Health Outcomes in China: Analysing the Effect of Health Care Reform and Rural-Urban Migration

A thesis submitted by

Yunrui Bai

in partial fulfillment of the requirements for the degree of

Master of Science

in

Economics

Tufts University

May, 2023

Advisor: Professor Jeffrey Zabel

Abstract

This thesis focuses on urban and rural health outcomes that develop differently because of exogenous factors such as policy reform and local development. With DID estimation, this thesis highlights that the New Rural Cooperative Medical Insurance holders, who are typically rural residents without urban employment, would see a 0.115 increase in the likelihood of reporting good objective health measure in six years after the reform. This increase would correspond to the semi-elasticity of 0.171. The study also considers a migration model to establish an upper bound of the impact of rural-to-urban migration on health outcomes. An OLS estimation suggests that the rural-urban migration would be associated with negative change of 0.120 in the good objective health status on average controlling for individual fixed effect. And the economic significance is characterised by the semi-elasticity of -0.162. These results suggest that the medical insurance reform, designed to increase medical insurance benefits for rural residents, actually observed the effect; and that the worsen health status for rural migrants are worth noting for policy makers.

Contents

List of Tables

List of Figures

1 Introduction

The health conditions for urban and rural residents can be very different in many countries. As suggested by Vlahov and Galea (2002), it can be due to how urbanization affects people's health through the accessibility to health services and the environmental factors hypothetically. The accessibility to health services can be boiled down into the different national medical insurance and medical services' qualities and densities for urban and rural residents, respectively. And the urban environmental factors include air quality, population density, and hygiene conditions etc. In China, these disparities became especially interesting not only because of the steep growth in the urbanized population over the past decades, but also stemming from its unique system of basic medical insurance and the system of place of registered residence (also called hukou) closely associated with it.

This study explores the question of how the urbanization process is related to people' health. The Difference-in-Difference approach was taken, motivated by the staggered health care reform in China basic medical insurance system, and a migration model is used to take advantage of the steep growth in the urbanized population. The main hypothesis is that improved healthcare accessibility, manifested in the upgrade of medical insurance coverage, would increase the health condition of the benefited group. Since the policy reform primarily increase the coverage for previous NCMI holders, and not causing URMI and UEMI benefit to change, the hypothesized benefited group is NCMI holders. And the rural-to-urban migration, ideally being unrelated to the baseline health condition, would worsen the health of the migrants despite the marginal improvement in the health care accessibility. Therefore, the two hypotheses would constitute the analysis to how urbanization in the population would affect health and how public policy could help with narrowing the gap between that in urban and rural areas.

A panel data model is implemented to support the assumption that the medical effect is drastically causing the healthcare expenditure disparity between rural and urban areas. Medical resources could be one of the determinants of the local urbanization level, and the local urbanization level could also work the other way around to affect the health service resources and the policy aspect of basic medical insurance. To be more specific, the CHNS-specified urbanization index contains twelve measures, one of which is the medical service measures. As for the policy aspect of medical insurance that is affected by the urbanization level, New Rural Cooperative Medical Insurance (NCMI) holders would normally be eligible with 50%-70% of the reimbursement rate for only the hospitalization cost, while the other types of insurance, i.e. urban resident medical insurance (URMI) and urban employee medical insurance (UEMI), typically have the reimbursement rate of 70%-80% for not only the hospitalization cost, but also outpatient care and prescription drugs and so on. The medical insurance policy gap is only built on the dichotomy division of the urban and rural designation, and this paper is further interested in how health services utilization would be related to the continuous change on the rural-urban spectrum. Therefore, I conduct a regression of medical insurance utilization on the local urbanization level. This does not imply there is a causal relationship assumed, but rather with the aim of capturing their relationship. This result serves as a supplementary discussion of the potential mechanism that urbanization channeling through the medical insurance to affect residence health.

Analyzing the medical effect requires controls for the same residential environment for either the urban or rural designation, and then estimating the difference across people with different types of medical insurance. I implement the Difference-in-Difference (DID) method to model the effect from staggered and non-health-related policy changes. While for the environmental effect, ideally the people would enjoy the same sets of medical resources but live in different residential settings in terms of urban and rural areas. This is rather hard to model since with empirical findings using data from the China Migrants Dynamic Survey (CMDS) and anecdotal evidence, there is no consistent pattern of changing the insurance plan when it comes to one's migration, typically from rural to urban. From the regulation perspective, rural residents are free to choose to enroll in NCMI since they have rural hukou, or UEMI as long as they find a job at a company that contributes a portion of the premium cost. And urban residents will be enrolled in the UEMI or URMI according to their employment status. Hence there is no accurate prediction from the transition of hukou type and employment status to their residential places and insurance type. Therefore, health outcomes are inevitably driven both by environmental factors and medical insurance factors. These two factors are hypothesized to work in opposite directions. Taking the migration from a rural to an urban area for example, the urban environment is generally considered as a jeopardize to people's health, while the health services accessibility would promote the health of people (Kawachi et al., 1997). The accessibility, due to the hukou and medical insurance system before 2009, is considered only marginally raising for rural-to-urban migrants. As the rural hukou holders would not enjoy as much medical insurance coverage as the urban hukou holders if they are not empolyed in cities. Thus this paper turns to explore the relationship between rural-urban migration on health outcomes using a migration model.

Ultimately, this thesis highlights the finding that the medical insurance reform, designed to increase medical insurance benefits for rural residents, actually observed the effect in six years; and that there are worsened health status for the rural migrants.

2 Policy Background

The National medical insurance system in China is the primary medical insurance program that people participate in. Sponsored by local governments, the insurance system has also gone through multiple transformations. Currently, the medical insurance is divided into two categories according to employment status, which is the employed and the unemployed. The latter is also called the resident basic medical insurance. Before the reform in 2009, the resident basic medical insurance was further divided for urban and rural residents. Dong (2009) reviewed the evolution of the public medical insurance program before 2009. Basic social medical insurance (BSMI) is the countrywide government system that serves as the primary third-party payer and the backbone for healthcare financing. A summary of the policy is in Table 1.

Before the 2009 reform, there were three schemes that segment people into three groups: rural residents, urban employed, and unemployed urban residents. But the system of BSMI did not cover all the residents. The three segmented scheme also differed by regions within China as local governments have the autonomy to make adjustments. In the 2009 reform plan, the government proclaimed that universal coverage through BSMI would be achieved by 2011. (General Office of the State Council, 2009)

The basic medical insurance scheme for urban employees evolved from the Free Medical Service program and the Labor Medical Service program, which date back to the planned economy era. Founded in 1952, the Free Medical Service program, which was financed by the central and local governments, was mainly designed for civil servants. The Labor Medical Service program was founded in 1951, covering employees in industrial enterprises. And this was financed by employers. Pilot reforms of the basic medical insurance scheme establishment were in 1994 and the full transformation was completed throughout the country in 1998. The four-year trial was performed in Zhenjiang City, Jiangsu Province and Jiujiang City, Jiangxi Province.

For rural residents, a scheme called the traditional rural cooperative healthcare system (TRCHS) was based on the People's Commune System. This scheme shouldered the role of the medical insurance counterpart for rural residents until the collapse of the People's Commune System in the early 1980s. Then a new health care scheme only started with a pilot trial in 2003 in hundreds of counties in four provinces. This new scheme was designed for rural areas was called the new rural cooperative medical insurance scheme (NCMI).

Unlike urban employees and rural residents, urban residents who were not employed only enjoyed public medical insurance before the 1980s when free urban medical services were active. Not until 2007 when the pilot reforms for the basic social medical insurance for urban residents program took place did the unemployed urban people enjoy public health insurance.

In 2009, establishing a consolidated health insurance system by 2020 was set as one of the main goals in China's health system reform agenda. That is, to integrate the medical insurance for urban residents, called as urban residents basic medical insurance (URMI), with NCMI. However, the progress was slow, mainly due to a lack of an agreement on which a suitable governance structure. A few provinces in China have piloted consolidation of the rural and urban schemes since 2013 (Meng et al., 2015). After the State Council's announcement of the *Notice on Printing and Distributing the Plan for Deepening the Reform of the Medical and Health Care System during the 13th Five-Year Plan* in 2016, the reform that is the integration of the basic medical insurance system has officially started throughout the country.(General Office of the State Council, 2017)

In the time when there are three main insurance schemes, NCMI differs the URMI and UEMI in the reimbursement rate and the type of cost eligible for the reimbursement. NCMI holders would normally be eligible with 50%-70% of the reimbursement rate for only the hospitalization cost URMI and UEMI typically have a reimbursement rate of 70%-80% for not only the hospitalization cost but also outpatient care and prescription drugs and so on. This is illustrated in more details in Table 1. While the URMI and UEMI does not have too much difference expect for the premium cost and the length of the insurance[1].

Hukou system reformation over the years is also an important factor to consider. The very first wording of the hukou system arose in 1951, and the strict administration of the urban-rural duality system started in 1958. Hukou contains two pieces of information. One is the place of residence, and this is specific to the house address. The other is the type of hukou, either agricultural or non-agricultural, which, for simplicity, are also known as urban or rural hukou. People face a constraint in rural-to-urban migration as well as migration from small cities to large cities. From this establishment until 1998, one's hukou type is directly determined by his/her mother's hukou type at the time of birth. Transition from rural to urban hukou is limited within three channels, all of which are controlled by annual quotas from the local or central government. The opportunity to transfer hukou type was essentially considered as a reward to people with certain values. The relaxation of the strict restriction for hukou change gradually started in about 1993. By this year there was 23 permitted ways for hukou transformation compared to 9 in the 1950s. Milestones in the reform of hukou system were marked by allowing the child's

[1]UEMI typically covers the medical insurance after the retirement if the male or female participant pays cumulativelyfor 25 or 20 more years, URMI only covers the year the participants pay for.

hukou type to follow the mother's or father's in 1998 [2] and the abolishment of the division of agricultural and non-agricultural hukou starting from 2000. The cancellation of the dichotomy of hukou type happens at different times for different provinces. This process was fully completed by 2016. After the 1980s, rural or agricultural hukou possessors could transfer to the non-agricultural type if they did not economically depend on agricultural activities and have residency in a urban area. This mainly resonated with the intention to develop small cities or towns, and therefore created a bunch of non-migrant hukou changes when their hometown changed designation (Chan and Xu, 1985).

3 Literature Review

Studies have been contributed to demonstrate that health outcomes can be shaped by social determinants, which is generally defined as the factors apart from medical care. As the review by Braveman and Gottlieb (2014) suggested, socioeconomic factors such as income and education could be the fundamental causes of health outcomes. This provides a concrete list of control variables that I include in my model. Medical care, one of the focuses of this paper, is subject to policy changes, and constantly affects the health outcomes for urban and rural residents in different manners.

The health-seeking behavior might be one of the causes of the urban and rural gap. Liu et al. (2003) used China Health and Nutrition Survey (CHNS) data in 1991 and 1993 to show that a higher urbanization level increases insurance coverage significantly and thus increases access to health care. Conducting a difference-in-difference estimation with the CHNS data from 1989 to 2006, the study by He and Nolen (2019) takes advantage of the urban employee medical insurance reform to show that a health insurance reform in China increased coverage for workers in Non-State Owned Enterprises and also led to better health outcomes. However, further analysis using the instrumental variable approach did not show significant improvements in health outcomes, healthcare utilization or out-of-pocket medical expenses for those reporting more preventive care. Building on

[2] In Guangzhou city alone, after one and a half year of this policy announcement, 7600 children was switched to non-agricultural hukou.

this, this paper examines the impact of the basic medical health insurance policy reform with another data set of the China Family Panel Study which enables the usage of more modern data till 2018. The update of data would enable more insight on the lagged effect on the health insurance reform.

When analyzing the regional gap of health outcomes for urban and rural residents, the core issue is to deal with the endogeneity of one changing the residential regions between urban and rural. Some of the researchers solved this issue by avoiding the use of immigrants and using external policy change, others resorted to the hukou system to determine a plausible external migration.

Using the China Health and Retirement Longitudinal Study (CHARLS), Hou et al. (2019) identified urban non-migrants with a rural hukou as the people whose hometown at the time of their birth was a rural area that became urban in the process of urbanization during their lifetime. This constitutes what the author defined as in situ, involuntarily urbanized population. This population group is also addressed as urbanized rural residents. And the author argued that the differences between in situ urban-rural population and urban/rural population would be independent of selective migration to urban areas, but purely caused by urbanization. The result from OLS regression suggests that the urbanized rural residents enjoy health advantages and less depression controlling for demographic, early life, socioeconomic, psychosocial and behavioral factors, and to mitigate the concerns of subjecting categorical variable of healthy scale to OLS regression, this result is further tested robust to multinomial logit regressions.

The study by Song and Smith (2021) uses county-level data from the 2015 China Health and Retirement Longitudinal Study (CHARLS) with CHARLS 2014 life history data to analyze the impact of possessing an urban hukou on the psychological well-being. The key indicator is the residents' mental health. They compared rural-to-urban migrants (i.e., the hukou non-converters) with urban residents who had a hukou transition history from rural to urban (i.e. they were born in rural areas with a rural hukou who later obtained urban hukou); instead of directly comparing rural migrants with urban hukou. People who went through the hukou transformation were divided into several categories according to the timing and means of the hukou transformation. Means of the transformation are merit-based hukou converters, family-based conversion(marriage), and collective hukou conversion(policy-based). Using inverse probability–weighted regression adjustment (IPWRA) estimators, with demographic characteristics, family history and personal history controlled, this study shows that in a broad sample, i.e. both current rural migrants and former/returned rural migrants with rural hukou are taken as the reference group, merit- and policy-based conversion and "other" types of conversion show a positive effect on individual psychological well-being. In a more narrow sample where only urban residents with rural origins are taken into consideration, only policy-based hukou converters have an advantage in psychological well-being compared with rural migrants. Merit- and family-based converters have similar depressive symptoms as rural-urban migrants.As the timing and causes of urban-rural hukou switch can be endogenously related to health outcomes, the study by Liu et al. (2015) avoids this issue by focusing on the urban and rural gap in children's health outcomes. They used data from the China Health and Nutrition Survey 1993–2009 with a sample of 9616 children under the age of 18. Descriptive statistics showed that children with urban hukou have 0.25 higher height z-scores and 0.15 higher weight z-scores than children with rural hukou. Both results are statistically significant at 1% level. And this difference by urban vs. rural hukou status is larger than the difference in height and weight (0.23 and 0.09, respectively) by urban vs. rural residence.

The endogeneity issue is further considered through the lens of the healthy migrant effect. Being aware of the of healthy migrant effect for international migrants[3], researchers have shown that this is also relevant for internal migrant within China. Hesketh et al. (2008) showed with a questionnaire taken in Zhejiang Province in 2004 that compared to permanent urban residents, rural-to-urban migrants have less access to medical care, but reported better self-rated health and less acute or chronic diseases among those with the same age and education. Chen (2011) used household survey data from Beijing in 2009

[3] Healthy migrant e ect essentially contains three phenomenon: rst, immigrants are of better health condition than local residents both in the host or original places; second, this health advantage fades over time; and last, the immigrants would return to their place of origin. Zhang et al. (2015)

and reported diminishing physical health advantage of urban-to-urban migrants [4] than host city residents with interaction term of place of origin and period in Beijing; whereas rural-to-urban migrants does not see this pattern in their physical health advantage. Psychological health outcomes state that in contrast to the healthy migrant effect, migrants have higher mental distress. Zhang et al. (2015) divided the sample from the China Labor-force Dynamics Survey 2012 (CLDS) into four categories regarding the migration status:returned population, migrant population, urban residents, and rural residents; to test whether the effect applies to the migrants within the nation. This division is based on their hukou status and migration history. Multiple logistic regression supported that the unhealthy migrants would return by presenting the physical health of the returned population was significantly worse than the migrant population with various controls [5]. Although the discussion on the association of internal migration with health outcomes in China is well-established to some point, and suggests that there is a health-relative migrating decision that may cause estimation bias of the impact of urbanization on health outcomes, this paper tries to take a step further using the Instrumental variable approach to model an exogenous internal migration. Inspired by the finding on telecommunication increasing the likelihood of outmigration(Lu et al., 2016), this paper examines if telecommunication, as well as some other infrastructure changes on the community level could be used as an instrument for the migration decision, which affects the health outcome of the movers.

4 Data

This paper uses two longitudinal datasets, one is the China Health and Nutrition Survey (CHNS) and the other is the China Family Panel Survey (CFPS). The China Health and Nutrition Survey (CHNS) was conducted starting from 1989 with updates till 2015 [6]. The China Family Panel Survey (CFPS) conducts every other year, from 2010 to 2020,

[4] by urban-to-urban migrants, the author refers to those who move from townships/small cities to big cities either with or without changing their place of registration to the host city

[5] Zhang et al. (2015) included variables describes four health elds, they are demographic characteristics, working environment, socioeconomic status, health behaviors and healthcare services access.

[6] CHNS survey years: 1989,1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, 2015

on the individual level. Both CFPS and CHNS are collected at the individual level with community-level information. Essentially this paper uses individual by year panel data.

4.1 China Health and Nutrition Survey

The CHNS community-level data contains a calculated urbanization index and health services costs, which are the key variables of interest. Besides the dichotomy division of designation recorded, i.e. rural versus urban, CHNS constructed an urbanization index. According to Jones-Smith and Popkin (2010), the urbanization index is calculated across 12 factors of the local community and had been proven effective in capturing the degree of urbanization in different study communities. The 12 factors are population density, economic activity, transportation infrastructure, sanitation, communications, housing, education, diversity, health, and social services.[7] Most importantly, the entry of health services measures number and type of health facilities in or nearby (12 km) the community and number of pharmacies in community. And social services also include the availability of medical insurance. This matters in further discussion is made in the Model and Result Sections on the mechanism of a scale of urbanicity is correlated with the health index of the community. The scale ranges from 0 to 120, with a higher score indicating a

[7] The detail measures of the 12 elements are as described by Jones-Smith and Popkin (2010): "1. Population density: the total population of the community divided by community area, from o cial records. 2. Economic activity: typical daily wage for ordinary male workers (reported by community o cials) and percent of the population engaged in nonagricultural work. 3. Traditional markets: distance to the market (three categories), (1) within the boundaries of the community, (2) within the city but not in this community, or (3) not within the city/village/town); the number of days of operation for eight di erent types of markets (including food and fuel markets). 4. Modern markets: number of supermarkets, cafes, internet cafes, indoor restaurants, outdoor xed and mobile eateries, bakeries, ice cream parlors, fast food restaurants, fruit and vegetable stands, bars within the community boundaries. 5. Transportation infrastructure: most common type of road, distance to bus stop, and distance to train stop. (Distance is categorized as (1) within community, (2) <= 1 km from community, and (3) >= 1 km from community). 6. Sanitation: proportion of households with treated water and prevalence of households without excreta present outside the home. 7. Communications: availability (within community boundaries) of a cinema, newspaper, postal service, telephone service; and percent of households with a computer, percent of households with a television, and percent of households with a cell phone. 8. Housing: average number of days a week that electricity is available to the community, percent of community with indoor tap water, percent of community with ush toilets, and percent of community that cooks with gas. 9. Education: average education level among adults >21 years old. 10. Diversity: variation in community education level and variation in community income level. 11. Health infrastructure: number and type of health facilities in or nearby (12 km) the community and the number of pharmacies in the community. 12. Social services: provision of preschool for children under 3 years old, availability of (o ered in community) commercial medical insurance, free medical insurance, and/or insurance for women and children."

community is more urbanized on a continuum. A detailed list of the variable definitions and the summary statistics are given in Table 3 and 4.

Unlike the CFPS which covers all provinces and autonomous cities/districts in mainland China, the population of the CHNS is only drawn from 15 of them.[8]. Nevertheless, the surveyed regions are rather diverse and dispersed in terms of economic activity and geographic location, making the analysis using the CHNS still a reasonable supplement to the one using the CFPS in the later section. The final sample contains approximately 120,000 observations, out of which only 5711, 6174, 1733, and 2075 entries have valid records of the insurance-covered health services cost and the related percentage of the coverage.

4.2 China Family Panel Survey

In the years 2011, 2013, 2015, and so on, the pre-survey was carried out with a only small group of people, so the observations for these years were dropped from the sample. The survey contains a large range of variables, including household characteristics, demographic variables, and health and medical-related information. Three core aspects were extracted from the database, as shown in the Table 5 in three separate panels, individual characteristics, health outcome measures, and insurance and medical issues. When applying the difference-in-difference method, the sample used for estimation was limited to data with both pre- and post-treatment periods.

4.2.1 health outcome measures

CFPS provides both objective and subjective measures of individual health. The interviewer evaluated the respondents' health status on a one to seven scale.[9]. Other than this rating, whether the respondent was diagnosed with chronic disease in the past six months is also included as a health status measure. And the respondent was requested to give a self-estimation on physical health on a one to five scale. In the baseline (2010) survey, the

[8] Beijing, Chongqing, Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shaanxi, Shandong, Shanghai, Yunnan, and Zhejiang

[9] the scale in the questionnaire and raw data is one being very poor and seven being very good

self-estimate scale for health is different than in other years.[10] To maintain consistency, this thesis only use the binary variables generated from the raw data of interviewer-reported health status. "ob_health" is generated by combining the interviewer-reported health status into the 0 (below the average) and 1 (equal or above the average), and labeling the levels in a new binary variable as 1 to be in good objective health status and 0 to in bad objective health status.

Chronic disease diagnosis indicator called "chronic" in this analysis does not differentiate the type of chronic condition but only showing the presence of any chronic disease diagnosed by a doctor in the past six months.

4.2.2 basic medical insurance

From Table 2, it can be seen that an issue of duplicate enrollment in the basic insurance plan is prevalent. And in the data, the people who enrolled in multiple insurance plans can be identified. Since this duplication happens when the working population migrates, (possibly from urban to urban or rural to urban or in the opposite direction), these people are the core sample of interest. Therefore they are not dropped from the sample, instead, they are flagged with an indicator called multiins. And the type of insurance is marked with a value zero instead of missing. However, people could have at most 6 ways of combinations of multiple plans in the enrollment of the basic medical insurance plan before 2020 when the fourth option in insurance plans came in. This duplicate enrollment is not permitted by regulation. The lack of a national medical insurance information system gives rise to this phenomenon. People would choose to opt in more than one insurance plan because they feel more secure to be able to be reimbursed for medical cost both in their hometown (typically where their place of registered residence are) and their current residential city. Table 6 shows the number of observations who report themselves in multiple basic insurance plans. The total number of observations with inconsistent or duplicate enrollment for basic medical insurance is 1609. Compared with the whole sample size, this takes up only 8%. When insurance is included as an

[10] In 2010, from 1 to 5 the scale noted healthy-fair-relatively unhealthy-unhealthy-very unhealthy; while in other years, the scale was noted as excellent-very healthy-healthy-fair-unhealthy.

explanatory variable, the regression is run with the sample without people in multiple insurance plans and then supplemented by a regression using the whole sample to confirm the robustness of the result.

Consistent with the policy for the unification of urban and rural basic medical insurance reformation, in the 2020 survey data, the medical insurance type first provided the option of the unified one. The official name is Urban and Rural Resident Basic Medical Insurance, and this is designed to take the place of both urban resident and new rural cooperative insurance. The enrollment of this medical insurance is noted with a new separate binary variable called basic_ins.

4.2.3 the migration variable

With the purpose of constructing a migration model, I construct the indicator for rural-to-urban migration called "r2u" and define it to be 0 when one stays in the same place of designation, and 1 being the individual moving from rural to urban during the year of 2010 and 2014, i.e. the urban to rural migration is trimmed from the estimation sample. To avoid the bias of setting up different comparison group, the migration model is run with another definition, namely include the urban-rural migrants in the group of r2u=0. The community id information is extracted from 2010 and 2014 CFPS community survey, and used to identify the rural-urban movers. The limited availability of the community information restrained the length of the panel. The migration model is thus a two-year panel.

4.2.4 summary statistics

Summary statistics for the full sample and subgroups by Hukou status are provided in Table 9 and 10. Group means and standard errors in parenthesis are separately computed for urban and rural hukou groups. The third column in Table 10 records the result from a two-sample t-test between the rural and urban subgroups. [11] According to the reported t-statistics, the significance of the group mean difference is noted with star signs. Therefore

[11] Mean di erence is computed with the rural hukou group minus the urban hukou group. And numbers in parenthesis are the pooled standard errors for the two samples.

the mean for the urban and rural hukou groups are not different in most of these variables except for the years of education. This confirms the intuitive urban-rural disparity. Table 8 contains the time communities await till the specified services are built.

5 Model Specification

5.1 Panel Data Model

This analysis tries to find out how urbanization could work through different channels and ultimately affect one's health outcomes. This paper first approaches it by using a non-health-related change at the community level with the data of CHNS. The reason I resort to this dataset is that there is a unique variable characterizing the urbanization index for the local community. And this makes the observable changes in urbanization more visible than the designation change characterized by the urban or rural dichotomy status recorded in CFPS. This also enlarges the sample size in terms of more observations now having variation in the independent variable. The following equation, with the key outcome variable, EXP_{ict}, being the insurance-covered health services cost, explores the assumption of whether urbanization brings about higher healthcare services with time fixed effect and community fixed effect controlled.

$$EXP_{it} = \beta_1 urban\_index_{ct} + \lambda_c + \gamma_t + \varepsilon_{it} \tag{1}$$

5.2 Difference-In-Difference model

Since there are staggering changes in the medical insurance policy unification, a DID setup is intended for this analysis

$$Y_{itp} = \sum_{j=1}^{t_p \quad 2} \tau_j D_{i,j} + \sum_{j=0}^{T \quad t_p} \tau_{t_p+j} D_{i,t_p+j} + \lambda_i + \gamma_t + f(X_i, t) + \varepsilon_{itp} \tag{2}$$

where t stands for the survey year, i stands for the individual, and p denotes the province. t_p is the treatment year varying by province and is summarized in Table 11. Thus τ_j for

$j \in [1, t_p - 2]$ is the series of pre-treatment period effect, and τ_{t_p+j} for $j \in [0, T - t_p]$ is the series of parameters of post-treatment effect varying by the length of lags. The yet-to-be-treated group is those with t_{p+j} greater than or equal to $T - t_p + 1$ for all time periods j. With this expression, the comparison group is set to be individuals in the treatment group with one period lead from the treatment and the never treated group. $D_{i,j}$ and D_{i,t_p+j} are the treatment group dummy, $D_{ik} = 1$ if the individual is in the province where policy reform was completed in period k. As the observations are made every other year, one period here is defined as two years.

The DID estimator of treatment group t_p lagging n years is thus

$$\begin{aligned}\tau_{t_p+n} = {} & (E[Y_{i,t_p+n}|D_{i,t_p+n} = 1] - E[Y_{i,t_p+n}|D_{i,t_p+n} = 0]) \\ & - (E[Y_{i,t_p\;\;1}|D_{i,t_p\;\;1} = 0] - E[Y_{i,t_p\;\;1}|D_{i,t_p\;\;1} = 0]), n \in [1, T - t_p]\end{aligned}$$

In particular, this is the same as claiming all provinces with the same treatment period are of the same treatment effect. It is noteworthy that $D_{i,t_p\;\;1} = 0$ stands for the never treated group. In order to have the never treated group, the final period T is set to be 2018 in the estimation, and the year of 2020 is excluded. Common trends assumption can be tested with the F test among all pre-treatment period parameters.

Included control variables $f(X_i, t)$ are functions of baseline (2010) characteristics, i.e. log of income, years of education, gender, ethnicity, and the age of each individual interacted with time t. The interaction terms allow for personal characteristics to impact the health outcome differently across time. Again Y_{it} is the health outcome. The regression is performed with individual fixed effect λ_i and year fixed effect γ_t.

The nationwide transformation into this new healthcare scheme was marked by the year of 2020 when the last province completed implementing the unified rural and urban resident medical insurance. Not until then, CFPS updated the options to the medical insurance question, adding the option of the urban and rural residents basic medical insurance. Prior to the 2020's survey, the answer to the question had been limited to the old scheme before the policy implementation, and this may cause measurement errors with regard to basic medical insurance enrollment. To mitigate the concerns of discrepancy in reported and actual treatment implementation time, figure 1 was produced to show the trend of the proportion of self-claimed urban residential medical insurance possessors from whenever the data was first available to 2018. Hypothetically, people with urban residential medical insurance would see a general increase in percentage if the uniform policy is enacted, yet the survey sheet stayed unchanged. In the graph, it can be seen that except for Inner Mongolia, Heilongjiang, Zhejiang, Jiangxi, and Shandong, all other provinces exhibited non-downward trends at the time of policy implementation. Also, it is clear from the graph that there are only post-treatment period data for Xinjiang, Ningxia, and Tibet. These three provinces were excluded from the treatment effect analysis.

In order to make sure that the timing of the policy implementation is not related to health-related demographic factors, provincial-level data from China Statistics Yearbook in the year 2007 were used for the regression of policy timing (defined as treatment year t_p minus 2007). Specifically, the regression can be expressed as

$$timing = f(gdp, pop\_dens, region, illiterate, urbpop, govexp, tech, inst, beds) \tag{3}$$

Each of the variables on right-hand side stands for GDP per capita, population density, region index, the illiterate population aged 15 and over, the proportion of the urban population, provincial government health expenditure, no. of health technicians, no. of medical institutions, and number of beds in medical institutions. All these covariates are from the baseline year 2007, which is one year prior to the first treatment year. It is expected none of these independent variables predicts the timing of the reform.

The model could be further expanded into a triple DID model with an indicator $Z_{n,it}$ for the different types of insurance holders. Only for NCMI holders $Z_{1,it} = 1$, only for URMI holders $Z_{2,it} = 1$, and only for CEMI holders $Z_{3,it} = 1$, otherwise 0. This would allow the time-varying treatment effect to be different across the different groups of insurance participants.

$$
\begin{aligned}
Y_{itp} = & \sum_{j=1}^{t_p \; 2} \tau_j D_{i,j} Z_{1,it} + \sum_{j=0}^{T \;\; t_p} \tau_{t_p+j} D_{i,t_p+j} Z_{1,it} + \sum_{j=1}^{t_p \; 2} \tau_j D_{i,j} Z_{2,it} + \sum_{j=0}^{T \;\; t_p} \tau_{t_p+j} D_{i,t_p+j} Z_{2,it} \\
& + \sum_{j=1}^{t_p \; 2} \tau_j D_{i,j} Z_{3,it} + \sum_{j=0}^{T \;\; t_p} \tau_{t_p+j} D_{i,t_p+j} Z_{3,it} + Z_{1,it} + Z_{2,it} + Z_{3,it} + \lambda_i + \gamma_t + f(X_i, t) + \varepsilon_{itp}
\end{aligned} \tag{4}
$$

5.3 Migration Model

Moving from the analysis using a plausibly exogenous shock of policy change, I am shifting the focus toward exploring how migration may impact health outcomes. The migration model considers both the environmental and medical impact that would happen when one moves from a rural to an urban area for any reasons exogenous to health and compares it to the population that stays in the same place. That being said, the data used for this analysis is restricted by dropping individuals who migrate from urban to rural during this period. Due to the availability of migration information, the data is further trimmed down to two periods, 2010 and 2014.

$$
\text{Health Outcome}_{it} = \beta_2 \text{Migration}_{it}^{\text{rural-to-urban}} + \alpha_2 X_{it} + u_t \mu_{it} \tag{5}
$$

In the regression equation 5, $\text{Migration}_{it}^{\text{rural-to-urban}}$ indicates whether the people have migrated from a rural to an urban area since 2010 or 2014, it equals 1 if, 0 otherwise. X_{it} are the control variables that contain gender, annual income, years of education, age, and ethnicity, v_t and u_i stand for time fixed effect and individual fixed effect respectively, ε_{it} and μ_{it} are the error terms.

The instrumental variable, Infrastructure is a list of variables that stands for the number of years the community has to wait before 2014 to have a series of services installed. The list of changes is the installation of electricity, cable radio, cable/satellite TV, telephone, and cell phone service; the establishment of postal service, the first private enterprise; and the availability of paved roads, railways, tap water, and pipeline gas.

6 Results

6.1 Panel Data Model

This session discusses the result from model 1 in order to lay an empirical foundation for the previously assumed statement that urbanization level is related to the policy aspect of medical insurance. First I will examine the individual level of health service expenditure, the fixed effect model described as equation 1 shows that there is a statistically significant increase at 1% level in the percentage of treatment costs covered by insurance given an increase in the urbanization level as shown in column 2 of Table 12. For every one more standard deviation of the urbanization index, there will be a 0.272 [12] standard deviation increase in the percentage of the insurance coverage rate. However, these statistical and economic significances do not sustain across other types of outcome variables, namely the insurance-covered treatment costs, percentage of treatment costs covered, and insurance-covered preventative service costs [13], as shown in Table 12 column 1 to 3. This result supplements an empirical evidence to the policy evidence described previous and in Table 1 that the for urbanized area, the medical insurance coverage is indeed more generous.

6.2 DID estimation

Prior to the DID estimation, the regression using equation 3 was performed on the province level and reported with robust standard error in Table 13. It can be seen that there are no significant determinants to the timing of policy implementation. Individual level baseline control group comparisons are reported in Table 14. Estimations from equation 2 are reported in Table 15 and in Figure 2. By estimating the unified lead and lag of treatment effect parameters f_k and l_k, it is assumed that there is no heterogeneity in the different treatment cohorts.

The numeric result from the main model is shown in Table 15, coefficients of f2 to f4 measure the pre-treatment effect from 2-period-lead to 4-period-lead, coefficient of L0, L1 to L3 estimate the post-treatment effect from the same year, 2-period-lag to 3-

[12] this calculation also follows the formula of standardized coe cient

[13] all expenditures here are surveyed and calculated based on the past four weeks

period-lag. The estimation is conducted using the full sample as long as the individual is in provinces where both pre-treatment and post-treatment period observations are available. It is assumed that this policy reform is most relevant to insurance holders with new cooperative medical insurance (NCMI). They are the ones who could only take part in the NCMI because of rural hukou and will benefit from substituting from NCMI to urban and rural resident medical insurance(NRRMI). The analysis using the main model with this subsample data is presented in Table 16 and Figure 2c and 2d.

Extending from equation 2, I implemented the triple DID model as in equation 4, similarly showing the separated effect for the three types of insurance holders. The result from Model 4 is displayed in Table 19 and Figure 3. From the graphical representations and the numeric ones, result from the two model specifications are about the same.

The DID result is presented in graphical form with the treatment year normalized to zero. Figure 2 presents the average treatment effects for all treatment cohorts with different insurance holder subsamples and outcome variables. Figure 2a shows the average treatment effect on the good objective health outcomes of the full sample, the statistical result is presented in Table 15 column 1. It is visible that for good objective health outcomes, the whole population sees an improvement after three periods, the same goes for people with NCMI, while there are no significant changes for urban residents. This corresponds with the fact that NCMI holders have been enjoying fewer benefits than URMI holders and also confirmed the assumption that the health insurance reform is an improvement to NCMI holders when not hurting the benefit of URMI holders. Table 16 column 1 shows that NCMI holders would enjoy a 0.115 increase in the good objective health measures after 3 periods lag, which is 6 years after the reform. Economic significance calculation following the formula of semi-elasticity is 0.171,i.e. after six years of the policy, the original NCMI holders experience a 0.171 increase in the likelihood of reporting good objective health status on average. And for chronic disease, all three estimations do not witness any significant change from the policy reform. This strikes the chord when chronic disease formation is determined more on living behaviors. This finding makes sense in terms of the medical insurance policy reform being short-term compared to chronic disease formation. F-tests for joint significance of the pre-treatment periods are reported in the last two rows of the event study model result table. For NCMI holders, we fail to reject the null hypothesis that the pre-trends are not significantly different from zero at 1% confidence level, therefore confirm there is no pre-trends for NCMI holders.

6.3 Migration Model OLS estimation

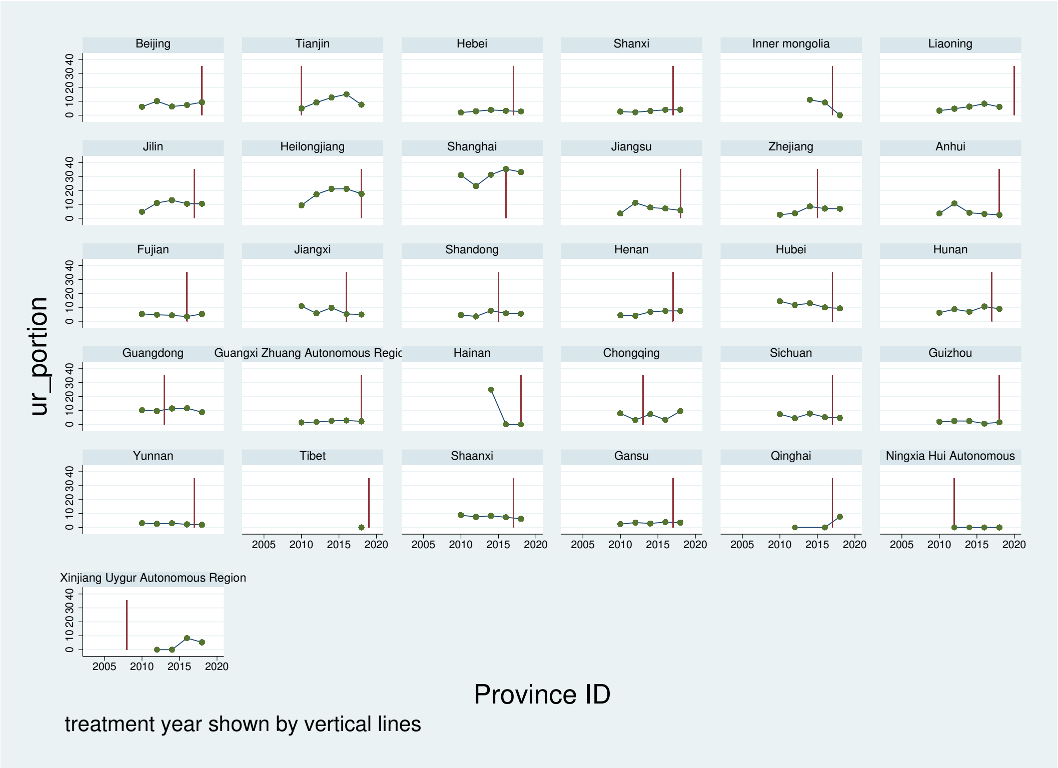
OLS estimation result from the migration model is presented in Table 20. Controlling for individual fixed effect, the rural-urban migration would be associated with negative change of 0.120 in the good objective health status on average. The economic significance is suggested by the semi-elasticity of -0.162, meaning that the likelihood of reporting good objective health status is decreased by 0.162 if one with population average objective health status moves from rural to urban. As r2u stands for the migration decision from rural to urban, with 0 being staying in the same place of designation, and 1 being the individual moving from rural to urban during the year of 2010 and 2014, i.e. the urban to rural migration is trimmed from the estimation sample; However, the result is similar when using r2u=1 if the individual moves from rural to urban areas and 0 otherwise. Therefore I consider the result robust to the different comparison groups. With healthy migrant effect considered, this OLS result could be deemed as an upper bound for the effect of rural-to-urban migration. Due to the limitation of data, further work could be done on finding more eligible data to exploit this issue [14].

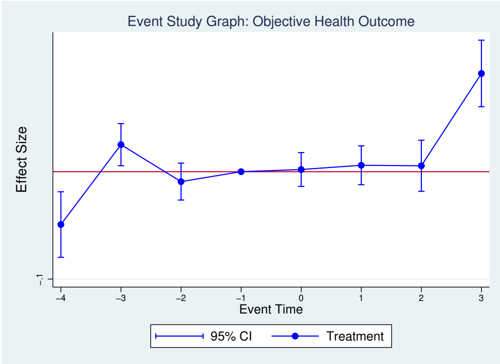
[14] I considered an IV approach which utilize the local infrastructure development as instrumental variables of the migration decision. The selection of an instrumental variable is primarily boosted by the nding that telecommunication increased the out-migration in rural villages in China (Lu et al., 2016). This paper extends from this argument and tests for various factors of local community signaling the convenience of transportation and telecommunication to decide on a valid instrument. It is assumed that living in a region with or without cable radio, for instance, will only a ect health outcomes through the decision to migrate. The information comes from CFPS community survey in 2010 and 2014, containing a lot of missing entries, which might be one reason for its invalidity in identifying the causal e ect. Though passing the weak instrument test, the main result from IV estimation is way too large for a binary outcome variable. It is suggested that the migration from rural to urban would causally and statistical-signi cantly at 1% decrease the likelihood of reporting good objective health outcomes by 3.90. Hence I did not include the analysis here.

7 Conclusion

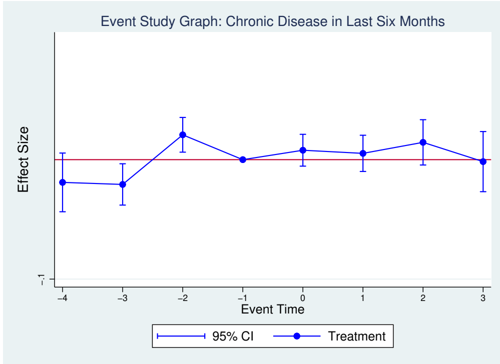
This paper explores the consequences of the recent rapid urbanization on residential health outcomes in China, approaching this both from the angle of healthcare accessibility and the migration from rural to urban areas. I first attempt to draw a linkage between the pattern of healthcare utilization with the urbanization index, and it turns out that the degree of urbanization for the residential area does have an impact on the percentage covered by health insurance. This further provokes the exploit into how health insurance system reform could affect urban and rural residents differently. It is suggested by the DID estimation that NCMI holders would enjoy a 0.115 increase in the good objective health measures after six years of the reform. This would translate into a 0.171 increase in the likelihood of reporting good objective health status measured at the average objective health level. Additionally, The pre-trends for NCMI holders are tested to not significantly different than zero, further confirming the validity of the effect. The estimation using good objective health outcomes correspond with the assumption that the basic medical insurance reform is providing benefit to the NCMI holders, who are essentially rural residents without urban employment. The chronic disease measure speaks differently than the good objective health measure that the medical insurance does not introduce a significant decrease in the chronic disease diagnosis. These findings suggest that part of the urban and rural health disparity could be solved by raising the medical insurance benefit for rural residents. Finally, a migration model is considered for the change in health outcomes caused by rural-to-urban migration. Taking the healthy migrant effect into consideration, the OLS estimation though biased, establishes an upper bound for the decreased health status related to the rural-urban migration. Controlling for individual fixed effect, the rural-urban migration would be associated with negative change of 0.120 in the good objective health status on average. The economic significance is suggested by the semi-elasticity of -0.162. These results shed light on the effectiveness of the medical insurance reform, designed to increase medical insurance benefits for rural residents; and on the worsened health status for rural migrants which are worth noticing for policy makers.

Figure 1: the trend in percent of urban residential medical insurance by province

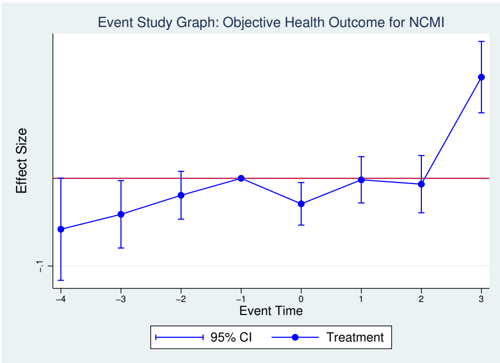




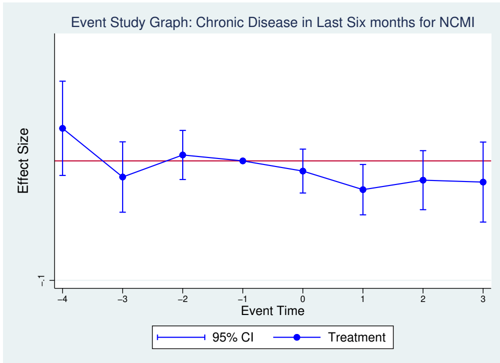
(a) Objective, Full Sample



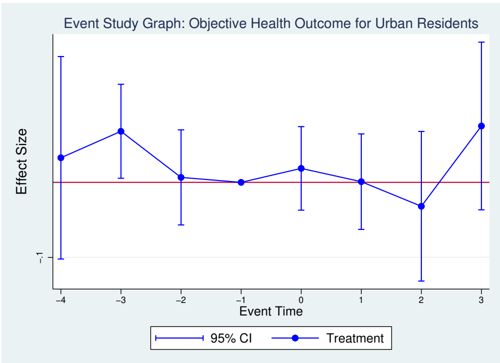
(b) Chronic, Full Sample



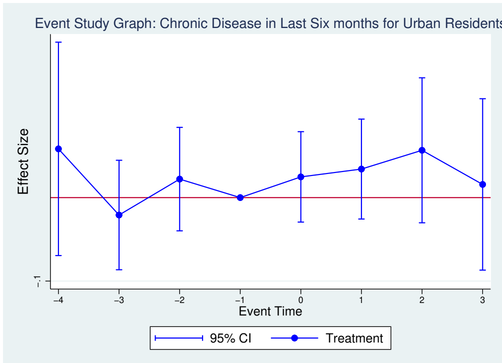
(c) Objective, NCMI



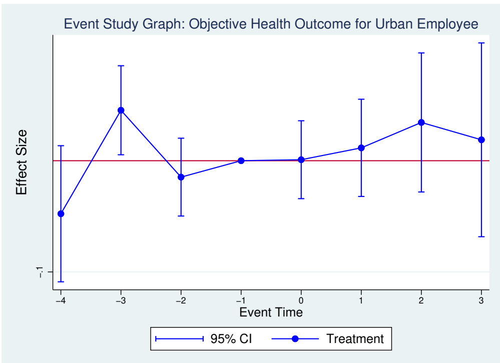
(d) Chronic, NCMI



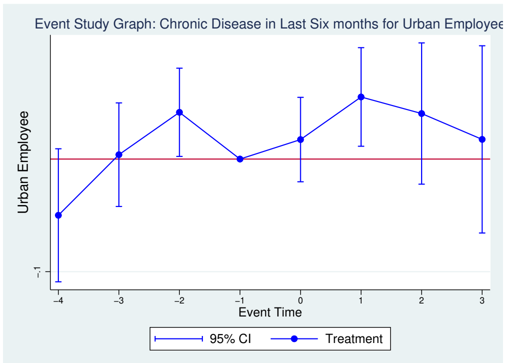
(e) Objective, URMI



(f) Chronic, URMI



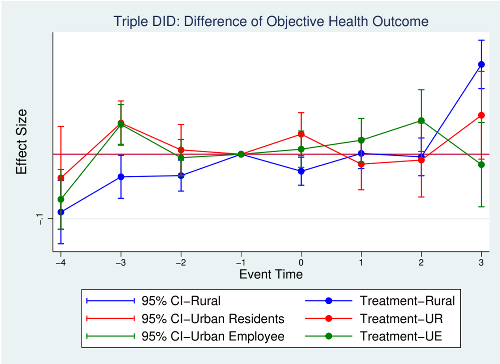
(g) Objective, UEMI



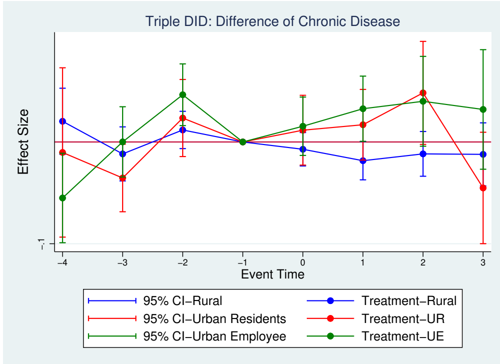
(h) Chronic, UEMI

Figure 2: Event Study Graph From Model 2

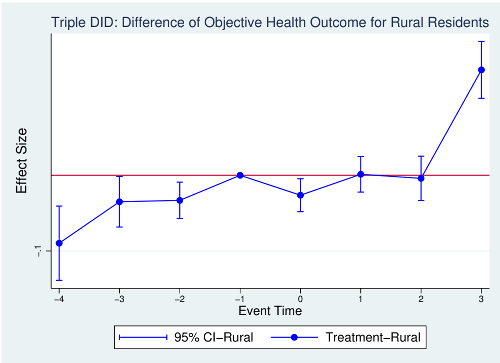
Note: 1. Objective health outcome measure is reported by the interviewer and categorized into a binary variable, 1 being in good health and 0 being in bad health. Chronic disease measures whether, in the last six months, the interviewee has been diagnosed with any chronic disease. 2. Besides using full sample, three kinds of insurance holders are separately measured using model 2 for their health reaction to the reform, and the three kinds are NCMI which is intended only for the rural residents, URMI, intended for urban residents, and nally UEMI, intended for urban employees.



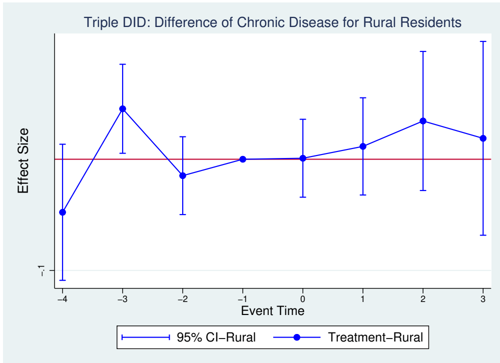
(a) Objective



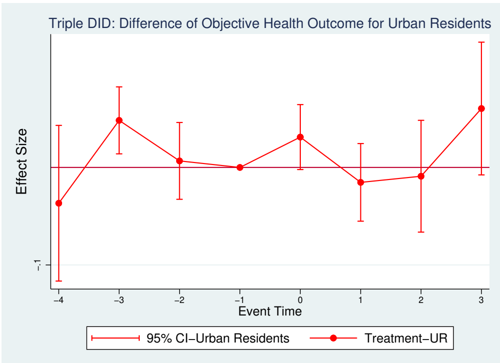
(b) Chronic



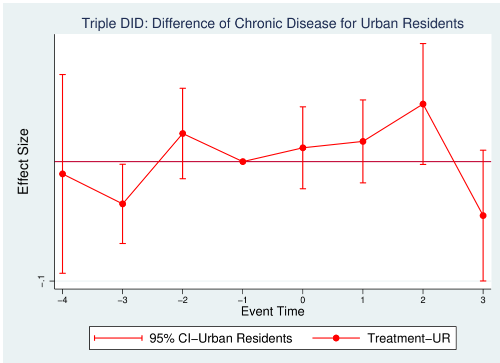
(c) Objective, NCMI



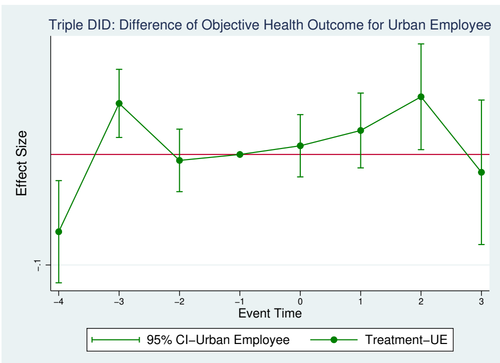
(d) Chronic, NCMI



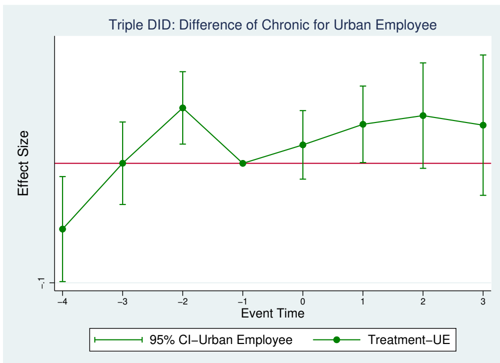
(e) Objective, URMI



(f) Chronic, URMI



(g) Objective, UEMI



(h) Chronic, UEMI

Figure 3: Event Study Graph From Triple DID Model 4

Note: 1. Objective health outcome measure is reported by the interviewer and categorized into a binary variable, 1 being in good health and 0 being in bad health. Chronic disease measures whether, in the last six months, the interviewee has been diagnosed with any chronic disease.
2. Besides using full sample, three kinds of insurance holders are plotted with the estimate from the triple DID model 4 for their health reaction to the reform, and the three kinds are NCMI which is intended only for the rural residents, URMI, intended for urban residents, and nally UEMI, intended for urban employees.

References

Braveman, P. and Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public health reports*, 129(1_suppl2):19–31.

Chan, K. W. and Xu, X. (1985). Urban population growth and urbanization in china since 1949: reconstructing a baseline. *The China Quarterly*, 104:583–613.

Chen, J. (2011). Internal migration and health: re-examining the healthy migrant phenomenon in china. *Social Science & Medicine*, 72(8):1294–1301.

Dong, K. (2009). Medical insurance system evolution in china. *China Economic Review*, 20(4):591–597.

General Office of the State Council (2009). Circular of the state council on printing and distributing the recent key implementation plans for the reform of the medical and health system (2009-2011). http://www.gov.cn/zwgk/2009-04/07/content_1279256.htm.

General Office of the State Council (2017). Notice on printing and distributing the plan for deepening the reform of the medical and health care system during the 13th five-year plan (guo fa [2016] no. 78). http://www.gov.cn/zhengce/content/2017-01/09/content_5158053.htm.

He, H. and Nolen, P. J. (2019). The effect of health insurance reform: Evidence from china. *China Economic Review*, 53:168–179.

Hesketh, T., Jun, Y. X., Lu, L., and Mei, W. H. (2008). Health status and access to health care of migrant workers in china. *Public health reports*, 123(2):189–197.

Hou, B., Nazroo, J., Banks, J., and Marshall, A. (2019). Are cities good for health? a study of the impacts of planned urbanization in china. *International journal of epidemiology*, 48(4):1083–1090.

Jones-Smith, J. C. and Popkin, B. M. (2010). Understanding community context and adult health changes in china: development of an urbanicity scale. *Social science & medicine*, 71(8):1436–1446.

Kawachi, I., Kennedy, B. P., Lochner, K., and Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American journal of public health*, 87(9):1491–1498.

Liu, G. G., Wu, X., Peng, C., and Fu, A. Z. (2003). Urbanization and health care in rural china. *Contemporary Economic Policy*, 21(1):11–24.

Liu, H., Rizzo, J. A., and Fang, H. (2015). Urban-rural disparities in child nutrition-related health outcomes in china: The role of hukou policy. *BMC Public Health*, 15(1):1–11.

Lu, Y., Xie, H., and Xu, L. C. (2016). Telecommunication externality on migration: Evidence from chinese villages. *China Economic Review*, 39:77–90.

Meng, Q., Fang, H., Liu, X., Yuan, B., and Xu, J. (2015). Consolidating the social health insurance schemes in china: towards an equitable and efficient health system. *The Lancet*, 386(10002):1484–1492.

Song, Q. and Smith, J. P. (2021). The citizenship advantage in psychological well-being: An examination of the hukou system in china. *Demography*, 58(1):165–189.

Vlahov, D. and Galea, S. (2002). Urbanization, urbanicity, and health. *Journal of Urban Health*, 79(1):S1–S12.

Zhang, L., Liu, S., Zhang, G., and Wu, S. (2015). Internal migration and the health of the returned population: a nationally representative study of china. *BMC Public Health*, 15(1):1–9.

Table 1: summary of the insurance scheme before 2009 (Dong, 2009)

Name	Insured people	Time first established	Payment scheme	Reimbursement scheme	More to note
The basic medical insurance scheme for urban employees (UEMI)	Urban employees	In 1994, pilot cities being Zhenjiang City and Jiujiang City for a 4 year trial; in 1998, launched throughout the country	Premiums are paid by employers and employees proportional to the employees' payroll; Premiums are collected and divided into social pooling funds and individual account.	The individual account pays for outpatient expenses, emergency services, and drug costs. And the social pooling account pays for inpatient's costs within a predefined band, which is above the deductible line but below the ceiling. The deductible for the social pooling account is 10% of the local average payroll and is capped at four times that amount.	
The basic social medical insurance for urban residents (URMI)	Urban unemployed residents, including children	In 2007, with 79 pilot cities	Adults pay the premiums for themselves (if not employed) and their children; the government subsidize around 40 yuan per year, depending on the local economy condition.	For hospitalization fees and medical expenses incurred in outpatient clinics for critical diseases	Some cities establish individual accounts, others uses a family-based scheme
The new rural co-operative medical insurance scheme (NCMI)	Rural residents	In 2003, hundreds of pilot counties in for provinces	Individual contribution is relatively limited, premium is subsidized by both the local and central governments.	Critical disease/hospitalization oriented	voluntary enrollment for individuals; installment are at discretion of local governments

Table 2: Medical Insurance Rate Participation by Year and Kind

year	population	Urban employed population	Total population covered by public medical insurance	Public insurance participation rate	Total number of participants in the basic social medical insurance scheme for urban employees: incumbent employees	Total number of participants in the basic social medical insurance scheme for urban employees: retired employees	The number of participants covered by medical insurance for urban/rural residents	The number of participants in the new rural cooperative medical insurance system (NRCMIS)
2004	13	2.65	2.04	15.7	0.9	0.34	-	0.8
2005	13.08	2.73	3.17	24.23	1	0.38	-	1.79
2006	13.14	2.83	5.67	43.16	1.16	0.42	-	4.1
2007	13.21	2.94	9.49	71.83	1.34	0.46	0.43	7.26
2008	13.28	3.02	11.33	85.33	1.5	0.5	1.18	8.15
2009	13.35	3.33	12.34	92.5	1.64	0.55	1.82	8.33
2010	13.41	3.47	12.69	94.61	1.78	0.59	1.95	8.36
2011	13.47	3.59	13.05	96.89	1.89	0.63	2.21	8.32
2012	13.54	3.71	13.41	99.07	1.99	0.66	2.72	8.05
2013	13.61	3.82	13.73	100.88	2.05	0.69	2.96	8.02
2014	13.68	3.93	13.33	97.49	2.1	0.73	3.15	7.36
2015	13.75	4.04	13.36	97.18	2.14	0.75	3.77	6.7
2016	13.83	4.14	-	-	2.17	0.78	4.49	-
2017	13.9	4.25	-	-	2.23	0.8	8.74	-
2018	13.95	4.34	13.45	96.36	2.33	0.84	8.97	1.3
2019	14	4.42	13.54	96.71	2.42	0.87	10.25	-
2020	14.12	4.63	13.61	96.39	2.54	0.9	10.17	-

Note:
1. Populations and number of participants are in 100 million.
2. Only data after 2003 was recorded in this table since the NRCMIS starts in 2003.
3. The basic social medical insurance for urban residents started in 2007, hence there is missing data in 2004-2007 for urban residents.
4. Column 8 shows the number of participants covered by basic medical insurance for urban residents only in years before 2019. NRCMIS and the basic social medical insurance for urban residents was officially incorporated into one, i.e. the basic social medical insurance for urban and rural residents in 2019. Numbers of participants shown in column 8, leaving column 9 to be zero.
5. In 2013 total participation rate exceed 100%, this is due to the duplicate participation of the public medical insurance when people could have one medical insurance accounts in mulitple places (though it is not in compliance with the regulation).
6. The data was originally presented in the China Statistical Yearbook.

Table 3: Variable Definition for CHNS

variable name	definition
index	urbanization index
index10	urbanization index divided by 10
cost	insurance-covered treatment costs
cost_prevent	insurance-covered preventative service costs
per_cost	percentage of treatment costs covered by insurance
per_cost_prevent	percentage of preventative service costs covered by insurance

Table 4: Summary statistics for CHNS

	Mean	p50	SD	Min	Max	N
index	59.45	58.03	20.82	14.26	106.5	127,719
index10	5.94	5.8	2.08	1.43	10.65	127,719
cost	105,588.23	0	614,298.19	0	16,000,000.00	5,714
cost_prevent	1,710.46	0	8,441.30	0	98,901.00	1,733
per_cost	33.84	0	39.06	0	100	6,177
per_cost_prevent	25.13	0	41.34	0	100	2,075

Table 5: Variable Definition for CFPS

variable name	definition	type
Panel A: individual characteristics		
urban	live in urban area (according to Census Bureau's definition)	binary
hukou	hukou status with 0=rural, 1=urban	binary
r2u	rural to urban migration with 0= stay in the same designation and missing for urban to rural migration	binary
gender	gender with 0=female, 1=male	binary
lnincome	natural log of income	continuous
educ_yrs	years of education	continuous
age	age	continuous
ethnicity	ethnicity	categorical
panel B: health outcome measures		
ob_health	good objective health rating by interviewer, generated from report_health	binary
report_health	interviewer-report health status with 1 being very good 7 being very poor	categorical
chronic	Any doctor-diagnosed chronic disease in the past 6 months	binary
Panel C: insurance		
multiins	indicator for multi-basic insurances with 1=enrolled in two or more of the three basic insurances at the same time, 0 otherwise	binary
insur_ue	binary indicator if the people is enrolled in urban employee medical insurance p	binary
insur_ur	binary indicator if the people is enrolled in urban resident medical insurance p	binary
insur_r	binary indicator if the people is enrolled in new rural cooperative medical insu	binary

Table 6: no. of duplicate enrolled observations

		insur_r=0	insur_r=1	total
insur_ur=0	insur_ue=0	42,827	119,081	161,908
	insur_ue=1	21,999	750	22,749
insur_ur=1	insur_ue=0	13,733	354	14,087
	insur_ue=1	465	40	505
	total	79,024	120,225	199,249

[a] Note: hence the number of observations with inconsistent or duplicate enrollment for basic medical insurance is 750+354+465+40=1609.

Table 7: Variable Definition for IV candidates (CFPS)

variable name	definition
electricity	no. of year before 2010 that electricity was installed
radio	no. of year before 2010 that cable radio was installed
TV	no. of year before 2010 that cable/satellite TV was installed
postal	no. of year before 2010 that postal service was established
telephone	no. of year before 2010 that telephone was installed
cellphone	no. of year before 2010 that cell phone service was installed
paved	no. of year before 2010 that paved road was available
railway	no. of year before 2010 that railway was available
tap	no. of year before 2010 that tap water was available
gas	no. of year before 2010 that pipeline gas was available

Table 8: Summary Statistics for IV candidates (CFPS)

	Mean	p50	SD	Min	Max	N
electricity	32	32	13	0	65	28,409
radio	28	32	16	1	61	14,048
TV	11	10	8	0	39	27,232
postal	38	37	16	2	70	17,485
telephone	19	15	12	0	61	29,484
cellphone	11	11	6	0	32	30,531
paved	23	18	18	1	80	26,077
railway	40	40	29	1	110	2,676
tap	16	13	13	0	61	21,959
gas	10	7	8	0	38	4,297

Table 9: Summary Statistics of CFPS

	Mean	p50	SD	Min	Max
Panel A: individual characteristics					
urban	0.47	0	0.5	0	1
hukou	0.26	0	0.44	0	1
migrant	0.06	0	0.24	0	1
hukou3	0.15	0	0.36	0	1
hukou12	0.16	0	0.37	0	1
gender	0.49	0	0.5	0	1
lnincome	9.51	9.88	1.46	0	16.15
educ_yrs	7.68	9	4.71	0	24
age	45.36	45	17.9	9	110
Panel B: health outcome measures					
ob_health	0.7	1	0.46	0	1
report_health	2.6	2	1.28	1	7
chronic	0.15	0	0.36	0	1
Panel C: insurance and medical cost					
multiins	0.01	0	0.09	0	1
insur_ue	0.12	0	0.32	0	1
insur_ur	0.07	0	0.26	0	1
insur_r	0.6	1	0.49	0	1
basic_ins	0.01	0	0.09	0	1

Table 10: Comparison by Hukou Subgroup

	rural	urban	comparison	p-value
hukou	0.067	0.485	-0.418	0.000
	(0.001)	(0.002)	(0.002)	
migrant	0.032	0.095	-0.063	0.000
	(0.001)	(0.001)	(0.002)	
hukou3	0.015	0.304	-0.288	0.000
	(0.001)	(0.003)	(0.003)	
hukou12	0.018	0.320	-0.302	0.000
	(0.001)	(0.003)	(0.003)	
gender	0.497	0.483	0.013	0.000
	(0.002)	(0.002)	(0.002)	
lnincome	9.105	9.835	-0.730	0.000
	(0.009)	(0.006)	(0.011)	
educ_yrs	6.383	8.971	-2.588	0.000
	(0.015)	(0.016)	(0.022)	
age	45.804	45.759	0.044	0.589
	(0.056)	(0.059)	(0.082)	
ob_health	0.582	0.637	-0.054	0.000
	(0.002)	(0.002)	(0.002)	
report_health	2.751	2.434	0.317	0.000
	(0.005)	(0.005)	(0.007)	
chronic	0.154	0.156	-0.002	0.179
	(0.001)	(0.001)	(0.002)	
multiins	0.004	0.013	-0.008	0.000
	(0.000)	(0.000)	(0.000)	
insur_ue	0.033	0.216	-0.183	0.000
	(0.001)	(0.001)	(0.001)	
insur_ur	0.018	0.139	-0.121	0.000
	(0.000)	(0.001)	(0.001)	
insur_r	0.788	0.415	0.373	0.000
	(0.001)	(0.002)	(0.002)	
basic_ins	0.005	0.010	-0.005	0.000
	(0.000)	(0.000)	(0.000)	

Table 11: Year of Each Province Completing the Healthcare reform

Province code	Province	Year of policy implementation
65	Xinjiang Uygur Autonomous	2008
12	Tianjin	2010
64	Ningxia Hui Autonomous	2012
44	Guangdong	2013
50	Chongqing	
33	Zhejiang	2015
37	Shandong	
31	Shanghai	2016
35	Fujian	
36	Jiangxi	
13	Hebei	2017
14	Shanxi	
15	Inner mongolia	
22	Jilin	
41	Henan	
42	Hubei	
43	Hunan	
51	Sichuan	
53	Yunnan	
61	Shaanxi	
62	Gansu	
63	Qinghai	
11	Beijing	2018
23	Heilongjiang	
32	Jiangsu	
34	Anhui	
45	Guangxi Zhuang Autonomous	
46	Hainan	
52	Guizhou	
54	Tibet	2019
21	Liaoning	2020

Table 12: Panel Data Model with CHNS data

VARIABLES	(1) insurance-covered treatment costs	(2) percentage of treatment costs covered	(3) insurance-covered preventative service costs	(4) percentage of preventative service costs covered
index	1,457.03	0.51***	309.09	0.04
	(4,150.51)	(0.16)	(231.49)	(0.54)
1991.wave	265,474.22	1.59	6,361.27	-31.66*
	(390,529.79)	(5.30)	(5,336.36)	(16.61)
1993.wave	495,314.44	-13.77*		-17.10
	(447,641.69)	(7.74)		(19.56)
1997.wave	539,028.89	-17.67**	-2,205.85	-37.07
	(517,544.19)	(8.12)	(3,258.42)	(23.70)
2000.wave	1,194,127.60**	-26.96***	-6,901.23	-97.95***
	(606,385.08)	(7.99)	(6,551.86)	(15.86)
2004.wave	681,104.37	-37.73***	-7,809.74	-104.80***
	(538,422.72)	(7.76)	(5,751.58)	(17.84)
2006.wave	743,771.86	-37.20***	-6,619.98	-100.39***
	(544,915.32)	(7.95)	(6,140.20)	(22.16)
2009.wave	782,016.52	-35.76***	-7,752.78	-82.37***
	(556,726.36)	(8.08)	(6,861.70)	(18.59)
2011.wave	787,885.89	-36.26***	-6,019.18	-102.78***
	(564,110.62)	(8.25)	(7,629.27)	(19.24)
2015.wave	913,817.76	-32.60***	-6,695.31	-102.34***
	(575,428.85)	(8.43)	(6,334.46)	(18.66)
1993o.wave			-	
Constant	-760,092.22	28.87***	-14,726.21	111.51***
	(517,647.35)	(9.63)	(11,706.21)	(26.74)
Observations	5,711	6,174	1,733	2,075
R-squared	0.06	0.05	0.05	0.33
Number of idind	4,892	5,222	1,673	1,979

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Determinant of Reform Timing

VARIABLES	(1) timing	(2) timing
gdp_07	-0.0001	-0.0002
	(0.0001)	(0.0002)
pop_dens_07	7.4866	17.1211
	(13.5717)	(16.6934)
region_07	-0.7770	-1.1512
	(0.6381)	(0.7974)
illiterate_07	0.0001	0.0000
	(0.0003)	(0.0006)
urbpop_07	-0.0302	-0.0048
	(0.1277)	(0.1446)
govexp_07		0.0393
		(0.0560)
tech_07		-0.0000
		(0.0000)
inst_07		0.0002
		(0.0001)
beds_07		-0.0000
		(0.0001)
Constant	12.8382***	13.5123**
	(4.5915)	(5.1080)
Observations	31	31
R-squared	0.1013	0.1670

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Mean comparison of covariates by ever-treated and never-treated group

VARIABLES	t-test result
gender	-0.0065
	(0.0038)
lnincome	0.0312
	(0.0185)
educ_yrs	0.8015***
	(0.0380)
age	1.8505***
	(0.1357)
ethnicity	0.1340***
	(0.0445)

Table 15: Result from Event Study Model

VARIABLES	(1) ob_health	(2) ob_health	(5) chronic	(6) chronic
f4	-0.0212*	-0.0492***	-0.0178**	-0.0190
	(0.0115)	(0.0156)	(0.0086)	(0.0125)
f3	0.0326***	0.0252**	-0.0070	-0.0207**
	(0.0076)	(0.0100)	(0.0065)	(0.0088)
f2	-0.0172***	-0.0092	0.0244***	0.0208***
	(0.0065)	(0.0088)	(0.0055)	(0.0074)
l0	0.0109*	0.0020	0.0106**	0.0079
	(0.0058)	(0.0080)	(0.0049)	(0.0068)
l1	0.0126*	0.0060	0.0037	0.0053
	(0.0068)	(0.0092)	(0.0057)	(0.0077)
l2	-0.0137	0.0055	0.0119*	0.0145
	(0.0087)	(0.0121)	(0.0071)	(0.0097)
l3	0.0730***	0.0915***	0.0033	-0.0016
	(0.0112)	(0.0158)	(0.0091)	(0.0128)
Constant	0.6944***	1.1582***	0.1328***	-0.0669
	(0.0029)	(0.3031)	(0.0024)	(0.1659)
Time-varying controls	No	Yes	NO	Yes
Individual and Year fixed effect	Yes	Yes	Yes	Yes
Observations	160,978	70,058	145,966	68,100
R-squared	0.0255	0.0350	0.0093	0.0124
Number of pid	52,034	18,258	46,856	18,256
F-test for the pre-trends	18.02	16.27	9.80	13.36
p-value	0.0000	0.0000	0.0001	0.0000

Note:
1. Column 1 and 2 are estimated using good objective health outcome, column 3 to 4 are estimated using chronic disease indicator;
2. Coefficient of f2 to f4 measures the pre treatment effect from 2-period-lead to 4-period-lead, coefficient of l0 to l3 estimate the post treatment effect from the same year to 3-period-lag;
3. Control variables are time interaction terms with gender, log of annual income, years of education, age, ethnicity, gender; all controls are baseline value at year 2000; estimation is conducted with fixed year effect and province fixed effect;
4. Estimation is conducted using the full sample as long as the individual is in provinces where both pre-treat and post-treat period observation is available, sample size after adding control variables trims down a lot from the missing values in income information.
5. F-test statistics for the joint significance of the pre-trend, i.e. f-test for f2=f3=f4=0, is reported in the last two rows.
6. Robust standard error is reported in parenthese, with *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 16: Result from Event Study Model with NCMI Subsample

	(1)	(2)
(3)		
VARIABLES	ob_health	chronic
f4	-0.0580*	0.0273
	(0.0297)	(0.0201)
f3	-0.0411**	-0.0135
	(0.0196)	(0.0150)
f2	-0.0194	0.0050
	(0.0139)	(0.0105)
l0	-0.0291**	-0.0085
	(0.0124)	(0.0094)
l1	-0.0017	-0.0241**
	(0.0135)	(0.0108)
l2	-0.0066	-0.0161
	(0.0167)	(0.0126)
l3	0.1154***	-0.0177
	(0.0208)	(0.0171)
Constant	0.7024	-0.4588
	(0.7179)	(0.5045)
Observations	41,745	40,715
R-squared	0.0285	0.0139
Number of pid	11,886	11,816
F-test for the pre-trends	1.28	2.12
p-value	0.2791	0.1205

Note:
1.The results are estimated from model 2 using the subsample of individuals with New Rural Cooperative Medical Insurance (NCMI).
2. Coefficients of f2 to f4 measure the pre-treatment effect from 2-period-lead to 4-period-lead, coefficients of l0 to l3 estimate the post-treatment effect from the same period to 3-period-lag;
3. All columns are estimated with control variables interacted with year dummy, identifying the time-varying effect of the baseline characteristics at year 2000, which are gender, log of annual income, years of education, age, ethnicity, gender; estimation is conducted with individual fixed effect and year fixed effect;
4. F-test statistics for the joint significance of the pre-trend, i.e. f-test for f2=f3=f4=0, is reported in the last two rows;
5. Robust standard error is reported in parenthese, with *** p<0.01, ** p<0.05, * p<0.1.

Table 17: Result from Event Study Model with Urban Employee Subsample

VARIABLES	(1) ob_health	(2) chronic
f4	-0.0477	-0.0523*
	(0.0312)	(0.0300)
f3	0.0454**	0.0016
	(0.0204)	(0.0234)
f2	-0.0148	0.0409**
	(0.0179)	(0.0200)
l0	0.0009	0.0164
	(0.0179)	(0.0191)
l1	0.0116	0.0537**
	(0.0223)	(0.0223)
l2	0.0344	0.0387
	(0.0319)	(0.0319)
l3	0.0188	0.0148
	(0.0445)	(0.0423)
Constant	0.7923	-2.9466***
	(0.5528)	(0.7591)
Observations	11,919	11,824
R-squared	0.0644	0.0203
Number of pid	5,270	5,234
F-test for the pre-trends	7.89	4.98
p-value	0.0004	0.0069

Note:
1.The results are estimated from model 2 using the subsample of individuals with Urban Employee Medical Insurance(UEMI).
2. Coefficients of f2 to f4 measure the pre-treatment effect from 2-period-lead to 4-period-lead, coefficients of l0 to l3 estimate the post-treatment effect from the same period to 3-period-lag;
3. All columns are estimated with control variables interacted with year dummy, identifying the time-varying effect of the baseline characteristics at year 2000, which are gender, log of annual income, years of education, age, ethnicity, gender; estimation is conducted with individual fixed effect and year fixed effect;
4. F-test statistics for the joint significance of the pre-trend, i.e. f-test for f2=f3=f4=0, is reported in the last two rows;
5. Robust standard error is reported in parenthese, with *** p<0.01, ** p<0.05, * p<0.1.

Table 18: Result from Event Study Model with Urban Resident Subsample

VARIABLES	(1) ob_health	(2) chronic
f4	0.0328	0.0584
	(0.0688)	(0.0652)
f3	0.0681**	-0.0209
	(0.0319)	(0.0334)
f2	0.0066	0.0222
	(0.0323)	(0.0316)
l0	0.0186	0.0248
	(0.0284)	(0.0276)
l1	0.0009	0.0342
	(0.0325)	(0.0305)
l2	-0.0319	0.0567
	(0.0509)	(0.0443)
l3	0.0752	0.0158
	(0.0570)	(0.0524)
Constant	3.0843	0.3082
	(2.0871)	(3.5292)
Observations	6,524	6,469
R-squared	0.0598	0.0354
Number of pid	3,741	3,717
F-test for the pre-trends	1.88	1.56
p-value	0.1535	0.2095

Note:
1.The results are estimated from model 2 using the subsample of individuals with Urban Resident Medical Insurance(URMI).
2. Coefficients of f2 to f4 measure the pre-treatment effect from 2-period-lead to 4-period-lead, coefficients of l0 to l3 estimate the post-treatment effect from the same period to 3-period-lag;
3. All columns are estimated with control variables interacted with year dummy, identifying the time-varying effect of the baseline characteristics at year 2000, which are gender, log of annual income, years of education, age, ethnicity, gender; estimation is conducted with individual fixed effect and year fixed effect;
4. F-test statistics for the joint significance of the pre-trend, i.e. f-test for f2=f3=f4=0, is reported in the last two rows;
5. Robust standard error is reported in parenthese, with *** p<0.01, ** p<0.05, * p<0.1.

Table 19: Triple DID result

VARIABLES	(1) ob_health	(2) chronic
f4r	-0.0897***	0.0203
	(0.0250)	(0.0166)
f3r	-0.0350**	-0.0117
	(0.0170)	(0.0136)
f2r	-0.0331***	0.0118
	(0.0123)	(0.0093)
l0r	-0.0263**	-0.0072
	(0.0111)	(0.0084)
l1r	0.0013	-0.0184*
	(0.0120)	(0.0096)
l2r	-0.0041	-0.0117
	(0.0149)	(0.0112)
l3r	0.1393***	-0.0122
	(0.0192)	(0.0158)
f4ur	-0.0367	-0.0103
	(0.0407)	(0.0424)
f3ur	0.0483***	-0.0354**
	(0.0175)	(0.0169)
f2ur	0.0067	0.0235
	(0.0201)	(0.0193)
l0ur	0.0311*	0.0116
	(0.0170)	(0.0175)
l1ur	-0.0153	0.0169
	(0.0203)	(0.0177)
l2ur	-0.0090	0.0482*
	(0.0292)	(0.0258)
l3ur	0.0604*	-0.0452
	(0.0347)	(0.0279)
f4ue	-0.0699***	-0.0550**
	(0.0236)	(0.0224)
f3ue	0.0462***	0.0001
	(0.0157)	(0.0176)
f2ue	-0.0054	0.0464***
	(0.0144)	(0.0155)
l0ue	0.0079	0.0155
	(0.0144)	(0.0146)
l1ue	0.0217	0.0327**
	(0.0173)	(0.0163)
l2ue	0.0522**	0.0400*
	(0.0244)	(0.0225)
l3ue	-0.0161	0.0320
	(0.0334)	(0.0300)
insur_r	0.0772***	0.0039
	(0.0077)	(0.0059)
insur_ur	0.0437***	0.0014
	(0.0097)	(0.0086)
insur_ue	0.0327***	-0.0002
	(0.0079)	(0.0075)
Constant	1.0645***	-0.0670
	(0.2897)	(0.1641)
Observations	70,058	68,100
R-squared	0.0387	0.0129
Number of pid	18,258	18,256

Note:
1.The results are estimated from model 4 using the full sample;
2. Coefficients of f2r to f4r measure the pre-treatment effect from 2-period-lead to 4-period-lead for NCMI participants, coefficients of l0r to l3r estimate the post-treatment effect from the same period to 3-period-lag for NCMI participants; the same applies to suffixes of ur and ue, which are referring to the URMI and UEMI participants;
3. All columns are estimated with control variables interacted with year dummy, identifying the time-varying effect of the baseline characteristics at year 2000, which are gender, log of annual income, years of education, age, ethnicity, gender; estimation is conducted with individual fixed effect and year fixed effect;
4. Robust standard error is reported in parenthese, with *** p<0.01, ** p<0.05, * p<0.1.

Table 20: Result from Migration Model with Simple OLS

VARIABLES	(1) ob_health	(2) ob_health	(3) chronic	(4) chronic
r2u	-0.0319	-0.1203***	0.0240	0.0229
	(0.0359)	(0.0452)	(0.0303)	(0.0381)
baseline_gender	0.0097**		-0.0330***	
	(0.0046)		(0.0042)	
baseline_lnincome	0.0236***		-0.0032**	
	(0.0018)		(0.0015)	
baseline_educ_yrs	0.0127***		0.0032***	
	(0.0007)		(0.0007)	
baseline_age	-0.0051***		0.0058***	
	(0.0002)		(0.0002)	
baseline_ethnicity	-0.0033***		-0.0020***	
	(0.0009)		(0.0007)	
Constant	0.6989***	0.8080***	-0.0854***	0.1454***
	(0.0171)	(0.0002)	(0.0151)	(0.0002)
individual fixed effect	No	Yes	NO	Yes
Observations	33,511	33,511	33,185	33,185
R-squared		0.0006		0.0000
Number of pid	19,960	19,960	19,956	19,956

Note:
1. The results are calculated with equation 5 using simple OLS method. r2u stands for the migration decision from rural to urban, with 0 being staying in the same place of designation, and 1 being the individual moving from rural to urban during the year of 2010 and 2014, i.e. the urban to rural migration is trimmed from the estimation sample; However, the result is similar when using r2u=1 if the individual moves from rural to urban areas and 0 otherwise;
2.Column (2) and (4)are regressed with individual fixed effect and thus omitted of the baseline controls;
3. Robust standard error is reported in parenthese, with *** p<0.01, ** p<0.05, * p<0.1.