On the Role of Health in Climbing the Income Ladder: Evidence from China*

Gordon G. Liu[†], Franklin Qian[‡] & Xiang Zhang[§]

March 23, 2021

Abstract

This paper uses two large panel data sets in China to study the effects of a health shock on household income mobility from 1991 to 2016. We compare outcomes of households with a member who receives a health shock with comparable households that do not receive any health shocks. To do so, we match on demographic and worker characteristics of household members. At the aggregated level, a health shock lowers the probability of "getting out of the low-income trap" by 8.4 percentage points. At the household level, a health shock lowers household income per capita by 13.1% and income position by 3.9 percentiles. We show that the reduced labor supply at the extensive margin and decrease in labor productivity, measured in hourly wage, explain the majority of households' income loss. Households who become poor due to a health shock continue to exhibit lower income mobility after three years.

JEL Classifications: D31, H0, I14

Keywords: Income mobility, health shock, low-income trap, China.

*We thank Rebecca Diamond, Matthew Gentzkow, Han Hong, Petra Persson, Jesse Shapiro, and Alessandra Voena for very constructive comments. We thank seminar and conference participants at the CEA meeting, International Symposium of Contemporary Labor Economics, Peking University, and NBER-CCER summer institute for very helpful feedback and discussions. Liu thanks the National Natural Science Foundation of China and the China Medical Board for support. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the National Natural Science Foundation of China or the China Medical Board. We thank Debopam Bhattacharya and Bhashkar Mazumder for sharing their code. All errors are our own.

†National School of Development, Peking University. Email: gordonliu@nsd.pku.edu.cn

‡Corresponding author, Department of Economics, Stanford University. Email: zqian1@stanford.edu

§Department of Economics, Princeton University. Email: xiangzhang@princeton.edu

1 Introduction

There has been a growing realization that income mobility could serve as a long-run equalizer of income inequality (Shorrocks 1978). However, little is known about what are the factors that determine income mobility. Motivated by the fact that major illness is the most sizable shock to the economic opportunities in the developing countries (Gertler and Gruber 2002), in this paper, we focus on the role of health human capital in determining household intra-generational income mobility. This paper attempts to demonstrate the extent to which a decline in health human capital prevents impacted households from moving upward in the income distribution. Specifically, we provide empirical evidence for the contemporaneous and lasting effect of worsened health status on households' intra-generational income mobility at both aggregate and household levels. We find that a severe decline in health human capital has a significant negative impact on the treated households' income mobility and hurts their prospects of climbing the income ladder.

This paper uses two large panel data sets in China, the China Health and Nutrition Survey (CHNS) and the China Family Panel Studies (CFPS). These two data sets span from 1991 to 2016 and provide rich information on household demographics, socioeconomic, and health status. They allow us to track household income mobility in a relatively long time window and estimate the impact of a severe reduction in health human capital on household income mobility. To conduct our analysis, we construct two samples based on the CHNS and CFPS data, respectively. In each sample, for every treated household – a household with one household member who experiences a severe decline in health human capital – we match it with three control households that do not experience any reduction in health human capital. We compare outcomes of treated households with control households and estimate the impacts of health human capital on income mobility.

In the first part of the paper, we focus on estimating the impact of a severe deterioration in health human capital on aggregate-level income mobility. We proxy a severe decline in health human capital using a health shock, defined as either a severe deterioration in the self-reported health status or a hospitalization record. We characterize income mobility based on intra-generational transition probability and upward mobility. These two measures quantify household income mobility based on changes in their income rank in the national income distribution. The first measure highlights the probability of surpassing a certain income percentile in the income distribution, and the second measure emphasizes the success rate of a household moving a positive amount in the income distribution relative to its initial income position. We match each treated household to three control households that have similar household and worker characteristics. Using different health shock definitions and samples, we find that a household in the treatment group consistently exhibits lower income mobility than the control households, which do not experience any health shock. On the one hand, the treated households have a lower probability of surpassing a given position in the national income distribution. For example, we find that the treated households' probability of "getting out of the low-income trap," defined as the probability of transiting out of the bottom quintile of the income distribution, conditioning on starting in the bottom quintile, is 8.4 percentage points, or 13.06% lower than that of the control group in the CHNS sample. On the other hand, the probability of "climbing up the income ladder," defined as the probability of surpassing its initial income position, is 9.5 and 4.6 percentage points lower on average for a household that experienced a health shock in the CHNS and CFPS sample, respectively.

What are the household-level changes in income and income percentile caused by a health shock that correspond to the differences in aggregate-level mobility between the treatment and control groups that we observe? And what are the welfare consequences of decreased income mobility after a health shock? In the second part of the paper, we provide household-level evidence rationalizing the effects of a health shock on aggregate-level income mobility by evaluating the impact of experiencing a health shock on household income per capita and income percentile in the income distribution. Under two definitions of a health shock, our estimates suggest that receiving a health shock significantly lowers a household's income per capita and income percentile. We find that, relative to the control group, a health shock causes a 13.1% and a 14.2% drop in household income per capita for the treatment group in the CNHS sample and CFPS sample, respectively. Additionally, we document a 3.94 percentile and a 3.14 percentile drop in income percentile caused by a health shock in the two samples, respectively. These estimated treatment effects quantify the average income loss and distance that a household moves in the national income distribution when receiving a health shock. These estimates help rationalize the aggregate-level effects by revealing the changes in income position that corresponds to the decreased probability of making upward movements and quantifying the welfare loss. Taking our estimates at the aggregate and household levels together, we see that a health shock sharply lowers a household's probability of improving its income and income percentile and increases its risk of transiting into poverty.

Further, we explore possible channels through which a health shock affects household income mobility by looking at household labor supply responses to a health shock. Using an event study approach, we find that individuals who experience a health shock reduce their labor supply at the extensive margin. Additionally, in both samples, we obtain statistically insignificant but economically large estimates suggesting that the yearly working hours drop by about 4.5% for the treated individuals following a health shock. Most importantly, we find that a health shock imposes a sizable negative impact on the treated individuals' and their household members' labor productivity, measured by the hourly wage. Using a back-of-envelope calculation, we conclude that the reduced labor supply at the extensive margin and the drop in labor productivity explain the majority of household income loss after a health shock.

We next analyze treatment heterogeneity along several dimensions. First, we provide evidence suggesting that, compared with urban households, rural households experience more substantial income loss when facing a health shock. In the CHNS sample, a rural treated household loses 9% more income per capita than an urban treated household. In the CFPS sample, a rural treated household loses 19% more income per capita than an urban treated household. These findings are consistent with the idea that rural households earn income mostly from manual labor and thus are more vulnerable to sudden income loss when receiving a health shock. Moreover, these estimates suggest a widening urban-rural gap in recent years as the CHNS data ranges from 1991 to 2015 while CFPS data covers 2010 – 2016 only. Second, we provide suggestive evidence that a health shock impacts households with different health insurance differently. Finally, we document that a health shock imposes different effects on households at different income positions in the national income distribution.

Finally, we track the medium-run income mobility of the treated households after the arrival of a health shock using the CFPS sample. We present two main empirical findings. First, we provide suggestive evidence that the treated households have a lower transition probability relative to the control households up to three years after the health shock. That is, the treated households have a lower probability of reaching a given income percentile compared with control households. For example, a treated household's probability of "getting out of the low-income trap" is 57.6%. In contrast, a control household in the bottom quintile has a 64.9% probability of succeeding. Second, we do not find a significant difference in upward mobility between the treatment and control groups after the health shock, suggesting that they have a similar ability to make upward movements. Combining the two findings, we show that households who experience a health shock have a lower probability of making long-distance upward movements and recover to their original income position in up to three years after a health shock. They still enjoy fewer economic opportunities and are more likely to be caught in the low-income trap in the medium-run. We also estimate the medium-run impact of a health shock on household income per capita and household out-of-pocket medical expenditure per capita. Based on these results, we estimate that the value of insurance covering both income loss and the additional medical expenditure has a certainty equivalent value of at least 4,944 RMB (760 USD) and an upper bound of 20,366 RMB

(3,133 USD) per person.

This paper is related to three bodies of literature. First, it contributes to the literature that tries to understand how health human capital is related to income mobility, by presenting new causal evidence on the impact of a health shock on intra-generational income mobility. Previous research has discussed the impact of health human capital on income mobility under an inter-generational context. Through direct and indirect transmission, health human capital introduces parent-induced inequality of economic opportunities among offspring (e.g., Corak 2013; Qin et al. 2014). However, to the best of our knowledge, our paper provides the first causal evidence of a health shock on household income mobility in an intra-generational context. Second, our paper is related to a nascent literature on the driving forces of income mobility. In the inter-generational context, several papers have documented factors that correlate with or drive high-income mobility.[1] At the macro level, Mayer and Lopoo (2008) suggest that higher government expenditure could reduce the investment gap in human capital between rich and poor children and, thus, generate higher inter-generational income mobility. Chetty et al. (2017) find that two factors – lower economic growth rate and more unequal income distribution – are responsible for the declining income mobility in the United States, with the latter factor playing a more critical role. At the micro level, Ermisch et al. (2006) suggest that assortative mating plays an important role in explaining inter-generational income mobility. Chetty et al. (2014) find that high income mobility is highly correlated with less residential segregation, less income inequality, better primary schools, greater social capital, and greater family stability in a commuting zone. Chetty and Hendren (2018a) and Chetty and Hendren (2018b) provide causal evidence that neighborhoods affect income mobility through the childhood exposure effect. More recently, Chetty et al. (2020) find that access to colleges varies greatly by parents' income, and the rates of upward mobility differ substantially across colleges. Third, our paper adds to the literature on intra-generational income mobility in the developing country context. For example, Khor and Pencavel (2006) and Chen and Cowell (2017) document China's high intra-generational income mobility. Some other papers document factors that are correlated with high income mobility in China. For example, Sun et al. (2007) attribute high income mobility in China to improvement in educational attainment, expansion of the labor force, and more migrant workers within a household. Zhang et al. (2007) find that the implementation of the land circulation policy, is also relevant.[2]

Our paper distinguishes from past works in three important dimensions. First, to the best

[1] Some earlier attempts failed to draw a clear causal inference. See Black et al. (2011) for a survey.

[2] Fan et al. (forthcoming) provide a comprehensive analysis of the rising income mobility and its correlates in China in an inter-generational context.

of our knowledge, our paper is the first to draw causal inference on the effects of a decline in health human capital on household income mobility in the intra-generational context. By evaluating the consequence of a health shock on income mobility, we quantify how a health shock affects a household's economic opportunities over time in the same generation. Second, apart from analyzing patterns of income mobility at the aggregate level, we try to provide micro-founded evidence that rationalizes the estimated aggregate-level income mobility. On the one hand, at the aggregate level, we estimate the wedge in transition probability and upward mobility between the treatment and control groups and show that households that experienced a health shock exhibit much lower income mobility. On the other hand, we take an event study approach and estimate the economic consequences of the worsened health status of one household member on the household income and income percentile. By estimating the average treatment effect of a health shock on household income and income percentile, we quantify the welfare loss and how a household moves in the national income distribution when receiving a health shock. Third, a novel aspect of our paper is that we do not limit our analysis to the contemporaneous effects of a health shock. Instead, our paper adds to a relatively smaller strand of literature by looking at the lasting impact of a health shock on households intra-generational income mobility, by tracking the medium-run income mobility of the treated households after the health shock.

The rest of the paper is organized as follows. Section 2 introduces our aggregate-level income mobility measures. Section 3 describes the data and sample construction. Section 4 presents our empirical framework, in which we introduce our identification strategy and econometric model. Section 5 presents our main results on the effects of a health shock on household income mobility and explores households' labor supply responses when faced with a health shock. In section 6, we provide evidence suggesting that a health shock imposes a long-lasting impact on household income mobility. Section 7 concludes. Additional details on the data and our methodology, as well as extensive sensitivity analysis, are presented in the Online Appendix.

2 Aggregate-Level Income Mobility Measures

In this section, we introduce two rank-based income mobility measures that quantify a household's income mobility in the national income distribution based on the changes in their income ranks, which we call aggregate-level income mobility.

2.1 Transition Probability

Following Bhattacharya and Mazumder (2011), we propose an intra-generational version of transition probability, which measures the probability that a household is above the sth quantile at time t_1, conditioning on being between quantiles s_1 and s_2 at time t_0. Let $F_0(\cdot)$ and $F_1(\cdot)$ donate the c.d.f. of the overall income distribution at times t_0 and t_1, respectively. Then the probability that the household is above the sth quantile of F_1, conditional on it being between quantiles s_1 and s_2 of $F_0(\cdot)$ is denoted by:

$$\theta(s, (s_1, s_2)) = \Pr[F_1(Y_1) > s, s_1 \leq F_0(Y_0) \leq s_2 | s_1 \leq F_0(Y_0) \leq s_2] \tag{1}$$

For an easier interpretation, we consider the situation where $s_1 = 0$ and $s_2 = s$. Thus, the transition probability simplifies to:

$$\theta(s) = \Pr[F_1(Y_1) > s, F_0(Y_0) \leq s | F_0(Y_0) \leq s]$$

Using this measure, we directly measure the probability of a household ending up at a position higher than the sth quantile, conditioning on starting at a position lower than or equal to the sth quantile. Specifically, when $s = 20$, it measures the household's probability of "getting out of the low-income trap."

2.2 Upward Mobility

Following Bhattacharya and Mazumder (2011) and Chetty et al. (2017), we propose an intra-generational version of upward mobility based on the household's position in the national income distribution. This upward mobility measure quantifies the probability that a household's income position at time t_1 exceeds its income position at time t_0 by a fixed amount. Let $F_0(\cdot)$ and $F_1(\cdot)$ donate the c.d.f. of the overall income distribution at times t_0 and t_1, respectively. Then the upward mobility for a household that is located above the s_1th and lower than or equal to the s_2th quantile of $F_0(\cdot)$ is denoted by:

$$v(\tau, s_1, s_2) = \Pr[F_1(Y_1) - F_0(Y_0) > \tau | s_1 < F_0(Y_0) \leq s_2] \tag{2}$$

where τ governs the distance of the movement in the income distribution.

In the main text of this paper, we consider the situation where $\tau = 0$. Thus, upward mobility is simplified as:

$$v(s_1, s_2) = \Pr[F_1(Y_1) - F_0(Y_0) > 0 | s_1 < F_0(Y_0) \leq s_2]$$

Using this measure, we directly measure the probability of a household ending up at a position higher than its previous position, conditional on the starting position higher being higher than the s_1th, and lower than or equal to the s_2th quantile. In other words, this measure quantifies a household's probability of "climbing the income ladder."

2.3 Contrasting Transition Probability with Upward Mobility

The upward mobility measure is distinct from the transition probability, in that it includes a household's short-distance upward movements in the income distribution from its original position, which is ignored by the transition probability measure (Bhattacharya and Mazumder 2011). The upward mobility measure takes into account those who have made upward movements but do not exceed a given threshold s. Thus, although the transition probability highlights the probability of reaching a certain threshold, the upward mobility measure emphasizes the household's movement relative to its original position.

These two measures complement each other when we compare income mobility between households with and without a health shock. On the one hand, for understanding differences in the absolute economic opportunities, defined as success rate of achieving specific economic goals, the transition probability measure works better. On the other hand, households with or without a health shock may be at different income positions.[3] Thus, one could expect that households with a health shock need to achieve higher gains in income to reach a specific income position. The upward mobility measure better fits the goal of measuring the relative movement of a household from its original income position.

3 Data and Sample Construction

Our primary data source is nine waves of the CHNS from 1991 to 2015, and four waves of the CFPS from 2010 to 2016.[4] The CHNS is jointly conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention. It uses a multistage, random cluster process to draw a sample of about 4,400 households in nine

[3]In our main analysis, households with or without a health shock have a similar income rank before the health shock. In our medium-run analysis in Section 6, a household that experiences a health shock is relatively lower in income position after a health shock. We will discuss it in more detail in the following sections.

[4]The nine waves of CHNS were conducted in 1991, 1993, 1997, 2000, 2004, 2006, 2009, 2011, and 2015. The four waves of the CFPS were conducted in 2010, 2012, 2014, and 2016.

provinces across China.[5] The average gap between each CHNS survey is three years. The CFPS is a nationally representative, bi-annual longitudinal survey of Chinese communities, families, and individuals. The CFPS has been carried out since 2010 by the Institute of Social Science Survey of Peking University, China. The target sample includes 16,000 households and all members over age 9 in each sampled household, for a total of more than 30,000 individuals. The CHNS data speaks more about the facts in 1990s and 2000s as most of the surveys were conducted before 2010. On the contrary, the CFPS started in 2010 and speaks about the facts in the post-2010 period.

3.1 Definitions of a Health Shock

In this paper, we use a health shock to proxy for a severe deterioration in health human capital. We define a health shock to a household as a health shock to any household member. We now describe our definitions of a health shock at the individual level, using both severely deteriorated self-reported health status and hospital admission.

3.1.1 Severe Deterioration in Self-Reported Health Status

We define an individual-level health shock as a sharp decline in one's health status from relatively good to poor based on self-reported health status. In the CHNS, the survey asks questions "How do you rate your health status?" If the respondent reports "excellent," "good," or "fair" health in wave t, and "poor" or "very poor" in wave $t+1$, we record a health shock between waves t and $t+1$[6].

We do not just use worsened self-reported health status to proxy for a health shock because of potential endogeneity problems. The first possible endogeneity problem is reverse causality. Using changes in self-reported health status may bias the estimates of the effects of a health shock on household income or income mobility, as individuals who experience an income shock for other reasons may attribute it to worsened health status[7]. The second possible endogeneity problem comes from measurement errors. On the one hand, the respondents are required to choose from four or five options. Nevertheless, it might be difficult for them to tell the differences between different health statuses in practice. For example, what distinguishes "good" from "excellent" or "poor" from "very poor"? On the other hand, a fundamental assumption using self-reported health status changes to proxy for a health

[5]Three mega cities joined the survey since 2011, and three more provinces joined since 2015. However, we do not include them in the sample. Please refer to section 3.2 for more details.

[6]Please refer to Appendix A.1 for more details on the definition of self-reported health status variable.

[7]For example, people may justify their decisions of leaving their jobs for non-health-related reasons to worsened health status (see, for example, Bound 1991).

shock is that an individual can assess her self-reported health using a consistent standard during a long period. However, this assumption is overly strong because it is hard to believe that people's standards of evaluating health status remain constant over a long time. To address these two concerns, we record a health shock if and only if a household member's health changes from "fair" or better to "poor" or worse. There are three advantages to doing so. First, an individual may report worse health when her income lowers, but it is less likely for her to report "poor" health if she is not really in that status. Second, an individual may not be able to distinguish among the five options, but likely to be able to tell the difference between "good" and "poor," so there will be less measurement error under this definition of a health shock. Finally, though people may not evaluate their health status consistently over time, it is more credible to believe that their definitions of "good" and "poor" health status are more stable. Hence, the endogeneity problem of self-reported health can be alleviated using our definition.

However, there might be one additional shortcoming of using a sharp decline in self-reported health status to define a health shock. Self-reported health can be regarded as an assessment of an individual's current health status: she may assess her health status based on her recent health conditions. If an individual's health status changed from "good" to "poor," but the individual subsequently recovered before the survey was conducted, we may observe no health shock in the survey, although one actually happened.

Considering the possible disadvantages of using self-reported health status to define health shocks, we use an event study design to estimate the effect of a health shock, which allows us to test the plausibility of parallel trends in the outcomes before a shock. Further, we propose another definition of a health shock using hospital admission records. Doing so allows us to compare the estimation results using different definitions of a health shock and test for the plausibility of using severe deterioration in self-reported health status to identify a health shock. We obtain robust results using the two definitions of a health shock.

3.1.2 Hospital Admission

We also use hospital admission records to identify health shocks, following Wagstaff (2007) and Dobkin et al. (2018). Of the two data sets, the CHNS only documents hospital admission records in the past four weeks before the survey. Hence, only a small proportion of the respondents report that they have been hospitalized. That makes it impossible to estimate the effects of a health shock, since there could be many individuals who had a hospital admission earlier than four weeks before the survey but do not report an admission. By contrast, the CFPS asks "whether you had an in-hospital treatment during the past year," which allows us to track health shocks in a year-long window. Thus, we use hospital admissions from the CFPS to detect severe health shocks and estimate their economic consequences. It would be helpful to look at the reasons for hospitalization to better understand the nature of a health shock defined by a hospitalization. However, the CFPS does not disclose information on the reasons for hospitalization. In section A.3, we provide detailed information on the total number of days in the hospital to validate the severity of the health shocks. We find that most individuals (89.53%) who have at least one hospitalization record spend at least four days in the hospital. This suggests that most hospitalization records are not simply overnight emergency visits but are caused by underlying health conditions and need adequate in-hospital treatment.

3.2 Sample Construction

CHNS Sample Construction To construct our sample from the CHNS, we first construct several subsamples that consist of households that appear in three consecutive surveys.[8] Given that we have nine waves of CHNS data, we can construct seven such subsamples. In each subsample, we select households that reported positive income in all three waves of the surveys. In the second step, any households that experienced a health shock in the first two waves are dropped, and households in which more than one household member experienced a health shock in the third wave are also dropped. In the third step, we do propensity score matching within each subsample. Note that there is one treated individual in each treated household. For each treated individual, we match three "placebo-treated individuals" who appear similar to the treated individuals but did not receive a health shock.[9] Specifically, we use a propensity score matching method with ten observable characteristics in the first two waves to estimate the propensity score.[10] Among these characteristics, eight are at the individual level, including age, gender, years of education, whether having health insurance, whether currently employed, yearly working hours, hourly wage, and self-reported health status; two are household-level characteristics, including urban/rural status, and household size. Because of the individual-level matching, each treated household is matched with three "placebo-treated households" that did not experience a health shock, which are henceforth the control households in our subsample. These control households have a household member

[8]Three consecutive periods observed for each individual is the minimum length required for our econometric specification. We do not use four or more consecutive waves due to sample attrition.

[9]These "placebo-treated individuals" are chosen from households that did not experience a health shock.

[10]We match each treated individual with three nearest "placebo-treated individuals" using the linearized propensity score, following Imbens and Rubin (2015). We regress whether an individual receives a health shock on ten observable characteristics in the first two waves to estimate the propensity score using a logit regression. Given observables X_i, we denote the conditional probability of receiving a health shock as x_i. We compute the linearized propensity score as $l(x_i) = \ln[\frac{e(x_i)}{1-e(x_i)}]$ and match based on this variable.

who appear similar to the individual who experiences a health shock in the treated households, and they have similar household characteristics as well. The final step of constructing the CHNS sample is to pool all the subsamples together after propensity score matching.

CFPS Sample Construction Similarly, we construct the CFPS samples, where a hospitalization record identifies a health shock. The only difference in the sample construction process for the CFPS data is that we additionally include the household sampling weights when doing the matching. We also explored constructing a CFPS sample that uses the sharp decline in the self-reported health status to identify health shocks. However, CFPS frequently changes the measurement of self-reported health status in early surveys. Thus, the respondents can not consistently assess their health status, which raises severe measurement errors. Appendix A.17 provides a more detailed discussion.

Timeline In Figure A2, we plot the timeline of our analysis. In each subsample, we label the survey wave before the health shock as wave $t = -1$ and label the survey wave after the health shock as wave $t = 0$. By our definition, a health shock happens somewhere between wave $t = -1$ and wave $t = 0$. The expected gap between when a health shock happens and wave $t = 0$ is one and a half years in the CHNS sample and half a year in the CFPS sample.

3.3 Variable Definitions

CHNS Household Income The household income variable is the net household income constructed by the CHNS. Conceptually, household income is the sum of all sources of income and revenue minus expenditures. Nine income sources - business income, farming income, fishing income, gardening income, livestock income, non-retirement wages, retirement income, subsidies, and other income - are included. The CHNS inflated household income to 2011 RMB.

CFPS Household Income The household income variable is the net household income constructed by CFPS, and we inflate it to 2011 RMB using the National Consumer Price Index.[11] Household income is the sum of five sources: wage income, business income, transfer income, property income, and other income.[12]

[11]The data can be found at http://data.stats.gov.cn/easyquery.htm?cn=C01.

[12]How CFPS asks about household income is slightly different across the four waves, and the composition of household income is slightly different across different waves. To make the household income variable comparable across all four waves, we make use of the CFPS-provided variable "household income comparable to 2010" in CFPS, which guarantees that the sources of household income stay the same.

Employment, Yearly Working Hours, and Hourly Wage In our paper, we define an individual as "employed" if she reports she is currently working. According to the definition, farmers who work in the agricultural sector or self-employment are all regarded as "employed". The yearly working hours is defined as the total working hours of all jobs. The hourly wage is computed by dividing individual wage income over yearly working hours. Please see mode detailed variable definitions of employment, yearly working hours, hourly wage, and household medical expenditure in Appendix A.2.

CFPS Household Consumption In the CFPS data, household total expenditure is the sum of spending in ten categories: food, clothing, medical, transportation, durable and non-durable goods (excluding food and clothing), entertainment and education, mortgage, taxation, donation, and others. In our analysis, we construct a household consumption variable that excludes household taxation, donation, and mortgage expenditure. We also construct a household non-medical consumption variable that further excludes household out-of-pocket medical expenditure.[13]

3.4 Summary Statistics

The total numbers of treatment households included in our CHNS and CFPS samples are 517 and 1,436, respectively. Table 1 reports the summary statistics for the two samples for treatment and control households separately. All of the covariates of the two groups are perfectly balanced after matching. Moreover, two variables that we do not implicitly match on – the household income per capita and the household income percentile – are perfectly balanced between the treatment and control group before the health shock. To further assess the quality of matching, we present additional summary statistics for household medical expenditure and non-medical consumption, as well as seven covariates describing housing characteristics that we do not match on in Table A4. We see that all but one covariates are perfectly balanced between the treatment and control groups. Table A3 reports the sample summary statistics before matching. Before matching, we see that many covariates are imbalanced between the treated and control households, highlighting the importance of matching.

[13]We do not do analysis using the CHNS data because we do not see consumption by category in the survey.

4 Empirical Framework

The impact of a health shock on household income mobility can be evaluated by comparing the outcomes of households that experienced a shock with those that did not. We use an event-study design similar to that used in Jaravel et al. (2018) and estimate the following equation:

$$y_{it} = \gamma_t + \alpha_i + \sum_{\tau=-2}^{\tau=0} \delta_\tau^{\text{All}} \mathbb{1}[D_{it}^{All} = \tau] + \sum_{\tau=-2}^{\tau=0} \delta_\tau^{\text{Treat}} \mathbb{1}[D_{it}^{Treat} = \tau] + \varepsilon_{it}, \tag{3}$$

where y_{it} denotes the outcome of household i in year t (the outcomes of interest include household income per capita and household income percentile); γ_t denotes calendar year fixed effects; and α_i denotes household fixed effects. D_{it}^{Treat} is the interaction term of the treatment dummy (whether a household member experienced a health shock) and a dummy for the wave relative to the health shock for household i in year t, where τ represents the relative wave (event-time). D_{it}^{All} is a dummy for the wave relative to a health shock for household i in year t. ε_{it} stands for the error term. We normalize the coefficient of the wave before the health shock, δ_{-1}^{Treat}, to zero. That is, when our identification assumption described below holds, the coefficient δ_0^{Treat} is the coefficient of interest, as it measures the causal effects of a health shock on the outcomes of interest.

First, our model permits the treated households to differ from the control households in levels of the outcome, by adding household fixed effects α_i. Second, following Jaravel et al. (2018), we add a set of leads and lags around the health shock, which are common to the control and treated households (δ_τ^{All}). By adding these leads and lags, we directly address the concern that the household fixed effects, as well as the controls, may not fully account for the trends in household income around the health shock. For example, previous literature finds a slight decrease in household income before a health shock (e.g., Dobkin et al. 2018). Moreover, the relative time effects for the treatment group δ_τ^{Treat} in the model have some important implications. In our model, the coefficients δ_τ^{Treat} in front of D_τ^{Treat} measure the average difference between the treatment and control households in wave τ relative to the health shock. The specification allows us to test whether there is any difference in trends of the outcome between the treatment and control households before the health shock, by examining whether the coefficient δ_{-2}^{Treat} differs from zero. This test could help check the validity of our identification assumption. Formally, similar to the spirit of Jaravel et al. (2018), our econometric model assumes the following conditional expectation function:

$$\mathbb{E}[y_{it}|D_\tau^{All}, D_\tau^{Treat}, t, i] = \gamma_t + \alpha_i + \delta_\tau^{All} + \delta_\tau^{Treat},$$

To interpret the coefficient δ_0^{Treat} as the causal effect of a health shock on the outcome, we need the following identification assumption:

$$\mathbb{E}[\mathbb{1}(D_{it}^{All} = \tau)\varepsilon_{it}|D_{it}^{All}, D_{it}^{Treat}, t, i] = 0 \quad \forall(t, \tau),$$

which says that the time relative to the health shock is not correlated with the error term for all households. In practice, our identification strategy may not be credible due to potential endogeneity concerns, in which case the time relative to the health shock is not strictly exogenous. One may hypothesize that it is not the health shock that results in the change in the outcome variable; instead, it is the changes in the outcome variable that cause the health shock. Take one outcome variable, for example, household income per capita. If we observe that there was a significant decline in income per capita before the health shock (that is, if we observe a significant positive δ_{-2}^{Treat}), we may be concerned that households reported poorer health due to decreasing income, and hence we observe a health shock.[14]

5 Empirical Results

In this section, we present the results on the effects of a health shock on household income mobility and consider possible explanations for why health shocks matter. We begin with the analysis of the effects of health shock on aggregate-level income mobility. Then, we provide household-level evidence to rationalize the observed effects of a health shock on aggregate-level mobility. In our regression analysis, we first show the effects of a health shock on income per capita, followed by the effects on household income percentile. Next, to examine why a health shock affects a household's income mobility, we evaluate the effects of a health shock on household labor supply decisions. Finally, we explore whether a health shock has different economic implications for different subgroups, by doing heterogeneity analysis along several dimensions.

5.1 Effects on Transition Probability and Upward Mobility

We begin by analyzing how a health shock affects a household's transition probability and upward mobility. We estimate the transition probability and upward mobility between period -1 (one wave before the health shock) and period 0 (the wave after the health shock) in

[14] Although the zero coefficient of δ_{-2}^{Treat} cannot fully alleviate this concern since the pre-trend may not be identified due to the statistical power (e.g., Freyaldenhoven et al. 2019), it is the best practice we can take due to the lack of IVs to instrument for the confounds. In Freyaldenhoven et al. (2019), the authors propose a method to identify the causal effect of a treatment in the event-study setting by using a 2SLS approach.

our two samples for the treatment and control households, respectively. Our results suggest that a health shock lowers household transition probability and upward mobility significantly.

Table 2 reports the transition probability for the treatment and control households in our two samples. There is no evidence of differential trends in transition probability between the treatment groups and the control group prior to the health shock.[15] After the health shock, there is clear evidence that households who experienced a health shock have much lower transition probability than those that did not, at all cutoff values s. These results suggest that a health shock plays a vital role in determining households' transition probability, regardless of the household's initial position in the income distribution.

In Table 2, columns (1)-(3) report the transition probability estimates for the CHNS sample, where a health shock is defined as a severe deterioration in self-reported health status. At all cutoff values, treated households consistently show a lower transition probability, compared with the control households. Moreover, as a central focus of policy interest, we focus on whether a health shock imposes a significant impact on the probability of "getting out of the low-income trap." Given that a household is in the bottom 20 percentiles of the income distribution before the health shock ($t = -1$), the chance of transiting out of the the bottom quintile after the health shock ($t = 0$) is 64.3% if the household does not suffer a health shock. However, the chance decreases by 8.4 percentage points to 55.9% if the household experienced a health shock. Considering the 64.3% transition probability for the control group, this 8.4 percentage points decrease reflects a 13.06% lower transition probability for households that experienced a health shock in the bottom 20 percentiles of the national income distribution.

Estimates using the CFPS sample tell a similar story with slight differences. In Table 2, columns (4)-(6), we report the transition probability for the treatment and control groups using the CFPS sample, where a health shock is defined as a hospitalization. Again, households with a health shock exhibit lower transition probability at all cutoff values. However, we only find small and statistically insignificant different in the chance of "getting out of the low-income trap" between two groups. In the control group, 56.2% control households and 54.5% treatment households that are initially in the bottom quintile move out of the bottom quintile in the next period. The difference in the probability of "getting out of the low-income trap" is 1.7 percentage points, which is much smaller than what we find in the CHNS sample (8.4 percentage points). A possible explanation for this is the expanded health insurance, unemployment insurance and Minimum Living Standard Scheme (MLSS) cover-

[15] In Figure A3 and A4, we present transition probability and upward mobility for the treatment and control households before the health shock. We do not observe any significant differences in transition probability and upward mobility.

age in recent years.[16] All these insurances may help low-income households better insure against health shocks and maintain income mobility. Overall, Table 2 shows that households with a health shock exhibit lower probability of successfully transiting from below or equal to the sth quantile to above the sth quantile, regardless of the cutoff value, s.[17]

In the upward mobility analysis, we divide households into five groups based on their income position in the period before the health shock, each of which represents 20 percentiles in the national income distribution. Table 3 reports the upward probability for households in each of these groups separately. Overall, we observe that experiencing a health shock lowers a household's ability to climb the income ladder, namely, the ability to surpass its previous income percentile.

Using the CHNS sample in columns (1)-(3), households in the treatment group have a 9.5 percentage points lower probability of climbing the income ladder compared with the control group. Compared with the mean of the control group, this estimate suggests an 18.73% lower upward mobility for households that experienced a health shock. The upward mobility estimates using the CFPS sample are reported in columns (4)-(6). Households in the treatment group have 4.3 percentage points lower upward mobility on average. These estimates reflect an 8.81% lower upward mobility for the treatment group compared with the control group.

The findings also show that a health shock imposes heterogeneous effects on the upward mobility of households at different income positions. In the CHNS sample, which speaks more to the trends in the 1990s and 2000s, upward mobility drops the most for households in the bottom and top two quintiles, while households that were initially between percentiles 20 and 60 are not significantly affected. In the CFPS sample, which emphasizes trends in the 2010s, health shocks impose relatively larger impacts on households that were initially between percentiles 20 and 60, while we do not see economically large difference in upward mobility for households in other income groups. In line with our estimates of the transition probability, these results suggest that the poor households are better insured against health shocks in recent years. Moreover, the top income-earners have been more likely to maintain

[16] A large proportion of households in the CHNS sample are taken from the surveys in early years (before 2006), while the CFPS began since 2010. Thus, the estimates using CHNS sample speaks more about the facts in earlier years, while the estimates using CFPS sample speak about the recent trends. China started expanding its urban MLSS in 2002, and its rural MLSS in 2006. The MLSS is a social assistance program through which government transfer money to poor or long-term unemployed households with an annual income below certain threshold. Also, China began expanding its health insurance coverage for rural residents in 2002 and reached 90% coverage in 2008. In 2007, China started providing basic health insurance coverage for elders, children, and non-working urban residents in 2007.

[17] Figure A5 plots the transition probability between period -1 (one wave before the health shock) and period 0 (the wave after the health shock) for the control and treatment groups separately, in intervals of 5 percentile points varies from 5 to 95.

their ability to move upward, while it has been less likely for the middle-income households to climb the income ladder in the face of a health shock.

5.2 Effects on Household Income

In section 5.1, we see that at the aggregate-level, a health shock lowers household income mobility. But how should we understand these results? More specifically, in the following two subsections, we answer the questions "How large are the changes in household income rank that corresponds to the decreased income mobility that we observe?", and "How large are the welfare loss that correspond to the decreased income mobility".

Table 4 presents the effects of a health shock on log household income per capita. Our estimates suggest that a health shock leads to a substantial amount of income loss, which directly measures household welfare. Using the CHNS sample in column (1), we find that a health shock results in a 13.1% decrease in household income per capita for households that experienced a health shock, relative to the control group. This estimate reflects a 1,301 RMB decrease in household income per capita for the treatment group relative to the control-group mean of 9,939 RMB[18]. Using the CFPS sample in columns (2), we document that a health shock results in 14.2% lower income per capita for the treated households, relative to the control households. Compared with the post-shock control group mean of 12,447, this 14.2% relative change reflects a 1,767 RMB decrease in household income per capita.[19]

Comparing our estimates with the 8.2% and 6.5% annual household income increase in China for urban and rural households, respectively, during 1991 and 2012, it is clear that a household experiencing a health shock suffers a substantial welfare loss and had a much larger likelihood of falling into poverty.[20] Our estimates suggest that a health shock introduces contemporaneous income loss which is equivalent to almost two years' of household

[18]1 U.S. dollar $\approx$ 5.5 RMB from 1991 to 1993, 8.3 RMB from 1998 to 2004, and gradually depreciated to 6.2 RMB at the end of 2015.

[19]We notice that a recent paper using the CHNS data gets different results as our paper. In Liu (2016), the author finds that a household can perfectly insure their income against a health shock. There are several reasons why we get different results from that paper's. Most importantly, we define a health shock in different ways. Specifically, in that paper, the author relies on the CHNS question "whether the individual, during the past four weeks, had been sick, injured, or suffering from a chronic or acute disease" to identify an individual-level health shock. Then, the author defines a household-level health shock based on changes in the health status of the household head and her spouse. In this way, any households experiencing a health shock beyond the four-week window are identified as "control households" that do not experience any health shocks. However, in our setting, we define a health shock as either a severe decline in the self-reported health status (in CHNS data), or a hospitalization record (in CFPS data). We can identify health shocks in a relatively long time, making it much less likely that we classify a household that experiences a health shock as a control household.

[20]Data on urban and rural household income per capita are from the National Bureau of Statistics of China at http://data.stats.gov.cn/easyquery.htm?cn=C01.

income growth. Further, the estimated effects of a health shock to household income per capita are similar using different samples, speaking to the fact that health always plays an essential role in determining household income throughout our sample period. Finally, the small and statistically insignificant pre-trends under different definitions of a health shock and using different samples mitigate the reverse-causality concern and lend credence to our identification strategy.

5.3 Effects on Household Income Percentile

We next analyze the effects of a health shock on the household's income percentile. By evaluating the effects of a health shock on income rank, we are answering the following question: "What are the consequences of a change in health status on a household's relative position in the income distribution?"

There are three advantages of using household income percentile as the outcome variable in our analysis. First, estimating the effects of a health shock on a household's income percentile provides micro-level evidence on how a health shock affects a household's movement in the income distribution, which explains our estimates of the effects of a health shock on aggregate-level income mobility in section 5.1. Recall that upward mobility quantifies the proportion of households that successfully move up a given distance in the income distribution, and the transition probability measures the proportion of households who successfully transit to a specific income position. However, these rank-based aggregate-level income mobility measures do not speak to the distance that a household moves when it experiences a health shock. By calculating the average treatment effect of a health shock on household income position, we directly provide micro-level estimates explaining the aggregate-level patterns we observe.

Second, the national income distribution changes rapidly in China during the sample period due to the rapid economic growth and fast-rising income inequality. From 1990 to 2016, China's income per capita grew more than twenty times, and the income inequality, measured in GINI coefficient, grew from around 0.35 to 0.55 (Xie and Zhou 2014). Thus, on the one hand, the same percentage points income loss corresponds to a larger drop in income position in the 1990s, relative to in the 2010s, as households income are more spread in the latter period. On the other hand, the same percentage points income loss introduces larger welfare loss in the early years than in recent years, as more households earn low income in the early years. By using the income position as the outcome variable, we provide an alternative way of assessing welfare cost of a decreased income mobility.

Finally, there has been a strand of literature suggesting that relative income, instead of absolute income, affects individuals' subjective well-being. For example, Clark et al. (2009) suggest that individuals are rank-sensitive: they are more satisfied as their percentile ranking improves; Card et al. (2012) use an experiment to demonstrate that job satisfaction depends on relative pay comparisons. Thus, changes in the position in the income distribution directly affect a household's subjective well-being, and analyzing the effects of a health shock on household income position is of great normative importance.

Table 5 presents the results on the effects of a health shock on household's income percentile. We find that a health shock significantly lowers the household income position in the national income distribution. Using the CHNS sample in column (1), where a health shock is defined as a severe deterioration in self-reported health status, we find that a health shock causes a 3.94 percentile drop in the income distribution. We observe similar patterns using the CFPS sample in Column (2). The estimates suggest that a health shock lowers income position by 3.14 percentiles.

5.4 Effects on Household Medical Expenditure

In section A.10, we study the impact of a health shock on household out-of-pocket medical expenditure. In table A5, we analyze the impact of a health shock on annual out-of-pocket medical expenditure per capita. Using our CFPS sample, we see that a health shock significantly increases household medical expenditure.[21] Our estimates suggest that a health shock results in a 94.6% increase in annual out-of-pocket medical expenditure per capita for the treated households, relative to the control households. This increase in medical expenditure, along with the decreased household income, account for 25.75% of the pre-treatment household income per capita for the treated households (2,660 RMB, or around 400 USD).

5.5 Effects on Household Non-medical Consumption

In section A.11, we study household non-medical consumption responses after a health shock. In table A6, we analyze the impact of a health shock on household non-medical consumption per capita. Using our CFPS sample, we see that households can perfectly smooth consumption when receiving a health shock. That is, a household's non-medical expenditure is not lowered by the income shock after a health shock. Our finding is in line with previous empirical evidence that households can smooth their consumption when facing a non-fatal health shock, in the developing country context.[22] Two possible channels through

[21]The CHNS sample does not provide information on household medical expenditure, so we cannot use the CHNS sample for analysis.

[22]See, for example, Wagstaff (2007) on Vietnam, Islam and Maitra (2012) on Bangladesh, and Liu (2016) on China. Some other papers suggest that a health shock lowers household consumption. For example,

which a household insure against the health shock are that, first, they can borrow from their relatives, and second, they may use their savings to smooth consumption.

5.6 Household Labor Supply Responses

From the analysis so far, we conclude that a health shock causes a substantial decrease in household income and significantly lowers households' income percentile. But why does a health shock introduces such effects? In this section, we discuss households' labor supply responses to a health shock.

There are three reasons why a health shock could affect household income through the labor supply channel. First, a health shock may directly change the labor supply decision of the individuals who experienced a health shock, on both the extensive and intensive margins. On the extensive margin, individuals who experienced a health shock may become unemployed in the labor market, or exit the labor market. On the intensive margin, individuals who experienced a severe health shock may reduce their hours of labor supply or exit from full-time work. Under both scenarios, household income would be lower due to the unearned income of those individuals who receive a health shock. Second, for the individuals who experienced it, a health shock lowers their labor productivity, especially for those who work in manual labor jobs. In a competitive labor market, the wage rate equals the marginal productivity of labor. Thus, a health shock would lower the treated individual's hourly wage, even holding her labor supply constant. Third, a health shock on one household member could impact the labor supply decisions of other household members, although the response is theoretically ambiguous. On the one hand, past research has emphasized that self-insurance is an essential channel through which households insure themselves against income shocks. When an income shock happens to one household member, the labor supply of other household members will increase to compensate for the income loss (Ashenfelter 1980, Cullen and Gruber 2000). On the other hand, there are several reasons why a negative health shock could lower the labor supply hours or hourly wage of other household members. First, theoretically, all household members jointly provide the public goods to be consumed within a family (Blundell et al. 2005, Cherchye et al. 2012).[23] Thus, when a household member receives a health shock, the other household members need to spend relatively more time producing public goods for the household. Second, it can be readily seen that a health shock

Gertler and Gruber (2002) suggest that a health shock lowers household non-medical consumption by 19.5%. However, they measure a health shock using changes in activities of daily living scores (ADL). A drastic change in the ADL score is often a symbol of disability (nearly 80% individuals in their sample drop out of the labor market when receiving such health shock). Thus, the measured health shocks in their paper are much severe and are different from those in our paper.

[23]Examples of the provision of public goods include caring for children and doing housework.

may increase the marginal utility of leisure (Coile 2004). For example, treated individuals may require assistance with activities. In this case, if other household members care for the ill individual, the time spent on caring for these treated individuals may create higher marginal utility for leisure.

Considering these channels, we analyze the treated individuals', as well as their household members' labor supply responses to a health shock. To do the analysis, we construct two sets of individual-level samples. The first set consists of individuals who experienced a health shock, and all individuals in the control group that is, all individuals in households that have no member experiencing any health shocks; the second sample consists of household members of the treated individuals and all individuals in the control group. We estimate Equation 3 and switch the outcomes to the labor supply measures of interest (including whether employed, yearly working hours, and hourly wage).

Extensive Margin Responses: In Figure 1A, we examine whether a health shock leads to any significant impact on the labor supply at the extensive margin. In the CHNS sample, we find that those who experienced a health shock are 2.2% more likely to be unemployed, compared with the individuals in the control group, though the point estimate is not statistically significant at 10% significance level. The household members of the treated individuals do not adjustment their labor supply at the extensive margin, as the point estimate is only -0.001 and statistically insignificant. In the CFPS sample, a health shock significantly reduces the treated individual's probability of being employed: receiving a health shock lower the probability of being employed by 9.2%. At the meantime, we do not see any statistically significant or economically large impact on other household members' probability of being employed.

We see a stronger response in labor supply at the extensive margin for the treated individuals in more recent years in the CFPS sample. We think the following two reasons play an important role. First, we think this could be related to the urbanization progress in China in recent years. The share of urban households increased from 21% in our CHNS sample to 50% in our CFPS sample. Thus, a much larger proportion of individuals are working in the urban sector instead of in the agricultural sector. By our definition, people who work in the agricultural sector are thought to be "employed" if they work any positive hours during the past year.[24] Thus, it is reasonable that the labor supply responses at the extensive margin should be much weaker in years where more people live in the rural areas and work in the agricultural sector. In addition, the stronger response in more recent years

[24]In China, farmers work on their own land, and there is no formal employment in this case. We treat people who work any positive hours in their land as being "employed" following the literature.

could be attributed to the market liberalization. In the early years, the urban labor market was highly regulated and dominated by state-owned enterprises. In recent years, market liberalization contributed to a higher job replacement rate, and we can see more responsive labor supply adjustments at the extensive margin for those who experienced a health shock.

Intensive Margin Responses: In Figure 1B, we plot the treatment effects on yearly working hours for those who are employed both before and after the health shock. In both samples, we find some weak evidence that a health shock lowers the treated individuals working hours. The point estimates suggest that, conditional on being employed, individuals who experience a health shock reduce their yearly working hours by 75 and 116 hours in the CHNS and the CFPS sample, respectively. Though not statistically significant at 10% significance level, these point estimates correspond to an economically large 4.5% of the control group's average yearly working hours. Meanwhile, in the CHNS sample, we do not find any statistically significant or economically large evidence suggesting that the treated individuals' household members respond to a health shock by adjusting their labor supply hours. In the CFPS sample, however, we see statistically insignificant but economically large effects of a health shock on household members' yearly working hours. After a health shock, the household members increase their yearly working hours which by about 4% of the control group's average yearly working hours. The increased hours of labor supply of household members can almost offset the reduced working hours of the treated individuals if they all continue working after the health shock.

Comparing the estimates of the household members' labor supply responses in the two samples, we find weak evidence suggesting a stronger labor supply responses in recent years at the intensive margin. We think there are two reasons related to the urban premium that rationalize this change through increased share of urban households. On the one hand, the labor hours required in producing crops are relatively fixed (so there's limited room for an increase in working hours), whereas people are more likely to find ways to increase their working hours in the urban sector. On the other hand, residents in urban areas have better access to find people caring for ill individuals, which allows them to work more. In summary, these two different urban premiums, along with the increased urban share in recent years, might help explain the pattern that we observe.

Besides, these labor supply responses of household members at the intensive margin in recent years suggest that there is still inadequate insurance in China. In Fadlon and Nielsen (2015) and Dobkin et al. (2018), the authors document that in the U.S. and Nordic countries, where there is better health and unemployment insurance, the spousal labor supply does not respond to non-fatal health shocks.

Effects on Hourly Wage: In Figure 1C, we report the impact of a health shock on labor productivity of the individuals who experienced a health shock. In the CHNS sample, a health shock lowers the hourly wage of the treated individuals by 15.3%, compared with the individuals in the control households. In the CFPS sample, the hourly wage of the treated individuals decreases by 6.9%, compared with individuals in the control group once a health shock occurs. Meanwhile, in both the CHNS and CFPS samples, we see a decrease in hourly wage among household members of the treated individuals. The hourly wage of the treated individuals' household members decrease by 7.6% and 8% in the CHNS and CFPS sample, respectively. Some of these point estimates are not statistically significant but are economically large.

Households in China earn most of their income through the wage income.[25] A simply back-of-envelope calculation suggests that the labor supply responses at the extensive margin and decreased labor productivity together explain the majority of a household's income loss due to a health shock. If we assume that 90% of a household's income comes from the wage income, the treated individual and her household members earn equal wage income and are fully employed before the health shock. Then we see that the drop in labor productivity explains 77.5% and 47.2% of the loss in income in the CHNS and the CFPS sample, respectively.[26] The labor supply responses at the extensive margin explain an additional 7.4% and 29.2% of the income loss in the two samples, respectively. These two forces jointly explain 84.9% and 76.4% of the income loss in the two samples, respectively.

5.7 Treatment Effect Heterogeneity and Implications

The estimated average treatment effects mask economically interesting heterogeneity. To analyze whether a health shock has different economic implications for different subgroups,

[25]We refer to the farm income and wage income for farmers as their "wage income". For farmers, we compute their hourly wage as their individual income divided by yearly working hours. The individual income for farmers is generally composed of their farm income, and their wage income if they have a job.

[26]In the CHNS sample, the household income per capita drops by 13.3% after a health shock. Then the decreased labor productivity explains $(15.3\% + 7.6\%)/2 \times 90\%/13.3\% = 77.5\%$ of the income loss. In the CFPS sample, the household income per capita drops by 14.2% after a health shock. Then the decrease in labor productivity explains $(6.9\% + 8.0\%)/2 \times 90\%/14.2\% = 47.2\%$ of the income loss.

we estimate the following equation:

$$
\begin{aligned}
y_{it} =& \gamma_t + \alpha_i + Z_i\theta + \sum_{\tau=-2}^{\tau=0} \delta_\tau^{\text{All}} \mathbb{1}[D_{it}^{All} = \tau] + \sum_{\tau=-2}^{\tau=0} \delta_\tau^{\text{Treat}} \mathbb{1}[D_{it}^{Treat} = \tau] \\
&+ \sum_{\tau=-2}^{\tau=0} \delta_\tau^{All,Z_i} \mathbb{1}[D_{it}^{All,} = \tau] Z_i + \sum_{\tau=-2}^{\tau=0} \delta_\tau^{Treat,Z_i} \mathbb{1}[D_{it}^{Treat} = \tau] Z_i + \varepsilon_{it}
\end{aligned}
\tag{4}
$$

where y_{it} denotes the outcome of household i in year t (the outcomes of interest include income per capita and the household's income percentile); γ_t denotes calendar year fixed effects, and α_i denotes household fixed effects; D_τ^{Treat} is the interaction term of the treatment dummy (whether a household member experienced a health shock) and a dummy for the wave relative to the health shock, where τ represents the relative wave (event-time). D_τ^{All} is a dummy for the wave relative to a health shock. Z_i is a dummy variable for the heterogeneity that we are interested in, which can be either urban/rural status or initial income group, in the wave prior to the health shock. ε_{it} stands for the error term. In this specification, the coefficient of interest is δ_τ^{Treat,Z_i}. We normalize treatment effects for the reference group to zero.[27]

By Urban/Rural Status We first explore whether a health shock imposes heterogeneous impacts on urban and rural residents. Figure 2 plots the differences in treatment effects between urban and rural households, using the treatment effects for urban households as the reference. In the CHNS sample, we find suggestive evidence that, compared with an urban treated household, a rural treated household suffers an additional 9% loss in household income per capita, and an additional 4.41 percentile drop in household income percentile, after experiencing a health shock. These results are economically large, though we cannot precisely estimate the coefficients at the 10% significance level. In the CFPS sample, we document a 19% additional loss in household income per capita (significant at 10% significant level), and an additional 2.76 percentile drop in household income percentile for a treated rural household, compared with a treated urban household.

These are suggestive evidence that, compared with an urban household, a rural household suffers more after receiving a health shock. This finding is in line with the idea that rural households earn income mostly by providing manual labor. In this way, rural households are more vulnerable to income loss after receiving a health shock. In addition, we see that the estimated urban-rural gap is larger in the CFPS sample. Since the CFPS sample speaks

[27]In our analysis, we treat urban households, or households initially between the 40th and 60th percentiles, as our reference group.

more about the trends in the 2010s, our estimates suggest that the urban-rural gap is larger in recent years.

By Health Insurance Status In section A.12, we discuss the heterogeneous impact of a health shock on household income and income percentile by health insurance status. In Figure A6, we explore the heterogeneous treatment effects by health insurance type. We find economically large but statistically insignificant evidence that rural residents who enrolled in the New Rural Cooperative Medical Scheme (NRCMS) suffer relatively more after receiving a health shock, compared with urban residents who are covered by the Urban Employee Basic Medical Insurance (UEBMI), or Urban Residents Basic Medical Insurance (URBMI).

By Initial Income Groups In section A.13, we investigate whether a health shock has different impacts on households who start in different income positions. Our estimates using two different samples resonate with our aggregate-level findings. First, the households that are initially in the bottom quintile lose less in recent years, potentially due to better MLSS coverage. Second, the high-income households are more capable of maintaining their income position after a health shock in recent years.

6 Medium-Run Income Mobility after a Health Shock

The estimates in Sections 5.1 to 5.3 report how a health shock affects transition probability, upward mobility, income, and income position for households that experience the shocks. However, these estimates only reflect the contemporaneous impact of a health shock on household income mobility. In this section, we track the medium-run income mobility of the treated households three years after the health shock. We see that, compared with the households in the control group, although the treated households have similar upward mobility, they have much lower transition probability. We also find suggestive evidence that households that experience a health shock have lower income three years after the health shock. These results suggest that a health shock imposes not only contemporaneous but also long-lasting impacts on household income mobility. Meanwhile, we see that the negative impacts on income mobility are larger for low-income households, suggesting that poorer households are more likely to get caught in the low-income trap after a health shock.

6.1 Sample Construction and Timeline

Recall that we construct several sub-samples and then pool them together when building the CHNS and CFPS samples. In each sub-sample, all households report positive income for three consecutive surveys. For analyzing the medium-run effects, we follow the income of households in each of the sub-samples for a fourth survey and track their medium-run income mobility.[28]

Figure A8 visualizes the timeline of the estimated medium-run income mobility. In our setting, a health shock happens between the survey wave $t = -1$ and $t = 0$, and we track household medium-run income mobility between survey wave $t = -1$ and $t = 1$. This allows us to see whether the treated households can regain their income mobility in up to three years after the health shock.

Inevitably, panel survey data are subject to sample non-responses. Thus, some households exit the survey in the fourth wave. In section A.15, we test whether the attrition in the fourth wave is random, therefore testing whether the new extended samples are representative of the original samples. Because of the non-random sample attrition issue of the CHNS extended sample, we analyze medium-run effects using only the CFPS extended sample.

6.2 Results

6.2.1 Effects on Income Mobility

Table 6 presents the medium-run transition probability for the treatment and control groups after the health shock, respectively.[29] The table delivers two main messages. First, we find suggestive evidence that, after the health shock, households that experienced a health shock still have lower transition probability compared with those that did not. Second, we see the difference in transition probability between the treatment and control groups becomes very small when the cutoff value is 0.8, suggesting that after receiving a health shock, richer households are more likely to maintain their income mobility.

In columns (1)-(3), we report the transition probability for the treatment and control groups for the CFPS sample. At the cutoff values $s = 0.2$, $s = 0.4$, and $s = 0.6$, a treated household has lower transition probability, compared with its control counterpart.

[28] We pool seven sub-samples to construct the CHNS baseline sample, and we pool two sub-samples when building the CFPS baseline sample. Six of the seven CHNS sub-samples have the fourth wave following them, except for the 2009-2011-2015 sub-sample because 2015 is the final survey wave. Similarly, we could extend only one CFPS sub-sample, as 2016 is the last survey wave. Once we have these extended sub-samples, we pool them together to form our extended CHNS and CFPS samples.

[29] Figure A9 plots the transition probability between period 0 (one wave after the health shock) and period 1 for the control and treatment groups, separately, in intervals of 5 percentile points varies from 5 to 95.

The differences are economically large, though some of them are statistically insignificant. Moreover, we see that households with a health shock are more prone to becoming caught in the low-income trap. In this sample, the chance of moving out of the bottom quintile is 64.9% for a household in the control group, but is 57.6% for a household in the treatment group. This 7.3 percentage points wedge reflects a 11.2% higher probability of being trapped for the treated households. Furthermore, we see that richer households are more likely to maintain their income position when receiving a health shock. Table 6 shows that, at the cutoff value $s = 0.8$, the difference in transition probability between the treated and control households is economically small and statistically insignificant.

Table 7 reports the medium-run upward mobility for the treatment and control groups after the health shock, respectively. We see that the treated and control households experience similar upward mobility. These upward mobility estimates suggest that the treated households have a similar ability to surpass their original income position in the medium-run, compared with the households in the control group.

Taking our findings for transition probability and upward mobility together, we see that the treated households continue to enjoy fewer economic opportunities compared with other households at the same income position before the health shocks. A possible interpretation of these observed facts is that the decline in health human capital imposes a lasting constraint on household labor supply. An alternative interpretation is that, since we observe that a health shock affects household labor supply at the extensive margin, the associated job displacement can lower a household's permanent income (see, for example, Lachowska et al. 2020). The lack of economic opportunities in the years after the health shock highlights that, apart from compensating for the out-of-pocket medical expenditures, continued financial support is needed to address the lasting plight of households that receive a health shock.

6.2.2 Effects on Household Income and Medical Expenditure

In this section, we evaluate the medium-run effects of a health shock on households' income per capita and out-of-pocket medical expenditure per capita. We estimate the following equation similar to Equation (3) using our CFPS extended sample.

$$y_{it} = \gamma_t + \alpha_i + \sum_{\tau=-2}^{\tau=1} \delta_\tau^{\text{All}} \mathbb{1}[D_{it}^{All} = \tau] + \sum_{\tau=-2}^{\tau=1} \delta_\tau^{\text{Treat}} \mathbb{1}[D_{it}^{Treat} = \tau] + \varepsilon_{it}, \tag{5}$$

where the only difference with Equation (3) is the range of the event-time dummy (now from $\tau = -2$ to $\tau = 1$, instead of from $\tau = -2$ to $\tau = 0$).

Table 8 reports the medium-run effects of a health shock. In the first column, we provide suggestive evidence that a health shock lowers household income per capita by 15.1%. This estimate reflects a 2,284 RMB decrease in household income per capita for the treatment group relative to the control-group mean of 15,124 RMB. The point estimate is economically large but statistically marginally insignificant at the 10% significance level. Column (2) reports the estimated effects of a health shock on household out-of-pocket medical expenditure. We see that a health shock has no economically large or statistically significant impact on household out-of-pocket medical expenditure in the medium-run.

6.2.3 Certainty Equivalent Value of Insurance

If there exists insurance that compensates for the income loss and the additional medical expenditure caused by the health shock, then what is the certainty equivalent value of it per person?

We can estimate a lower bound of the certainty equivalent value by only considering the contemporaneous and the medium-run effects of a health shock, and we assume there is no income loss or additional medical expenditure in the future. In this case, the certainty equivalent value of insurance is 4,944 RMB (760 USD) per person in 2011 RMB. On the other hand, we consider a scenario where the 2,284 RMB income loss three years after the health shock is permanent. Because on average, the household members receive their health shocks when they are 51 and stop working at 60, we assume that the households live for ten years after the health shock. Assuming an annual inflation rate of 3%, we get an upper bound of 20,366 RMB (3,133 USD).

7 Conclusion

Income mobility could serve as an equalizer of long-term income inequality. Thus, understanding the key factors that help households maintain high income mobility is of great importance.

This paper uses two large survey data sets in China to study the effects of a health shock on household income mobility. At the aggregate level, we find that households that experience a health shock exhibit significantly lower transition probability and upward mobility. At the household level, we take an event study approach and document that a health shock produces a substantial income loss for the treated households and lower their income percentile. Specifically, we see that the income loss and the increased medical expenditure are equivalent to about a quarter of the pre-treatment household income. We also explore household labor supply decisions in response to a health shock. Most importantly, we document that the reduced labor supply at the extensive margin and decreased labor productivity account for the majority of a household's income loss. Finally, we track household income mobility in the medium-run after a health shock. We see that households that receive a health shock still exhibit lower income mobility.

The main lesson of our analysis is that health human capital plays an important role in determining households' income mobility. Additionally, we see a widening urban-rural gap in the ability to insure against a health shock in recent years. We also see that a health shock imposes a long-term impact on household income mobility. These findings could be important for some policy-making, such as the design of unemployment and health insurance in China. These policy tools can play an essential role in reducing urban-rural inequality and improving general welfare. Our analysis throughout this paper is primarily reduced-form. In order to draw inferences on the optimal health insurance or social insurance design, additional assumptions would be required to estimate the marginal utility of consumption and the welfare costs of government transfers. However, we hope that the tools in, for example, Blundell et al. (2016), Chetty and Finkelstein (2013), and Low and Pistaferri (2015) will help surmount these complications and shed light on these research questions.

References

Ashenfelter, Orley, "Unemployment as Disequilibrium in a Model of Aggregate Labor Supply," *Econometrica*, 1980, pp. 547–564.

Bhattacharya, Debopam and Bhashkar Mazumder, "A Nonparametric Analysis of Black–White Differences in Intergenerational Income Mobility in the United States," *Quantitative Economics*, 2011, *2* (3), 335–379.

Black, Sandra E, Paul J Devereux et al., "Recent Developments in Intergenerational Mobility," *Handbook of Labor Economics*, 2011, *4*, 1487–1541.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw, "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 2016, *84* (5), 1705–1753.

_ , Pierre-André Chiappori, and Costas Meghir, "Collective Labor Supply with Children," *Journal of Political Economy*, 2005, *113* (6), 1277–1306.

Bound, John, "Self-Reported Versus Objective Measures of Health in Retirement Models," *Journal of Human Resources*, 1991, *26* (1), 106–138.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez, "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 2012, *102* (6), 2981–3003.

Chen, Yi and Frank A Cowell, "Mobility in China," *Review of Income and Wealth*, 2017, *63* (2), 203–218.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen, "Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information," *American Economic Review*, 2012, *102* (7), 3377–3405.

Chetty, Raj and Amy Finkelstein, "Social Insurance: Connecting Theory to Data," in "Handbook of Public Economics," Vol. 5, Elsevier, 2013, pp. 111–193.

_ and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

_ and _ , "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates," *Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.

— , **David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang**, "The Fading American Dream: Trends in Absolute Income Mobility Since 1940," *Science*, 2017, *356* (6336), 398–406.

— , **John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Income Segregation and Intergenerational Mobility across Colleges in the United States*," *Quarterly Journal of Economics*, 02 2020. qjaa005.

— , **Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Clark, Andrew E, Niels Westergård-Nielsen, and Nicolai Kristensen, "Economic Satisfaction and Income Rank in Small Neighbourhoods," *Journal of the European Economic Association*, 2009, *7* (2-3), 519–527.

Coile, Courtney C., "Health Shocks and Couples' Labor Supply Decisions," Working Paper 10810, National Bureau of Economic Research October 2004.

Corak, Miles, "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," *Journal of Economic Perspectives*, 2013, *27* (3), 79–102.

Cullen, Julie Berry and Jonathan Gruber, "Does Unemployment Insurance Crowd out Spousal Labor Supply?," *Journal of Labor Economics*, 2000, *18* (3), 546–572.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo, "The Economic Consequences of Hospital Admissions," *American Economic Review*, 2018, *108* (2), 308–352.

Ermisch, John, Marco Francesconi, and Thomas Siedler, "Intergenerational Mobility and Marital Sorting," *Economic Journal*, 2006, *116* (513), 659–679.

Fadlon, Itzik and Torben Heien Nielsen, "Family Labor Supply Responses to Severe Health Shocks," Working Paper 21352, National Bureau of Economic Research July 2015.

Fan, Yi, Junjian Yi, and Junsen Zhang, "Rising Intergenerational Income Persistence in China," *American Economic Journal: Economic Policy*, forthcoming.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M Shapiro, "Pre-event Trends in the Panel Event-Study Design," *American Economic Review*, 2019, *109* (9), 3307–3338.

Gertler, Paul and Jonathan Gruber, "Insuring Consumption against Illness," *American Economic Review*, 2002, *92* (1), 51–70.

Imbens, Guido W and Donald B Rubin, *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, 2015.

Islam, Asadul and Pushkar Maitra, "Health shocks and consumption smoothing in rural households: Does microcredit have a role to play?," *Journal of Development Economics*, 2012, *97* (2), 232–243.

Jaravel, Xavier, Neviana Petkova, and Alex Bell, "Team-Specific Capital and Innovation," *American Economic Review*, April 2018, *108* (4-5), 1034–1073.

Khor, Niny and John Pencavel, "Income Mobility of Individuals in China and the United States," *Economics of Transition*, 2006, *14* (3), 417–458.

Lachowska, Marta, Alexandre Mas, and Stephen A Woodbury, "Sources of displaced workers' long-term earnings losses," *American Economic Review*, 2020, *110* (10), 3231–66.

Liu, Kai, "Insuring against health shocks: Health insurance and household choices," *Journal of Health Economics*, 2016, *46*, 16–32.

Low, Hamish and Luigi Pistaferri, "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *American Economic Review*, 2015, *105* (10), 2986–3029.

Mayer, Susan E and Leonard M Lopoo, "Government Spending and Intergenerational Mobility," *Journal of Public Economics*, 2008, *92* (1), 139–158.

Qin, Xuezheng, Tianyu Wang, and Castiel Chen Zhuang, "Intergenerational Transfer of Human Capital and Its Impact on Income Mobility: Evidence from China," *China Economic Review*, 2014.

Shorrocks, Anthony, "Income Inequality and Income Mobility," *Journal of Economic Theory*, 1978, *19* (2), 376–393.

Sun, Wenkai, Jiangyong Lu, and Chongen Bai, "An Analysis of Income Mobility in Rural China," *Economic Research Journal (in Chinese)*, 2007, *8*, 43–57.

Wagstaff, Adam, "The Economic Consequences of Health Shocks: Evidence from Vietnam," *Journal of Health Economics*, 2007, *26* (1), 82–100.

Xie, Yu and Xiang Zhou, "Income Inequality in Todays China," *Proceedings of the National Academy of Sciences*, 2014, *111* (19), 6928–6933.

Zhang, Qi, Jianwei Mi, and Jikun Huang, "Income Mobility and Income Distribution: Evidence from Rural China," *Economic Research Journal (in Chinese)*, 2007, *11*, 123–138.

Table 1: Summary Statistics

	CHNS Sample		CFPS Sample	
	Control (1)	Treatment (2)	Control (1)	Treatment (2)
Individual Characteristics				
Age	52.33	51.70	51.49	51.13
	(14.67)	(14.57)	(13.59)	(13.23)
Female	0.41	0.41	0.47	0.45
	(0.49)	(0.49)	(0.50)	(0.50)
Years of Education	5.15	5.16	7.20	7.21
	(4.16)	(4.13)	(4.50)	(4.36)
Health Insurance	0.36	0.37	0.89	0.91
	(0.48)	(0.48)	(0.31)	(0.29)
Currently Working Pre-shock	0.74	0.73	0.62	0.63
	(0.44)	(0.44)	(0.49)	(0.48)
Currently Working Post-shock	0.63	0.60	0.73	0.62
	(0.48)	(0.49)	(0.45)	(0.49)
Yearly Working Hours Pre-shock	1,825	1,745	2,520	2,590
	(981)	(946)	(938)	(1,006)
Yearly Working Hours Post-shock	1,652	1,455	2,508	2,538
	(1,058)	(1,003)	(1,072)	(1,083)
Hourly-wage Pre-shock	3.63	3.13	9.51	9.71
	(5.03)	(2.81)	(10.89)	(12.69)
Hourly-wage Post-shock	7.79	5.84	10.92	10.12
	(17.97)	(8.57)	(14.25)	(12.15)
Household Characteristics				
Household Size	3.92	3.88	3.79	3.89
	(1.59)	(1.56)	(1.69)	(1.70)
Urban	0.21	0.22	0.50	0.49
	(0.41)	(0.41)	(0.50)	(0.50)
Income per capita pre-shock	6,578	6,164	10,315	10,329
	(7,160)	(7,531)	(10,751)	(10,579)
Income per capita post-shock	9,939	7,843	12,447	11,649
	(12,266)	(10,558)	(12,244)	(12,436)
Number of Households	1,551	517	4,308	1,436

Notes: This table reports the summary statistics of the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status, and the CFPS sample, where a health shock is identified by a hospitalization record, for the treatment and control groups. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. Income per capita and hourly wage have been adjusted to 2011 RMB.

Table 2: Transition Probability by Treatment Status

	CHNS Sample			CFPS Sample		
Percentile	**Control** (1)	**Treatment** (2)	**Difference** (3)	**Control** (4)	**Treatment** (5)	**Difference** (6)
20%	0.643	0.559	-0.084	0.562	0.545	-0.017
	(0.026)	(0.047)	(0.054)	(0.023)	(0.040)	(0.046)
40%	0.388	0.343	-0.045	0.384	0.321	-0.063
	(0.019)	(0.030)	(0.039)	(0.015)	(0.027)	(0.031)
60%	0.239	0.200	-0.039	0.224	0.168	-0.056
	(0.013)	(0.021)	(0.025)	(0.011)	(0.017)	(0.021)
80%	0.132	0.062	-0.070	0.090	0.081	-0.009
	(0.011)	(0.012)	(0.015)	(0.006)	(0.011)	(0.013)

Notes: This table reports the transition probability for the treatment and control groups in our two samples. Columns (1)-(3) report the estimates for the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Columns (4)-(6) report the estimates of the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The transition probability measures the probability that a household ends up at a position higher than the sth quantile in wave t, conditioning on starting at a position lower than or equal to the sth quantile in wave $t-1$. The health shock happens between times $t-1$ and t. The "Control" and "Treatment" columns report the transition probabilities for the treatment and control groups, respectively. The "Difference" columns report the difference in the transition probabilities between the treatment and control groups. Bootstrapped standard errors are in parenthesis.

Table 3: Upward Mobility by Treatment Status

	CHNS Sample			CFPS Sample		
Income Group	**Control** (1)	**Treatment** (2)	**Difference** (3)	**Control** (4)	**Treatment** (5)	**Difference** (6)
Bottom 20%	0.815	0.724	-0.091	0.782	0.747	-0.035
	(0.021)	(0.042)	(0.046)	(0.019)	(0.035)	(0.035)
20%-40%	0.573	0.574	0.001	0.624	0.509	-0.115
	(0.026)	(0.047)	(0.058)	(0.022)	(0.039)	(0.043)
40%-60%	0.466	0.442	-0.023	0.517	0.398	-0.119
	(0.027)	(0.052)	(0.054)	(0.023)	(0.035)	(0.040)
60%-80%	0.385	0.181	-0.204	0.317	0.373	0.056
	(0.031)	(0.041)	(0.053)	(0.021)	(0.039)	(0.047)
Top 20%	0.296	0.139	-0.157	0.213	0.198	-0.015
	(0.026)	(0.035)	(0.044)	(0.020)	(0.033)	(0.033)
Average	0.507	0.412	-0.095	0.491	0.445	-0.046

Notes: This table reports the upward mobility for the treatment and control groups in our two samples by initial income group. We group households into five income groups based on their initial income position, with each income group representing 20 percentiles in the national income distribution. Columns (1)-(3) report the estimates for the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Columns (4)-(6) report the estimates for the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. Upward mobility measures the probability that a household ends up in an income group higher than its original one. The "Control" and "Treatment" columns report upward mobility for the treatment and control groups, respectively. The "Difference" columns report the difference in upward mobility between the treatment and control groups. Bootstrapped standard errors are in parenthesis.

Table 4: Effects of a Health Shock on Log Household Income per Capita

	CHNS Sample	CFPS Sample
	(1)	(2)
δ_{-2}^{Treat}	0.010	-0.017
	(0.046)	(0.052)
δ_{0}^{Treat}	-0.131	-0.142
	(0.049)	(0.053)
Control Mean Post-treatment	9,939	12,447
Household FE	Yes	Yes
Observations	6,204	17,232
R-squared	0.696	0.620

Notes: This table reports the estimated effects of a health shock on log household income per capita using Equation (3). Columns (1) report the estimates using the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Columns (2) report the estimates using the CFPS sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated treatment effects. "Control mean post-treatment" reports the mean household income per capita of the control group in event period 0 (where the health shock happens between periods -1 and 0). Standard errors are clustered at the household level.

Table 5: Effects of a Health Shock on Household Income Percentile

	CHNS Sample	CFPS Sample
	(1)	(2)
δ_{-2}^{Treat}	0.180	-0.264
	(1.590)	(1.187)
δ_{0}^{Treat}	-3.939	-3.141
	(1.593)	(1.099)
Control Mean Post-treatment	49.42	49.65
Household FE	Yes	Yes
Observations	6,204	17,232
R-squared	0.594	0.670

Notes: This table reports the estimated effects of a health shock on household income percentile using Equation (3). Columns (1) report the estimates using the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Columns (2) report the estimates using the CFPS sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated treatment effects. "Control mean post-treatment" reports the mean household income percentile of the control group in event period 0 (where the health shock happens between periods -1 and 0). Standard errors are clustered at the household level.

Table 6: Medium-run Transition Probability after the Health Shock by Treatment Status

	CFPS Sample		
Percentile	Control (1)	Treatment (2)	Difference (3)
20%	0.649	0.576	-0.073
	(0.040)	(0.067)	(0.080)
40%	0.427	0.345	-0.082
	(0.029)	(0.045)	(0.053)
60%	0.275	0.218	-0.056
	(0.019)	(0.030)	(0.036)
80%	0.101	0.095	-0.006
	(0.012)	(0.019)	(0.023)

Notes: This table reports the transition probability after a health shock for the treatment and control groups in our two samples. Columns (1)-(3) report the estimates for the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The transition probability measures the probability that a household ends up in a position higher than the sth quantile conditioning on starting at a position lower than or equal to the sth quantile. The "Control" and "Treatment" columns report the transition probabilities for the treatment and control groups, respectively. The "Difference" columns report the difference in the transition probability between the treatment and control groups. Bootstrapped standard errors are in parenthesis.

Table 7: Medium-run Upward Mobility after the Health Shock by Treatment Status

	CFPS Sample		
Initial Income Group	**Control** (1)	**Treatment** (2)	**Difference** (3)
Bottom 20%	0.823	0.886	0.063
	(0.032)	(0.0037)	(0.049)
20%-40%	0.628	0.609	-0.018
	(0.041)	(0.070)	(0.081)
40%-60%	0.515	0.390	-0.124
	(0.072)	(0.058)	(0.066)
60%-80%	0.334	0.295	-0.039
	(0.032)	(0.055)	(0.063)
Top 20%	0.189	0.196	0.007
	(0.028)	(0.046)	(0.056)
Average	0.498	0.475	-0.022

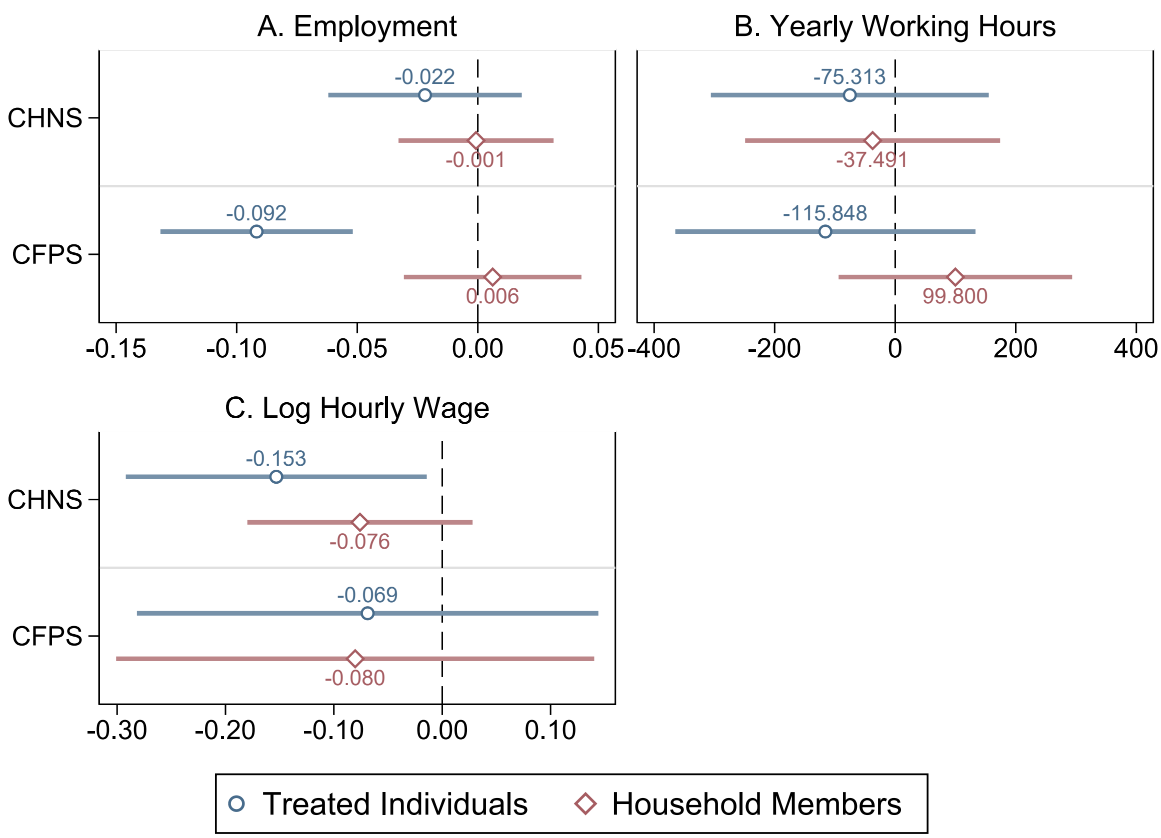
Notes: This table reports the upward mobility after the health shock for the treatment and control groups in our two samples by initial income group. We group households into five income groups based on their initial income position, with each income group representing 20 percentiles in the national income distribution. Columns (1)-(3) report the estimates for the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. Upward mobility measures the probability that a household ends up in a income group higher than its original one. The "Control" and "Treatment" columns report upward mobility for the treatment and control groups, respectively. The "Difference" columns report the difference in upward mobility between the treatment and control groups. Bootstrapped standard errors are in parenthesis.

Table 8: Medium-run Effects of a Health Shock

	CFPS Sample	
	Income per Capita	Medical Exp. per Capita
	(1)	(2)
δ_{-2}^{Treat}	0.041	-0.071
	(0.078)	(0.113)
δ_{0}^{Treat}	-0.184	0.908
	(0.093)	(0.109)
δ_{1}^{Treat}	-0.151	0.146
	(0.093)	(0.114)
Control Mean Medium-run	15,124	1,542
Household FE	Yes	Yes
Observations	8,798	8,785
R-squared	0.571	0.565

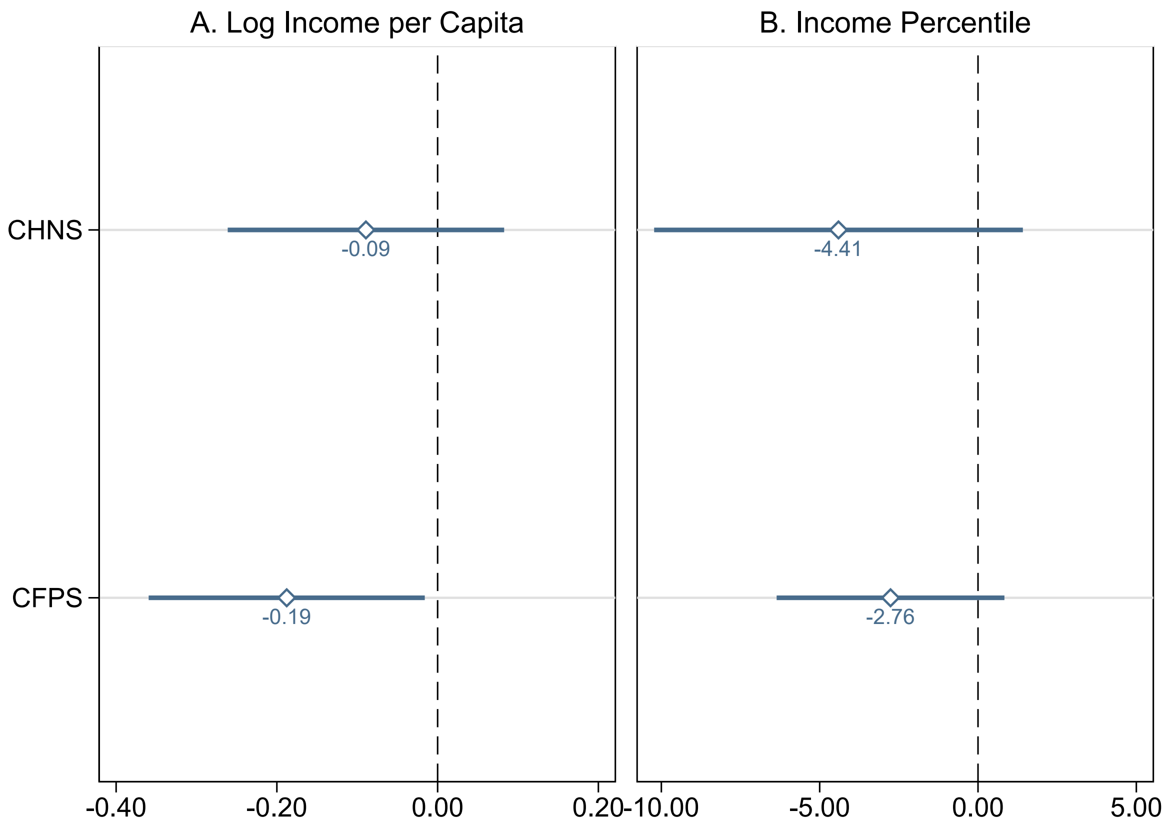
Notes: This table reports the estimated effects of a health shock on household log income per capital and log households out-of-pocket medical expenditure per capita using Equation (5). All results are estimated using the CFPS medium-run sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated short-run treatment effects, and the coefficients δ_{1}^{Treat} are the estimated medium-run treatment effects. The "Control Mean Medium-run" in columns (1) nd (2) are the control group mean household income per capita and out-of-pocket medical expenditure per person in the survey wave 1 (two waves after the healt shock). Standard errors are clustered at the household level.

Figure 1: Labor Supply Responses to a Health Shock



Notes: This figure plots the labor supply responses of treated individuals, as well as their household members, by estimating Equation (3). Figure A presents the estimated effects of a health shock on the probability of being employed. Figure B plots the estimated effects of a health shock on yearly working hours, conditional on the individuals being employed before and after the health shock. Figure C plots the estimated effects of a health shock on the log hourly wage, conditional on the individuals being employed both before and after the health shock. In the CHNS sample, a health shock is identified by a severe deterioration in self-reported health status. In the CFPS sample, a health shock is identified by a hospitalization record. The horizontal line represents the 90% confidence interval. Standard errors are clustered at the individual level.

Figure 2: Heterogeneity in Treatment Status: By Urban/Rural Status



Notes: We estimate the differences in the treatment effects between urban and rural households using Equation (4) (using treatment effects on urban households as reference). Figure A presents the difference in the estimated effects of a health shock on the household income per capita. Figure B plots the difference in the estimated effects of a health shock on the household income percentile. In the CHNS sample, a health shock is identified by a severe deterioration in self-reported health status. In the CFPS sample, a health shock is identified by a hospitalization record. The horizontal line represents the 90% confidence interval. Standard errors are clustered at the household level.

A Appendix

A.1 Details Self-reported Health Status Information

Self-Reported Health Status The CHNS ask questions related to "How do you rate your health status?" In the CHNS, there were four levels of health status before 2009: "excellent," "good," "fair," and "poor." In the 2009, 2011, and 2015 surveys, there was an additional choice, "very poor," and the respondents were required to choose one from the five options.

Table A1: Self-reported Health Status in CHNS

Choice	CHNS SRHS: 1991-2006	CHNS SRHS: 2009-2015
1	Excellent	Excellent
2	Good	Good
3	Fair	Fair
4	Poor	Poor
5	N.A.	Very Poor

Change in Health Status We define change in the health status following Gertler and Gruber (2002). We consider a household member to have "worse health status" if her self-reported health status changes from "excellent" or "good or fair" to "poor." In the CHNS, "excellent" health is defined as "excellent" self-reported health status, "good or fair" health is defined as "good" or "fair" self-reported health status, and "poor" refers to "poor" or "very poor" health status.

A.2 Additional Variable Definitions

In this section, we introduce how we define individual employment status, and how we compute individual yearly working hours and hourly wage.

Employment Status We define one individual's employment status

1. CHNS Data: The CHNS asks a question "Are you presently working?" in the adult questionnaire for all adults age 18 and older. According to the CHNS's definition, individuals who are actively seeking a job are also regarded as "unemployed", but farmers who work in the agricultural sector is thought of being "employed". The corresponding variable is B2 in the codebook "CHNS Variable Index for the Household/Individual Surveys through 2015".

2. CFPS Data:

   - In the 2010 and 2012 CFPS survey, they ask, "Do you currently have a job? (Including self-employment, or work in the agricultural sector)". We mainly use this variable to identify one individual's employment status. The corresponding variables are QG3 in the 2010 codebook and QG101 in the 2012 codebook. In 2014 and 2016 surveys, people who work in the agricultural sector or are self-employed but currently not working (because of the slack season) are also treated as employed. To have a consistent definition of employment, we adjust one's employment status based on the questionnaire's corresponding questions in the 2012 survey. The 2010 survey does not ask questions about whether it is because of the slack season so that you are not working now, so we cannot adjust employment in 2010.
   - In 2014 CFPS survey, we rely on the CFPS-constructed variable "employ2014" to identify one's employment status. This variable documents the employment status of the survey respondents.
   - In 2016 CFPS survey, we rely on the CFPS-constructed variable "employ" to identify one's employment status. This variable documents the employment status of the survey respondents.
   - It is obvious that there are inconsistencies in the definitions of employment across different surveys. The definition of employment in 2010 indeed results in a lower employment rate in that year. This explains why in the summary statistics Table 1, we see a lower "pre-shock" employment rate.

Yearly Working Hours We define the yearly working hours as the total number of hours of all jobs.

Hourly Wage In the CHNS sample, we compute the hourly wage as the total individual income divided by the yearly working hours. In the CFPS sample, we compute the hourly wage as the total wage income divided by the yearly working hours.

Medical Expenditure In the CFPS sample, the annual household medical expenditure variable is the out-of-pocket medical expenditure constructed by the CFPS. Based on this variable and household size, we compute the household medical expenditure per capita.

A.3 Characteristics of Hospitalization

The CFPS does not ask about the reasons or length of hospitalization in 2012, 2014, and 2016 surveys. Meanwhile, though it asks about the reasons for hospitalization in the 2010 survey, it does not disclose them in the data for privacy issues. In this section, we look at the length of hospitalization in the 2010 CFPS data to assess the severity of health shocks defined by a hospitalization record.

In Table A2, we provide summary statistics for the total number of days in the hospital, conditional on having at least one hospitalization record. In the 2010 CFPS survey, among 33,600 adults surveyed, 2,637 have at least one hospitalization record. Conditional on having at least one hospitalization record, the mean and median days in the hospital are 16 and 10, respectively. The share of people who spent only 1, 2, and 3 days in the hospital is 2.24, 3.00, and 5.23, respectively. Thus,the majority of individuals who have been hospitalized in the past year spend at least four days in the hospital (89.53%). Some 2.16% individuals spent at least two months in hospital.

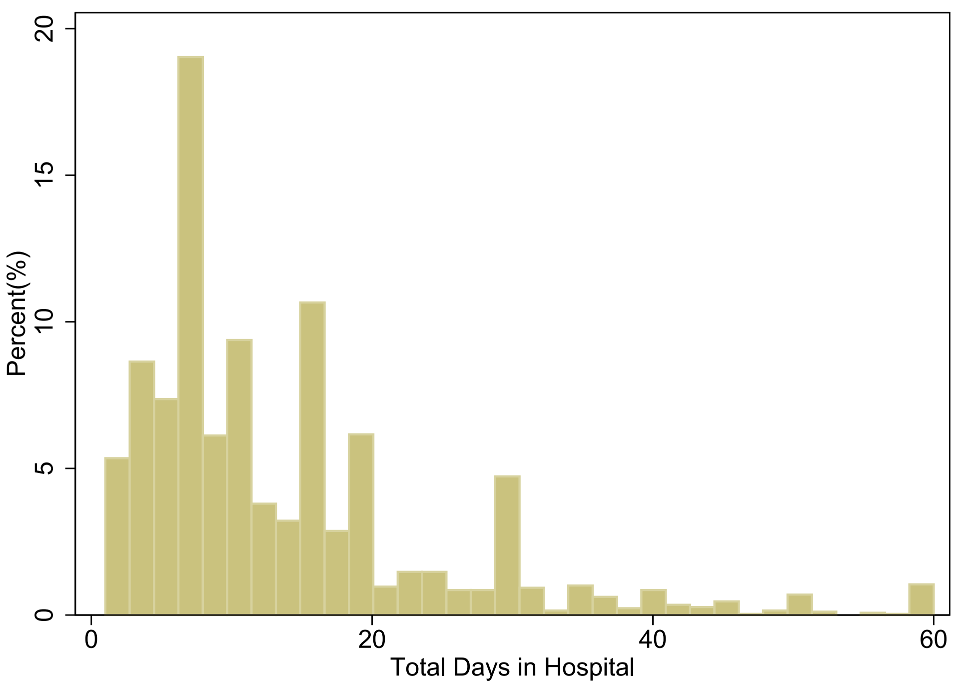
Table A2: Summary Statistics of the Total Number of Days in the Hospital

Mean	16.03
Median	10
Max	365
Min	1
25 Percentile	7
75 Percentile	19
Days = 1	2.24%
Days = 2	3.00%
Days = 3	5.23%
Days > 60	2.16%
No. of Individuals	2,637

In Figure A1, we plot the distribution of the total length of hospitalization for individuals who spent less than 60 days in the hospital. We see that most of them spent less than 20 days in the hospital, with a peak at seven days.

To sum, these statistics suggest that the majority of the health shocks, defined by a hospitalization record, are not simply overnight emergency visits. Instead, most of them are caused by underlying health conditions and need in-hospital treatments. Although we cannot directly observe reasons for hospitalization, these findings help validate our approach using a hospitalization record to define a health shock.

Figure A1: Number of Days in the Hospital



A.4 Summary Statistics before Matching

Table A3 reports the summary statistics of the CHNS and CFPS sample before matching. We see that before matching, covariates are largely imbalanced between the control and treatment group.

Table A3: Summary Statistics before Matching

	CHNS Sample		CFPS Sample	
	Control (1)	Treatment (2)	Control (1)	Treatment (2)
Individual Characteristics				
Age	41.94	51.70	46.22	51.13
	(17.29)	(14.57)	(13.98)	(13.23)
Female	0.50	0.41	0.51	0.45
	(0.50)	(0.49)	(0.50)	(0.50)
Years of Education	7.15	5.16	7.94	7.21
	(4.02)	(4.13)	(4.28)	(4.36)
Health Insurance	0.45	0.37	0.88	0.91
	(0.50)	(0.48)	(0.33)	(0.29)
Currently Working Pre-shock	0.77	0.73	0.67	0.63
	(0.42)	(0.44)	(0.47)	(0.48)
Currently Working Post-shock	0.69	0.60	0.79	0.62
	(0.46)	(0.49)	(0.41)	(0.49)
Yearly Working Hours Pre-shock	1,965	1,745	2,534	2,590
	(907)	(946)	(907)	(1,006)
Yearly Working Hours Post-shock	1,853	1,455	2,604	2,538
	(950)	(1,003)	(1,014)	(1,083)
Hourly-wage Pre-shock	4.85	3.13	9.84	9.71
	(7.14)	(2.81)	(12.23)	(12.69)
Hourly-wage Post-shock	8.45	5.84	10.90	10.12
	(13.85)	(8.57)	(14.45)	(12.15)
Household Characteristics				
Household Size	3.98	3.88	3.93	3.89
	(1.52)	(1.56)	(1.61)	(1.70)
Urban	0.25	0.22	0.49	0.49
	(0.43)	(0.41)	(0.50)	(0.50)
Income per Capita Pre-shock	7,907	6,164	10,278	10,329
	(9,109)	(7,531)	(10,485)	(10,579)
Income per Capita Post-shock	11,654	7,843	12,485	11,649
	(13,572)	(10,558)	(12,366)	(12,436)
Number of Households	21,778	517	16,115	1,436

Notes: This table reports the summary statistics of the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status, and the CFPS sample, where a health shock is identified by a hospitalization record, for the treatment and control groups before matching. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. Income per capita and hourly wage have been adjusted to 2011 RMB.

A.5 Additional Summary Statistics for the CFPS Sample

Table A4: Additional Summary Statistics for the CFPS Sample

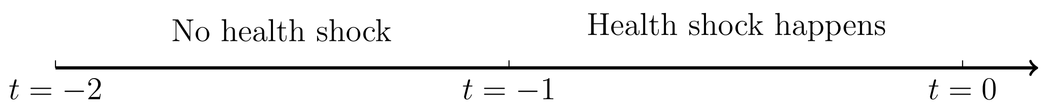
	CFPS Sample			
	Control (1)	Treatment (2)	T-Stat (3)	Normalized Difference (4)
Household Expenditure				
Medical Expenditure per Capita Pre-shock	782	840	0.72	0.02
	(2,825)	(3,134)		
Medical Expenditure per Capita Post-shock	927	2,595	8.45	0.34
	(0.32)	(0.33)		
Non-medical Consumption per Capita Pre-shock	8,276	8,714	1.58	0.45
	(8.965)	(10,109)		
Non-medical Consumption per Capita Post-shock	11,410	11,285	-0.16	-0.01
	(24,877)	(16,832)		
Housing Characteristics				
House Ownership	0.88	0.87	-1.16	-0.26
	(0.32)	(0.33)		
House Area (in m^2)	129.51	136.61	2.84	0.06
	(99.10)	(109.02)		
House Value (in 10,000 RMB)	25.75	25.69	-0.03	-0.00
	(50.04)	(112.97)		
House Type: Apartment	0.20	0.21	1.34	0.03
	(0.40)	(0.41)		
House Type: Townhouse	0.24	0.24	0.48	0.01
	(0.43)	(0.43)		
Access to Tap Water	0.66	0.68	1.37	0.03
	(0.47)	(0.47)		
Access to Flushing Toilet	0.44	0.46	1.47	0.04
	(0.50)	(0.50)		
Number of Households	4,308	1,436		

Notes: This table reports additional summary statistics of the CFPS sample, where a health shock is identified by a hospitalization record, for the treatment and control groups before matching. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock.

A.6 Timeline: Main Identification

Figure A2 plots the timeline of our main analysis. Each subsample consists of households that participated in three consecutive waves of the survey. We label the first survey wave in each subsample as wave $t = -2$, the second as wave $t = -1$, and the last as wave $t = 0$. By our sample construction criteria and definitions of a health shock, a health shock happens between $t = -1$ and $t = 0$. The expected gap between the time when a health shock happens and wave $t = 0$ is one and a half years in the CHNS sample, and half a year in the CFPS sample.

Figure A2: Timeline for Estimating the Contemporaneous Effects

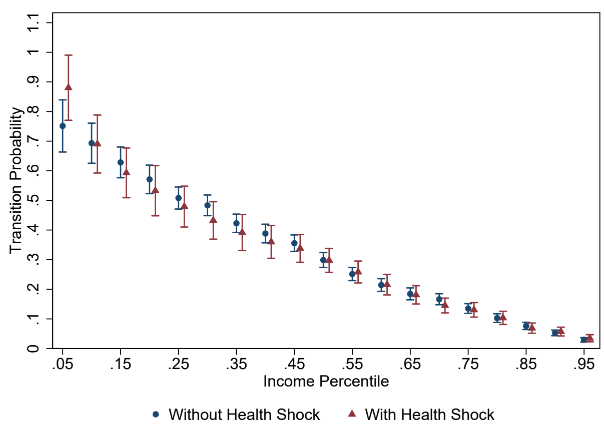


A.7 Pre-Shock Transition Probability Figures

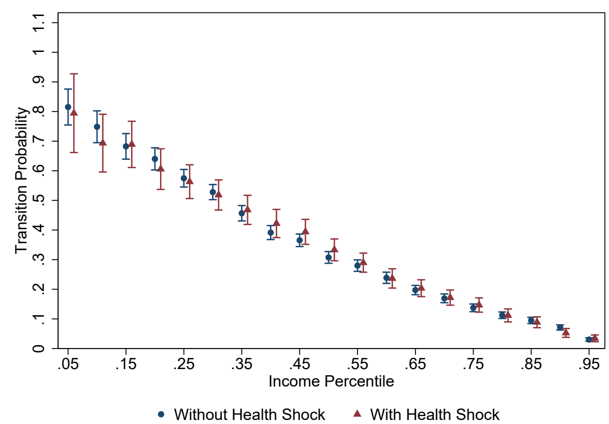
In this section, we plot the the transition probability between period -2 (two waves before the health shock) and period -1 (the wave before the health shock) in our two samples (CHNS sample and CFPS sample) for the treatment and control households, respectively, in intervals of 5 percentile points from 5 to 95. We do not see any statistically significant difference in the transition probability between the treatment and control groups before the health shock.

Figure A3: Transition Probability before the Health Shock

(a) CHNS Sample



(b) CFPS Sample



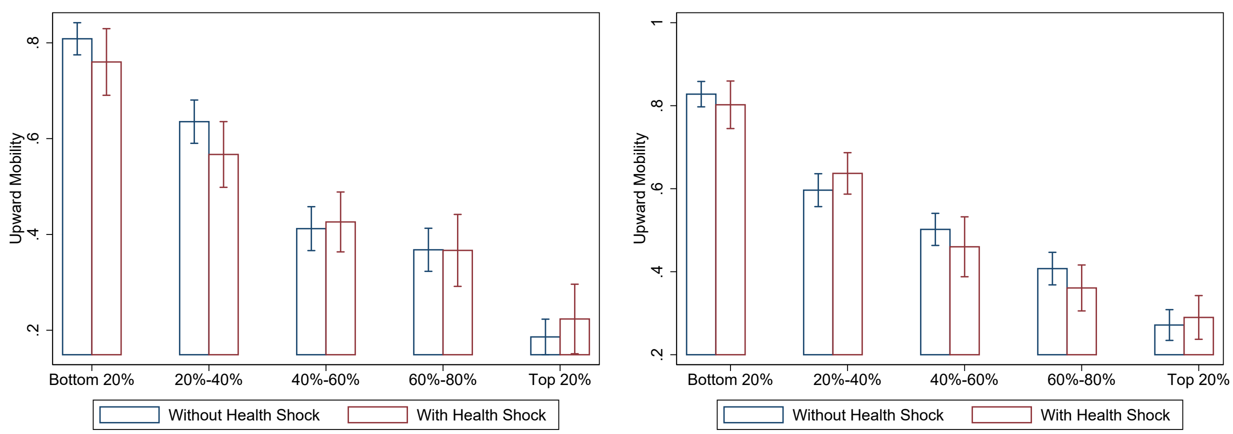
Notes: This figure plots the transition probability before the health shock for the treatment and control groups in our two samples. Figure (a) presents the transition probability of the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Figure (b) plots the transition probability of the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The vertical line represents the 90% confidence interval computed based on the bootstrapped standard error.

A.8 Pre-Shock Upward Mobility Figures

In this section, we plot the upward mobility between period -2 (two waves before the health shock) and period -1 (the wave before the health shock) in our two samples (CHNS sample and CFPS sample) for the treatment and control households, respectively. We do not see any statistically significant difference in the upward mobility between the treatment and control groups before the health shock.

Figure A4: Upward Mobility before the Health Shock

(a) CHNS Sample (b) CFPS Sample



Notes: This figure plots upward mobility before the health shock for the treatment and control groups in our two samples by initial income group. We group households into five income groups based on their initial income position, with each income group representing 20 percentiles in the national income distribution. Figure (a) presents the upward mobility of the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Figure (b) plots the upward mobility of the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The vertical line represents the 90% confidence interval computed based on the bootstrapped standard error.

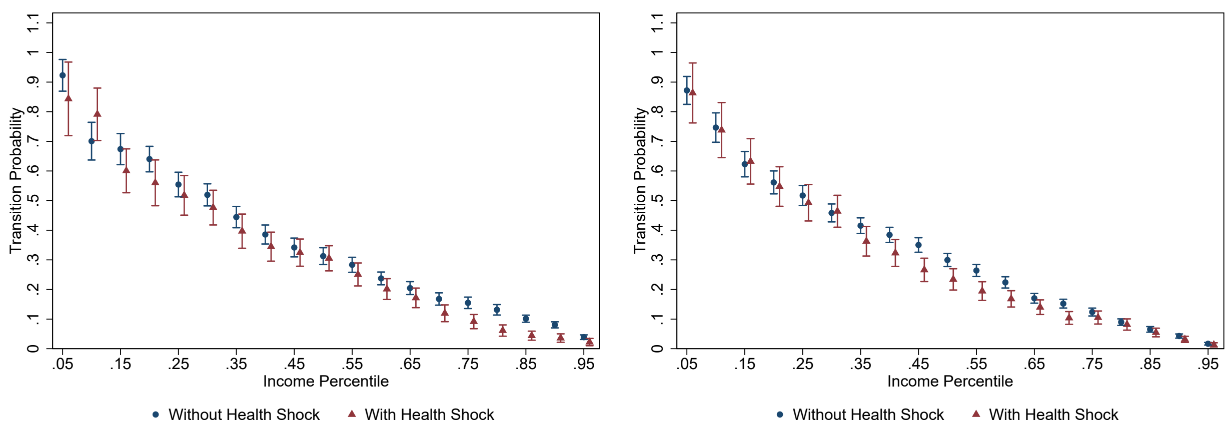
A.9 Health Shock and Transition Probability

In this section, we plot the the transition probability between period -1 (one wave before the health shock) and period 0 (the wave after the health shock) in our two samples (CHNS sample and CFPS sample) for the treatment and control households, respectively, in intervals of 5 percentile points from 5 to 95.

Figure A5: Health Shock and Transition Probability

(a) CHNS Sample

(b) CFPS Sample



Notes: This figure plots the transition probability for the treatment and control groups in our two samples. Figure (a) presents the transition probability for the CHNS sample, where a health shock is identified by a severe deterioration in self-reported health status. Figure (b) plots the transition probability for the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The vertical line represents the 90% confidence interval computed based on the bootstrapped standard error.

A.10 Effects of Health Shock on Medical Expenditure

In this section, we explore the effects of a health shock on annual out-of-pocket medical expenditure per capita. In column (1) of Table A5, we see that a health shock increases the household out-of-pocket medical expenditure per capita by 94.6% of treated households, relative to a control one. Considering the post-treatment mean out-of-pocket medical expenditure per capita of 927 RMB for the control households, this 94.6% increase in the medical expenditure translates to an additional 893 RMB out-of-pocket medical expenditure per capita for the treated households.

Based on our estimates, we see that a health shock lowers household income per capita and introduces additional medical expenditure. Taking our estimates together (1,767 RMB decrease in household income per capita and 893 RMB increase in medical spending per capita), we see that a health shock reduces the money available for expenditure by 2,660 RMB per capita. This number corresponds to 25.75% of the pre-treatment household income per capita for the treated households.

Table A5: Effects of a Health Shock on Household Medical Expenditure per Capita

	CFPS Sample
	(1)
δ_{-2}^{Treat}	-0.078
	(0.071)
δ_{0}^{Treat}	0.946
	(0.073)
Control Mean Post-Treatment	927
Household FE	Yes
Controls	No
Observations	15,353
R-squared	0.617

Notes: This table reports the estimated effects of a health shock on household medical expenditure per capita using Equation (3). Columns (1)-(2) report the estimates using the CFPS sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated treatment effects. "Control mean post-treatment" reports the mean household medical expenditure per capita of the control group in event period 0 (where the health shock happens between periods -1 and 0). Standard errors are clustered at the household level.

A.11 Effects of Health Shock on Household Consumption

In this section, we explore the effects of a health shock on household non-medical consumption per capita. We see from Table A6 that a health shock imposes no impact on household non-medical expenditure per capita. That is, a household is able to perfectly smooth their consumption when receiving a health shock.

We see from our previous estimates that a health shock lowers household income per capita significantly. There are several channels through which a household insure themselves against an income shock. In China's context, we hypothesize that the most possible explanation is that a household can borrow from their relatives.

Table A6: Effects of a Health Shock on Household Consumption per Capita

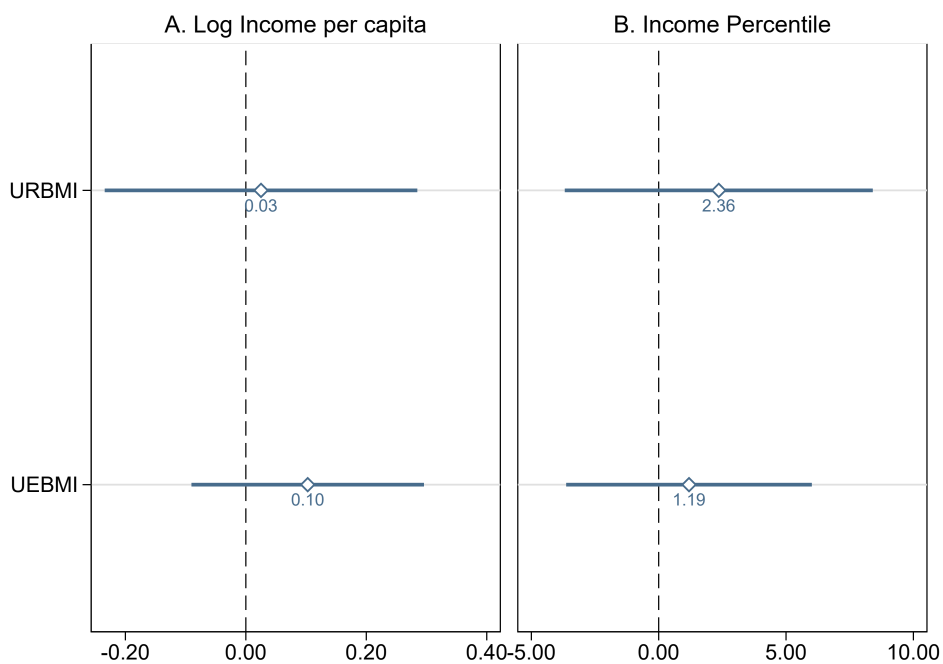
	CFPS Sample
	(1)
δ_{-2}^{Treat}	0.007
	(0.037)
δ_{0}^{Treat}	−0.001
	(0.034)
Control Mean Post-Treatment	11,410
Household FE	Yes
Observations	16,193
R-squared	0.709

Notes: This table reports the estimated effects of a health shock on household consumption per capita using Equation (3). Columns (1) report the estimates using the CFPS sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated treatment effects. "Control mean post-treatment" reports the mean household medical expenditure per capita of the control group in event period 0 (where the health shock happens between periods -1 and 0). Standard errors are clustered at the household level.

A.12 Treatment Heterogeneity by Health Insurance Type

In this section, we explore the heterogeneous treatment effects by health insurance types using the CFPS sample. In China, the insurance status is largely determined by the urban/rural status. A rural resident in China is often enrolled in the New Rural Cooperative Medical Scheme (NRCMS). Residents who live in the urban area are often covered by the Urban Employee Basic Medical Insurance (UEBMI) or Urban Residents Basic Medical Insurance (URBMI). We see that compared with people who are covered by the NRCMS, people with UEBMI or URBMI suffer less in the face of a health shock, but the difference is not statistically significant.

Figure A6: Heterogeneity in Treatment Status: By Health Insurance Type



Notes: We estimate the differences in the treatment effects between households using Equation (4) (using treatment effects on households participating New Rural Cooperative Medical Scheme as reference). Rural Households generally participate in the New Rural Cooperative Medical Scheme (NRCMS). Urban households can participate in either Urban Residents Basic Medical Insurance (URBMI), or Urban Employee Basic Medical Insurance (UEBMI). Figure A presents the difference in the estimated effects of a health shock on the household income per capita. Figure B plots the difference in the estimated effects of a health shock on household income percentile. In the CFPS sample, a health shock is identified by a hospitalization record. The horizontal line represents the 90% confidence interval. Standard errors are clustered at the household level.

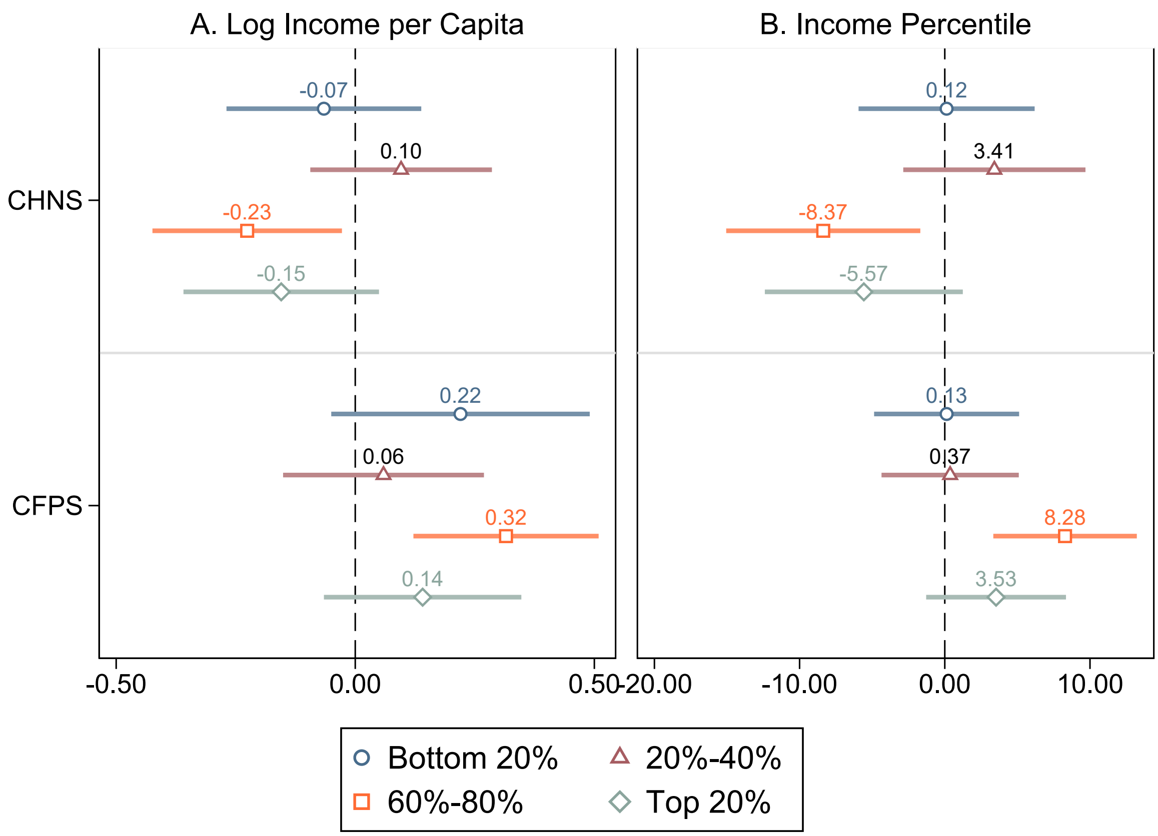
A.13 Treatment Heterogeneity by Initial Income Groups

In this section, we investigate whether a health shock has different impacts on households who start in different income positions. We group households into five income groups based on their income position in the wave before the health shock, each of which represents 20 percentiles in the national income distribution. Figure A7 plots the heterogeneity in the treatment effects between different income groups, using households that are initially between the 40th and 60th percentiles (the medium group) as our reference group.

In the CHNS sample, we see that a health shock imposes economically large impacts on households that are above the 60th percentile in the national income distribution, relative to the households in the medium group. Compared with a treated household in the medium income group, a treated household that are initially in the 60-80 percentiles and 80-100 percentiles in the income distribution loses 23% and 15% more income per capita, or 8.37 and 5.57 income percentile, after receiving a health shock. Meanwhile, we do not observe statistically significant and economically large differences in the treatment effects between the households in the medium group and households in the bottom 40 percentiles. We observe different patterns in the CFPS sample. We see an economically large difference in the treatment effects between the households in the medium income group and households in the bottom quintile and top 40 percentiles. Compared with a treated household in the medium income group, a treated household that are initially in the bottom 20 percentile, 60-80 percentile and 80-100 percentile in the income distribution loses 22%, 32% and 14% less income per capita after receiving a health shock.

These estimates using two different samples resonate with our aggregate-level findings. First, the households that are initially in the bottom quintile lose less in recent years, potentially due to better MLSS coverage. Second, the high-income households are more capable of maintaining their income position after a health shock in recent years.

Figure A7: Heterogeneity in Treatment Status: By Income Group



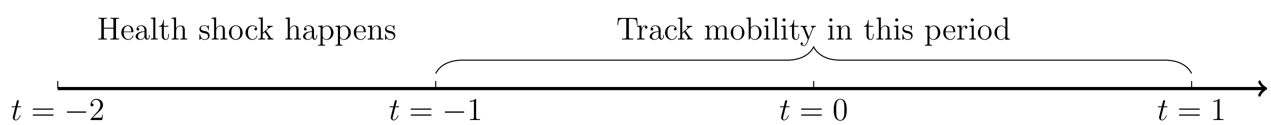
Notes: We estimate the differences in the treatment effects between households starting from different income groups using Equation (4) (using treatment effects on households between 40th and 60th percentiles of the national income distribution as reference). We group households into five income groups based on their initial income position, with each income group representing 20 percentiles in the national income distribution. Figure A presents the difference in the estimated effects of a health shock on the household income per capita. Figure B plots the difference in the estimated effects of a health shock on household income percentile. In the CHNS sample, a health shock is identified by a severe deterioration in self-reported health status. In the CFPS sample, a health shock is identified by a hospitalization record. The horizontal line represents the 90% confidence interval. Standard errors are clustered at the household level.

A.14 Timeline: Medium-run Effects

Figure A8 plots the timeline of our medium-run analysis. Each subsample consists of households that participated in four consecutive waves of the survey, and a health shock happens between the first and the second wave. We label the first survey wave in each subsample as wave $t = -2$, the second as wave $t = -1$, the third as wave $t = 0$, and the last as wave $t = 1$. The expected gap between the time when a health shock happens and wave $t = -1$ is half a year in the CFPS sample. In the CFPS sample, $t = 1$ is 2-3 years after a health shock.

In our medium-run effects analysis, we compare household income mobility between $t = -1$ and $t = 1$. Thus, we could measure whether a health shock affects household income mobility in the medium-run.

Figure A8: Timeline for Estimating the Medium-run Effects



A.15 Sample Attrition in the Extended Samples

A very important issue that we need to address when extending our sample is sample attrition. Inevitably, panel survey data are subject to sample non-responses. However, to make sure our extended samples have the same representatives as our baseline samples, we want the sample attrition to be random. Specifically, we want to ensure that the households that receive a health shock have the same attrition patterns as control households.

We use two approaches to verify the random attrition in the extended samples. First, we run a regression using a dummy variable indicating whether we lose the household when extending the sub-sample to a fourth wave as the outcome variable. On the regression's right-hand side, we include a dummy indicating the treatment status and match-pair fixed effects.[30] Table A7 shows the results of this regression. We see from Column (1) that the CHNS extended sample suffers a severe non-random attrition issue. Our estimates suggest that being in the treatment group is associated with a 7.7 percentage points higher probability of attrition. Thus, the CHNS extended sample is incomparable with the CHNS baseline sample, and we decide not to use it to analyze the medium-run effects. Column (2) shows the regression results for the CFPS extended sample. We see that the coefficient on the treatment dummy is economically small and statistically insignificant, suggesting that the treatment status has no effect on whether we will lose a household in the extended sample. Second, given that the sample attrition in the CFPS extended sample is random, we should be able to recover the contemporaneous effects of a health shock on household income by estimating Equation 3 using the extended sample. We show in section 6.2.3 that this is indeed the case for the CFPS extended sample, which further verifies our random attrition claim.

[30] In each sub-sample, we match one treated household with three control households. These four households form one match-pair. Our regression results do not change much if we exclude these match-pair fixed effects.

Table A7: Random Attrition Test

	CHNS Sample	CFPS Sample
	(1)	(2)
Treated Dummy	0.073	-0.006
	(0.026)	(0.018)
Match-pair FE	Yes	Yes
Observations	1,860	2,472
R-squared	0.281	0.422

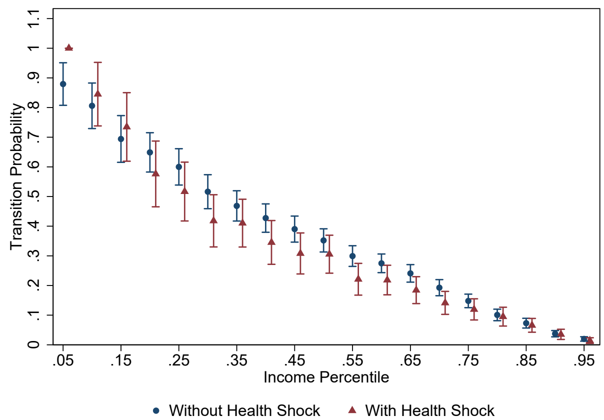
Notes: This table reports the results of a regression where we use a dummy variable indicating whether we lose the household when extending the sub-sample to a fourth wave as the outcome variable. We include a dummy indicating the treatment status and match-pair fixed effects on the right-hand side. All standard errors are clustered at the household level.

A.16 Post-Shock Transition Probability Figures

In this section, we plot the the transition probability between period -1 (the wave before the health shock) and period 1 in our CFPS sample for the treatment and control households, respectively, in intervals of 5 percentile points from 5 to 95.

Figure A9: Health Shock and Transition Probability

(a) CFPS Sample



Notes: This figure plots the transition probability after the health shock for the treatment and control groups in our two samples. Figure (a) plots the transition probability of the CFPS sample, where a health shock is identified by a hospitalization record. The treatment group refers to households that experienced a health shock. The control group refers to households with no health shock. The vertical line represents the 90% confidence interval computed based on the bootstrapped standard error.

A.17 Comparison of Regression Results under Different Health Shock Definition: CFPS Sample

The CFPS revised its measure of self-reported health status significantly during 2010 and 2016. Table A8 lists the self-reported health status choices in the CFPS survey (with share of people choosing the corresponding choice in the bracket). Due to the lack of consistency in self-reported health measurements, we are not able to identify a health shock using self-reported health status in the CFPS.

Table A8: Self-reported Health Status in CFPS

Choice	CFPS 2010	CFPS 2012	CFPS 2014-2016
1	Good (47.32%)	Excellent (10.26%)	Excellent (14.11%)
2	Fair (35.74%)	Good (20.49%)	Good (19.83%)
3	Somewhat Poor (6.14%)	Quite Good (33.18%)	Quite Good (34.43%)
4	Poor (8.78%)	Fair (not read out) (18.34%)	Fair (15.96%)
5	Very Poor (1.98%)	Poor (17.62%)	Poor (15.74%)

Alternatively, the CFPS asks about "How do you rate your health status now compared with that of last year". Three choices are available: better, same , or worse. Based on this variable, we define a health shock as a self-reported worsening health status. This measure is a loose definition of "health shock" because a worsening health status does not necessarily correspond to a health shock, and some endogeneity concerns may arise (see section 3.1.1 for a discussion). But still, we provide the results using this definition for completeness.

Table A9 reports the effects of a health shock on log household income per capita, household income percentile, and log household medical expenditure per capita. Columns (1)-(3) report the estimated effects using the CFPS SRHS sample, where a health shock is identified as a self-reported worsened health. We see mild decrease in household income per capita and household income percentile after the health shock. The estimated effects are small in magnitude and statistically insignificant. Meanwhile, we see a mild increase in health expenditure (23.3%, compared with 94.9% in the CFPS hospitalization sample). Columns (4)-(6) report the estimated effects using the CFPS hospitalization sample, where a health shock is identified as a hospitalization record (which are the same as the results in our main text).

Table A9: Effects of a Health Shock on Household Outcomes

	CFPS SRHS Sample			CFPS Hospitalization Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
δ_{-2}^{Treat}	0.016	-0.065	-0.007	-0.020	-0.013	-0.074
	(0.079)	(1.597)	(0.108)	(0.052)	(1.186)	(0.071)
δ_{0}^{Treat}	-0.017	-1.542	0.233	-0.142	-3.052	0.949
	(0.070)	(1.548)	(0.106)	(0.053)	(1.096)	(0.073)
Observations	10,188	10,200	8,633	17,232	17,232	15,353
R-squared	0.662	0.706	0.597	0.621	0.671	0.619
Household FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the estimated effects of a health shock using Equation (3). Columns (1)-(3) report the estimates using the CFPS SRHS sample, where a health shock is identified by a self-reported worsening health status. Columns (4)-(6) report the estimates using the CFPS hospitalization sample, where a health shock is identified by a hospitalization record. The coefficients δ_{-1}^{Treat} are normalized to zero. The coefficients δ_{0}^{Treat} are the estimated treatment effects. The outcome of columns (1) and (4) is the log household income per capita; the outcome of columns (2) and (4) is the household income percentile; the outcome of columns (3) and (6) is the log household medical expenditure per capita. Standard errors are clustered at the household level.