Working Paper in Economics

20188

November 2018

Estimating the Impact of Social Medical Insurance Schemes on Children's Health and Hospital Use: The Chinese Case

Jing Guana
J.D. Tena

Estimating the Impact of Social Medical Insurance Schemes on Children's Health and Hospital Use: The Chinese Case

Jing Guan[a], J.D. Tena[b]

[a]School of Insurance and Economics, University of International Business and Economics, Beijing, China

[b]Management School, University of Liverpool, Liverpool, United Kingdom and DISea, Università di Sassari, Italy.

Abstract: This study investigates the causal impact of acquiring social medical Insurance on hospital utilization and health status for children under 16 years old in China from 2010 to 2016. We consider the China Family Panel Studies (CFPS), a longitudinal database which allows us to control for the effect of unobserved individual heterogeneity by means of difference-in-difference regressions combined with matching regression techniques. Our findings suggest that participating in social medical insurance schemes significantly increases children's yearly hospital use, especially for low income and rural children. Moreover, this increase is not significantly different for people who were not previously sick. It is also found that social medical insurance schemes have no effect or even a marginally negative effect on children's health status in some cases. We discuss some potential explanations for this result.

Keywords: China; Social Medical Insurance; Health Outcomes; Difference-in-difference; Propensity Score Matching.

JEL codes: I130: Health Insurance.

1. Introduction

Expanding social medical insurance schemes to uninsured children can be deemed as a very interesting policy strategy, since its beneficial effects are expected to be more persistent in younger people compared to adults. For example, Moav (2005) proposes a theoretical model to explain the persistency of poverty as a result of a lack of investment in children. The importance of children as economic assets in developing economies has been also discussed to a large extent in other papers; see Galor (2005) and references therein. It is surprising that, in spite of this concern, the investigation of the causal impact of insurance policies on children's health has received relatively little attention in empirical research, especially for developing countries.

Some of the most relevant studies about children's health are focused on the US. They show that Medicaid expansion in child population made access to health care easier (Miller 2012), increased the probability of hospitalization (Currie & Gruber 1996; Dafny & Gruber 2000) and improved access to primary care (Miller 2012; Kaestner et al., 2001). As for children's health status, it increased due to the expansion of Medicaid, which was measured by looking at mortality rates (Currie & Gruber 1996) and self-rated health (Miller 2012). According to these analyses, it is possible to conclude that Medicaid expansion had a positive effect on children's health outcomes.

However, studies are scarcer when the attention is turned to a developing country like China. Chen and Jin (2012) represent an important exception to this concern. They consider a large cross-sectional database to estimate the impact of insurance for Chinese children in rural areas for the year 2006. Given the absence of longitudinal information, their empirical analysis was conducted by comparing the health status in counties where the policy was and was not applied to similar households with high and low probabilities of being insured. This methodology, denoted as propensity score matching with difference-in-difference estimation (PSM with DID) is a highly insightful way to estimate the causal impact of insurance on children's health status for cross-sectional data. However, as the authors indicate, a remarkable problem of this estimation is that the selection of similar individuals can only be based on observables, while key non-observable variables could be very different by individuals in regions affected and not affected by the insurance policy. In this respect, a longitudinal database circumvents this problem, as it allows for the identification of the same individual before and after the policy takes place.

Other relevant studies based on Chinese background have focused on the total population. Liu et al. (2002) found that there was a significant increase in outpatient visits by lower socioeconomic groups in response to the pilot experiment of the Urban Employees Basic Medical Insurance Scheme (UEBMI). Similarly, a study by Lei and Lin (2009) showed that enrolling in the New-Type Rural Cooperative Medical Scheme (NCMS) increased the probability of preventive care utilization in rural China. More recently, Li and Zhang (2013) took a closer look at the impact of different kinds of insurance systems on the health outcomes of Chinese senior citizens in the Zhejiang and Gansu provinces. They found that people with UEBMI and URBMI tended to use more medical services, and people with NCMS did not increase utilization of outpatient and inpatient services. However, there is little reason to think that conclusions for adults can be extrapolated to children. Unlike adults, children usually do not take their own health decisions. Their demand for health attention could also be different, as they are more likely to be affected by common childhood illnesses and, in general, their health status depend to a large extent on their genetic features and direct care offered by their parents.

This paper seeks to fill this gap in the literature by examining the impacts of health insurance schemes on health care utilization and health status among Chinese children who are under sixteen years old. To do so, we use a longitudinal database from the China Family Panel Studies (CFPS) conducted in 2010, 2012, 2014 and 2016 in order to estimate the causal impact of medical insurance on four health outcomes: hospital visit frequency in the previous year, hospital visit frequency in the last month, frequency of sickness in the last month and current self-rated health status. There are at least four important features of our database to conduct this analysis. First, comparing to prior studies, our database has a separate questionnaire for children under 16 years old, which includes some rich, individual-level information covering demographic and economic characteristics as well as education, social welfare and health outcomes. A second advantage is that dealing with a longitudinal database enables us to control time-invariant, individual characteristics, as we can compare the same individual before and after the treatment policy was implemented. A third relevant aspect is that it covers twenty-five provinces, which allows us to have a more comprehensive idea of the impact of the insurance schemes on health outcomes. Finally, a continually updated database permits us to look into the interaction between insurance schemes and health some years after the establishment of schemes. A follow-up study based on recent database is relevant because more individuals are covered by the social medical insurance schemes currently, and previous results cannot be generalised.

The econometric analysis is conducted under four different econometric methodologies, namely individual fixed effect (FE), 2-stage-residual-inclusion (2SRI), difference-in-difference (DID), and propensity score matching with difference-in-difference (PSM with DID) estimations. To preview, it is found that social medical insurance schemes significantly increased hospital visit frequency in the previous year, especially for children who come from rural areas. However, we did not find evidence to support that participating in social medical insurance schemes improves children's health status.

This paper proceeds as follows. The next section describes the Chinese social medical insurance system. Section 3 presents our database and the variables considered in the paper. Section 4 discusses the econometric models considered in the paper. Main results are displayed and analysed in Section 5, Section 6 presents a more extended analysis and discussion of these results, and some concluding remarks follow in Section 7.

2. Background

The reform of China's social medical insurance schemes started in 1998 in order to deal with the influence of economic reforms which took place in the 1980s. In urban China, before the reform of urban social medical insurance, children were covered by social medical insurance packages offered by their parents' employers. Government officials and workers of state-owned enterprises were eligible for the Government Insurance Program (GIP) and the Labour Insurance Program (LIP), respectively. Their health insurances were paid by employers and medical expenses were reimbursed from the employers' pre-tax income (Gordon G et al., 1999). In addition to this, half of their children's medical expenses were also reimbursed by the employers (W. Chen et al., 2009). However, only 51% of urban population were beneficiaries of the insurance schemes of which 7% and 43% were covered by the GIP and the LIP, respectively, by the end of the 1990s (Gordon G et al., 1999). A large number of children whose parents were not covered by the two types of insurance programs had to pay for their health care out of their own pockets. Regarding rural China, farmers could join the Cooperative Medical Insurance Scheme (CMS) with money from the local collective welfare fund and individual monthly premium payments before the insurance reform (Hsiao, 1984). The insurance scheme varied widely among different places, and most of the time, children were excluded from it.

In spite of this, the three insurance schemes mentioned above played an important role in improving children's health status and relieving household financial burden. However, health care costs increased sharply, barefoot doctors were exodus from rural medical service system and urban enterprises were also challenged by poor financial performance due to the economic reforms (Gordon G et al. 1999; Hsiao 1984). Farmers in rural China found it more and more difficult to afford medical expenses for themselves and their childrens (Feng et al., 1995). Employers in urban China were no longer able to cover half-medical expenses for their employee's children. Even the coverage of employees could not be guaranteed (Y. Liu, 2002). Therefore, a new insurance scheme, called the Urban Employees Basic Medical Insurance (UEBMI), was established in 1998 to cover medical costs for all the urban employees (State Council, 1998). However, children were excluded, which caused a sharp increase in the number of uninsured children.

After some time, the Chinese government started to notice the importance of children's health care and two independent social medical insurance schemes, the New-Type Rural Cooperative Medical Scheme (NCMS) and the Urban Residents Basic Medical Insurance (URBMI), were established in rural China and urban China in 2003 and 2007, respectively. These schemes cover children and other uninsured individuals (State Council, 2003 & State Council, 2007). Before the establishment of these two insurance schemes, provinces would sometimes offer insurance guidance for children. However, the organizers, coverages and insurance premiums varied dramatically (W. Chen et al., 2009), and no nationwide social insurance schemes were available at that time.

Currently, NCMS and URBMI are the social medical insurance schemes obtainable to children, and both of them are considered in our analysis. They have some features in common. Firstly, both of the two schemes are voluntary programs that are funded by enrolee's premiums and by subsidies from central and local governments. Secondly, they both require full household participation in principle, which means children are either included or excluded from the program depending on their parents' participation (Li & Zhang, 2013).[1] More specifically, the household insurance premium is family size multiplies premium per capita. Thirdly, territory based insurance schemes require that only local residents are included. In addition to this, local designated hospitals usually offer relatively more convenient access to medical care and an easier reimburse process compared to local non-designated hospitals, as well as hospitals in other areas.

Government contributions and individual premiums of both schemes differ depending on the region's economic status and each individual's economic situation. At the beginning of the establishment of the schemes, it was required by the central government that the total government subsidies for each NCMS enrolee and URBMI enrolee should not be less than 20 RMB and 40 RMB, respectively.[2] Then the subsidies from various levels of government have increased considerably to about 450 RMB in 2017, and the premiums paid by enrolees increased to about 180 RMB at the meanwhile (Yizhou, 2017; Zongli & Long, 2017). It should be noted that children can get extra government subsidy, which may cover full premium in two special cases. One is for children who are in Dibao Program, which ensures minimum living standard for poor households; the other is for children with severe physical disabilities.

The reimbursement rates of both schemes vary according to the level of care. There are primary, secondary and tertiary care levels in the health care system, and among them primary care levels offer the highest reimbursement rate, while tertiary care levels offer the lowest. Similar to the enrolee's premium and government subsidy, reimbursement rates are also dissimilar in different regions of China. The coverages of NCMS's related to inpatient and outpatient service medical costs are approximately 70% and 50% respectively, while the coverage of URBMI's related inpatient

[1] Partial participation is also observed in reality due to household members migration or different registration types((Y. Chen & Jin, 2012).
[2] The exchange rate between the Chinese currency (RMB) and the US dollar at the study period was roughly 6.3 RMB for 1 dollar.

service medical costs is about 70% (The State Council Information Office of PRC, 2017). The URBMI also covers some outpatient chronic or fatal diseases (Li & Zhang, 2013).

There has been a considerable increment in the number of people who have taken medical insurance since the reform of the social medical insurance system, which started in 1998. Over 1.3 billion Chinese, which is 95% of the total population, have taken part in social medical insurance schemes (The State Council Information Office of PRC, 2017). Children account for 17% of the whole Chinese population and most of them also covered by the insurance schemes (The National Bureau of Statistics of PRC, 2016b).

Overall, the Chinese government has given high priority to children’s health care in the last 20 years. Besides social medical insurance expansion in children population, China also established a childcare management system in 2001 in order to offer disease screening for new-born babies (State Council, 2001). The rates of child care management under 3 years old and 7 years old were 91.1% and 92.4%, respectively by 2016 (The National Bureau of Statistics of PRC, 2016c). In addition to this, the number of health care institutions per 1000 people has more than doubled since 1998. More specifically, there were 74 hospital beds per 1000 people in 2016, in contrast to 31 in 1998. Maternal and childcare service centres offer specific treatment for women and children. There were 3021 maternal and childcare service centres in 2016, and hospital beds increased from 0.6 per 1000 people in 1998, to 2 per 1000 people in 2016.

Given the previous discussion, we focus on NCMS and URBMI which are available for Chinese children. The following section provides detailed information of our database.

3. Data

This study uses data from the China Family Panel Studies (CFPS) conducted by the Institute for Social Science Survey (ISSS) of Peking University. It officially launched its baseline survey in April 2010 and full-scale follow-up interviews took place every other year, with the last one happening in 2016. It is representative of all family members of households in 25 provinces in Mainland China, which accounts for 94.5% of the total Chinese population.

The CFPS provides information at individual, family and community levels. Here, we restrict our attention to individuals younger than sixteen years old. It is a longitudinal database, as the same individual can be identified in different years, but some of them were dropped from the sample due to deaths, migration of members as well as moving towards the adult database when their ages were over 16. In addition to this, each year there were new individuals included in the sample for reasons such as marriage or divorce happening in their family. The CFPS sample is self-renewing, based on the natural changes of the baseline Chinese families. Thus, it is in the ideal situation to avoid being subject to attrition over time. More detailed information about data collection on the CFPS database can be found in Xie & Hu (2014).

The four response variables taken into consideration are individual hospital visit frequency in last year, individual hospital visit frequency in last month, individual frequency of sickness in last month and self-rated health status. The yearly and monthly hospital visit frequency are measurements of health care utilization, they include hospital visits due to illness, and exclude vaccinations, routine physical examination, or other things alike. Monthly sick frequency and self-rated health status are measurements of health status. Self-rated health status seizes individual's own assessment of health. It takes values 1 to 5 which means individual's self-rated health status is excellent, very good, good, fair and unhealthy, respectively. We do not consider self-rated health status in 2010 because the meaning of their values was inconsistent with the other waves. As shown in Figure 1, the distributions of the four variables are clearly not normal as there is a large proportion of zeroes in yearly hospital visit frequency, monthly hospital visit frequency and monthly sick frequency which are 49.5%, 22.6% and 71.3%, respectively. Moreover, the mass of the distribution is concentrated in just a few numbers of discrete cases. As it will be discussed in Section 4, these two issues together suggest that a traditional linear OLS might not be the most suitable procedure for the analysis of these four variables.

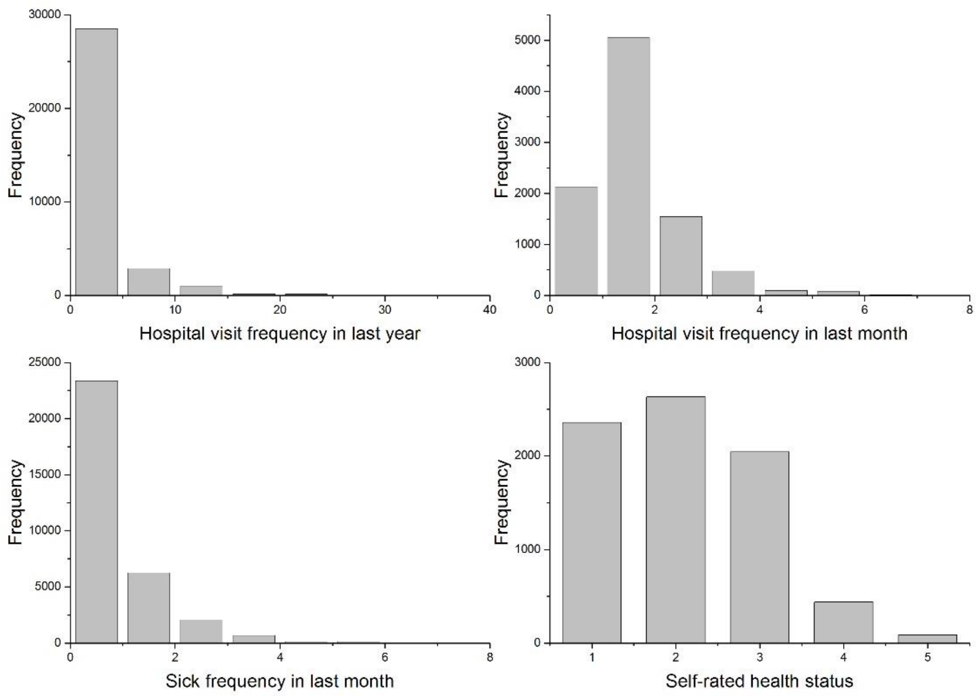

Figure 1 Distribution of response variables

Our treatment variable takes values 1 and 0 depending on whether or not the children have already participated in any of the two types of social medical insurance schemes: NCMS and URBMI. A figure of the dynamic evolution of the overall take-up rate is not shown for the sake of brevity.

However, it can be mentioned that it shows a clear upward trend along time with the lowest take-up rate in 2012 (50.9%) and the highest in 2016 (85.3%) with an average value of 65.7%.

Control variables are split into two types: predisposing and enabling. Regarding the first group, some individuals are inclined to use medical care more than others, and this can be captured by the following individual characteristics: Age, Squared Age, Gender, Han (Ethnicity), Grade, Height and Weight. More specifically, the inclusion of Age can be justified because younger children may have a poor immune system, which may result in more visits to the hospital. Moreover, parents may take different insurance purchase decisions according to their children's age. Squared Age is used to model more accurately the effect of age, since there is a chance it could be non-linear. Regarding the inclusion of Gender, males are generally involved in more risky behaviour and females' mortality rate may be higher due to environmental disadvantages in remote areas (Waldron, 1983). Therefore, in principle, the influence of Gender on health is ambiguous. Regarding Ethnicity, although there are 56 ethnic groups in China, one of these groups, Han people, has the largest population. All the remaining groups are called minorities and they usually live in underdeveloped provinces with relatively primitive medical facilities which may lead to less health care utilization, and a poorer health status. According to this, we consider a simple binary ethnicity variable that takes value 1 for Han people and 0 otherwise. Grade is a categorical variable for children's education stage with values 1 to 5 which means nursery education, primary education, lower secondary education, upper secondary education and tertiary education, respectively. Students usually participate in some insurance schemes offered by their schools, having on average higher insurance coverage than children who have dropped out of school. Weight and Height are traditionally important indicators for children's health.

Enabling variables are related to the availability of medical services. We include Income, Education Cost, Registration Type (Hukou) and Urban Area in this variable list. Income is measured by family's total income from the previous year. Parents with higher incomes have the financial ability to pay for their children's insurance premiums and health related service's fees. This means their children cannot only have more opportunities to get insurance products but also may have better health outcomes because of efficient health care treatments. Education Cost is measured by children's education cost per year. Typically, schools with high tuition fees offer special commercial insurances for their students. More importantly, Education cost also relates to family's income and they must be included together in the analysis. This variable could interact with treatment in two possible ways. First, a higher education cost may increase the financial burden of the families; therefore they may reduce their expenditure on insurance products. A second possibility is that a higher education cost indicates that children come from a higher social status and therefore they have the financial ability to get more insurance coverage. Registration Type (Hukou) is measured by the child's current household registration type which includes agriculture registration and non-agriculture registration based on whether the household origin is rural or urban, respectively. This variable takes value 1 for agricultural registration and 0 for non-agricultural registration. Urban Area is defined by the children's current living area without considering their family origin. It equals to 1 when a child is currently in an urban area and 0 otherwise. Hukou and Urban Area can identify the original area and migration of our sample. They are relevant as there are different social insurance schemes in urban and rural China. In addition to this, many inhabitants who originally come from rural areas and have agricultural registration come to the urban area seeking their fortune because of urbanization. Their children, who then attend school in an urban area, usually have no limitations of registration type when participating in social medical insurance schemes.

The common issue of microdata is the existence of missing values in different observations for different regressors. It imposes a serious limitation in the degrees of freedom of the regression. Due to this problem, we apply the EM algorithm to tackle data irregularities (Graham, 2009). It includes two steps: the expectation step and the maximization step. In the expectation step, every variable with missing values is regressed on all other ones restricted to individuals with the observed variable. Year dummies and province dummies are also included in the model for a more precise imputation. Furthermore, the missing values are substituted with the estimations from the regression model. This regression model varies with the type of variables with missing data. For continuous variables such as Income, Age, Squared Age, Weight, Height and Education Cost, we use linear regression models to impute the missingness. For binary variables like Urban Area, Hukou and Han logit models are applied. Finally, for Grade, a categorical variable, an ordered logistic model is applied. We follow Von Hippel (2007) for excluding response variables from the imputation model, since artificial correlation between the control variables and response variables can contaminate our results. In the maximization step, the missing values are substituted again by repeating the regression. This process is repeated iteratively for every variable until the likelihood ratio reaching convergence.

Table 1 shows the main statistics of the variables by insurance before and after the EM imputation procedure has been applied. The 'Uninsured', 'Insured' and 'All' columns show the mean of the different variables for uninsured people, NCMS and NRBMI participants, and the whole sample respectively. It is obvious that due to the presence of missing values, the number of observations is not the same in all the variables before the imputation. This problem is especially severe for Grade, Expenditure Cost, Han and Household Income. Regarding the remaining variables besides the response ones, the number of gaps is very small and always below 5% of the total sample. In general, imputation increases the number of observations without having a significant impact on variables features. For this reason, imputed data are considered in the preferred analysis developed in the following sections. However, we also study the robustness of our results to this transformation.

Table 1 Statistic summary by insurance and imputation

	Before Imputation				After Imputation			
	Uninsured_b	Insured_b	All_b	#Observations_b	Uninsured_a	Insured_a	All_a	#Observations_a
Panel 1: Utilization								
#Hospital visit in last year	1.5(0.027)	2.1(0.026)*	1.9(0.019)	32,222	1.6(0.027)	2.1(0.025)*	1.9(0.019)	32,789
#Hospital visit in last month	1.1(0.020)	1.1(0.012)	1.1(0.011)	9,198	1.1(0.02)	1.1(0.012)	1.1(0.01)	9,418
Panel 2: Health status								
#Sick in last month	0.5(0.009)	0.4(0.006)	0.4(0.005)	32,222	0.5(0.009)	0.4(0.006)*	0.4(0.005)	32,789
Self-rated health status	2.1(0.021)	2.1(0.013)	2.1(0.011)	7,535	2.1(0.021)	2.1(0.013)	2.1(0.011)	7,551
Panel 3: Independent variables								
Insurance	0 (0.000)	1(0.000)	0.7(0.003)	32,222	0(0.000)	1(0.000)	0.7(0.003)	32,789
Age	6.8(0.045)	7.8(0.029)*	7.5(0.025)	32,213	6.8(0.044)	7.8(0.029)*	7.5(0.025)	32,789
Gender[1]	0.5(0.005)	0.5(0.003)	0.5(0.003)	32,222	0.5(0.005)	0.5(0.003)	0.5(0.003)	32,789
Grade	2.1(0.010)	2.2(0.009)*	2.2(0.007)	20,404	1.7(0.008)	2(0.007)*	1.9(0.006)	32,789
Han	0.9(0.003)	0.9(0.002)*	0.9(0.002)	28,393	0.9(0.003)	0.9(0.002)*	0.9(0.002)	32,789
Height (cm)	111.7(0.334)	119.6(0.215)*	117(0.183)	30,634	110.4(0.323)	118.9(0.21)*	116(0.178)	32,789
Weight (0.5 kg)	47.4(0.273)	53.0(0.192)*	51.1(0.158)	31,296	47.7(0.267)	53.1(0.188)*	51.2(0.155)	32,789
Hukou[2]	0.8(0.004)	0.8(0.003)*	0.8(0.002)	32,148	0.8(0.004)	0.8(0.003)*	0.8(0.002)	32,789
Urban Area	0.4(0.005)	0.4(0.003)	0.4(0.003)	31,982	0.4(0.005)	0.4(0.003)	0.4(0.003)	32,789
Household Income (10M RMB)	4.3(0.061)	5.2(0.052)*	4.9(0.041)	30,436	4.3(0.057)	5.2(0.051)*	4.9(0.039)	32,789
Education Cost (10M RMB)	0.2(0.004)	0.2(0.004)*	0.2(0.003)	22,201	0.2(0.004)	0.3(0.003)*	0.3(0.003)	32,789

Standard error of mean in parentheses.
* indicates the difference between insured group and uninsured group is significant at 5% level.
Response variables are not included in the imputation model, the reason for the difference of response variables before and after imputation is due to the missingness of 'Insurance'.
[1] Gender takes value 1 for male, and takes value 0 for female.
[2] Hukou takes value 1 for agriculture hukou, and takes value 0 for non-agriculture hukou.

4. Methodology

Our purpose is to estimate the causal impact of participating in social medical insurance on health care utilization and health status. It should be noted that unobservable individual characteristics, like risk preference and time preference, could affect individuals' insurance enrolment decisions, their health care utilizations and health status. Therefore, failing to control it could result in bias estimation. For robustness, here this issue is considered under four alternative methodologies. The first estimation (FE henceforth) is based on the following regression model

$$Y_{it} = \alpha_0 + \alpha_1 INS_{it} + \alpha_2 X_{it} + T_t + \gamma_i + \gamma_p + \varepsilon_{it} , \tag{1}$$

where Y_{it} is the response variable, either hospital visit times in last year; hospital visit time in last month; or sick frequency in last month; or self-rated health status, in year t for individual i; INS_{it} is a dummy variable that takes value 1 if individual i took the treatment in year t and zero otherwise; X_{it} is a *11x1* vector including the control variables which are predisposing variables and enabling variables defined in the data section, T_t is a year fixed effect; γ_i and γ_p are individual and province fixed effects; ε_{it} is the error component and α_i, for i=0, 1 and 2 are parameters to be estimated. Our focus estimation is a_1 which explains the impact of participating in social medical insurance schemes on our response variables.

Although FE already controls for omitted variable bias caused by unobservable time-invariant individual characteristics, however, there might be some remaining time-varying reverse causality between insurance and children's health utilization as well as health status. An instrument variable method is a common way to deal with this issue. Given that our potential endogenous variable, INS, is a dummy variable, a 2-stage-residual-inclusion (2SRI) is more desireable than a 2-stage-least-square (2SLS) (Terza et al. 2008). In the first step, we estimated INS using the usual controls plus the percentage of children population with insurance in each province as instrumental variable using logit models. This last variable was proved to be a strong instrument but it can be assumed it does not explain changes in individual health expectations. In the second step, we add the residual getting from the first step in the FE model.

A difference-in-difference (DID) approach is especially desirable in this context, as it mimics an experimental research design by comparing the effect of treatment on a treated group versus a control group. This approach has become increasingly popular in studying the impact of insurance on health outcomes; see for example Kaestner et al. (2001) and Liu et al. (2002). DID estimation is based on the difference between the response variables for treated and control units before and after the intervention. It can be obtained from the following regression analysis. In our baseline estimation, we consider the impact of insurance between each two consecutive waves. Therefore, an individual belongs to the treatment group if she/he is not insured in the first wave, but insured in the second wave, while individuals in the control group are not insured in any of the two waves.

An alternative to this approach is propensity score matching with difference-in-difference method (PSM with DID), which is to estimate the DID regression on similar individuals from treated and control groups based on observables. PSM with DID method has been used in empirical works by, for example, Lei & Lin (2009) and Chen & Jin (2012). Similar individuals in the treated and control groups can be obtained by using different matching methods, which include nearest neighbour matching, caliper matching, kernel matching, etc. Here, we use the kernel matching method because it achieves a lower variance compared to other alternatives as more information is used (Baser, 2006). Kernel matching is a non-parametric matching estimator that uses weighted averages of all individuals in the control group to construct the counterfactual outcome. Comparing to other matching methods, kernel method does not match observations in control to any given treatment observation, but rather constructs a weighted average of all observations in the control group of the sample as a hypothetical comparison observation.[3] The weights are determined by the distance to the treatment observation: closer comparison observations always receive larger weights (Mensah et, al., 2010). The precise nature of the weighting is determined by the form of the kernel and, more importantly, its bandwidth: a larger bandwidth tends to lead to lower standard errors, but endangers the identification assumption of conditional independence (Silverman, 1986). Based on Heckcan et al. (1997), the bandwidth we choose is 0.06 which optimises the trade-off between variance and bias.

Note that, although in the PSM with DID approach treatment and control groups are chosen based on observables, this methodology removes at least any bias based on time-invariant non-observable individual characteristics.

[3] By applying nearest neighbour matching technique, one or more individuals from the control group are chosen to match a treated individual that is closest regarding propensity score. Caliper matching further imposes a tolerance level on the maximum propensity score distance. However, both of them only consider limited information of the control group.

It is noteworthy that the four approaches are defined for the linear case, however, it is easy to adjust these models for the case in which the regressions are non-linear. This is relevant because as it has been shown, the response variables take discrete values and neither of them follows a normal distribution. However, it is straightforward to adjust equation (1) and our 2SRI, DID and PSM with DID to allow for a categorical dependent variable such as a logit model. Using this type of specification, it is possible to estimate the marginal effect of participating in social medical insurance schemes on children's health outcomes.

5. Results

Table 2 shows the estimated causal effects of social medical insurance on the different indicators of health care utilization, and health status of children described in the data section. The first four columns report estimations of the impact of our treatment variable for each response variable, considering the FE, 2SRI, DID and PSM with DID approaches applied to the imputed database. The last four columns show similar results using the non-imputed database. These initial estimations do not take consideration of the discrete nature of our response variables previously discussed. However, we start with this standard approach for comparison with the previous literature (Lei & Lin, 2009; Chen & Jin, 2012; Li & Zhang, 2013).

It can be noticed that both imputed and non-imputed databases yield similar qualitative results. Based on the fact that estimations with the imputed database are more precise due to the use of more observations, we will carry out our analysis with the imputed database. However, subsequent results in this paper are robust to the consideration of the non-imputed database.[4]

[4] These estimations are available from the authors upon request.

Regarding the impact of insurance on health care utilization, all four estimations consistently show that participating in social medical insurance schemes significantly increases health care utilization when it is observed for a long time period. More specifically, being insured increases yearly hospital visit frequency on average by a range of 0.2-0.8 times. However, its impact on monthly hospital visit frequency is not significant. Regarding health status, social medical insurance schemes significantly increase sick frequency in the previous month. However, this significance is only marginal when the PSM with DID method is used. In addition to this, we could not find a significant causal effect on children's self-rated health status. An explanation of this perverse effect could suggest the presence of potential endogeneity problems of the treatment variable. That is, individuals who can foresee they will suffer in the future are more likely to buy a health insurance. However, it is important to note that individual fixed effects already takes into account time invariant unobserved individual characteristics. Moreover, our instrumental variable estimation provides similar results. Thus, three more plausible explanations for the lack of a positive effect of insurance on health status are: (1) an increase of health expectations by insured individuals, (2) the fact that medical treatment only have an effect on health in the long-run and (3) the presence of a moral hazard problem for insured people because of a reduction in future costs associated to illness (Ehrlich & Becker, 1972; Bates et al. 2010). These issues will be discussed in the following section.

Table 2 Impact of the social medical insurance schemes on health outcomes

	Imputed				**Non-Imputed**			
	(1)	(2)	(4)	(4)	(5)	(6)	(7)	(8)
	FE	2SRI	DID	PSM with DID	FE	2SRI	DID	PSM with DID
#Hospital visit in last year	0.219***	0.776**	0.431***	0.408***	0.080	0.011	0.268**	0.299**
	(4.27)	(2.05)	(4.33)	(3.99)	(1.14)	(0.02)	(2.21)	(2.12)
R-Square	0.594	0.594	0.040	0.005	0.740	0.740	0.036	0.003
#Observations	32,789	32,789	14,450	14,444	12,661	12,661	5,169	5,220
#Hospital visit in last month	0.015	0.468	0.016	0.0004	-0.013	1.095	-0.192	-0.096
	(-0.33)	(1.19)	(-0.24)	(-0.01)	(-0.11)	(0.95)	(-1.60)	(-0.79)
R-Square	0.754	0.754	0.023	0.002	0.906	0.906	0.037	0.011
#Observations	9,418	9,418	4,162	4,162	2,725	2,725	1,179	1,115
#Sick in last month	0.016	0.037	0.061**	0.057*	-0.009	0.018	0.041	0.038
	(-1.11)	(0.36)	(-2.04)	(-1.84)	(-0.37)	(0.08)	(-0.92)	(-0.85)
R-Square	0.528	0.528	0.037	0.004	0.665	0.665	0.024	0.003
#Observations	32,789	32,789	14,450	14,444	12,661	12,661	5,169	5,220
Self-rated health status	-0.018	0.368	0.047	0.122	0.013	-0.160	0.010	0.094

	(-0.44)	(1.44)	(0.55)	(1.16)	(0.14)	(-0.18)	(0.07)	(0.69)
R-Square	0.734	0.734	0.013	0.005	0.899	0.899	0.020	0.004
#Observations	7,551	7,551	2,598	1,883	4,052	4,052	1,039	920

T statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

As discussed before, neither of the dependent variables follows normal distribution. More specifically, all our dependent variables, besides self-rated health status, take a few positive discrete values and zeroes while self-rated health status is measured by an ordinal 5-point scale, which implies that the distance between a 1 (excellent) and 2 (very good) has the same meaning as the distance between a 4 (fair) and 5 (pool). Therefore, a proper regression analysis should not be based on a standard OLS estimation, but on an alternative methodology that takes into account these data features.

According to this discussion, we split the different discrete values that the four response variables may take into ordered groups and specify a bivariate logit model for the probability to belong to each of these groups compared to the lowered order group. In particular, for hospital visit frequency in the previous year, the following four subgroups are defined: low hospital visit (LHV_y); fair hospital visit (FHV_y); high hospital visit (HHV_y) and higher hospital visit (HrHV_y) for individuals who went to hospital 1 , 2 , 3 and more than 3 times, respectively throughout the course of the past year. For example, in the LHV_y group, the response variable is a dummy variable which takes value 0 and 1 for individuals who did not go to hospital and went to hospital 1 time in last year, respectively. In the FHV_y group, the dummy response variable takes value 0 and 1 for individuals who went to hospital less than 2 times and 2 times in last year, respectively. Definitions of all the other groups are alike. Similarly, the following four subgroups are defined for hospital visit frequency in last month: low hospital visit (LHV_m); fair hospital visit (FHV_m); high hospital visit (HHV_m) and higher hospital visit (HrHV_m) for individuals who went to hospital 1 , 2 , 3 and more than 3 times, respectively in last month.

For sick frequency in last month, four subgroups are defined: low sick frequency (LSF_m); fair sick frequency (FSF_m); high sick frequency (HSF_m) and higher sick frequency (HrSF_m) for individuals who got sick 1, 2, 3 and more than 3 times, respectively. Regarding self-rated health status, we follow Simon et al. (2017) and dichotomize it into three indicators: “excellent (E),” “very good or better (VG),” and “good or better (G),” for self-rated health status values of 1, 2 and 3, respectively.

Note that as estimation results under these four proposed econometric approaches are similar, for the sake of brevity and also because it is the most accurate estimation method, we carry on the analysis only showing results under the PSM with DID approach from now on. Table 3 shows the marginal effects at mean values of PSM with DID method of each sub-group. It can be noted that social medical insurance schemes significantly increase health care utilization in LHV_y, HHV_y and HrHV_y groups. This indicates that having insurance increases the probability of using hospitals per year by 4.3%, 2.2% and 3.6% for the LHV_y FHV_y and HrHV_y groups, respectively. This is particularly interesting, as it suggests the insurance schemes not only encourage children to start using medical services, but also enable them to take more medical treatments when necessary, as frequent hospital visits are typically associated with medical treatments. However, being insured does not increase the monthly hospital use in each sub-group which can be due to the fact that a significant change of health care utilization may not be observable during such a short time period.

Regarding sick frequency, participating in the insurance schemes significantly increases sick frequency in LSF_m group. Noted this is a marginal significance and only happened in low sick frequency group which suggests that there is not strong evidence about a deterioration of health status caused by social medical insurance. This result is consistent with the aggregate analysis under the PSM with DID method reported in Table 2.

Regarding the dichotomized self-rated health status, participating in the insurance schemes does not have significant influence on the probability of reporting ‘very good’ or ‘good’. However, we observe a significant decrease in the probability of reporting ‘excellent’ of 10.8% which seems to indicate that insurance only has a significant negative effect on self-rated health status for those who already have an excellent health condition.

Table 3 Impact of the social medical insurance schemes on health outcomes in different heath status groups. Marginal impacts evaluated at mean values.

Variable	(1)	(2)	(3)	(4)
	LHV_y	FHV_y	HHV_y	HrHV_y
#Hospital visit in last year	0.043**	0.019	0.022*	0.036***
	(2.40)	(1.31)	(1.78)	(2.65)
Pseudo R-Square	0.002	0.002	0.002	0.004
#Observations	9,621	11,111	12,293	14,444
	LHV_m	FHV_m	HHV_m	HrHV_m
#Hospital visit in last month	0.006	0.014	-0.015	0.008
	(0.18)	(0.52)	(-1.12)	(0.63)
Pseudo R-Square	0.003	0.001	0.003	0.017
#Observations	3,166	3,847	4,063	4,162
	LSF_m	FSF_m	HSF_m	HrSF_m
#Sick in last month	0.027*	0.015	-0.002	0.005
	(1.72)	(1.50)	(-0.41)	(1.12)
Pseudo R-Square	0.001	0.002	0.003	0.018
#Observations	13,058	13,979	14,288	14,444
	E	VG	G	
Self-rated health status	-0.108**	-0.030	0.008	
	(-2.41)	(-0.57)	(0.31)	
Pseudo R-Square	0.007	0.002	0.002	
#Observations	1,883	1,883	1,883	

Z-statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

6. Extended Analysis

It is also particularly relevant to study the differential impact of insurance schemes on high and low income families. In fact, Currie and Gruber (1996) and Kaestner et al. (2001) indicate in their highly influential papers that low income children are the main target of social insurance expansion in many countries, and their corresponding health outcomes are worth investigating. As mentioned in the background section, individuals living under the national poverty line (NPL) can get government subsidies to cover the entire insurance premium. Therefore, it is apparent that the

Chinese government wants more financially disadvantaged children to get access to insurance schemes benefits. However, the efficiency of this affirmative action needs to be further examined.

In order to do this, we classify children into two groups and include individuals in the below-NPL and above-NPL groups depending on whether their household per capita income is less or higher than the NPL which is 2300 RMB per capita, respectively.

Table 4 presents PSM with DID estimations of the impact of social medical insurance on health outcomes of children in different income sub-groups by using an ordered logit model. Only one estimation, ignoring the discrete nature of the response variables for each of the two groups, is presented because, due to the small number of observations, it is impractical to perform different estimations for each of the different discrete values of the response variable. It can be seen that insurance participation significantly increases yearly hospital visit frequency in both income groups but consistently with the previous analysis, no improvement for monthly hospital visit times is observed. Self-rated health status is not affected by insurance treatment in neither of the groups. However, consistently with our previous results, a significantly positive impact of insurance on monthly sick frequency is observed in the estimation for the low income population.

Table 4 Impact of the social medical insurance schemes on health outcomes in different income groups

	(1)	**(2)**
	Income_above NPL	Income_below NPL
#Hospital visit in last year	0.241***	0.504***
	(3.29)	(3.07)
Pseudo R-Square	0.002	0.003
#Observations	11,843	2,601
#Hospital visit in last month	-0.027	0.219
	(-0.20)	(-0.76)
Pseudo R-Square	0.000	0.002
#Observations	3,454	708
#Sick in last month	0.070	0.617***
	(-0.82)	(-3.36)
Pseudo R-Square	0.001	0.007
#Observations	11,843	2,601
Self-rated health status	0.214	0.934
	(1.00)	(1.64)

Pseudo R-Square	0.002	0.009
#Observations	1,580	303

Z-statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

There is another important distinction in regards to the different impact of insurance in rural and urban areas. Children's health outcomes, and also the possibility of delivering a proper health treatment, may vary according to their living environments, healthcare facilities and socio-economic factors. In addition to this, social medical insurance schemes of rural and urban China have different target populations. Therefore, some researchers study urban insurance and rural insurance schemes separately (Li & Zhang, 2013). The estimations of ordered logit model using PSM with DID method for rural and urban areas are shown in Table 5. It can be seen that yearly hospital visit frequency significantly improves among rural population, but not among the urban population. However, no evidence is found for the improvement of monthly hospital visit, monthly sick frequency and self-rated health status of both groups after participating in the insurance scheme.

Table 5 Impact of the social medical insurance schemes on health outcomes in different geography groups

	(1)	**(2)**
	Rural	Urban
#Hospital visit in last year	0.413***	0.121
	(4.81)	(1.14)
Pseudo R-Square	0.002	0.001
#Observations	8,987	5,457
#Hospital visit in last month	-0.022	0.148
	(-0.14)	(-0.72)
Pseudo R-Square	0.002	0.002
#Observations	2,550	1,612
#Sick in last month	0.155	0.205
	(-1.60)	(-1.64)
Pseudo R-Square	0.002	0.002
#Observations	8,987	5,457
Self-rated health status	0.288	0.234
	(1.15)	(0.71)
Pseudo R-Square	0.004	0.001
#Observations	1,197	686

Z-statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Our different analyses consistently show that social medical insurances increase the use of hospitals, however, the increase of hospital use may occur either because of a better access to medical services or the increase of illness (Lei & Lin, 2009). In order to figure out this issue, we include a dummy variable for sickness, which takes value 1 and 0 for those who were sick and who were not sick in the previous month, respectively. We check whether the effect of insurance on hospital use is different for these two groups by using propensity score matching with triple differences method (PSM with DDD).

Column (1) of Table 6 shows the PSM with DDD estimation by using ordered logit model for the aggregate analysis and followed by the marginal impacts evaluated at mean values of PSM with DDD method by using logit model for four sub-groups which are defined as in Table 3. It shows that hospital use does not vary significantly between the two groups, which means better access to medical services and increase of illness are not significantly different to explain the increase of hospital use. This result is inconsistent with Lei & Lin (2009), who found that uninsured people accessed hospital more when sick. However, they considered this result as unexpected, which could be due to the small number of observations.

Table 6 Impact of the social medical insurance schemes on yearly hospital use in different sick groups by using triple differences method.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Variable	#Hospital visit in last year	LHV_y	FHV_y	HHV_y	HrHV_y
Insurance	-0.02	-0.031	-0.007	0.038	-0.017
	(-0.14)	(-1.04)	(-0.25)	(1.32)	(-0.86)
Pseudo R-Square	0.051	0.072	0.028	0.035	0.105
#Observations	14,444	9,621	11,111	12,293	14,444

Z-statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Some of the reasons to explain the absence of a positive effect of insurance on health status require a more detailed analysis. The possibility of a moral hazard problem is not, in principle, a main concern in our case because many risky health behaviours like smoking and drinking are not common in children population. Moreover, many aspects of their life styles are mainly decided by their parents. Despite these considerations, we explore the possibility that an insured family could have less incentives to lead a healthy live by focussing our attention on Body Mass Index (BMI) which is a common measurement of moral hazard (Simon et al., 2017). In particular, we estimate the causal impact of buying an insurance on BMI using PSM with DID method.

Given that it can be argued that the impact of BMI on health is not linear, but it is only a serious concern when it surpasses the threshold of overweight, we define a dummy variable denoted by Overweight which takes value 1 and 0 depending on whether an individual's BMI is more than 25 or not and estimate the impact of insurance schemes on this variable. This is a reasonable concern as overweight is associated with various risky health behaviours which include lack of physical activity and unhealthy eating patterns (Middleman et al., 1998). Thus. Table 7 presents the results of these estimations. There is no evidence showing that participating in the insurance schemes leads to a significant increase either of BMI or overweight. This is consistent with Simon et al.(2017).

Table 7 Impact of the social medical insurance schemes on BMI and overweight

Variable	**(1)**	**(2)**
	BMI[1]	Overweight
Insurance	-0.127	0.001
	(-0.51)	(0.12)
Pseudo R-Square	0.003	0.006
#Observations	14,444	14,444

T-statistic in parenthesis in column (1) and z-statistics in parenthesis of column (2).
* p<0.1; ** p<0.05; *** p<0.01.
[I] BMI is the ratio of weight in kilograms to height in square metres.

Another alternative explanation for the lack of impact of insurance on health status is that insurance may need more than two years to have a positive effect. We deal with this problem considering an additional estimation considering three waves given that estimating causal effects based on two consecutive waves could be regarded as too short period for this decision to take effect on health outcomes. Therefore, individuals belonging to the treatment group in the new estimation are those who got a medical insurance in the last two waves, but did not get insurance in the first one. Control group includes individuals who did not get insurance in any of the three waves. By doing this, we are able to detect the change of health outcomes after being insured at least four years. Table 8 presents the PSM with DID estimation by using ordered logit model. It shows that participating in social medical insurance can increase the yearly hospital visit frequency in long term. However, it does not have a significant influence on monthly hospital visit frequency and health status. This new estimation is also applied to different income and geography groups.[5] All of these results agree with our previous estimations.

Table 8 Impact of social medical insurance schemes on long-term health outcomes

	(1)	(2)	(3)	(4)
	#Hospital visit in last year	#Hospital visit in last month	#Sick in last month	Self-rated health status
Insurance	0.454***	0.168	0.209	-0.232
	(3.40)	(0.65)	(1.33)	(-0.26)
Pseudo R-Square	0.002	0.002	0.009	0.003
#Observations	4,828	1,457	4,828	212

Z-statistics in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Overall, it can be concluded from this analysis that participating in social medical insurance schemes significantly increases the access to medical services among children population, which is measured by yearly hospital use. In addition to this, we find evidence of a marginal deterioration of health status. However, this deterioration is not consistent with the adoption of more risky food habits. A potential explanation of this deterioration could be that individuals are inclined to have more health expectations after getting insurance.

Regarding hospital use, our results are consistent with existing studies which find that social medical insurance schemes increased hospital use (Li & Zhang, 2013) while inconsistent with Lei & Lin (2009) who find that social medical insurance did not affect formal medical services. However, an important difference with our research relates to the fact that the focus here is on children's population rather than the whole population. Our results are also consistent with the previous literature regarding the impact of insuranc on health status. In this respect, Chen and Jin (2012) found that there is no impact of social medical insurance on mortality rate in children population

[5] These estimations are available from the authors upon request.

under a cross sectional database. However, mortality rate is an extreme event for children, compared to sick frequency, to measure the health status of children.

By disaggregating our analysis into different income sub-groups, we found that social medical insurance schemes increase the yearly use of hospitals in both income groups, while significantly decreasing the health status among children whose family income is under NPL. It can be seen that the issue of rising health expectation is especially common among financially challenged individuals. According to Maslow's Hierarchy of Needs, only individuals who have already met the most basic needs start to think about improving their own health condition. In general, people under NPL struggle more about basic needs compared to people who above NPL. Therefore, being insured enables the poorer to consider more minor health problems.

When we disaggregate the analysis between rural and urban areas, it is found that the increase of hospital use is only significant in the former, while health is unaffected by insurance. This is an interesting result as rural population is, at least in principle, more constrained in the use of health facilities.

7. Concluding remarks

Social medical insurance schemes have experienced a rapid expansion in the latest years. However, the ways in which these schemes work, especially for children in China, have not been an issue of great concern in previous literature. This paper tries to fill this gap by examining the effect of social medical insurance schemes on 0-15-year-old Chinese's health care utilization and health status.

Our results clearly indicate that health insurance exerts a positive effect on hospital utilization especially for children living in rural areas. This is particularly relevant as children mortality rates in rural areas more than double that in urban areas (The National Bureau of Statistics of PRC, 2016a). Therefore, more hospital use in rural areas might have a positive influence in decreasing the gap of mortality rate between rural and urban China.

However, we do not find evidence of a positive impact of insurance on health status similar to the one found by Chen and Jin (2012). We have discussed different possible explanations for this result such as moral hazard, the impossibility to observe the long-run effect, and the possibility that individuals become more demanding after they get the insurance. Logic and some evidence suggest that, in principle, the last one seems the most plausible explanation which is consistent with the important decrease of the Chinese children population during the analysis period (UNICEF, 2016). However, this paper cannot provide a completely non-speculative answer to this question. Therefore, further research is warranted.

Acknowledgements

This research was completed while Jing Guan visited the Management School at Liverpool University. We want to thank participants at the XXI Applied Economics Meeting and seminar participants at the University of Liverpool. The usual disclaimer applies.

References

Baser, O., 2006, Too Much Ado About Propensity Score Models? Comparing Methods of Propensity Score Matching. Value in Health, 9(6): 377-385.

Bates, L. J., K. Mukherjee, and R. E. Santerre, 2010, Medical Insurance Coverage and Health

Production Efficiency, *The Journal of Risk and Insurance*, 77(1): 211–229.

Chen, W., H. Jiang, and Y. Huang, 2009, Evolution and Current Situation of Urban Children Health Security, *Chinese Journal of Health Policy*, 2(2): 18–23.

Chen, Y., and G. Jin, 2012, Does Health Insurance Coverage Lead to Better Health and Educational Outcomes? Evidence from Rural China, *Journal of Health Economics*, 31(1): 1–14.

Ehrlich I., and G. S. Becker, 1972, Market Insurance, Self-Insurance, and Self-Protection, *Journal of Political Economy*, 80(4): 623–648.

Feng, X., S. Tang, B. Gerald, M. Segall, and X. Gu, 1995, Cooperative Medical Schemes in Contemporary Rural China, *Social Science and Medicin*e, 41(8): 1111–1118.

Galor, O., 2005, The Demographic Transition and the Emergence of Sustained Economic Growth, *Journal of the European Economic Association*, 3(2): 494–504.

Graham, J. W., 2009, Missing Data Analysis: Making it Work in the Real World, *Annual Review of Psychology*, 60(1): 549–576.

Hsiao, W. C., 1984, Transformation of Health Care in China, *New England Journal of Medicine*, 310(14): 932-936.

Lei, X., and W. Lin, 2009, The New Cooperative Medical Scheme in Rural China: Does More Coverage Mean More Service and Better Health?, *Health Economics*, 18: 25–46.

Li, X., and W. Zhang, 2013, The Impacts of Health Insurance on Health Care Utilization Among the Older People in China, *Social Science and Medicine*, 85: 59–65.

Liu G. G., R. Cai, Z. Zhao, P. Yuen, X. Xiong, S. Chao, B. Wang, 1999, Urban Health Care Reform Initiative in China: Findings From Its Pilot Experiment in Zhengjiang City(1), *International Journal of Economic Development*, 1(4): 504–526.

Liu, G. G., Z. Zhao, R. Cai, T. Yamada, and T. Yamada, 2002, Equity in Health Care Access to: Assessing the Urban Health Insurance Reform in China, *Social Science and Medicine*, 55(10): 1779–1794.

Liu, Y., 2002, Reforming China's Urban Health Insurance System, *Health Policy*, 60(2): 133–150.

Mensah, J., J. R. Oppogng, and C. M. Schmidt, 2010, Ghana's National Health Insurance Scheme in the Context of the Health Mdgs: An Empirical Evaluation Using Propensity Score Matching, *Health Economics*, 19: 95–106.

Middleman, A. B., I. Vazquez, and R. H. Durant, 1998, Eating Patterns, Physical Activity, and Attempts to Change Weight Among Adolescents, *Journal of Adolescent Health*, 22(1): 37–42.

Miller, S., 2012, The Impact of the Massachusetts Health Care Reform on Health Care Use Among Children, *American Economic Review*, 102(3): 502–507.

Moav, O., 2005, Cheap Children and the Persistence of Poverty, *The Economic Journal*, 115(500): 88–110.

Silverman, B., 1986, Density Estimation for Statistics and Data Analysis. CRC Press. Available at https://doi.org/10.2307/2347507

Simon, K., A. Soni, and J. Cawley, 2017, The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions, *Journal of Policy Analysis and Management*, 36(2): 390–417.

Terza, J. V., A. Basu, and P. J. Rathouz. 2008, Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling, Journal of Health Economics, 27(3): 531–43.

The National Bureau of Statistics of PRC., 2016a, 2016 Statistical Report. Available at http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A0O0I&sj=2016

The National Bureau of Statistics of PRC., 2016b, Census. Available at http://data.stats.gov.cn/easyquery.htm?cn=C01

The National Bureau of Statistics of PRC., 2016c, Statistical Monitoring Report of "Chinese Children's Development Program (2011-2020), Available at http://www.stats.gov.cn/tjsj/zxfb/201710/t20171026_1546618.html

The State Council Information Office of PRC., 2017, Development of China's Public Health as an Essential Element of Human Rights. Available at http://www.xinhuanet.com/politics/2017-09/29/c_1121747583.htm

UNICEF., 2016, Population Status of Children in China in 2015 Facts and Figures. Available at http://www.unicef.cn/cn/uploadfile/2017/0927/20170927095135855.pdf

Von Hippel, P. T., 2007, Regression With Missing Y's: an Improved Strategy for Analyzing Multiple Imputed Data, *Sociological Methodology*, 37(1): 83–117.

Waldron, I., 1983, Sex Differences in Human Mortality: the Role of Genetic Factors, *Social Science and Medicine*, 17(6): 321–333.

Xie Y., and J. Hu., 2014, An Introduction to the China Family Panel Studies (CFPS). *Chinese Sociological Review*, 47(1): 3–29.