HEALTH INSURANCE, HEALTH STATUS, RETIREMENT AND WORK: EVIDENCE FROM CHINA

by

Weiyi Li

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
September 2024

Contents

List of Tables

List of Figures

Abstract

This dissertation contains three empirical chapters that examine the health and labour market outcomes across different public health insurance plans for the middle-aged and elderly population in China. Chapter 4 examines self-ranked health state inequalities between individuals under two main public health insurance plans. Using an ordered probit model and three waves of the China Health and Retirement Longitudinal Study (CHARLS), this chapter empirically investigates the association between the type of health insurance plan and self-ranked health states. The findings indicate that the more generous employee insurance plan is associated with better current health states and positive health transitions over time, with notable sex and regional disparities. Chapter 5 compares work-to-retirement transitions between individuals under two main public health insurance plans, while controlling for pension plans. The analysis, using a probit model with lagged independent variables, reveals that the more generous employee health insurance plan is associated with a higher likelihood of transitioning to retirement, particularly for females with an employee pension plan and males without one. These results suggest that individuals without adequate health insurance protection may extend their working years. Chapter 6 focuses on the working population, comparing the number of missed workdays due to health issues between individuals under two main public health insurance plans in different sectors. Using a logit model and a heaped generalized Poisson model, the chapter finds that individuals with the more generous employee health insurance plan have fewer missed workdays due to health issues, particularly those in the agricultural sector as well as those employed in both the agricultural and the non-agricultural sectors. Overall, this dissertation underscores the importance of equitable health insurance in enhancing health and labour market outcomes for the middle-aged and elderly in China, advocating for continued reforms to address existing disparities.

Glossary

ADL Activity of daily life

BOAI Basic Old Age Insurance

CHARLS China Health and Retirement Longitudinal Study

NRCMS New Rural Cooperative Medical System

NRP New Rural Resident Pension

p probability

PEP Public Employee Pension

UEBMI Urban Employee Basic Medical Insurance

URBMI Urban Resident Basic Medical Insurance

URP Urban Resident Pension

Acknowledgements

I would like to express my gratitude to all those who have contributed to the successful completion of this PhD dissertation. This journey has been a culmination of years of dedication, perseverance, and unwavering support from various individuals and institutions.

First and foremost, I extend my deepest appreciation to my supervisor, Talan Iscan, for his invaluable guidance, mentorship, and insightful feedback throughout every stage of this research. He has been not only an exceptional supervisor but also a perfect role model in life. His dedication to both work and personal life, along with his encouragement to face challenges, has profoundly influenced my approach to work, research, and life. I am also profoundly grateful to my co-supervisor, Yulia Kotlyarova, for her dedication to work and students, which has inspired me to aspire to become a professor one day. I am thankful to my dissertation committee member, Lars Osberg, for his time, expertise, and constructive criticisms, which have enriched the quality of this work. Additionally, I thank my committee member, Teresa Cyrus, for her generous kindness and support. Moreover, thank you, Casey Warman, for facilitating my defence.

I would also like to extend my gratitude to Professors Barry Lesser, Christos Ntantamis, Norovsambuu Tumennasan, Shelley Phipps, Kuan Xu, and all the other professors in the Department of Economics who have provided invaluable help and inspiration. Additionally, I am grateful to the staff at the Department of Economics for their support. Special thanks to Jodi for assisting with the paperwork, Gina for helping me reserve the meeting room, and Marilin for her assistance with numerous tedious tasks.

My journey would not have been possible without the support and assistance of my friends and family. I am especially thankful to Sarah, Zongming, Min, Polly, Joseph, Ali, and all my friends who participated in my PhD studies. I cherish our coffee breaks, where we discussed economics and shared our ideas about various interesting phenomena in life. Moreover, I would like to thank my dear friends Xiaoxi, Yaqi, Yiwen, Fanlin, and Yifan, who accompanied me during the lonely final year of my PhD studies. Their support and companionship were invaluable. Finally, I am deeply grateful to my parents for their unwavering financial and emotional support. Their belief in me has been my foundation and my motivation.

Chapter 1

Introduction

Ensuring equitable healthcare access and promoting public health are fundamental for the economic stability of a society. Health insurance systems play a crucial role in achieving these goals. This dissertation consists of three chapters that examine the health and labour market outcomes of different public health insurance plans for the middle-aged and elderly population in China.

Countries like Canada, the UK, and Germany, with comprehensive public health insurance schemes providing universal coverage, demonstrate better health outcomes, lower infant mortality rates, and higher life expectancy (Holstein et al., 2004; Muldoon et al., 2011; Ranabhat et al., 2018). China's health insurance system has undergone significant reforms since 1998, establishing three main plans: the Urban Employee Basic Medical Insurance (UEBMI) for urban employees, the New Rural Cooperative Medical Insurance (NRCMI) for the rural residents, and the Urban Resident Basic Medical Insurance (URBMI) for the unemployed and other urban residents, achieving a 95% coverage rate (Yu, 2015; Administration, N.H.S, 2021). Despite the notable achievement of attaining widespread coverage, China's public health insurance system remains segmented, with varying degrees of benefits and financial protection. Given the current context of unequal health insurance benefits among the workforce, new reforms continued after the attainment of widespread coverage, and the goal of the continued reform has been to improve the benefits of the less generous health insurance plans. Due to the similarities in terms of both design and benefits, NRCMI and URBMI have been integrated into one plan that targets both the rural and the urban resident population. This integration of the two health insurance plans began in several pioneer cities and has been progressively implemented in other

provinces. In 2020, all provinces in China completed the integration. Since then, China's public health insurance system has two main plans: one targeting employees and the other targeting urban and rural residents (National Healthcare Security Administration, 2020).

Both the new and old series of reforms have attracted significant attention from scholars, leading to extensive research on health insurance. These studies include evaluations of specific health insurance reforms, such as the New Cooperative Medical Scheme policy, and investigations into whether the merging of rural and urban health insurance plans has increased healthcare resource utilization and improved the health status of the target population. However, the disparity in health and labour market outcomes between individuals under the two main public health insurance plans has been largely overlooked. This dissertation evaluates inequalities in health and labour market outcomes among the middle-aged and elderly population with different health insurance plans in China, highlighting the need for a more equitable health insurance system. The advantage of focusing on the middle-aged and elderly population is that, for a specific cohort within a survey dataset targeting this group, all three types of health insurance plans were introduced after major life decisions, such as career and marriage choices, had already been made. Therefore, the introduction of health insurance for this group can be viewed as a quasi-natural experiment.

Chapter 2 provides an introduction to the background of China's health insurance and pension systems.

Chapter 3 describes the China Health and Retirement Longitudinal Study (CHARLS) and explains the process of constructing the dataset used in the three empirical chapters. This chapter also presents the detailed definitions and summary statistics of the variables.

Chapter 4 examines the self-ranked health state inequalities between individuals under two main public health insurance plans. I use an ordered probit model and three waves of the CHARLS dataset to empirically examine the association between the type of health insurance plan and self-ranked health states. I find that the type of health insurance plan is not only associated with respondents' current self-ranked health states but also with transitions of health

states over time. Having the more generous employee insurance is associated with a lower probability of reporting poor health and a higher probability of remaining in good health and transitioning into better health. There are also sex and regional disparities in self-ranked health states. Females are disadvantaged in terms of both current health and health transitions. Individuals living in richer provinces are more likely to remain in good health and less likely to transition into fair and poor health.

Chapter 5 compares the work-to-retirement transitions between individuals under two main public health insurance plans, while controlling for individuals' pension plans. Using the same dataset as in chapter 4, as well as a probit model with lagged independent variables, I find that the more generous employee health insurance plan is associated with a higher likelihood of transitioning from work to retirement, regardless of individuals' pension plan. However, the employee pension plan has an add-on effect. The association between health insurance and the probability of transitioning from work to retirement is larger for females with an employee pension plan and for males without an employee pension plan. These results imply that individuals without adequate protection provided by health insurance may extend their working years, even if they no longer desire to work or their physical condition is not suitable for working. This can significantly diminish the quality of their elderly life.

Chapter 6 focuses on the working population. I compare the number of missed workdays due to health issues between the individuals under two main public health insurance plans in different sectors. First, I use a logit model with lagged independent variables to estimate the association between individuals' health insurance plans and their likelihood of experiencing any missed workdays due to health issues. Then, using the same dataset, I employ a heaped generalized Poisson model to estimate the association between the type of health insurance plan and the exact number of missed workdays due to health issues. I find that having the more generous employee health insurance plan is associated with both a smaller probability of experiencing any missed workdays due to health issues and fewer missed workdays overall. However, these associations are driven by those employed solely in the agricultural sector as

well as those employed in both the agricultural and the non-agricultural sectors.

Chapter 2

Health Insurance and Pension Plans in China: Background

Before evaluating the inequalities in health and labour market outcomes among individuals covered by different types of health insurance plans, it is crucial to understand the background of China's health insurance and pension systems to comprehend the broader context in which this dissertation's findings are situated. This chapter offers a comprehensive overview of the historical developments, policy changes, and existing inequalities within China's healthcare system, setting the stage for the subsequent empirical analyses presented in this dissertation.

2.1 Public health insurance system

The extensive health insurance coverage under a variety of occupation and residence-linked insurance plans dates back to 1960s in China. Figure 2.1 shows the major changes in China's health insurance plans since 1970. By 1970, China had already achieved widespread health insurance coverage organized under three main plans for (1) government employees and college students (including military personnel); (2) employees in state-owned enterprises; and (3) rural residents. The distinguishing feature of these health insurance plans was their different benefits (Xing-Yuan and Sheng-Lan, 1995; Qingyue and Shenglan, 2013; Yu, 2015). Following the introduction of the Household Responsibility System in 1978, however, the system gradually eroded and eventually collapsed. Rural cooperatives that formed the basis of rural resident health insurance were immediately dissolved. By the late 1990s, state-owned enterprises had been restructured following many bankruptcies of them.

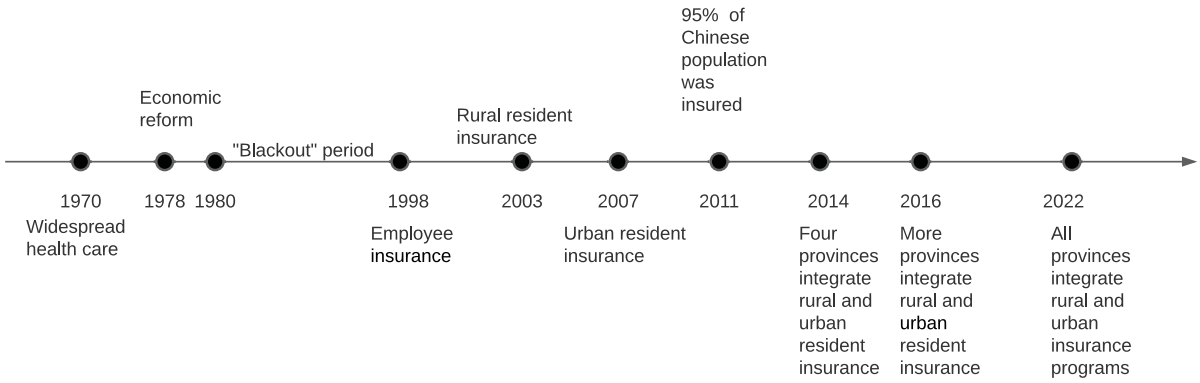


Figure 2.1: Changes in Health Insurance Plans in China since 1970

Source: Compiled by the author using the information from Xing-Yuan and Sheng-Lan (1995); Qingyue and Shenglan (2013); Yu (2015).

The erosion of access to public health insurance from 1980 to 1998 is known as the "blackout period." During this period, most households paid directly for the medical services they received. With central government funding accounting for only 10 percent of the revenues of public hospitals, inadequate funding for services was acute. Cash-strapped hospitals turned to seek alternative revenue sources. Patients were routinely over-prescribed drugs and diagnostic tests (Yip et al., 2012). Prescribed drugs ended up accounting for about half of China's total health spending (Yu, 2015). While starting in the 1990s, several better-off regional municipal governments implemented their own supplementary health insurance plans, the financing of the existing health insurance plans was in disarray. Faced with an unsustainable trend and a rapidly ageing population, the central government initiated a comprehensive health insurance reform that is ongoing to this day.

As part of the health insurance reforms, from 1998 to 2007 the central government progressively introduced three health insurance plans following the template

that existed before the blackout period: urban employee, rural household, and urban resident health insurance plans[1]. These three types of health insurance plans each target different groups of people. The urban employee health insurance plan is designed for urban employees, civil servants, and retirees[2]. It is worth noting that the employee health insurance plan is exclusively for employees; their spouses and children cannot participate. The urban resident health insurance plan covers individuals who are unemployed, self-employed, children, students, and seniors. The rural household health insurance plan targets all rural residents.

By 2011, 95 percent of the population was covered by one of these three health insurance plans (Yu, 2015). However, considerable disparities regarding quality and effectiveness across the three health insurance plans remain, with the urban employee health insurance plan being the most generous and comprehensive (Figure 2.2). For example, the employee health insurance covered both outpatient and inpatient health services. The rural resident health insurance covered both services in about 70 percent of the counties, and the other 30 percent only covered inpatient care. The urban resident health insurance only covered inpatient services. All three plans had drug coverage, but there was a big discrepancy across plans in the number of drugs being covered (Qingyue and Shenglan, 2013).

Moreover, the three plans were funded differently. The employee health insurance plan was mandatory and was financed by a payroll tax, with employer contributions set at 6 percent of the monthly payroll of an employee and employee contributions

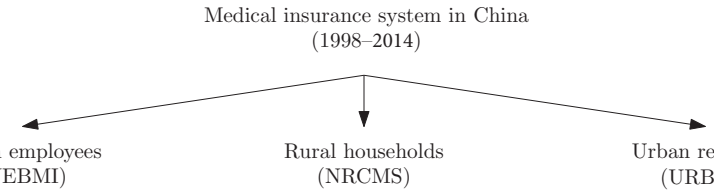
[1]Urban and rural refer to hukou status.

[2]The term "retirees" here refers to individuals who have undergone mandatory retirement. In this dissertation, I distinguish between mandatory (or administrative) retirement and economic retirement. Mandatory (or administrative) retirement refers to the requirement that employees of a government institution or a formal sector employer covered under the Employee Pension program must leave their original position, process their retirement, and start collecting their pension at the mandatory retirement age, although they can find another job and stay in the labour force. In Chapter 5, I study the effect of health insurance plans on an individual's retirement decision. In this context, retirement refers to the situation when individuals cease all work activities. Retired (mandatory) employees with a certain minimum years of service continue to be covered under the employee health insurance, though they do not make any contributions (Hou and Zhang, 2017). (See details in section 2.2.)

at 2 percent. The employee health insurance plan consists of two main parts: a social pooling account and an individual health account (Dong, 2009). An employee's monthly payment plus 30 percent of the contributions from their employer accumulate in an individual health account[3], which can be used for outpatient expenses, emergency services and medications. In several cities, the individual health account was eligible even for groceries. The rest of the premiums were deposited in the social pooling account, which was used for inpatient costs and other reimbursements. Table 2.1 presents the data on coverage and financing from 2008 to 2018. In contrast, the urban and rural resident health insurance plans require low premiums from insured individuals, as they are also funded by local and central governments.

The urban and rural resident health insurance plans have high coverage rates of 97% and 93%, respectively. Although these two plans require low premiums from insured individuals, they are less generous compared to the employee health insurance plan. The urban and rural insurance plans are quite similar in terms of their benefits; since 2010 they have been gradually integrated by local governments. Starting in 2010 with Tianjin and then followed by Guangdong, Shandong and several cities in Zhejiang, there has been a gradual policy change toward the integration of the rural and urban resident health insurance plans. This new wave of reform received a further political boost in February 2015 by the State Council, which identified four pilot provinces (Jiangsu, Anhui, Fujian and Qinghai) to evaluate the effectiveness of the integration of rural and urban health insurance plans. In May 2016, a further group of provinces (municipalities) was selected: Shanghai, Zhejiang, Hunan, Chongqing, Sichuan, Shanxi and Ningxia. Integration came with economic incentives, as integrated plans tend to have more federal funding per capita than those plans that

[3]The individual health account does not have an insurance component. The funds in individual accounts accumulate over time and can earn interest. Unused funds can roll over from year to year and can be inherited, which makes the accounts similar to savings accounts in their functionality (JIN et al., 2014). However, it is not possible to empirically identify the separate effects of health insurance and benefits from a health spending account, as the health spending account is an integral feature of the employee health insurance plan.



	Urban employees (UEBMI)	Rural households (NRCMS)	Urban residents (URBMI)
Target pop. coverage (2011):	Urban employees and civil servants and retirees since 1998, 92%	All rural residents since 2004, 97%	Urban residents who are unemployed, the self-employed, children, students, seniors since 2007, 93%
Premiums:	Payroll tax: 2% employee, 6% employer	Insured (small %); local government; central government (central and western regions)	Insured; subsidized by local and central government
Benefits:	Comprehensive coverage	Inpatient services and severe diseases	Limited inpatient services
Design:	Individual health spending account (100% of employee contributions + 30% of employer contributions) Pooled account (70% of employer contributions)	County level two models • Pooled accounts • Pooled + individual accounts	Pooled at the city level
Variation:	Generous coverage for civil servants; no contribution by retirees with minimum years of contribution	Benefits, deductables, reimbursement rates vary by region and model with some generous pooled accounts	Benefits, deductables, reimbursement rates vary by city

Figure 2.2: Comparison of Three Health Insurance Plans

Notes: the premium rates and reimbursement rates in this table are the national average. These rates can vary by county, city or province, depending on the pooling unit. Within a pooling unit, and within a health insurance plan, the rates can vary by the type of hospital. These rates incorporate a tiered reimbursement structure, where the reimbursement rate decreases for costs exceeding a cutoff point. Additionally, for employee health insurance, reimbursement rates and deductibles vary based on employment status, offering higher reimbursement rates and lower deductibles for retired employees compared to those still employed.
Source: Compiled by the author using the information from Dong (2009); Yip et al. (2012); Qingyue and Shenglan (2013); Hou and Zhang (2017).

remain separate (see Table 2.2 for data on government fundings for urban resident, rural resident and integrated resident health insurances). However, the difference in benefits between the integrated resident health insurance and the non-integrated resident health insurance was negligible when they are compared with the benefits of the employee health insurance plan.

For both the employee health insurance plan and the resident health insurance plan, reimbursement rates are less than 100% and vary based on the type of hospital[4],

[4]Hospitals in China are categorized into three levels–primary, secondary, and tertiary–based on their capacity to provide medical care, medical training, and conduct medical research (Li et al., 2008).

the services provided, and the expenditure amount, with progressive adjustments. Additionally, for the employee health insurance plan, the rates are also influenced by the individual's employment status. Since the type of hospital visited significantly impacts the reimbursement rate, section A.1 in Appendix A provides a detailed classification of different hospital types in China. Section A.2 in Appendix A provides a more detailed overview of the employee health insurance plan and the integrated resident health insurance plan.

In provinces where the urban resident health insurance and the rural resident health insurance are still separate, over our study period, the rural resident health insurance had an additional coverage for catastrophic illnesses. Both are voluntary with annual enrolments (see Table 2.3 for data on enrolment from 2008 to 2018 in these plans). Both are subsidized by local and central governments. The pooling unit for rural resident health insurance is a county and that of the urban resident health insurance is a city. The most basic and common urban resident health insurance only covers inpatient expenses, and benefits tend to vary across cities. For example, several cities have individual health spending accounts (Cao et al., 2012). Urban and rural insurance plans receive similar subsidies from the central government, and subsidies from local governments vary depending on their fiscal capacity. Premiums tend to be higher for rural resident health insurance compared to those of urban resident health insurance. Nevertheless, the premium for employee health insurance is approximately 10 times higher than that of both urban and rural resident health insurance plans (Yip et al., 2012).

In addition, the rural resident health insurance plan is a household insurance plan and has three distinct policies. Given that the premiums and funding are pooled at the county level, the local government has discretion over their choice of the policy (Cao et al., 2012). The most common policy has a social pooling account and provides partial reimbursement for inpatient services and catastrophic diseases. The

second policy resembles the employee health insurance with a pooling account and a household account. The third policy, which has the most comprehensive coverage, has a pooling account only, which covers both outpatient and inpatient costs (Cao et al., 2012). Reimbursement rates for rural patients also depend on the designated tier of the servicing hospital.Nevertheless, the benefits are limited, compared with the employee health insurance, whatever the policy is.

In addition to significant disparities in funding across types of health insurance, there are also substantial disparities across regions and provinces. Better-off provinces and municipalities typically have more funding per capita for their health insurance plans, as detailed in Table 2.4. Despite these regional disparities, from an equitable health policy perspective, the variation in benefits between the employee health insurance plan and the resident health insurance plan remains a pressing concern. This disparity highlights underlying issues in the allocation and generosity of benefits that warrant further investigation and policy adjustment to ensure fairer health coverage across different societal segments.

Table 2.1: Enrollment, Premiums, and Outlays for the Urban Employee Basic Medical Insurance

Year	Persons insured (1)	Percent employees (2)	Premiums (3)	Outlays (4)	(3)/(4)
2008	1.99	75.0	2,885.5	2,019.7	1.43
2009	2.19	74.8	3,671.9	2,797.4	1.31
2010	2.37	75.0	3,955.4	3,271.6	1.21
2011	2.52	75.1	5,539.2	4,431.4	1.25
2012	2.64	75.0	6,061.9	4,868.5	1.25
2013	2.74	74.7	7,061.6	5,829.9	1.21
2014	2.82	74.4	8,037.9	6,696.6	1.20
2015	2.88	73.9	9,083.5	7,531.5	1.21
2016	2.95	73.5	10,273.7	8,286.7	1.24
2017	3.03	73.5	12,278.3	9,466.9	1.30
2018	3.16	73.6	13,537.8	10,706.6	1.26

Notes: The number of insured is in 100 million persons. Premiums and outlays (benefit payments) are in 100 million yuan. A value less than one in (3)/(4) (the last column) indicates that disbursements by the employee health exceed the total premiums paid by the insured and their employers during the plan year. The source file does not specify whether the employees are part-time or full-time.
Source: Compiled by the author using the information from the National Health and Family Planning Commision (2018)

Table 2.2: Comparison of Integrated, Urban and Rural Resident Medical Insurances

	Total funding (100 million yuan)			Per capita funding (yuan)		
Year	Urban	Urban and Rural	Rural	Urban	Urban and Rural	Rural
2014	1,494.5	–	3,074.9	453.3	–	417.2
2015	2,085.1	–	3,197.5	530.7	–	483.6
2016	696.4	2,220.6	3,230.6	570.2	620.4	551.4
2017	282.6	5,472.3	999.8	647.0	646.1	612.9
2018	200.4	6,653.1	695.4	695.7	723.2	654.6

Notes: Urban figures for 2014 and 2015 include integrated urban and rural resident insurance funding in several provinces. The original data source does not specify whether the values are in nominal or real terms. Using 2014 as the base year (with a price index of 100), the price indices for subsequent years are 101, 103, 105, and 107 (The World Bank, 2018).
Source: Compiled by the author using the information from the National Health and Family Planning Commision (2018).

Table 2.3: Urban Resident and New Rural Cooperative Medical Insurances: Enrolment

	Persons insured			Rural cooperative			
Year	Total	Urban	Rural	Counties	Coverage rate, %	Premiums per person	Outlays
2008	9.33	1.18	8.15	2,729	91.5	96.3	662.3
2009	10.15	1.82	8.33	2,716	94.1	113.3	922.9
2010	10.31	1.95	8.36	2,678	96.0	156.5	1,187.8
2011	10.53	2.21	8.32	2,637	97.4	246.2	1,710.1
2012	10.76	2.71	8.05	2,566	98.2	308.5	2,408.0
2013	10.98	2.96	8.02	2,489	98.7	370.5	2,909.2
2014	10.50	3.14	7.36	–	98.9	410.8	2,890.4
2015	10.46	3.76	6.70	–	98.8	490.3	2,933.4
2016	7.23	4.48	2.75	–	99.3	559.0	1,363.6
2017	10.06	8.73	1.33	–	100.0	613.4	754.2
2018	10.06	8.73	1.30	–	–	654.6	839.0

Notes: The number of insured are in 100 million persons. Per capita contributions are in yuan, and outlays (benefit payments) are in 100 million yuan. Tianjin, Shandong, Guangdong and parts of Zhejiang Provinces were not included in 2014 data. The changes in the covered persons by the rural health insurance program from 2014 to 2016 are due to the fact that several provinces eliminated the rural resident insurance and began integrating the rural and urban resident insurance plans."–" indicates missing data due to the integration of urban and rural resident insurance plans. Source: Compiled by the author using the information from National Health and Family Planning Commision (2018).

Table 2.4: Subsidies to the Rural Resident Health Insurance, 2013

Region	City/Province	Funding per capita	Income per capita
East	Beijing	893.9	94,648
	Hebei	346.2	38,909
	Fujian	350.2	58,145
Middle	Hunan	350.6	36,943
	Anhui	368.1	32,001
	Hubei	365.1	42,826
West	Xinjiang	371.7	37,553
	Xizang	366.7	26,326
	Guangxi	344.2	30,741
Northeast	Heilongjiang	354.3	37,697
	Liaoning	354.0	61,996
	Jilin	362.8	47,428

Notes: Funding and income per capita are in current yuan.
Source: Compiled by the author using the information from the National Health and Family Planning Commision (2013).

2.2 Health insurance and mandatory retirement

In China, the individuals covered by the employee health insurance plan do not lose their insurance after mandatory retirement[5] as long as they have completed the minimum required years of contribution[6]. Those who have reached the mandatory retirement age but do not qualify for the post-retirement employee health insurance due to inadequate years of contribution can maintain their employee health insurance by paying a lump-sum fee or by paying monthly premiums until satisfying the minimum years of contribution. By contrast, resident health insurance does not require a minimum number of years of contribution, but enrolled individuals must pay the premiums continuously (annually) to keep the insurance active. Interruptions in payment can affect the ability to claim medical expenses and access insurance benefits. If there is an interruption in premium payments, the outstanding premiums usually need to be paid before the insurance can be reactivated. Specific policies for handling missed payments may vary by region, with some areas possibly requiring a late fee.

The minimum years of contribution[7] for employee health insurance in China exhibit three primary characteristics. First, they vary at the city and province levels but are standardized within several provinces. Second, this requirement has been increasing in some provinces. Third, the requirement differs by gender, generally being

[5]Mandatory retirement, as used here, is a direct translation of the term adopted by the Chinese government to define the age at which employees with pension plans must retire from their current job or position and begin collecting pension payments. However, individuals can remain in the labour force through through other employment. In the Chinese literature, this is alternately referred to as mandatory retirement or administrative retirement. "Mandatory retirement" is distinct from "retirement" when it is used alone. "Retirement" refers to the economic concept whereby an individual who was previously employed ceases to engage in wage-earning or self-employment activities. (Giles et al., 2023)

[6]It refers to the required period during which an individual must have made contributions to the employee health insurance plan in order to be eligible for full benefits after retirement.

[7]Unless otherwise specified, "minimum years of contribution" refers to the cumulative years of contributions, which include both actual years and equivalent years of contribution. The latter are calculated based on employees' working years prior to the enactment of employee health insurance (The People's Government of Beijing Municipality, 2023).

longer for males than for females due to later mandatory retirement ages. For instance, in Guangdong province, the minimum years of contribution previously varied among cities like Guangzhou, Shenzhen, and Jiangmen. In Guangzhou, it was 15 years for both genders until 2022. In Jiangmen, it was 20 years for those enrolled before 2018 and reduced to 15 years thereafter. In Shenzhen, those who joined the employee health insurance after 2014 needed 25 years of contribution. However, since 2022, policies across these cities have been aligning, with an anticipated increase to 30 years for males and 25 years for females (Healthcare Security Administration of Guangdong Province, 2022; 51 Shebao, 2023).

Although the minimum years of contribution differ at the city and province levels, there are two commonly used policies: 30 years for males and 25 years for females, or 25 years for males and 20 years for females. Additional policies also vary by province. Typically, the actual years of contribution accumulate when an individual moves from one province to another. However, some provinces, such as Hunan, impose a requirement for a minimum of 10 years of contribution to be specifically in that province to maintain health insurance after retirement. In a few provinces, certain cities even mandate that these contributions be continuous (51 Shebao, 2023).

There is a possibility that a firm may go bankrupt. When a firm goes bankrupt in China, employees may face an interruption in their health insurance coverage due to the cessation of premium payments by the employer. Local governments may offer temporary support, and individuals can switch to resident health insurance or purchase private insurance to maintain coverage. If they find new employment, the new employer can enroll them in their health insurance scheme (Jaelani et al., 2023).

2.3 China's current pension system and mandatory retirement age

The current pension system has three components, with the first component being the public pension plans. There are four public pension plans, each one targeting a

specific population. The Basic Old Age Insurance (BOAI) covers employees in enterprises. The Public Employee Pension (PEP) covers civil servants and employees in government institutions, such as schools and cultural and health facilities. These two plans are mandatory, while the other two plans are voluntary. The voluntary Urban Resident Pension (URP) and New Rural Resident Pension (NRP) cover, respectively, urban and rural residents who are not working for the previously mentioned institutions.

Figure 2.3 shows the major changes in China's public pension system since 1951. In 2014, the State Council announced the integration of URP and NRP into a uniform Resident Pension plan. Immediately after, in 2015, the PEP was merged into BOAI to create a uniform Employee Pension plan. By 2017, approximately 29% of the total population in China had the employee pension plan, and about 37% of the total population was covered by the resident pension plan (Fang and Feng, 2018). Individuals are not allowed to participate in both plans as they are mutually exclusive.

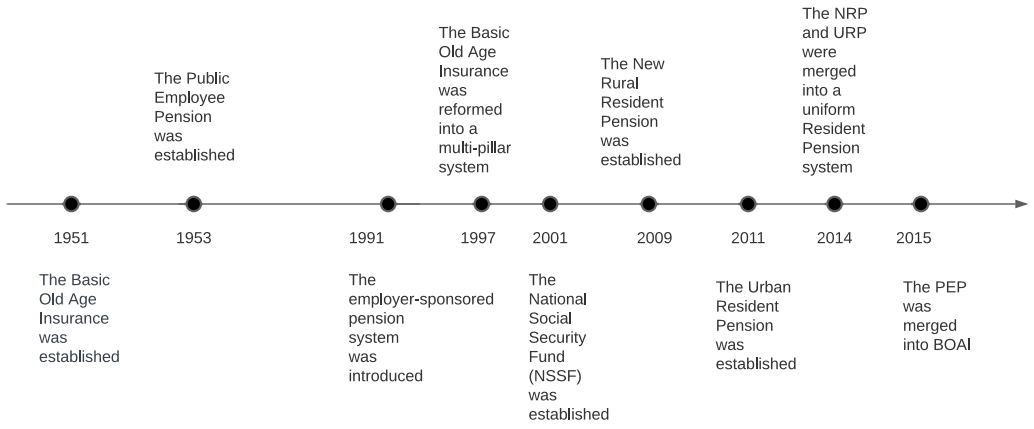


Figure 2.3: Changes in the Pension System in China since 1951

Source: Compiled by the author using the information from Fang and Feng (2018).

There are also two voluntary components in the Chinese public pension system: the employer-sponsored annuity plan, tailored for employees, and the household

savings-based annuity insurance plan, intended for individuals not covered by the employee pension plan. These components function as supplements to the public pension plans and, as such, do not receive fiscal subsidies from the government. The number of participants in these additional plans is limited. Only less than 0.5% of all firms in China offer an employer-sponsored annuity plan. Besides, both plans are still in the initial development stage (Fang and Feng, 2018). Given that only a small number of individuals have these supplementary plans, this dissertation solely focuses on the public pension system.

The employee pension plan in China primarily aligns with a defined benefit (DB) plan but includes features similar to a defined contribution (DC) plan through its employer-sponsored annuity component. This plan ensures a fixed monthly payment upon retirement, calculated based on the employee's average salary during the final years of their career, their total years of contribution, and a specific benefit accrual rate set by the government. Both employers and employees must contribute to the pension fund. Employee contribution rates typically range from 8% to 10% of their monthly salary, while employers contribute at a significantly higher rate. The retirement benefits are determined using a formula that accounts for the average monthly salary over the last few years of employment, the number of years of contribution, and a coefficient factor (typically 1%). This formula ensures a predictable payout, which is a characteristic of DB plans.

In China, the mandatory retirement age is a crucial exogenous factor influencing retirement decisions for individuals enrolled in the employee pension plan. Only non-agricultural workers are eligible for the employee pension plan, but not all non-agricultural workers are covered by it. Only certain jobs offer the employee pension plan. Others who do not have the employee pension plan can choose to opt in to the resident pension plan, which does not have a mandatory retirement age (Figure 2.4).

The mandatory retirement age is 60 for males, 50 for women working at "blue-collar" jobs, and 55 for women working "white-collar" jobs. However, only about 7% of women employees are eligible for the latter (Fang and Feng, 2018). These age limits have not changed since the retirement plan's inception in the 1950s. The minimum contribution period is 15 years. In the event of early retirement, pension benefits are pro-rated. The employee pension plan allows early retirement for health reasons or hazardous work conditions. There are also those who report "internal retirement", who continue to receive a salary from their employer without working and are below the mandatory retirement age. Administrative approval rates for early retirement and internal retirement vary across regions. For example, internal retirement is not approved in well-off cities.

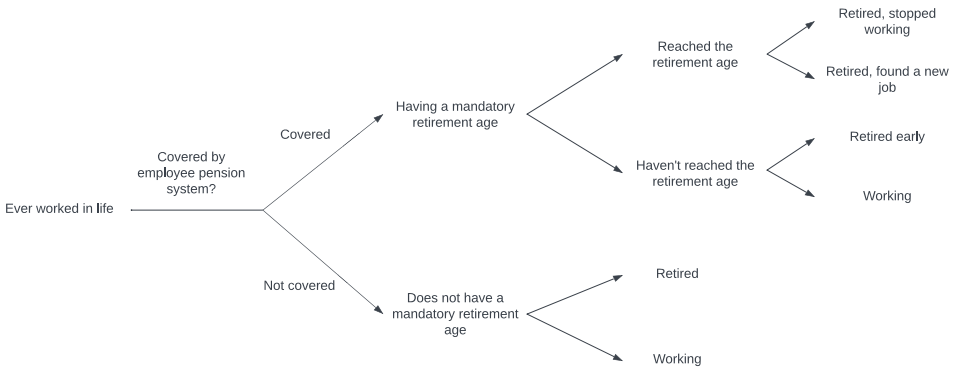


Figure 2.4: Pension Plan, Retirement Age and Employment

Notes: Individuals who are not covered by the employee pension plan include those covered by other pension plans and individuals who do not have any pension plan. Mandatory retirement is a direct translation of the term adopted by the Chinese government. "Retirement", when it is used alone, refers to the economic concept of retirement, which is when an individual who was previously employed ceases to engage in wage-earning or self-employed activities.

There are also individuals who continue to work after reaching their mandatory retirement age. They can draw their pensions, and re-enter the workforce. There is no penalty for re-entering the workforce, and their employers do not pay pension premiums. Individuals who are not covered by the employee pension plan can contribute

to the resident pension plan or a private pension plan.

The resident pension plan in China combines features of both defined benefit (DB) and defined contribution (DC) plans, but it more closely resembles a defined contribution framework. Participation in this plan is voluntary, allowing individuals to set their contribution levels based on their financial ability, a defining characteristic of DC plans. Contributors have personal accounts where their funds accumulate, and these accounts directly influence the pension benefits they receive, similar to DC plans where the retirement payout is determined by the accumulated contributions and investment returns. Participants can choose their contribution amounts within the limits established by the government, emphasizing the flexibility typical of DC plans. Additionally, the government often provides subsidies or top-ups to individual accounts, especially for lower-income residents, introducing social welfare elements into the pension scheme.

Despite its primary alignment with DC plans, the resident pension plan includes a DB element through a basic pension benefit that supplements individual accounts. This guaranteed minimum payout helps reduce the retirement income uncertainty that is common with pure DC plans.

Although the resident pension plan includes government subsidies, the benefits it offers are significantly lower than those of the employee pension plan. In 2018, the median employee and resident pension payments were about 2880 and 95 yuan/month, respectively (Giles et al., 2023), while the poverty line was 250 yuan/month (National Bureau of Statistics, 2023). The resident pension plan is only 40% of the official poverty line. Hence, for those who are covered by the resident pension plan, it remains challenging to survive by relying on a resident pension alone.

Chapter 3

Data

For empirical analyses examining whether there are health and labour market outcome inequalities among individuals covered by different types of health insurance, I use the China Health and Retirement Longitudinal Study (CHARLS). This study is nationally representative and modeled after the U.S. Health and Retirement Study (HRS) [1] (Zhao et al., 2014). The target population of CHARLS is those who are aged 45 and over and their spouses. Following the baseline survey, a certain number of respondents exit and new respondents enter in each wave. The survey contains questions on health and socio-economic conditions. There are eight sections: demographic information; family information; health states and body function; health care and insurance; employment, retirement and pension; income, expenditures and assets, housing characteristics; and interviewer observations (Zhao et al., 2013). The initial national baseline survey was conducted in 2011, with 17,708 respondents from 28 provinces, autonomous regions, and municipalities. There are four waves of released survey data, conducted in every two years from 2011 until 2015, and then in 2018. It is worth noting the CHARLS study population, consisting of individuals aged 45 and older in 2011, was largely established in their careers by the time employee health insurance was introduced in 1998, at which point they were aged 32 and older. Consequently, most respondents had chosen their jobs and occupations without considering the availability of health insurance, allowing the insurance introduction to be viewed as a quasi-natural experiment.

[1] Each dataset has an official website with detailed descriptions. (The Regents of the University of Michigan, 2024)

Although there are four waves available, I only use the first three (2011, 2013 and 2015) waves. The reason is that after 2015, the survey frequency changed from two years to three years, making comparisons difficult. In three empirical chapters, I use lagged independent variables, which requires equal time intervals (Craig and Sendi, 2002). Moreover, in the 2015 wave, the coding of "type of health insurance plan" is highly problematic. When the 2015 wave was initially released, data on type of health insurance was entirely missing. Thus, I tried to get access to the type of health insurance data in 2015 using the "harmonized dataset " which was created by the CHARLS team as a user-friendly version of a subset of the CHARLS interviews. In the harmonized dataset, the coverage rate of public health insurance increases from 92 percent to 95 percent from 2011 to 2013, suddenly drops to 90 percent in 2015, before increasing to 96 percent in 2018. This sudden drop and rebound cannot be explained by any independent evidence on plan participation that I am aware of. To assess the reliability of the harmonized dataset, I compared the health insurance coverage rates between the original and harmonized datasets and found them to be consistent for the 2011, 2013, and 2018 waves. However, the 2015 data could not be verified due to its absence in the official release. I also checked the main questionnaire and the codebook in 2015. Unfortunately, given the published data, it is not possible to generate an accurate health insurance coverage rate for 2015. Therefore, I focus on the 2011 and 2013 waves of the CHARLS dataset, and several variables except for the type of health insurance variable in the 2015 wave.

I will present the sampling methodology of the CHARLS dataset in section 3.1 and the dataset construction in section 3.2. In section 3.3, I show how each variable used in the empirical chapters is defined and section 3.4 presents the summary statistics of these variables.

3.1 Sampling methodology

CHARLS uses a multi-stage probability sampling (Figure 3.1). In the first stage, 150 county-level units were randomly chosen using a probability-proportional-to-size sampling (PPS) technique (Zhao et al., 2014). The sample was stratified by region and, within region, by urban districts and rural counties and income per capita. The final sample fell within 150 counties among these 28 provinces. In the next state, the primary sampling units (PSUs), which are the lowest level of administrative units, were selected. In rural areas, this unit is a village *(cun)* and in urban areas it is a district *(shequ* or *juweihui)*. Collective dwellings such as military bases, schools, dormitories, nursing homes, and domestic helpers living in the homes of employers are excluded (Zhao et al., 2014). Although nursing homes were excluded from the sample, the number of residents in nursing homes was rather small when the surveys were conducted. Migrants[2] were sampled in their homes of origin as household members. Dwelling units with more than one household were surveyed based on whether each household contained individuals aged 39 and above, and if so, one age-eligible household was randomly selected. If the selected household had more than one age-eligible member, one member was randomly selected. If the selected person was between 39 and 45, then he or she was designated for inclusion in the future refreshment sample. If the chosen person was aged 45 or above, both the person and their spouse were interviewed, regardless of the spouse's age. However, if the selected respondent was unable to answer the questions (due to illness or being unreachable because of migration), a knowledgeable person in the household was allowed to respond instead ("a proxy"). However, the section on health states and body function which includes the self-ranked health states did not allow for proxy respondents.

[2]Migrants are individuals who move between provinces or between rural and urban districts within China.

3.2 Dataset construction

In the empirical analyses, I use a combination of current and lagged independent variables. This requires observing the same individual in at least two consequent waves of CHARLS. Figure 3.2 shows how respondents enter and exit the survey. The 2011 wave had 17,596 respondents, 15,039 of them were re-sampled in 2013, while 2,557 exited. The 2013 wave had 3,416 new entrants with a total of 18,455 respondents. Then, among these respondents to the 2013 wave, 16,140 of them were re-sampled in the 2015 wave, 2,315 exited, and 4,827 new respondents entered the survey. Among the 16,140 respondents who were re-sampled in the 2015 wave, 13,436 of them were in both 2011 and 2013 waves, and 2,703 of were only in the 2013 wave.

I am aware that sample attrition can lead to selective samples which introduce attrition bias into estimates of parametric. Usually, there are three types of attrition: attrition due to death, attrition due to not reporting health states, and attrition due to non-response (Jung, 2020). While death is a factor in survey attrition, it is not the primary reason. Specifically, in the 2011 wave, of the 2,557 individuals who exited the survey, only 16 percent did so due to death. However, other reasons for survey attrition are not documented. While using sampling weights or inverse probability weights can account for attrition on observables, attrition on unobservables would require selection models and exclusion restrictions which are often impossible to find[3]. Many studies, however, point to very mild attrition effects even in longitudinal surveys with high attrition rates (Fitzgerald, Gottschalk and Moffitt, 1998; Lillard and Panis, 1998; Alderman et al., 2001). Moreover, I limited the sample to individuals up to 85 years of age, which the attrition issue due to death is partly mitigated.

[3]Attrition on observables occurs when, given the explanatory variables, the attrition process is independent of the dependent variable. Conversely, attrition on unobservables arises when this conditional independence does not hold. A sample selection model can adjust for attrition on unobservables (Jung, 2020). However, it requires an exclusion restriction to ensure identification. This restriction takes the form of an instrumental variable that influences attrition but not the dependent variable (Hausman and Wise 1979; Ridder 1992).

Figure 3.3 shows the main steps involved in the construction of the basic dataset. After merging all the relevant survey sections for this study and deleting observations with errors, I was left with 17,538 respondents in the 2011 wave, and 18,356 respondents in the 2013 wave. Among these observations, there are 16,336 complete cases in 2011 and 16,244 complete cases in 2013. After combining these two panels with complete cases, the pooled dataset has 32,580 observations. This pooled cross-sectional dataset is used as the starting point for all the empirical analyses, followed by additional data selection criteria specific to each chapter, and will be merged with several variables in the 2015 wave to construct a final dataset for each chapter.

The CHARLS dataset provides household and individual weights, both of which include options for with and without non-response corrections. The non-response-corrected weights are derived by multiplying the uncorrected weights by an inverse probability weight, constructed through a logit regression on variables such as age, gender, and rural/urban status. Individual weights use the inverted household weights as a base, divided by the probability of an individual being sampled, conditional on the household being selected.

I do not use weights in this study for the following reasons: First, the inverse probability weighting factor is constructed based on individuals' self-reported data, which introduces a degree of subjectivity. Second, published studies have not reached a consensus on which weight should be used. For instance, there are studies that use individual weights with non-response adjustment (Lei et al., 2024), while others use individual longitudinal weights (Li et al., 2019). Third, most of the models use lagged independent variables, which limit analysis to individuals who participated in consecutive waves, making the inclusion of weights less meaningful. Nevertheless, I applied individual weights with non-response adjustment (chapter 4) as a sensitivity check, and the results remained consistent with those from the unweighted models. Li et al., 2019 also find such a consistency between weighted and unweighted.

3.3 Construction of variables

I use information on self-ranked health states, enrolment in health insurance plans, demographic characteristics, and socio-economic variables. One of the socio-economic variables, expenditure, is only used in chapters 4 and 5. In chapters 5 and 6, I also use the work, retirement and pension data. Moreover, I use information on children of the survey respondents in chapter 5. The number of respondents in each relevant section of the survey varies.

3.3.1 Health indicator: self-ranked health state

Self-ranked health state is used in all the empirical analyses. Self-ranked (also referred to as self-reported or self-assessed in the literature) health outcomes not only serve as predictive indicators of objective measures, such as mortality, but are also extensively used in the labour economics for their role in influencing labour market decisions (Kerkhpfs and Lindeboom, 1995; Lindeboom and Kerkhofs, 2009; and Kapteyn and Meijer, 2014). The rationale for and advantages of using self-reported health status are further discussed in detail in each of the empirical chapters.

In anticipation that self-ranked health states can be influenced by framing and salience, CHARLS asks respondents to self-rank their health states in two separate instances in the survey section Health States and Body Function—at the beginning and at the end of the section. In each instance, respondents are presented with a different scale to rank their health states. The first, "excellent scale", uses excellent, very good, good, fair, and poor as the health-state categories. The second, "very good scale", uses very good, good, fair, poor, and very poor. I adopt the "very good scale" in this dissertation. The main reason is that, in the 2011 wave, about 25% of the responses are missing for the "excellent scale." Respondents were randomly assigned to one of two groups, with one group answering the "excellent scale" at the beginning and answering the "very good scale" at the end, and the other group the

other way around. Although I have to use the "very good scale", I am aware that framing may lead to different distributions for self-ranked health. Thus, I compare the distributions of self-ranked health state under these two scaling methods. To make these two scaling methods comparable and to match the variables I use in regressions, I regroup several categories of self-ranked health state for both scaling methods. In the "excellent scale", I regroup excellent, very good, and good into good. In the "very good scale", I regroup very good and good into good, as well as regrouping poor and very poor into poor. Thus, each scaling method has three categories of self-ranked health state: good, fair, and poor. I find that the "very good scale" shifts the distribution downward, with smaller fractions of respondents grouped into the good and fair categories and a larger fraction of respondents grouped into poor (Table 3.1).

To assess whether one respondent will be grouped into the same health status under these two scaling methods, I compare the individual responses in these two scaling methods. Table 3.2 compares the individual responses to health status classifications using two different scaling methods: the "excellent scale" and the "very good scale". The column represents the reclassified groups from the "excellent scale", and the row represents the reclassified groups from the "very good scale". Both scales are re-classified into "good, fair, poor and missing". Each cell represents the number of individuals who fall into a specific category on both scales. For example, 2,878 individuals are classified as good on both the excellent and very-good scales, while 1,258 individuals are classified as fair on the excellent scale and good on the "very good scale". However, results of the marginal homogeneity test (Stuart-Maxwell) show that the distributions of the paired responses are statistically different (Table 3.2). Although these two distributions are statistically different, the excellent in the first scale and the very good in the second scale are small and do not necessarily affect the distribution of rescaled health state indicators. Moreover, under the "very good scale", the very good group (9%) and very poor group (5%) are small compared to

the other three groups, providing some justification for regrouping.

Using the "excellent scale" results in a higher fraction of individuals reporting good health, and a lower fraction of individuals reporting poor health. Hence, all countries need to use the same scaling method while doing international comparison. However, given the error in the "excellent scale " in China, only the "very good scale" can be chosen. In latter waves, only the very good scale is available. The reason I report Tables 3.1 and 3.2 here is that using the very good scale could potentially result in an underestimation of individuals' health outcomes.

In studies assessing self-ranked health, some utilize the original five-point Likert scale, ranging from very good to very poor (Lei et al., 2014; Liu et al., 2015; Jiang et al., 2019), while others employ scales such as excellent to poor (Jung, 2020). However, regrouping the five-point Likert scale into broader categories, "good, fair, and poor", is also a common approach (Gong et al., 2016; Li et al., 2019; Zhang et al., 2023), particularly to better capture trends or "general attitudes" (Robinson, 2018). Another reason for regrouping in the CHARLS data is the limited sample size within the "very poor" category. The "very poor" category includes approximately 5 percent of respondents, while the "fair" category accounts for about 50 percent. Accordingly, I use three combined categories of self-rated health in the regression analyses rather than the original five categories.

3.3.2 Health insurance

Health insurance plan is also used in all the empirical analyses. I consider two main types of health insurance, the urban employee health insurance and the urban and rural resident health insurance, as well as a "no health insurance" group. The survey collects information on many types of health insurance plans. Type of health insurance in the questionnaire varies across different waves. The 2011 survey contains nine types of health insurance: urban resident medical (health) insurance, rural medical

Table 3.1: Indicators of Health State

Categories	Excellent scale		Very-good scale	
	N	%	N	%
a) Original five categories				
Excellent	241	1.3	–	–
Very good	1,761	9.6	1,619	8.8
Good	2,498	13.6	2,517	13.7
Fair	9,201	50.1	9,092	49.5
Poor	3,820	20.8	3,383	18.4
Very poor	–	–	918	5.0
Missing	835	4.6	827	4.5
b) Combined three categories				
Good	4,500	24.5	4,136	22.5
Fair	9,201	50.1	9,092	49.5
Poor	3,820	20.8	4,310	23.4
Missing	835	4.6	827	4.5

Notes: The number of observations is 18,356. Panel b: under "excellent scale", good combines excellent, very good and good from the original categories in panel a. Other categories remain the same. Under "very-good scale", poor combines poor and very poor from the original categories in panel a. Other categories remain the same. "-" means not applicable. The result of the marginal homogeneity (Stuart-Maxwell) show that the distributions of the paired responses are statistically different (Table 3.2).

Source: Based on CHARLS 2013 wave. See Figure 3.3 for the construction of the 2013 wave.

Table 3.2: Individual Responses to Health State in Two Scaling Methods

Excellent scale	Very-good scale				
	Good	Fair	Poor	Missing	Total
Good	2,878	1,258	169	195	4,500
Fair	953	6,611	1,280	357	9,201
Poor	110	870	2,588	252	3,820
Missing	195	353	264	23	835
Total	4,136	9,092	4,301	827	18,356

Notes: p-value for the marginal homogeneity test (Stuart-Maxwell) test: 0.0000, where H_0 : There is no difference in the distribution between the paired responses.

Source: Based on CHARLS 2013 wave. See Figure 3.3 for the construction of the 2013 wave.

(health) insurance, employee medical (health) insurance, urban and rural resident medical (health) insurance, government medical (health) insurance, medical aid, private medical (health) insurance purchased by the respondent's union, private medical (health) insurance purchased by the individual, and other medical (health) insurance.

In the 2013 survey, urban non-employed person's health insurance is also included. Respondents can choose more than one type of health insurance in the survey.

The survey respondents rarely switch overtime between different health insurance plans and most changes in health insurance plan are from having no insurance to resident insurance (Table 3.3).

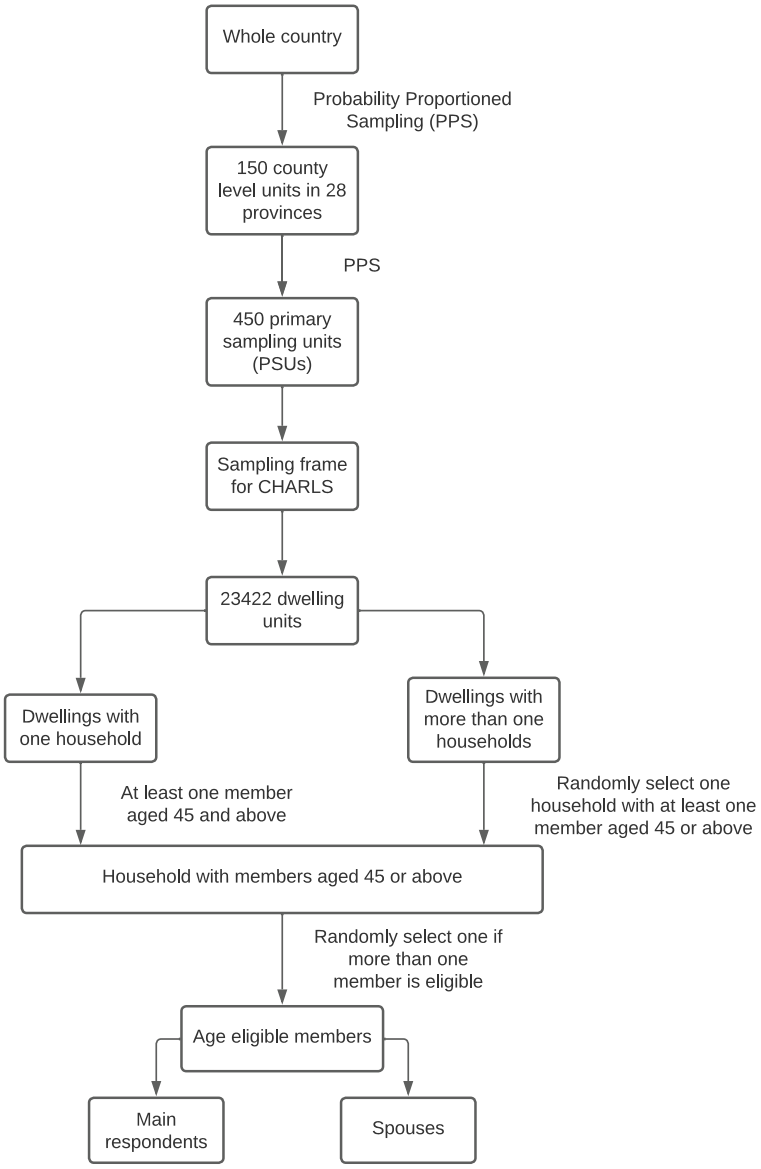


Figure 3.1: Sampling Procedure for Baseline Survey 2011

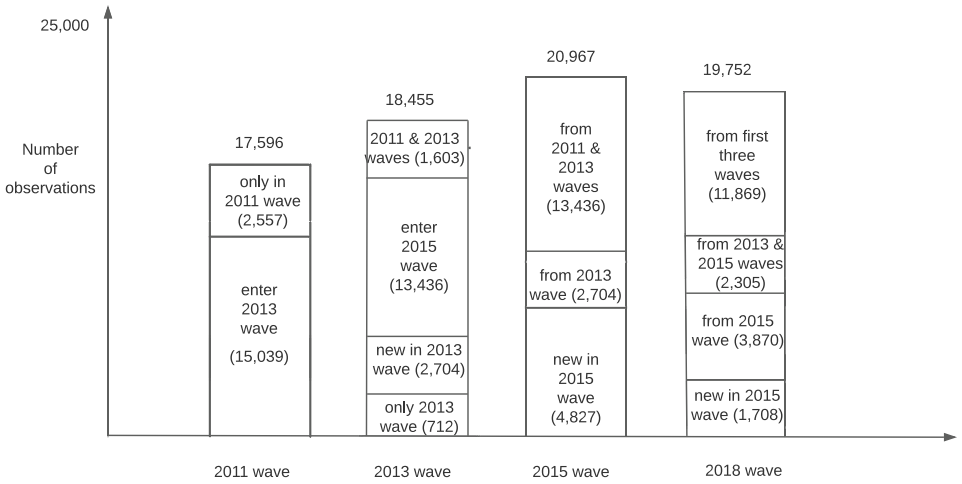


Figure 3.2: Number of Respondents by Survey Year (Wave)

Notes: This figure displays the number of observations entering and exiting the survey during each wave. Only the waves from 2011 to 2015 are used in the empirical analysis.

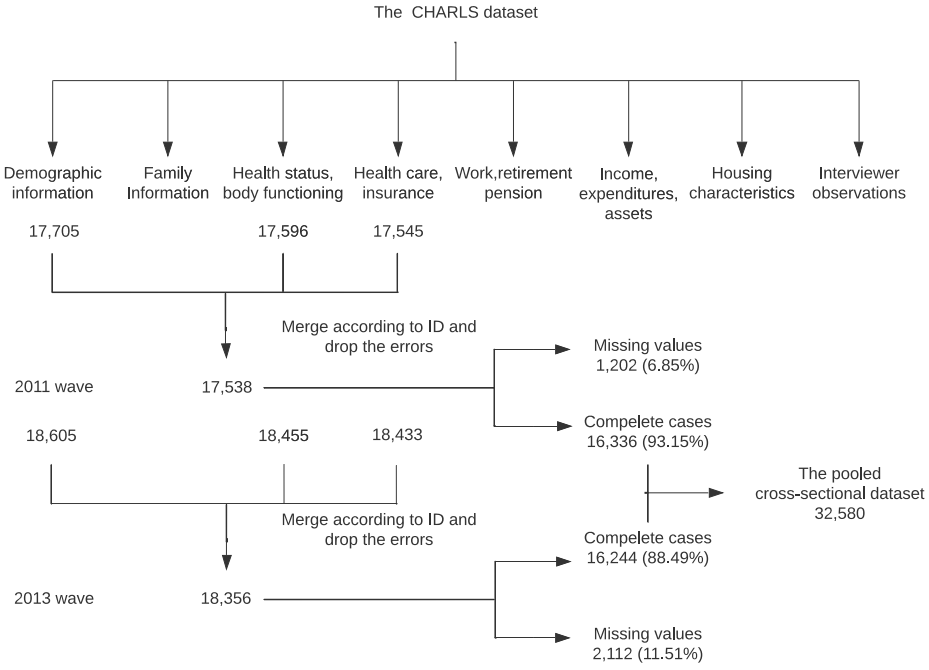


Figure 3.3: Dataset Construction

Notes: This figure shows the number of observations remaining in the sample prior to integration with the expenditure data. With initial counts of 17,538 observations for the 2011 wave and 18,356 for the 2013 wave, Figure 3.4 shows how the number of observations changes after merging the expenditure data.

Table 3.3: Incidence of Change in Insurance, 2011 and 2013

Insurance in 2011	Insurance in 2013			
	Employee	Resident	No insurance	Total
Employee	1,491	123	24	1,638
Resident	229	11,271	299	11,799
No insurance	65	633	193	891
Total	1,785	12,027	516	14,328

Respondents with multiple health insurance plans are only a small fraction of all respondents. In any case, to those with multiple insurances, I assigned the most generous health insurance.[4] Unfortunately, private insurances vary too much in benefits and premiums, and I was not able to obtain sufficiently detailed information on private health insurance plans. However, for the vast majority of the respondents, private health insurance does not appear to be the primary health insurance, and I did not include those who only had private medical (health) insurance, as well as those with medical aid, urban non-employed person's health insurance or other health insurance. Migrant workers may have one health insurance in their region of origin and another in the region where they work and reside. However, they are allowed to charge their claims only against one health insurance.

In several provinces, urban resident medical (health) insurance and rural medical (health) insurance have already been integrated into a single urban and rural resident medical (health) insurance, but before 2013, only Tianjing and Guangdong had integrated plans. In other provinces, these two insurances are not integrated, but they have similar benefits. Thus, I combine these three insurances into resident health insurance. I also combine employee medical (health) insurance and the government medical (health) insurance into employee health insurance for the same reason.

[4]Ranking of insurances by generosity from highest to lowest: government medical (health) insurance, urban employees (UEBMI), urban residents (URBMI), rural households (NRCMI), urban and rural resident medical (health) insurance, medical aid, private medical (health) insurance, urban non-employed person's health insurance and other medical (health) insurance.

3.3.3 Pension plan

Individual's pension plan is used in chapter 5. The employee pension plan comprises pension schemes from government, institutions, and basic pensions provided by firms. It is noted that the information on pension plans from the 2011 wave differs from that in later waves. The 2011 wave of the questionnaire does not include a direct question about coverage under specific pension plans; instead, it inquires whether a person is currently receiving payments from a certain type of pension plan. Consequently, information on pension plan coverage is not available for individuals who are still employed and have not yet begun to receive pension benefits. However, I can determine whether a person is covered by an employee pension plan through other questions included in the questionnaire.

In the 2013 and subsequent waves, the questionnaire lists 12 different types of pension plans, which include several supplementary pension options: government or institutional pension programs, basic corporate pensions, supplemental pension insurance from employers, rural pensions, residents' pensions, urban residents' pensions, new rural social pension insurance, endowment insurance for land-losing farmers, old-age pension allowances, commercial pensions, life insurance, and others, including an option for none.

From the 2013 wave onwards, having an employee pension plan is defined as being covered by at least one of the following: a government or institutional pension program, a basic corporate pension, or supplemental pension insurance from an employer. Resident pension plans comprise rural pension, residents' pension, urban residents' pension, new rural social pension insurance, endowment insurance for land-losing farmers, and old-age pension allowance. These plans are established by the State Council but managed by local governments, leading to variations in generosity across different cities and regions due to the differing financial situations of these governments. Private pension plans include commercial pensions, life insurance, and

others.

Although there are 12 different pension plans in the CHARLS dataset, individuals' pension plan is regrouped into only two categories: employee pension plan and resident or no pension plan. Ning et al. (2016a) found no evidence that the new rural social pension plan significantly induces the elderly to withdraw from the labour market. Furthermore, these resident pension plans are gradually established since 2011, and the replacement rates depend on the number of years of contribution. Due to the limited duration of coverage, the impact of these resident pension plans on individuals' retirement state remains limited (Giles et al., 2015; Ning et al., 2016a). Literature about the effectiveness of resident pension on individual's retirement decisions is limited. I propose that the policy is too new to observe the effect. Ning et al. (2016) studied the relationship between the rural pension plan and labour supply of the elderly, and found that having the rural pension plan is not a significant determinant of retirement for older adults in rural China. Cheng et al., 2017 studied the effect of pension income on elderly living arrangements[5] and found that this effect is economically small, and is only significant for a small group of people with certain demographic characteristics (see also chapter 5.).

3.3.4 Age, sex and objective measurement of health

Age and sex are used in all empirical chapters, and the objective measurement of health is only used in chapter 4. Respondents' age ranges from 45 to 84. These respondents are divided into 8 age groups with 5-year windows. Sex is captured by a dummy variable for females. I also consider respondents' objective health using two dummy variables: chronic disease[6] and disability[7].

[5]Living arrangement is measured by a discrete variable, which includes living in the extended household (i.e. with adult children or other adult members) and living independently.

[6]Chronic disease equals to 1 if the respondent has at least one type of chronic diseases from the fourteen types of chronic disease in the original questionnaire.

[7]Disability equals to 1 if the respondent has one of following: physical disabilities, brain damage or mental issues, vision problem, hearing problem and speech impediment.

3.3.5 Socio-economic variables

All three empirical chapters control for the socio-economic status of the respondents, including educational attainment, marital status, and *hukou* (household registration) status. The first two empirical chapters also control for indicators of income.

Income and expenditure

Unfortunately, the income data is incomplete in the CHARLS dataset. One variable that is expected to be correlated with personal income is per capita household consumption expenditures. Per capita household consumption expenditures are used in chapters 4 and 5. Notably, in these chapters, expenditure data has the advantage of being smoothed across the life cycle, unlike income data (Deaton, 1999; Lee et al., 2014; Zhang et al., 2017b). The age range in the sample extends from 45 to 85, indicating significant variation in income sources. For the working demographic, wage income predominates. In contrast, retired individuals experience varied income compositions, often marked by substantial reductions from pre-retirement levels. Consequently, in a sample encompassing both working and retired populations, expenditure data may provide a more reliable indicator of economic status than income data. However, the consumption expenditure data are not available for all households.

In CHARLS, the recall periods for household expenditure can be a week, a month, or a year, depending on the type of expenditure. There are three weekly expenditures, which include: consumption of food at home (purchased and self-produced); eating out expenditures; and alcohol and cigarettes.

There are seven types of monthly expenditures, including: fees for internet, telephone, and cell phone services; utility fees for water and electricity; gas, coal[8]; fees for housekeepers; local transportation; household and personal care products; and entertainment expenditures, which include expenses on books and newspapers.

[8]In China, households use gas for cooking. In rural areas, coal is used for heating.

There are 14 types of yearly expenditures in the original dataset, but I only include 12 of them. Two types of yearly expenditure are excluded for different reasons. The included yearly expenditures are: clothing; long-distance travel expenses; heating (in the city); furniture and consumption of durable goods, such as fridges and TVs; education and training; fitness; beauty; purchase, maintenance, and repair of vehicles (bikes), appliances, and communication products; taxes and other fees paid to the government; electronics (laptops, computers, and accessories); property management fees (including parking fees); donations. The excluded yearly expenditures are: medical expenditure and automobiles (cars). I exclude medical expenditure as it is endogenous. Expenditure on buying cars is also excluded since cars are expensive and can be used for several years.

I use the logarithm of per capita household annual expenditure, as it is the most frequently used indicator of expenditure in the published papers that used the CHARLS dataset (Zhang et al., 2017a; Tan et al., 2019). I constructed the logarithm of per capita household annual expenditure values based on the methods outlined in these studies.

Here is the the process of generating per capita household annual expenditure:

1. Annual total household expenditure = weekly expenditure × 52 + Monthly expenditure × 12 + Yearly expenditure

2. Per capita household annual expenditure = annual total household expenditure / number of household members[9]

3. Log form transformation.

Also, I will need to merge the logarithm of per capita household annual expenditure with the dataset that has been used in regressions. Figure 3.4 shows the process

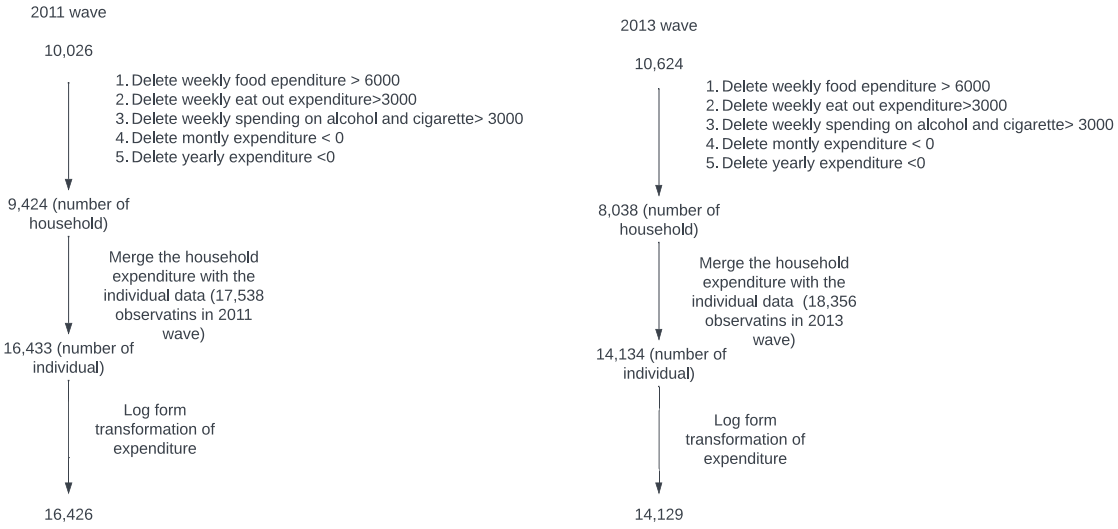
[9]Number of household members is defined as "the number of people usually ate meals together in the household in the past week, excluding guests".

and the change in the number of observations at each step. In CHARLS, household annual expenditure is collected at the household level, and all questions related to expenditure are answered by one representative from each household. In the CHARLS survey, both the main respondent and their spouse are interviewed to answer the questions on individuals, but only one person responds to the household expenditure question. Hence, the two individuals from the same household are those respondents and their spouses. But there are also individuals living in the same residence but are not married. In this case, these two individuals also come from the same household. Therefore, after deleting the outliers, I merge the expenditure data, which is at the household level, with all other variables, which are at the individual level. There is a probability that two individuals come from the same household. Hence, in the 2011 wave, when I merge the 9,424 household expenditure observations with the 17,538 individual observations, 16,433 observations are matched. In the 2013 wave, there are more households for which the expenditure data is missing. Using the same procedure, only 14,134 observations are matched, although there were 18,356 observations before merging with household expenditure data.

Outliers were excluded during the construction of the variable (Figure 3.4), where the cutoff criteria for dropping observations are based on the CHARLS questionnaire, aligning with the method used by Lei et al. (2014).

I use nominal annual expenditure data, and controls for economic regions do account for regional differences in prices. According to the National Bureau of Statistics in China (2014), consumer prices at the national level remained stable between 2011 and 2013, the period from which the expenditure data are drawn.

Figure 3.4: Change in the Number of Observations After Merging with the Original Dataset in Each Wave



Notes: these are two cross sectional datasets, which are not the datasets used in regressions. The cutoffs of dropping observations are based on the cutoffs provided in the CHARLS questionnaire. In the 2011 wave, after merging the household expenditure data by household ID with the initial 17,538 observations shown in Figure 3.3, the sample is reduced to 16,433 observations. Similarly, for the 2013 wave, following the same procedure, the number of observations decreases from 18,356 to 14,134.

Other social economic indicators

The survey has 11 levels for educational attainment: illiterate[10], illiterate but not completed primary school, home school *(sishu)*, elementary school, middle school, high school, vocational school, college, bachelor's degree, master's degree, and PhD. I regroup them into illiterate, literate but less than high school, and at least high school. For marital status, the survey has seven categories: married and living with the spouse, married but temporarily not living with the spouse, separated, divorced, widowed, never married, and cohabitating. I keep married and living with the spouse ("married, spouse present") and married but temporarily not living with the spouse ("married, spouse away") as separate categories, but combine those who are currently

[10]It means actually illiterate. Individuals who did not have formal schooling but were educated at home and can read and write were classified as attending "*Sishu*".

not married or without a partner within a single category, which I call "unmarried". In the CHARLS dataset, the proportion of each of these marital status categories is less than 1%, necessitating their combination.). *Hukou* status is measured by a dummy variable, which is 1 if the respondent's current *hukou* status is urban and 0 if the respondent's current *hukou* status is rural.

3.3.6 Employment variables

Chapters 5 and 6 use several variables related to work and retirement, including sector of employment, retirement status, and the number of missed work days. Individuals are defined as working in the agricultural sector if they engage in agricultural work (including farming, forestry, fishing, and husbandry for own family or others) for more than 10 days in the past year. They are defined as working in the non-agricultural sector if they worked at least one hour outside of the agricultural sector in the past week. Jobs in the non-agricultural sector include any wage-earning job, self-employment, and unpaid family business work. However, it does not include unpaid housework and unpaid voluntary work. I adopt the definition of "working" as specified in the CHARLS questionnaire. This definition is also consistent with those used in other countries, such as in Canada's Labour Force Survey.

Then, I classify individuals as working solely in the agricultural sector if they exclusively engage in agricultural work, working solely in the non-agricultural sector if they exclusively engage in non-agricultural work, and working in both sectors if they engage in both. In all remaining chapters, unless otherwise stated, "employed in the agricultural sector" means that an individual is exclusively employed in the agricultural sector, and the same applies to the non-agricultural sector. Thus, the sector of employment sector has three categories: agricultural sector, non-agricultural sector and both sectors.

Individuals are defined as "working" if they are employed in any of the three

sectors, and they are defined as "not working" or "retired"[11] if they are not currently employed but have worked in any of those three sectors for at least three months in their lifetime. Individuals who worked less than three months in their life are classified as "never worked", and are excluded from the dataset in chapters 5 and 6. According to these definitions, I construct an indicator for retirement status, which is 1 if individuals are retired, and 0 if they are working.

The variable "number of missed work days" is only used in chapter 6. In the questionnaire, respondents are asked about the number of days they missed work due to health problems in the past year, so the "number of missed work days" is considered as a continuous variable. For individuals who work in both sectors, the "number of missed work days" is determined by the greater number of missed work days from the two sectors. Only the working population (working = 1) reported this variable.

3.3.7 Life-style variables

Life-style factors include smoking and drinking, and are included only in chapter 4. Smoking is a dummy variable, which takes a value of 1 if the respondent has ever smoked. The reason that I do not consider the frequency of smoking is that among those who have ever smoked, about 80 percent kept this habit and among those who kept this habit, about 70 percent are heavy smokers[12]. Drinking is also measured by a dummy variable, which takes a value of 1 if the respondent drinks more than once a month, and 0 otherwise.

[11] The term "retired" or "retirement" used here should be distinguished from the normal meaning of retirement in the Chinese context. The latter typically refers to retirement from government, enterprises, and institutions.

[12] Heavy smoker is defined as a person who smokes more than 10 cigarettes per day.

3.3.8 Geographical factor

The geographical factor I employ in all three chapters is economic region. I take economic region into consideration because economic development varies in the four economic regions in China. These four economic regions are East, Middle, Northeast and West. The east is the wealthiest region among the four, while the West is the most deprived region. The Middle and the West are similar in terms of their economic conditions. Wealthier regions have more fiscal and medical resources, while poorer regions receive relatively more subsidies for health expenditures from the central government. Insurance patterns are similar for the East, Middle and West regions. The northeast has a larger fraction of respondents who are insured by the employee insurance but it also has a larger fraction of respondents who do not have any insurance. The distribution of respondents among the regions is quite even except for the Northeast. Only about 7% respondents are from the Northeast, but there are only three provinces in the Northeast (Table 3.4).

Table 3.4: Provinces, Autonomous Regions, and Municipalities by Economic Regions

ID	Name
East	
03	Fujian
06	Hebei
09	Beijing
11	Jiangsu
15	Guangdong
18	Shanghai
19	Tianjin
20	Zhejiang
29	Shandong
Middle	
07	Jiangxi
17	Shanxi1
26	Anhui
27	Hubei
32	Henan
33	Hunan
Northeast	
14	Heilongjiang
16	Liaoning
21	Jilin
West	
01	Yunnan
04	Qinghai
05	Sichuan
08	Xinjiang Uygur Autonomous Region
10	Inner Mongolia
12	Chongqing
13	Gansu
24	Guangxi Zhuang Autonomous Region
28	Shanxi3
34	Guizhou

Source: China Health and Retirement Longitudinal Study, 2011-2012 national baseline users' guide.
Notes: 1 and 3 represent tones.

3.3.9 Information on children

In chapter 5, I include information on the children of respondents, including the number of children alive, the highest education level of children, and an indicator of having any child in school. The number of children alive is defined as the total number of children alive, including biological children, children from current spouse if re-married, and adopted children. This variable is grouped into a categorical variable which has 6 values. It takes value 1 if the respondent does not have any child, 2 if the respondent only has 1 child, 3 if the respondent has 2 children, 4 if the respondent has 3 children, 5 if the respondent has 4 children, and 6 if respondent's total number of alive children is equal to or larger than 5. The highest education level of a respondent's children is controlled because transfers from children is the main source of income for the elderly rural population in China (Giles et al., 2023). Children's education level can be used as an indicator of children's social economics status (SES). It is defined as the education level of the child with the highest education level of a respondent. Children's education has 11 levels for educational attainment in the survey. I also regroup them into three categories: less than middle school diploma, high school diploma, and above high school, since the overall education level has improved in China over time. The reason of including children's education level is because transfers from children is the main source of income for the elderly rural population in China (Giles et al., 2023). Children's education level can be used as an indicator of children's SES.

Having children in school is included as a control variable, as compulsory public education in China is not free; parents are required to pay tuition fees for children in school up to university though parents typically continue to bear the financial responsibility for university tuition at both undergraduate and postgraduate levels. Overall, having children in school represents an additional financial burden that may influence an individual's retirement decision. It is measured as a dummy variable,

which takes a value of 1 of the respondent has at least one child in school (all levels of education), and 0 otherwise. The reason of including having children in school variable is that in China, compulsory public schools are not free. Parents must pay tuition fees if they have any children in school before university. For undergraduate and postgraduate studies, due to cultural expectations, university tuition is typically the financial responsibility of parents (Johnstone and Shroff-Mehta, 2003; Yang, 2015; Schwartzman et al., 2015). Overall, having children in school represents an additional financial burden that may influence an individual's retirement decision.

3.3.10 Living standards

Living standards include living standard and relative living standard. Both are self reported variables. In the survey, respondents are asked to rate their own standard of living, which has 5 ranks: very high, relatively high, average, relatively poor and poor. I combine them into 3 categories. Very high and relatively high are combined into high, and relatively poor and poor are combined into poor. This variable is only available in the 2011 wave. In the 2013 wave, respondents are instead asked for their living standard compared to the average living standard of people in their city or county. I define this question as relative living standard. In the survey, it has 6 values: much better, a little better, about the same, a little worse, much worse and do not know. I regroup them into 4 categories. Much better and a little better are combined into better. A little worse and much worse are combined into worse. Each standard of living variable is available in only one of the CHARLS waves, so I do not include these two variables in the regression analysis. These variables can nevertheless facilitate a comparison of living standards across individuals with different types of health insurance plans. Therefore, they are included in the summary statistics section to provide context.

3.4 Summary statistics

In section 3.4.1, I report the summary statistics for the variables mentioned in section 3.3 by health insurance plan in order to get profiles for the individuals holding different health insurance plan. Similarly, in section 3.4.2, I report the summary statistics for the same variables by pension plan. Note that individuals' health insurance plan is highly correlated with their pension plan (Table 3.5). In 2013, among the individuals covered by the employee pension plan, about 78% had employee health insurance and 20% had resident health insurance. The rest of the respondents were more likely to have the resident health insurance.

These tables do not include the variable "number of missed work days", since only the working population reported this variable. The summary statistics for "number of missed work days" will be reported in chapter 6. Also, the summary statistics for expenditure are reported separately in section 3.5.1 as the sample size differs substantially for these analyses.

Table 3.5: Distribution of Pension and Health Insurance Plans, 2013

Pension plan	Health insurance plan			N (%)
	Employee	Resident	No health insurance	
Employee	2,212	583	54	2,849
	(77.64)	(20.46)	(1.90)	(100.00)
Resident	79	10,737	254	11,070
	(0.71)	(96.99)	(2.29)	(100.00)
Other	92	302	19	413
	(22.28)	(73.12)	(4.60)	(100.00)
No pension	217	2,954	401	3,572
	(6.08)	(82.70)	(11.23)	(100.00)
Total	2,600	14,576	728	17,904
	(14.52)	(81.41)	(4.07)	(100.00)

Notes: This table shows the correlation between pension plans and health insurance plans. Percentages in parentheses indicate the fraction of individuals with each type of health insurance plan, conditional on their specific pension plan.
Source: Author's calculations based on CHARLS 2013 wave.

3.4.1 Summary statistics of respondents by different health insurance plans

Individual characteristics vary for respondents with different health insurance plans. Those with employee health insurance have the largest fraction of males, and those without any health insurance have the largest fraction of females. Those with the employee health insurance have the best self-ranked health state, while the ones with resident health insurance have the worst. Although those with the employee health insurance have the best self-ranked health state, they have the largest proportion of respondents with a chronic disease, despite the fact that the disability rate for this group is the lowest. Those without any health insurance have a lower fraction of respondents who have a chronic disease, compared to those with resident health insurance. Overall, the individuals without any health insurance are healthier than the ones with resident health insurance in terms of both subjective and objective measurement of health. Since the resident health insurance is optional, the belief of being healthy might be a reason why the individuals without any health insurance do not opt in to the resident health insurance. In addition, contrary to the evidence in certain high-income countries, those with employee health insurance, who perceive themselves as the healthiest, have the largest fraction of respondents who drink regularly and used to smoke. Those without any health insurance have the lowest drinking and smoking rates. These data suggest the potential presence of selection bias. I will discuss possible selection bias in detail in section 3.4.1.

The overall education level is the highest for those with the employee health insurance, with 43% of the respondents having a high school degree or above. The overall education level is slightly better for those without any health insurance than those with resident health insurance. With 19% of the respondents being unmarried at the time of the survey, the group without any health insurance has a significantly larger fraction of unmarried respondents compared to the other two groups. The

respondents with the employee health insurance are most likely to work only in the non-agricultural sector or be retired. Those without any health insurance are more likely to work in the non-agricultural sector compared to those with the resident health insurance.

Having an employee pension is highly correlated with having employee health insurance. Eighty-eight percent of the respondents who have the employee health insurance also have the employee pension, while only 4% of the respondents with the resident health insurance have the employee pension plan. About 8% of the respondents who do not have any health insurance have the employee pension plan. *Hukou* status is also highly correlated with the health insurance plan. Nighty-two percent of the respondents with the employee health insurance have urban *hukou*, while only 9% of the respondents with resident health insurance have urban *hukou*. About 70% of the respondents who do not have any health insurance have urban *hukou*. The distribution of age is similar among the three groups, but the group with the employee health insurance has a larger fraction of respondents who are aged between 45 and 49. Fewer respondents with the employee health insurance reside in the Western region. The respondents with employee health insurance also have fewer children, but the overall education level is higher compared to the other two groups.

The living standard is the best for those with the employee health insurance, regardless of whether the absolute or relative living standard is considered. The relative living standard is defined as the living standard compared to the other respondents in the same county or village. Compared to those with the resident health insurance, the respondents without any health insurance have a worse living standard in terms of the absolute living standard, but their relative living standard is better.

In summary, compared to individuals with the resident health insurance plan, those with the employee health insurance plan are more likely to be male and to have the employee pension plan. They are also more likely to report better health, despite

being more likely to drink regularly and having a history of smoking. Additionally, they have a higher level of education, are more likely to be married and live with their spouse, have the urban *hukou*, work in the non-agricultural sector or be retired, are older, and are more likely to reside in the Eastern region. They have fewer children, but their children's highest education levels are higher, and they report better living standards.

In contrast to individuals with the resident health insurance plan, those without any public health insurance plan are more likely to be female but are also more likely to have an employee pension plan. They report slightly better health and are less likely to report having any chronic diseases. Additionally, they are slightly less likely to drink regularly and to have a history of smoking. They have a slightly higher level of education, are more likely to be unmarried in terms of their current marital status, and are more likely to have the urban *hukou*, work in the non-agricultural sector, and be retired. Regarding living standards, they exhibit larger variations in absolute living standards, with a higher percentage of respondents reporting both high and low standards. In terms of relative living standards, they have a larger percentage of respondents reporting better standards and a slightly smaller percentage reporting worse standards, but there is also a larger percentage of respondents who do not have an idea.

Overall, the profiles of individuals with different health insurance plans vary, especially for those with the employee health insurance plan and those with the resident health insurance plan. The socio-economic characteristics that differ significantly between individuals with resident health insurance and those without any insurance include sex, marital status, *hukou* status, and working sectors. Compared to individuals with the resident health insurance plan, those without any health insurance are healthier in terms of both subjective and objective health measurements, and they also have healthier lifestyles. However, since resident health insurance is voluntary,

the better health outcomes of individuals without any health insurance imply the possibility of selection bias.

These tables present summary statistics that align with other studies on China using the same dataset. In China, individuals with more generous employee health insurance plans show a higher prevalence of chronic disease and unhealthy habits, such as smoking and drinking, yet still report better overall health. This suggests that comprehensive health insurance and self-reported health are associated in complex ways status.

Table 3.6: Summary Statistics by Health Insurance

Variable	Employee health insurance N = 3,855		Resident health insurance N = 25,052		No health insurance N = 1,446	
	N	%	N	%	N	%
Gender						
Female	1,584	41.1	13,468	53.8	823	56.9
Male	2,271	58.9	11,584	46.2	623	43.1
Health status						
Good	1,075	27.9	5,498	22.0	360	24.9
Fair	2,166	56.2	12,468	49.8	697	48.2
Poor	614	15.9	7,086	28.3	389	26.9
Chronic disease						
Yes	2,834	73.5	17,882	71.4	897	62.0
No	1,021	26.5	7,170	28.6	549	38.0
Disability						
Yes	427	11.1	4,530	18.1	265	18.3
No	3,428	88.9	20,522	81.9	1181	81.7
Regular drinker						
Yes	1,119	29.0	6,433	25.7	313	21.7
No	2,736	71.0	18,619	74.3	1,133	78.4
Smoked						
Yes	1,711	44.4	10,286	41.1	572	39.6
No	2,144	55.6	14,766	58.9	874	60.4
Educational attainment						
Illiterate	190	4.9	7,601	30.3	459	31.7
Literate, less than high school	2,009	52.1	15,754	62.9	845	58.4
High school and above	1,656	43.0	1,697	6.8	142	9.8
Marital status						
Married, spouse present	3,328	86.3	20,592	82.2	1,059	73.2
Married, spouse away	147	3.8	1,544	6.2	111	7.7
Unmarried	380	9.9	2,916	11.6	276	19.1
Sector of employment						
Only in agriculture	298	7.7	12,561	50.1	484	33.5
Only in non-agriculture	936	24.3	2,812	11.2	280	19.4
Both sectors	142	3.7	2,785	11.1	100	6.9
Used to work	2,466	64.0	6,482	25.9	521	36.0
Never worked	13	0.3	412	1.6	61	4.2
Employee pension						
Yes	3,375	87.6	1,000	4.0	118	8.2
No	480	12.5	24,052	96.0	1,328	91.8
Urban hukou						
Yes	3,553	92.2	2,178	8.7	1,002	69.3
No	302	7.8	22,874	91.3	444	30.7
Age group						
45 - 49	361	9.4	4,182	16.7	258	17.8
50 - 54	523	13.6	4,078	16.3	232	16.0

Continued on next page

Table 3.6 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)
55 - 59	723	18.8	5,197	20.7	291	20.1
60 - 64	745	19.3	4,663	18.6	233	16.1
65 - 69	584	15.2	3,026	12.1	134	9.3
70 - 74	469	12.2	2,012	8.0	126	8.7
75 - 79	306	7.9	1,260	5.0	117	8.1
80 - 84	144	3.7	634	2.5	55	3.8
Region						
East	1,289	33.4	7,663	30.6	414	28.6
Middle	1,058	27.4	7,451	29.7	366	25.3
Northeast	524	13.6	1,490	6.0	189	13.1
West	984	25.5	8,448	33.7	477	33.0
Number of children						
0	-	-	2	0.0	1	0.1
1	1,360	35.3	2,816	11.2	227	15.7
2	1,331	34.5	9,196	36.7	503	34.8
3	627	16.3	6,416	25.6	332	23.0
4	320	8.3	3,608	14.4	184	12.7
5 or more	217	5.6	3,014	12.0	199	13.8
Highest education level of children						
Less than high school	509	13.2	14,823	59.2	851	58.9
High school	542	14.1	2,783	11.1	202	14.0
Above high school	2,804	72.7	7,446	29.7	393	27.2
Living standard, 2011						
High	101	6.0	295	2.6	26	3.1
Average	1,021	60.6	6,034	52.7	359	42.3
Poor	562	33.4	5,129	44.8	463	54.6
Relative living standard, 2013						
Better	139	8.7	280	2.6	14	3.5
Same	675	42.3	2,469	23.1	75	18.9
Worse	658	41.2	7,265	67.9	262	66.2
Have no idea	124	7.8	694	6.5	45	11.4

Notes: "-" means that no observations or data are available. This table includes more variables compared to the pooled cross-sectional dataset in Figure 3.3. Thus, the total number of observations is less than 32,580 due to missing values. The results of the Pearson Chi-Square test show that the differences in distributions of all categorical variables across health insurance types are statistically significant at 1% significance level.
Source: Based on CHARLS 2011 and 2013 survey waves (pooled observations). See Figure 3.3 for the construction of the dataset.

Selection issue

To further explore the selection issue, I focus on the group without any health insurance in the 2011 wave to see whether those who report poor health are more likely to opt for any health insurance in the 2013 wave. Table 3.7 shows that in the "no health insurance" group in the 2011 wave, the percentage of individuals opting for the resident health insurance plan in the next wave is higher for those who report poor or very poor health. However, the percentage of individuals opting for the employee health insurance plan does not vary significantly across groups, as the employee health insurance plan is determined by individuals' employment status and is not voluntary. This implies that health state could be one of the factors influencing the decision to opt for the resident health insurance plan.

Table 3.7: Health Insurance Plan in 2013 by Health State in 2011 for the Respondents without Insurance in 2011

Variable	No insurance in 2011 N = 919		Enrolled in employee insurance in 2013 N = 69		Opt in resident insurance in 2013 N = 644		Remain in no insurance in 2013 N = 206	
	N	%	N	%	N	%	N	%
Health state in 2011								
Very good	70	7.62	5	7.2	46	65.7	19	27.1
Good	152	16.54	11	7.2	103	67.8	38	25.0
Fair	446	48.53	34	7.6	312	70.0	100	22.4
Poor	200	21.76	15	7.5	145	72.5	40	20.0
Very poor	51	5.55	4	7.8	38	75.0	9	17.6

Notes: % under the column for 2011 are computed by column. The rest are computed by row relative to 2011.
Source: Based on CHARLS 2011 and 2013 waves. See Figure 3.3 for the construction of the 2011–13 panel.

3.4.2 Summary statistics of respondents with different pension plans

Individual characteristics also vary for respondents with different pension plans. The group with the employee pension plan has a larger fraction of males. They also have a better self-ranked health state. Although the group with the employee pension has a better self-ranked health state, it has a larger proportion of respondents with a chronic disease, despite the fact that the disability rate is lower. The group with the

employee pension also has a higher percentage of individuals who drink regularly and have ever smoked. Their overall education level is also higher. They are also more likely to be married and to live with their spouse. In terms of working characteristics, those with the employee pension are more likely to work only in the non-agricultural sector or to be retired.

Having the employee pension is highly correlated with having the employee health insurance. Seventy-five of the respondents who have the employee pension also hold the employee health insurance, while only two percent of the respondents who have the resident pension or do not have any pension hold the employee health insurance. *Hukou* status is also highly correlated with pension plan. Nighty-five percent of the respondents with the employee pension plan have urban *hukou*, while only nine percent of the respondents with the resident pension plan or do not have any pension plan have urban *hukou*. The respondents who either have the resident pension or have no pension are younger, with a larger fraction of respondents who are aged between 45 and 49. A smaller percentage of respondents with the employee pension plan reside in the western region. Those with the employee pension plan also have fewer children, but their children's highest education level is higher. The living standard is the best for those with employee pension, regardless of the absolute or relative living standard.

In summary, compared to individuals with the resident pension plan or those without any pension plan, individuals with the employee pension plan are more likely to be males and to have the employee health insurance. Although they report better health, they are also more likely to report chronic diseases, drink regularly, and have a history of smoking. They have a higher level of education, and are more likely to be married and live with their spouse, have an urban *hukou*, work in the non-agricultural sector or be retired. They are also older, are more likely to reside in the East, and have fewer children, but their children have higher levels of education. Additionally, they report a better living standard.

Moreover, the differences in profiles between individuals with the employee pension plan and individuals with the resident pension plan or without any pension plan are similar to the differences in profiles between individuals with the employee health insurance plan and individuals with the resident health insurance plan.

It is worth noting that some key variables are highly correlated. For example, males are more likely to have both employee health insurance and employee pension plans. The observed sex gap in self-ranked health status may partly result from these differences in health and pension plans, suggesting that males and females may need to be studied separately.

Table 3.8: Summary Statistics by Pension Plan

Variable	Employee pension plan N = 4,667		Resident or no pension plan N = 26,128	
	N	%	N	%
Sex				
Female	1,931	41.4	14,191	54.3
Male	2,736	58.6	11,937	45.7
Health status				
Good	1,310	28.1	5,751	22.0
Fair	2,555	54.8	12,982	49.7
Poor	802	17.2	7,395	28.3
Chronic disease				
Yes	3,474	74.4	18,437	70.6
No	1,193	25.6	7,691	29.4
Disability				
Yes	593	12.7	4,699	18.0
No	4,074	87.3	21,429	82.0
Health insurance plan				
Employee health insurance	3,375	75.1	480	1.9
Resident health insurance	1,000	22.3	24,052	93.0
No health insurance	118	2.6	1,328	5.1
Regular drinker				
Yes	1,349	28.9	6,605	25.3
No	3,318	71.1	19,523	74.7
Smoked				
Yes	2,103	45.1	10,623	40.7
No	2,564	54.9	15,505	59.3
Educational attainment				
Illiterate	1,726	37.0	8,079	30.9
Literate, less than high school	2,639	56.6	16,196	62.0
High school and above	1,726	37.0	1,853	7.1
Marital status				
Married, spouse present	4,013	86.0	21,302	81.5
Married, spouse away	168	3.6	1,659	6.4
Unmarried	486	10.4	3,167	12.1
Sector of employment				
Only in agriculture	443	9.5	13,002	49.8
Only in non-agriculture	1,068	22.9	3,043	11.7
Both sectors	233	5.0	2,809	10.8
Used to work	2,904	62.2	6,798	26.0
Never worked	19	0.4	476	1.8
Urban hukou				
Yes	3,974	85.2	2,414	9.2
No	693	14.9	23,714	90.8
Age group				
45 - 49	393	8.4	4,467	17.1

Continued on next page

Table 3.8 (*Continued*)

	(1)	(2)	(3)	(4)
50 - 54	591	12.7	4,305	16.5
55 - 59	854	18.3	5,433	20.8
60 - 64	932	20.0	4,779	18.3
65 - 69	732	15.7	3,060	11.7
70 - 74	595	12.8	2,068	7.9
75 - 79	380	8.1	1,349	5.2
80 - 84	190	4.1	667	2.6
Region				
East	1,572	33.7	7,936	30.4
Middle	1,228	26.3	7,734	29.6
Northeast	691	14.8	1,583	6.1
West	1,176	25.2	8,875	34.0
Number of children				
0	-	-	3	0.0
1	1,441	30.9	3,056	11.7
2	1,606	34.4	9,573	36.6
3	852	18.3	6,615	25.3
4	442	9.5	3,721	14.2
5 or more	326	7.0	3,160	12.1
Highest education level of children				
Less than high school	839	18.0	15,529	59.4
High school	683	14.6	2,898	11.1
Above high school	3,145	67.4	7,701	29.5
Living standard, 2011				
High	121	5.5	315	2.6
Average	1,326	60.5	6,213	51.8
Poor	745	34.0	5,473	45.6
Relative living standard, 2013				
Better	153	8.3	289	2.6
Same	724	39.3	2,542	23.1
Worse	829	45.0	7,445	67.6
Have no idea	138	7.5	739	6.7

Notes: "-" means that no observations or data are available. This table includes more variables compared to the pooled cross-sectional dataset in Figure 3.3. Thus, the total number of observations is less than 32,580 due to missing values. The results of the Pearson Chi-Square test show that the differences in distributions of all categorical variables across pension plan status are statistically significant at 1% significant level.

Source: Based on CHARLS 2011 and 2013 survey waves (pooled observations). See Figure 3.3 for the construction of the dataset.

3.5 Expenditure, income

Section 3.5.1 provides a detailed summary of the expenditure data, while section 3.5.2 explores the variations in income across different types of enterprises and residences at the macro level, employing data sourced from the National Bureau of Statistics of China. These sections collectively present a comprehensive overview of economic indicators, revealing differences in income associated with various health insurance plans.

3.5.1 Summary statistics of expenditure

Tables 3.9 and 3.10 provide summary statistics for per capita household annual expenditure and its logarithmic transformation across different cross-sectional waves. The data indicate that the average per capita household annual expenditure was higher in the 2013 wave compared to the 2011 wave.

Table 3.9: Summary Statistics of Per Capita Household Annual Expenditure (Yuan)

Wave	N	Mean	Std.dev.	Min	Max
2011 wave	16,291	8470.00	11335.63	36	249985.5
2013 wave	14,120	10147.79	13635.29	6	366680

Notes: In 2011, the exchange rate between the RMB and the US dollar was approximately 6.2 Yuan per 1 US dollar.

Table 3.10: Summary Statistics of the Logarithm of Per Capita Household Annual Expenditure (Yuan)

Wave	N	Mean	Std.dev.	Min	Max
2011 wave	16,291	8.64	0.88	3.38	12.43
2013 wave	14,120	8.84	0.86	1.79	12.81

Table 3.11 presents the per capita household annual expenditure by health insurance plan, demonstrating that individuals covered by employee health insurance have higher expenditures compared to those with resident health insurance. Additionally, the data show no statistically significant difference in the annual expenditures between

individuals with resident health insurance and those without any health insurance.

Table 3.11: Summary Statistics of Per Capita Household Annual Expenditure (Yuan) by Health Insurance Plan (Pooled data)

Health insurance plan	N	Mean	Median	Std.dev.	Min	Max
Employee	4,226	14,561	10,619	15,846.77	175	274,976.00
Resident	24,130	8,317	5,692	11,438.39	0	366,680.00
No health insurance	1,615	8,295	5,696	11,210.77	0	163,624.00

Notes: Results of a t-test show that the mean of per capita household annual expenditure is statistically different between individuals with employee health insurance and those with resident health insurance (p-vaule is 0.00). Results of a t-test show that the means are not statistically different between the resident health insurance group and the no health insurance group (p-value is 0.91).

Table 3.11 displays the per capita household annual expenditure by pension plan type. It reveals that individuals covered by the employee pension plan incur higher expenditures compared to those with other types of pension plans.

Table 3.12: Summary Statistics of Per Capita Household Annual Expenditure (Yuan) by Pension Plan (Pooled data)

Pension plan	N	Mean	Median	Std.dev.	Min	Max
Employee	4,965	13,653	9,785	16,493.81	0	366,680.00
No employee pension	25,411	8377.428	5,706	11316.33	0	366,680.00

Notes: Results of a t-test show that the mean of per capita household annual expenditure is statistically different between individuals with employee pension plan and those without (p-vaule is 0.00). The maximum values of the two groups are identical as they are top-censored.

Table 3.13 presents the per capita household annual expenditure across different groups of pension plans. The data illustrate that individuals with resident health insurance or those without any pension plan experience lower expenditures compared to those covered by an employee pension plan. Additionally, there is no statistically significant difference in the per capita household annual expenditure between those with employee pension plans and those with commercial or other private pension plans.

Table 3.13: Summary Statistics of Per Capita Household Annual Expenditure (Yuan) by Pension Plan (Four Types, 2013 Wave)

Health insurance plan	N	Mean	Std.dev.	Min	Max
Employee	2,388	14,059	16,795.53	0	366,680.00
Resident	8,470	8,809	12072.55	0	366,680.00
Commercial or other	352	13,561	13,296.54	300	114,866.70
No pension	2,792	10293.27	13,647.52	0	307,935.00

Notes: The employee pension plan encompasses various programs such as the pension program of the government or institutions, the basic pension provided by firms, and the firms' supplemental pension insurance. The resident pension plan includes the rural pension, residents' pension, urban residents' pension, new rural social pension insurance, and endowment insurance for land-losing farmers.Results of a t-test indicate that the mean per capita household annual expenditure is not statistically different between individuals with an employee pension plan and those with commercial or other pension plans (p-value is 0.92). In contrast, the same tests show that the mean per capita household annual expenditure of individuals with an employee pension plan is statistically different from that of the other two categories (p-values are 0.00).

3.5.2 Income at macro level

Although the income data in the CHARLS dataset are incomplete, I present income and wage data at the macro level to provide a broader overview of differences in income across employee types and regions.

First, there is an income gap among urban residents based on enterprise type, but the income gap varies depending on region. Figure 3.5 compares the average wages by enterprise types in four provinces or municipalities. Each province/municipality comes from an economic region. The state-owned, urban collectives, and cooperatives offer employee pensions, while other enterprises may not. Apart from those three, other enterprises also offer employee pension plans, though the benefits are lower. The three enterprise types mentioned above offer better pension plans and have higher average wages in the representative provinces in the Middle, Northeastern, and Western regions. However, in the Eastern region, the average wage in those enterprises that do not offer pension plans or offers less generous pension plans is higher. The figure illustrates wage variation across different jobs within a single region.

Figure 3.6 compares the annual per capita income of rural residents with that

of urban residents, and the difference also depends on region. Income is the highest in the Eastern region and the lowest in the Western region for both urban and rural residents. Meanwhile, the difference between the per capita income of rural residents and the per capita income of urban residents is the largest in the western region, with rural residents earning only 29% of what urban residents earn. Thus, rural residents residing in the western region are the most economically deprived group.

In summary, jobs that offer more generous health insurance and pension plans do not necessarily offer higher wages; it depends on the region. For example, in wealthier regions, incomes are higher in foreign-funded enterprises, but they may offer less generous health insurance and pension plans. However, in economically deprived regions, wages are also higher for jobs that provide health insurance and pension plans.

3.5.3 Discussion

The summary statistics suggest that individuals with employee health insurance have higher household expenditures per capita compared to those with resident health insurance, and a similar pattern is observed for pension status. However, these statistics need to be interpreted with caution. The current expenditure, or even income, if it is available, might not reflect their income level during their working years.

In the working years of the survey cohorts, the jobs that offer employee health insurance do not necessarily offer higher income. When employee health insurance was introduced in 1998, and even in the years afterwards, China experienced significant economic development. Many people working in private companies or who were self-employed during these years of rapid economic growth earned far more than those working for the government or in state-owned companies (Liu, 1998; Bramall, 2001; Bian and Zhang, 2002; Meng, 2004; Nee and Cao, 2005).

Unfortunately, the income data from the 1990s is not available from the Bureau

of Statistics in China. The expenditure data does not necessarily reflect the real economic condition of the respondents in their working years, because the average age in the sample is 60, with a large proportion of retired individuals. The individuals with employee health insurance, most likely, also have employee pensions, which are their main source of income post-retirement. In contrast, the individuals with resident health insurance typically have either no pension or have a pension that provides minimal income. While expenditure is somewhat smoothed, there were not many financial tools available in the 1990s to smooth consumption (Meng, 2003; Boyreau-Debray and Wei, 2004; Modigliani and Cao, 2004; Lardy, 2016).

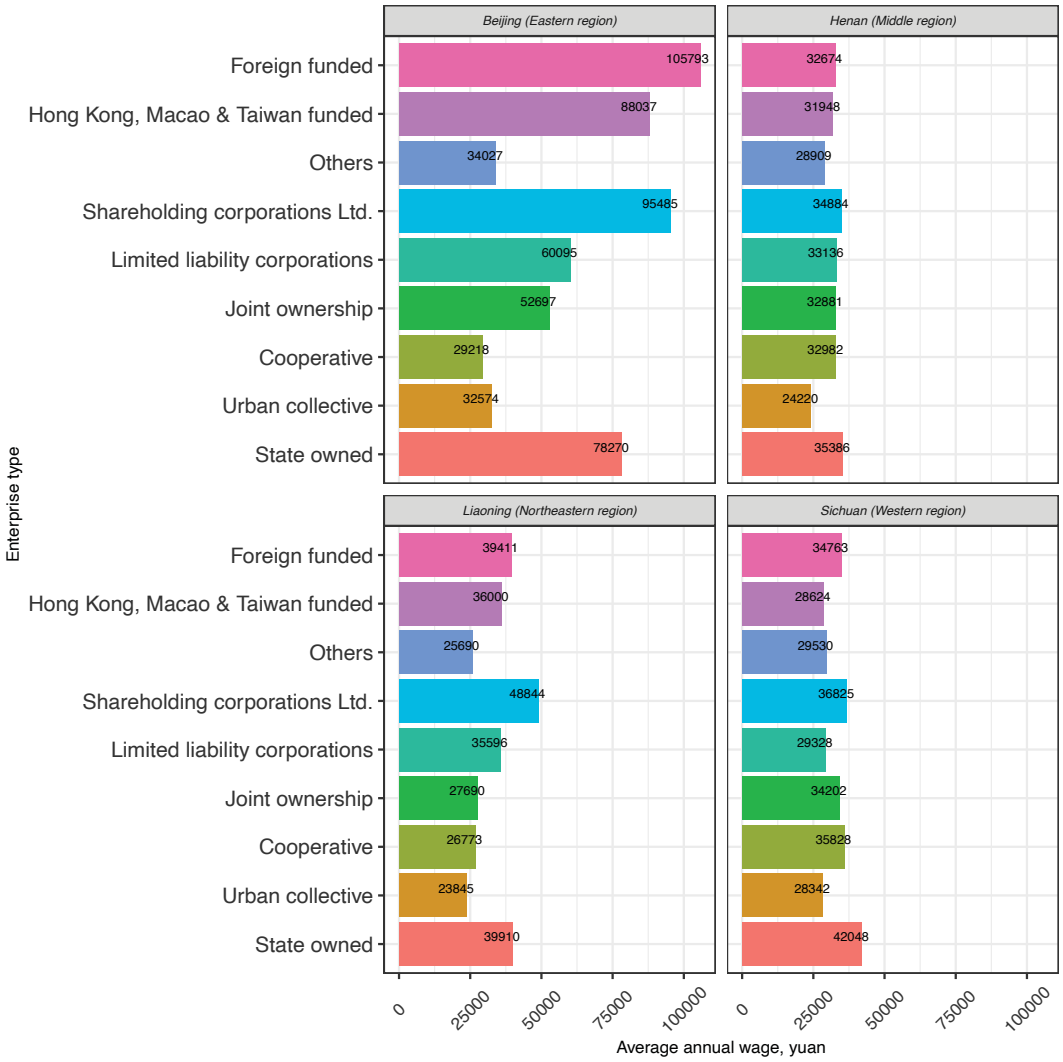


Figure 3.5: Average Wage by Enterprise Type, 2011

Notes: This figure shows the average annual wage by the enterprise type in four provinces/municipalities in 2013. Each province or municipality is located in one economic region. The figure illustrates wage variation across jobs within a single region, the difference in absolute expense or wages across regions is not the main focus.
Source: National Bureau of Statistics of China.

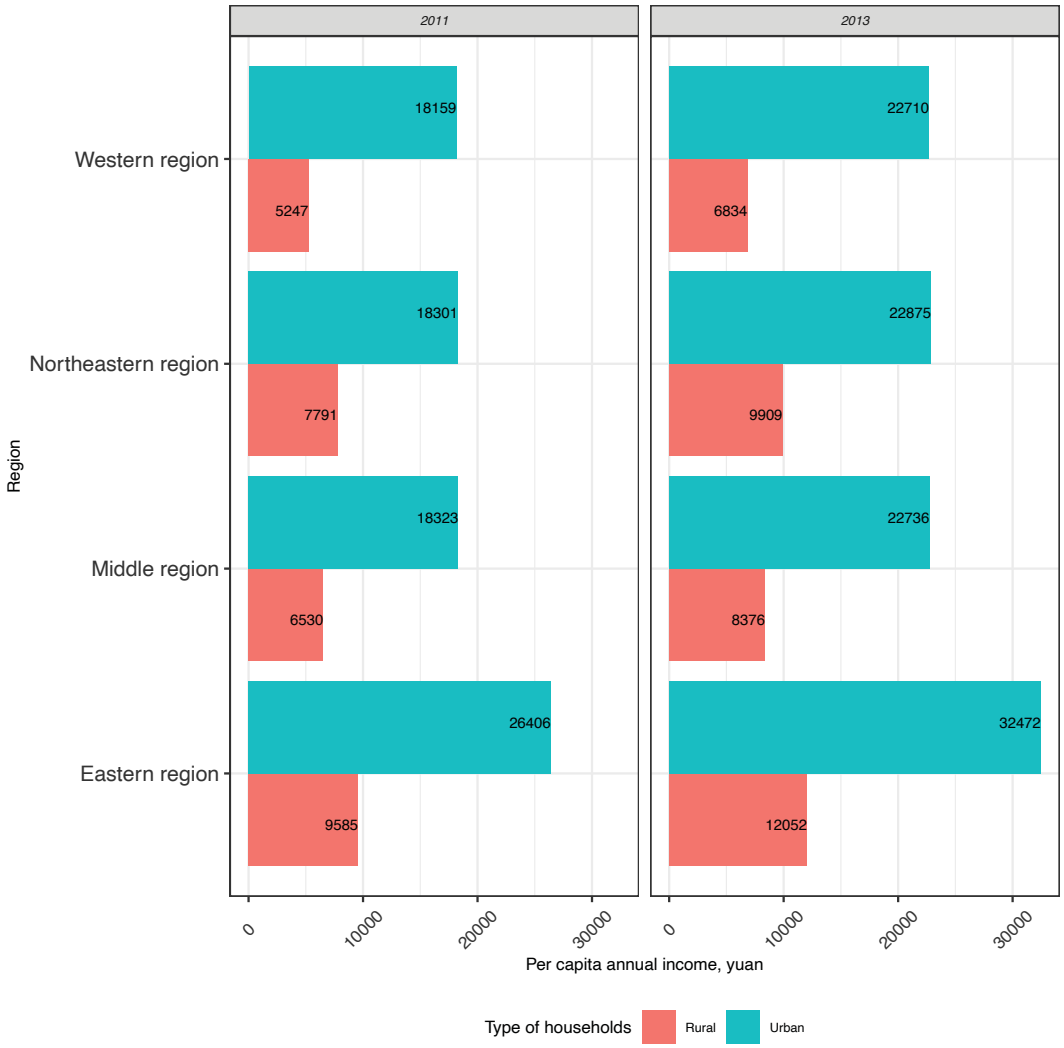


Figure 3.6: Rural-urban Income Gap by Region

Notes: Per capita income is net income for rural residents and disposable income for urban residents. For rural residents, only net income data is available from the Bureau of Statistics of China. However, net and disposable incomes are comparable for rural residents, as they receive minimal government transfers.
Source: National Bureau of Statistics of China.

Chapter 4

Health Insurance and Health Status

4.1 Introduction

Since 2011, nearly 95% of the Chinese population has been insured by some type of health insurance (see chapter 2.1). Despite the success in achieving widespread coverage, China's health insurance system remains segmented with varying degrees of benefits across health insurance plans and financial protection afforded to those under different plans. These health insurance plans are individually effective, either in improving health service utilisation or improving health outcomes of the insured (Liu et al., 2002; Chen and Jin, 2012; Cheng et al., 2015; Sun and Lyu, 2020). However, these health insurance plans target distinct groups and vary significantly in terms of premiums and benefits, potentially leading to disparities in health outcomes across their respective target populations. Moreover, there are significant variations in benefits across regions (see chapter 2.1). These differences in health insurance benefits across plans and regions have attracted the attention of researchers.

In the existing body of literature, it has been shown that there are large gaps in health service usage and health outcomes across different health insurance plans (Meng et al., 2015; Liu et al., 2015; Zhang et al., 2017b; Fan et al., 2019, 2020). Furthermore, among individuals covered by the same insurance plan, disparities in medical service usage and out-of-pocket cost still exist, and these disparities are associated with individuals' demographic characteristics (Zhang et al., 2017b; Ying et al., 2020). However, these studies focused on comparing the urban and rural resident

health insurance plans, which were sequentially integrated in various provinces between 2014 and 2020. Even before the integration, these two resident health insurance plans were already similar in terms of premiums and benefits. Since 2020, as a result of the integration, China has two main health insurance plans: the urban employee health insurance plan and the urban-rural resident health insurance plan. However, there has been no systematic comparison of health outcomes of those enrolled in the current more generous public employee health insurance and the less generous public resident insurance plans.

In this chapter, I assess the self-ranked (also referred to as self-assessed or self-reported in the literature) health state outcome gaps that can be attributed to the type of the two main public health insurance plans. Specifically, I aim to answer the following questions: 1) Is there a self-ranked health insurance gap between individuals covered by an employee health insurance plan and those covered by a resident health insurance plan? Do those with the more generous employee health insurance plan report better health states? 2) If a self-ranked health gap exists, does it widen or narrow over time?

I use the CHARLS dataset and an ordered probit model to answer both questions. Specifically, to answer the first question, I estimate the association between the health state ranked by survey respondents and their health insurance plan, while controlling for observable socio-economic characteristics. Subsequently, to answer the second question, I estimate age-specific dynamic transition matrices for self-ranked health state by health insurance plan.

I find that individuals with less generous resident health insurance plans have a disadvantage not only in their current health state but also in transitions between different health states. Additionally, individuals residing in less developed regions encounter further disadvantages regarding both their current health states and health transitions. My work provides a systematic comparison of self-ranked health states

under the current most popular public health insurance plans and across four major economic regions. It advances the understanding of the impact of health insurance by focusing on comparing individuals' health states across employee health insurance and resident health insurance, and estimating dynamic health transitions while controlling for regional disparities. An advantage of employing longitudinal data [1] and dynamic transition matrices is the ability to infer the association between type of health insurance plan and health states through observed differences in changes in self-ranked health states over time. Thus, my estimates remain unaffected by potentially systematic variations in self-ranked health states due to residence or occupation-based pooling inherent in the design of health insurance plans[2]. While dynamic health transition matrices have been used extensively in a variety of contexts (Holtz-Eakin, 1993; Jung, 2006; Molla and Madans, 2008), here I use them to evaluate the association between a range of health insurance designs and self-ranked health states.

My findings illuminate one of the potential sources of inequality in health outcomes, identifying significant disparities in health states among individuals covered by different health insurance plans. The estimation results show that individuals covered under the resident health insurance with lower benefits are significantly less likely to report good health. Furthermore, even if they report good health, it tends to be less persistent over time. For example, controlling for certain socio-economic characteristics, males aged 45 to 49 with employee health insurance plans are, on average, six percentage points more likely to report good health compared to their counterparts with resident health insurance plans. Specifically, the probability of reporting good health is 59% for those with employee health insurance compared to 53% for those with resident health insurance. In terms of transitions between different health

[1] Movers are not tracked and represent only approximately 3 percent of total respondents.

[2] Reporting heterogeneity in self-ranked health is present in the CHARLS dataset and similar datasets internationally (Etilé and Milcent, 2006; Tubeuf et al., 2008; Ziebarth, 2010; Schneider et al., 2012; Pfarr et al., 2012; Mu, 2014; Molina, 2016; O'Doherty et al., 2017; Baidin et al., 2021). Section 4.2 presents a comprehensive discussion of reporting heterogeneity.

states, individuals with resident insurance are associated with a lower likelihood of remaining in good health if they were initially healthy, and a higher likelihood of transitioning into fair and poor health. Specifically, for males who were initially in good health and living in the Eastern region, the probability of remaining in good health is about 65% if they have employee health insurance, compared to 54% for those with resident health insurance. Additionally, for those with employee health insurance, the probabilities of transitioning from good health to fair and poor health are 33% and 2%, respectively. In contrast, for those with resident health insurance, these probabilities are higher at 42% and 4%, respectively.

In addition to the statistically significant association of health insurance with current health and health transitions, I also observe significant differences based on sex and region. I find females are at a disadvantage in self-ranked health while controlling for other socio-economic characteristics. For example, among individuals with employee health insurance, while controlling for other characteristics, the probability of reporting good health is 55% for females compared to 59% for males. Moreover, even after controlling for health insurance types and region, females have significant higher vulnerability in transition probabilities. For example, in the Eastern region, healthy (reporting good health) females with employee health insurance have approximately a 6 percentage point lower likelihood of maintaining good health status compared to their male counterparts, with rates of 59% for females versus 65% for males. Meanwhile, females have a 5.5 percentage point higher probability of transitioning into fair health and a 0.5 percentage point higher probability of transitioning into poor health, respectively.

For the regional effect, my findings indicate that individuals living in lower-income regions are both less likely to report good health and more likely to transition into poor health. While controlling for other characteristics, compared to residing in the wealthier Eastern region, living in the poorer Western region, on average, is associated

with a six percentage point lower probability of reporting good health. In terms of health transitions, regardless of their health insurance plans, and for males who are initially in good health, those residing in the Western region have a nine percentage point lower probability of remaining in good health, compared to those who reside in the Eastern region. Additionally, those in the Western region have a seven percentage point higher probability of transitioning to fair health and a two percentage point higher probability of transitioning to poor health.

My findings suggest that even though China has achieved widespread coverage for health insurance, and the government has integrated the urban resident health insurance plan and the rural resident health insurance plan to mitigate the disparities across different health insurance plans, differences in health states between individuals covered by the employee health insurance plan and the resident health insurance plan remain. To mitigate health disparities among individuals with different types of health insurance and promote equity, further integration of health insurance plans may be necessary. In addition, the age-specific dynamic transition matrices I estimate in this chapter can be used to assess the general equilibrium implications of a potential health insurance plan reform with elevated benefits (Johansson (2000); Jung and Tran (2016); Lim (2016); Bairoliya et al. (2018); Frankovic and Kuhn (2023)).

4.2 Related Literature

There is a large body of literature that compares health insurance plans in China and studies the health equity among these plans. Health equity can be interpreted in two ways: one is "outcome-based" equality and the other one is "opportunity-based" equality (Liu et al., 2015). The "outcome-based" health equality focuses on mitigating the health gaps between different groups of people, for instance, making an effort to improve the health and well-being of vulnerable elderly people or those who are already in relatively poor health. In comparison, the "opportunity-based" equality

focuses on achieving equal benefits, equal health utilization of health resources and so on.

Most of the existing body of literature focuses on studying "opportunity-based" equality, by comparing utilization rates and out-of-pocket expenditures across different health insurance plans (Meng et al., 2015; Zhang et al., 2017b; Fan et al., 2020). There is limited research that specifically compares health outcomes across health insurance plans. Moreover, both the "opportunity-based" and "outcome-based" studies emphasize the comparison between the urban and rural resident health insurance, which have been already integrated. For example, using data from the China Health and Nutrition Survey, Fan et al. (2019) find that those covered by the urban resident health insurance plan tend to be in better health than those covered by the rural resident health insurance plan, and the association between the health state and type of health insurance plan is stronger for individuals living in with low incomes. There is one study that compares the outcome of elderly people across three health insurance plans, including the employee health insurance. Using a Nation Survey of the Aged Population in Urban/Rural China conducted in 2006 and 2010, as well as a Multiple Linear Regression, Liu et al. (2015) find that for individuals aged 60 years old or older, the employee health insurance (UEBMI) participants have better self-ranked health state, physical functions and psychological well-being than the urban resident (URBMI) health insurance and rural resident (NCMS) health insurance participants. Compared the study of Liu et al. (2015), I use a more up-to-date and nationally representative dataset, which includes a wider age group. I also employ an ordered probit model, which is more suitable for a discrete, self-ranked health variable.

Understanding inequalities in health service utilization under different health insurance plans helps to understand health state inequalities, as they are caused by the difference in generosity of these plans. Findings of the studies on health service utilisation are quite consistent. Individuals covered by more generous health insurance

plans have better access to health services (Li and Zhang, 2013; Tan et al., 2019; Lee et al., 2022). Integrating different health insurance plans is shown to be effective in mitigating the inequalities in health care utilisation, while controlling for the need of using health care services (Zhao et al., 2019). In the two provinces where the urban and rural resident health insurance plans were integrated before 2013 using the more generous urban health insurance plan as a reference, there was a substantial reduction in observed inequality in inpatient service utilization across urban and rural residents (Zhao et al., 2019). When urban and rural resident health insurance plans are integrated, individuals receive more benefits (Yang et al., 2018; Huang and Wu, 2020; Fan et al., 2021; Zhou et al., 2022).

Several studies highlight health and health insurance inequalities across regions (Tang et al., 2008; Zhang and Kanbur, 2009), as well as urban-rural disparities in health service utilization (Hu et al., 2008). Generally, rural residents face disadvantages in accessing healthcare. The variance in the generosity of health insurance plans significantly influences the disparities in health outcomes and access to health services between rural and urban residents.

I acknowledge the presence of reporting heterogeneity in self-ranked health, which stems from various socioeconomic and demographic characteristics. This suggests that identical objective health states may yield different self-ranked health outcomes due to these factors (Bago d'Uva et al., 2008; Shmueli, 2003). This issue is particularly pronounced in underserviced rural areas of China. For example, survey responses in these regions underestimate the prevalence of diabetes by 16% and hypertension by 38% (Ning et al., 2016b). Consequently, if individuals are unaware of their health conditions, such as these diseases, their self-estimation of health could be significantly skewed.

Although self-ranked health states have a disadvantage of reporting error, and are sensitive to various socioeconomic and demographic characteristics, they remain

the most widely used health indicator in empirical research. A substantial body of literature supports the reliability of self-ranked health as a comprehensive measure of overall health and a predictor of mortality, morbidity (Okun et al., 1984; McCallum et al., 1994; Idler and Kasl, 1995; Lundberg and Manderbacka, 1996), and medical service utilization (Wagstaff and Van Doorslaer, 2000). In contrast, a single health condition often fails to represent overall health. For instance, an individual without any reported health conditions may not necessarily be healthier than someone with controlled conditions (Lixin Cai, 2009). Moreover, a specific disease can impact individuals differently. For example, individuals with different characteristics have shown diverse responses to COVID-19 (Pradhan and Olsson, 2020; Pijls et al., 2021). As a result, certain health conditions may provide a narrower assessment of an individual's overall health compared to self-reported health.

Building on this body of evidence, most literature studying the relationship between health insurance and health outcome use self-ranked (self-rated or self-reported) health as an indicator of health outcome (Franks et al., 2003; Lei and Lin, 2009; Pan et al., 2016; Idler and Cartwright, 2018). Although the CHARLS dataset includes data on disease prevalence and biomarkers, there remains no consensus on how to integrate these metrics into a comprehensive measure of overall health (Culyer et al., 1971; Ware Jr et al., 1981; DeSalvo et al., 2006; Ye et al., 2019; Sokoya et al., 2022). Furthermore, many conditions like hypertension and diabetes are self-reported, particularly in areas with inadequate medical services, making these measures less objective.

The advantages of self-ranked health states are clear, although it has the disadvantage of reporting heterogeneity. Reporting heterogeneity can be mitigated by including objective health measures and accounting for health behaviors such as smoking and drinking (Pfarr et al., 2012). Additionally, while reporting heterogeneity exists, the variability in individual rankings due to unobservable characteristics is unlikely to change over time (Hernández-Quevedo et al., 2004; Benitez-Silva and Ni, 2008;

Mullins and White, 2019). For example, if an individual tends to overestimate their health in one year, they are likely to do the same in the following wave. Consequently, reporting heterogeneity does not pose a significant problem in estimating dynamic transitions across self-ranked health states when controlling for socioeconomic and demographic characteristics.

Regarding dynamic health transitions, the prevalent use of self-ranked health in research further underscores its significance. Self-ranked health states are commonly used to estimate these transitions. For example, Jung (2020) employ data from the Medical Expenditure Panel Survey (MEPS) and the Health and Retirement Study (Rand-HRS) to estimate health transitions, also leveraging self-ranked health states. Another example is Molla and Madans (2008), which use a different dataset. Using the same indicator with these studies facilitates international comparisons. Given these factors, self-ranked health remains a reliable indicator of an individual's overall health. Thus, it is appropriate and informative for the hypotheses I test to use self-ranked health states as the health outcome indicator in estimating individuals' current health and health transitions by health insurance plans.

Income can influence health through various channels such as nutrition, lifestyle choices, and access to healthcare services. Most studies examining the effects of health insurance on health outcomes control for income due to its significant impact (Smith, 1999; Pan et al., 2016; Chen et al., 2019). There is a substantial body of literature exploring the relationship between socioeconomic status (SES) and health, with income often employed as a key indicator of SES in these analyses (Liu et al., 2002; Wagstaff et al., 2009; Zhang et al., 2017a).In addition to the direct effects of SES on health, some research has identified that the impact of specific health insurance policies can vary across different SES groups. This heterogeneity suggests that policy effectiveness may differ based on socioeconomic status.

However, there are also reasons which state that income should not be included.

First of all, the income variable is endogenous. Although the aforementioned studies on health include income and find a correlation between the two, they acknowledge that this correlation alone should not be interpreted as causality. The reason is that a person who is in poor health also loses their full ability to earn income, and a person with low income may not receive sufficient treatment. Theoretically, the causality between health and SES can go either way. On one hand, low SES may cause poor health due to malnutrition, worse working environment, and less access to medical services. On the other hand, poor health can lead to low SES, since poor health reduces the working ability and increases medical expenses. Some researchers are interested in the direction of causality (Smith, 1999; Deaton, 2002; Meer et al., 2003; Fuchs, 2004; Smith, 2004; Frijters et al., 2005); the effect of income and wealth would be overestimated in a model that uses cross-sectional data and treats them as exogenous variables (Lixin Cai, 2009). Usually simultaneous equation models and instrumental variable models are utilized to deal with the endogeneity of income and wealth. However, the effect of income on health is not my main interest. I am interested in how health states change over time, instead of the mechanism of change in health. Thus, it is used as a control variable recognizing that no causal statements can be made.

Moreover, in the cohort I study, income was exogenous when the health insurance was introduced. Jobs that offer employee health insurance do not necessarily imply that they also offer higher income. When employee health insurance was introduced and even in the years afterwards, China experienced significant economic development. Many people working in private companies or self-employed during this boom earned far more than those working for the government or in state-owned companies (Liu, 1998; Bramall, 2001; Bian and Zhang, 2002; Meng, 2004; Nee and Cao, 2005). Another problem of the income data for this cohort is the current income does not reflect the real economic condition of the respondents in their work year, because the

average age in the sample is 60, with a large proportion of retired individuals. Those with employee health insurance likely also have employee pensions, which are their main source of post-retirement income. In contrast, individuals with resident health insurance may either not have a pension or have one that provides minimal income. Even though they may have earned more when they were younger, their income is significantly affected as they age. While expenditure is some-what smoothed, there were not many financial tools available in the 1990s to smooth consumption (Meng, 2003; Boyreau-Debray and Wei, 2004; Modigliani and Cao, 2004; Lardy, 2016).

Second, for the specific age group I study, the effect of income on an individual's health outcome is limited. Although lower income is associated with poorer health, this negative consequence of lower income affects young adults more, and fades away after approximately age 36 (Rözer and Volker, 2016). Similarly, Dalstra et al. (2006) find that in five out of 11 OECD countries in their study, income only has a modest effect on health when controlling for education for individuals aged from 60 to 79 years old. Similar effects are observed in House et al. (1994) and Pavalko and Caputo (2013). The reason that the effect of income on health diminishes as individuals age is attributed to various factors, including selective mortality of lower income individuals who may die younger. Another reason could be that although income can affect health in many ways, for middle aged and elderly people, getting access to healthcare is the channel by which income can affect health. Huguet et al. (2008) find that household income is significantly associated with health in the U.S. but not in Canada, highlighting the role of healthcare access in mitigating health inequality. Public provision of health insurance (a single payer system) is expected to break the link between ability to pay for health services and health outcomes. In those cases, whatever remains of the relationship running from income to health outcomes is mediated through primarily other social channels, such as low income households facing longer commute times to access health services or lower quality of health services

provided to low income households. In China, such differential access and quality issues are partly accounted for by urban versus rural hukou and residence in Eastern versus Western regions.

The fourth reason for not including income is that education, income and health insurance are highly correlated (Lixin Cai, 2009). Jin et al. (2016) find that both income and education have significant association with types of insurance coverage when they both are included in the distribution. Both income and education are included as categorical variables, and the effects of income and education are very similar. The significant effects are only observed in the highest income and education groups. The blending of indicators limits the value and policy applicability of the research (Fuchs, 2004). Many studies use education as an indicator of SES. Income question has many missing responses. I also use education as an indicator of SES.

In terms of the studies which estimated health transitions, some studies control for income (Rust, 1990; Carro and Traferri, 2009; Jung, 2020), while others do not (Briggs and Sculpher, 1998; Molla and Madans, 2008; Van Den Hout et al., 2009; Preston et al., 2014).

Overall, there are compelling reasons both for and against including income as a control variable when studying the effect of health insurance on health outcomes. On the one hand, income is a significant determinant of health, influencing factors such as access to healthcare, nutrition, and lifestyle choices, which suggests that it should be controlled for to isolate the specific effects of health insurance policies. On the other hand, income itself can be influenced by health status, potentially leading to endogeneity if included as a control. A discussion concerning the how I handle the indicator of income is elaborated in section 4.3.

4.3 Methods

Although estimating the dynamic transitions of health states is the main purpose of this study, it is worthwhile to observe the relationship between the type of health insurance and an individual's current health state. Here I first specify a model to test the hypothesis that a more generous health insurance plan is associated with better self-ranked health. For this purpose, I use two waves (2011, 2013)[3] of the CHARLS dataset and an ordered probit model to test whether those who have the employee health insurance have a larger probability of reporting a good health state, and have a smaller probability of reporting a poor health state. Then I specify a model of health transition matrices to test the hypothesis that type of health insurance plan affects the dynamic transitions of health states. In the model of estimating health transitions, I use the 2011 and 2013 waves of CHARLS datasets and the self-ranked health state in the 2015 wave.

In the model description, I adopt the notation and formulas outlined by Greene (2000). First, to test the hypothesis that an individual with the employee health insurance has higher probability of reporting a good health state, and has lower probability of reporting a poor health state, I consider a model of health state

$$y_i^* = X_i'\beta + \epsilon_i, \tag{4.1}$$

where y_i^* is a measure of individual i's health state, in latent form and X_i is a vector of determinants of health, including insurance plan. The dependent variable health state y_i^* and the covariants X_i are from the same time period. I test whether respondents' current self-ranked health state is correlated with the type of their health insurance plan. Since the effect of several variables on health can vary by sex, I estimate this

[3]See section 3.1 for the reason of not using more recent waves. The key constraint is the information on health insurance plan in the 2015 wave. However, several variables from the 2015 wave are used in the model for estimating health transition matrices.

model separately for males and females.

The self-ranked health state is measured on a scale of 1 to 3 for all models, with 1 being good and 3 being poor. In the models, the observed health state variable takes on values 1 to 3, depending on whether the latent health state lies between certain thresholds.

Thus, corresponding to equation (4.1), I transform y_i^* into self-ranked health state y_i as follows:

$$y_i = s \text{ if } \mu_{s-1} < y_i^* \leq \mu_s, \text{ for } s = 1, ..., S. \tag{4.2}$$

For the case of $S = 3$, I have

$$y_i = \begin{cases} 1 = \text{Good}, & -\infty < y_i^* < \mu_1, \\ 2 = \text{Fair}, & \mu_1 \leq y_i^* < \mu_2, \\ 3 = \text{Poor}, & \mu_2 \leq y_i^* < \infty, \end{cases} \tag{4.3}$$

where μ_1 is the threshold for being in good and fair health state, and μ_2 is that for being in fair and poor. These thresholds are estimated jointly with β. A positive coefficient estimate for β implies that an increase in the corresponding control variable X_i increases the probability of the highest-ordered health state (poor = 3).

To interpret the coefficients in the ordered probit model, I consider probabilities instead. Given X_i, the probability of being in each health state for individual i is

$$\begin{aligned} p(y_i = 1|X) &= F(\mu_1 - X_i'\beta), \\ p(y_i = 2|X) &= F(\mu_2 - X_i'\beta) - F(\mu_1 - X_i'\beta), \\ p(y_i = 3|X) &= 1 - F(\mu_2 - X_i'\beta), \end{aligned}$$

where F is the cumulative standard normal distribution function. For all probabilities to be positive, the thresholds need to satisfy $\mu_1 < \mu_2$. I estimate the probabilities

of being in each health state for respondents insured by employee health insurance, resident health insurance and those who are not insured. In addition, I report the marginal effects for each explanatory variable.

In the baseline model, X_i includes health insurance plan, age, marital status, drinking, smoking, chronic disease, disability, education, *hukou* status and region. The baseline model does not include a variable related to personal income or wealth.[4] I also estimate a model which excludes the indicators of objective measurement. Moreover, I allow for the possibility that the relationship between certain explanatory variables and self-ranked health state may have changed between 2011 and 2013. Thus, I also estimate this model separately for the 2011 and 2013 waves. Since the local health insurance budget also varies across provinces, I include province fixed effects instead of region fixed effects for the purpose of a sensitivity check. Income, despite being endogenous, could also be a determinant of health outcomes, given that health insurance plans in China allow for out-of-pocket expenses. However, in CHARLS, both personal income and wealth are poorly measured (see chapter 3). Per capita household consumption expenditures, which are expected to be correlated with personal income, can be a substitute for income. Thus, I also estimate this model by including per capita household annual expenditure as another sensitivity check. All models are estimated separately for males and females.

To test whether the type of health insurance plan is associated with the dynamic transitions of health states, I model the dynamic transitions of health state as

[4]The reasons are discussed in section 4.2. Nevertheless, given that health insurance plans in China allow for out-of-pocket expenses, it is not automatically possible to rule out income as a determinant of health outcomes. However, there is no satisfactory method to deal with the endogeneity of income when studying the effect of health insurance on health outcomes; the existing studies directly include income as control variables, except for the studies on the causal effects of income on individual health (Smith, 1999; Deaton, 2002; Meer et al., 2003; Fuchs, 2004; Smith, 2004; Frijters et al., 2005). Moreover, in CHARLS, both personal income and wealth are poorly measured (see chapter 3). One variable that is expected to be correlated with personal income is per capita household consumption expenditures, but they are not available for all households. Given all the conceptual and practical considerations, the baseline model does not include per capita household consumption expenditures. However, the section on sensivity analysis introduces per capita household consumption expenditures as an additional correlate of health state.

driven by past health state, so that an individual's true (latent) health states evolves according to the model

$$y^*_{i,t} = Y'_{i,t-1}\alpha + X'_{i,t-1}\beta + \epsilon_{i,t}, \tag{4.4}$$

where $y^*_{i,t}$ is individual i's current true (latent) health state, $Y'_{i,t-1}$ is a vector of factors for individual i's health state from the previous wave, and $X_{i,t-1}$ is a vector of determinants of health in the previous wave. The elements in $X_{i,t-1}$ are the same as in X_i from equation (4.1).

Alternatively, I could split the sample into three groups according to the observed previous health state $y_{i,t-1}$, and then specify the model as

$$y^*_{i,t} = X'_{i,t-1}\beta_s + \epsilon_{i,t}, \tag{4.5}$$

conditional on $y^*_{i,t-1} = s$.

Equation (4.4) is restrictive, as it assumes that the effects of lagged independent variables on the current health state are the same, regardless of an individual's lagged health state, while equation (4.5) allows for health-state specific β_s parameters. However, splitting the sample leads to less precise estimates. To avoid the loss in precision while also allowing some β_s parameters to be health-state specific, I consider a partially restricted model as

$$y^*_{i,t} = Y'_{i,t-1}\alpha + X'_{i,t-1}\beta + Z'_{i,t-1}\gamma + \epsilon_{i,t}, \tag{4.6}$$

where $X_{i,t-1}$ and $Y_{i,t-1}$ are the same as in equation (4.4). $Z_{i,t-1}$ contains the interaction terms of $Y_{i,t-1}$ and age, as well as the interaction terms of $Y'_{i,t-1}$ and insurance plan at $t-1$. Thus, while the β coefficients are restricted, the γ coefficients are health state specific, so that the impact of lagged insurance plan and age on the current health state is health-state specific.

I follow the same procedure in the context of equation (4.6), where I transform $y^*_{i,t}$ into self-ranked health state $y_{i,t}$.

For each $s = 1, 2, 3$, I report the transition probability by

$$p_i(s'|s) = \frac{v_i(s', s)}{\sum_{s''=1}^{S} v_i(s'', s)}, \tag{4.7}$$

where $v_i(s', s)$ is the joint probability of reporting a health state s' in period t and health state s in the previous period $t-1$. Given $X_{i,t-1}$, the probability of being in each health state at time t for individual i is

$$\begin{aligned}
p(y_i(s') = 1|s) &= F(\mu_1 - Y'_{i,t-1}(s)\alpha - X'_{i,t-1}\beta - Z'_{i,t-1}\gamma), \\
p(y_i(s') = 2|s) &= F(\mu_2 - Y'_{i,t-1}(s)\alpha - X'_{i,t-1}\beta - Z'_{i,t-1}\gamma) \\
&\quad - F(\mu_1 - Y'_{i,t-1}(s)\alpha - X'_{i}\beta - Z'_{i,t-1}\gamma), \\
p(y_i(s') = 3|s) &= 1 - F(\mu_2 - Y'_{i,t-1}(s)\alpha - X'_{i,t-1}\beta - Z'_{i,t-1}\gamma),
\end{aligned}$$

where $y_{i,t-1} = s$, F is the cumulative standard normal distribution function and s takes one of three values 1, 2 and 3.

Thus, one's probability of being in good, fair and poor health at t is conditional on their health state at $t-1$. The health transition matrix will be a 3×3 square matrix, where each cell $p_{s,s'}$ represents the probability of transitioning from the health state s at $t-1$ in the row to the health state s' at t in the column.

In the baseline model, $X_{i,t}$ includes insurance plan, sex, age, marital status, drinking, smoking, chronic disease, disability, education, *hukou* status and region; $Z_{i,t-1}$ includes the interaction terms of one' s health state at $t-1$ with the health insurance plan, sex and age, respectively. Additionally, I estimate this model by excluding sex and then estimate it separately for males and females. I also estimate this model for the 2011-13 and 2013-15 waves separately, as I allow for the possibility

that the relationship between certain explanatory variables and health transition may have changed over time.

I do not estimate a model that includes per capita household annual expenditure as a sensitivity check here. First, the expenditure measure does not include medical expenses, so higher expenditure does not imply that individuals are spending more on medical treatments. Second, even if medical expenses were included, the role of health insurance in covering treatment costs would likely minimize the impact of expenditure on healthcare access. The employee health insurance plays a more important role here. Moreover, dynamic health transitions are more frequently used in general equilibrium models. The inclusion of numerous individual characteristics, while comprehensive, complicates the delineation of effects applicable to a typical or average individual within such models. Consequently, for the sake of parsimony and to enhance the model's interpretability, I also estimate a model which is confined to essential variables: health insurance coverage, sex, age, and region, alongside their interaction terms.

4.4 Data

I use the baseline wave in 2011 and the subsequent wave in 2013 to investigate the determinants of self-ranked health state and whether they vary between health insurance plans. I use the pooled cross-sections described in Figure 3.3. Then, to investigate whether the health insurance plan is associated with the dynamic transitions of health states, I use the 2011 and 2013 waves, together with the health state variable in the 2015 wave. I merge the 16,336 complete cases in 2011 with the self-ranked health state in 2013, and 13,647 observations remain. About 16% of the respondents cannot be matched. I also merge the 16,244 complete cases in 2013 with the self-ranked health state in 2015, and 13,762 observations remain. About 15% of the respondents cannot be matched. Finally, after pooling, there are a total of

27,339 observations. While death is one of the reasons for exiting the survey, there is no information on death as a source of attrition, so it cannot be separated from sample stratification. As such, I do not consider death as one of the health states in estimating health state transitions, but I am aware that individuals in poor health may be more likely to transition to death from the one wave to the next. Thus, in the three state health matrices excluding death, the probability of staying in poor health may be overestimated due to attrition.

4.4.1 Pooled cross-sections: summary statistics

Table 4.1 represents the summary statistics of the pooled cross-sections which are used to investigate the association between self-ranked health state and health insurance plans.

Table 4.1: Pooled Data for Health State: Summary Statistics

Variables	Pooled data N = 32,580		2011 wave N = 16,336		2013 wave N = 16,244	
	N	%	N	%	N	%
Sex						
Female	16,764	51.5	8,337	51.0	8,427	51.9
Male	15,816	48.6	7,999	49.0	7,817	48.1
Health state						
Good	7,592	23.3	3,786	23.2	3,806	23.4
Fair	16,428	50.4	7,983	48.9	8,445	52.0
Poor	8,560	26.3	4,567	28.0	3,993	24.6
Health insurance plan						
Employee health insurance	4,466	13.7	2,148	13.2	2,318	14.3
Resident health insurance	26,445	81.2	13,119	80.3	13,326	82.0
No health insurance	1,669	5.1	1,069	6.5	600	3.7
Educational attainment						
Illiterate	8,574	26.3	4,429	27.1	4,145	25.5
Literate, less than high school	19,868	61.0	9,850	60.3	10,018	61.7
High school and above	4,138	12.7	2,057	12.6	2,081	12.8
Marital Status						
Married, spouse present	26,617	81.7	13,198	80.8	13,419	82.6
Married, spouse away	1,969	6.0	1,132	6.9	837	5.2
Unmarried	3,994	12.3	2,006	12.3	1,988	12.2
Regular drinker						
Yes	8,543	26.2	4,169	25.5	4,374	26.9
No	24,037	73.8	12,167	74.5	11,870	73.1
Chronic diease						
Yes	22,999	70.6	11,275	69.0	11,724	72.2
No	9,581	29.4	5,061	31.0	4,520	27.8
Disability						
Yes	5,540	17.0	2,897	17.7	2,643	16.3
No	27,040	83.0	13,439	82.3	13,601	83.7
Smoking						
Yes	13,545	41.6	6,598	40.4	6,947	42.8
No	19,035	58.4	9,738	59.6	9,297	57.2
Urban *hukou*						
Yes	7,068	21.7	3,501	21.4	3,567	22.0
No	25,512	78.3	12,835	78.6	12,677	78.0
Age group						
45 - 49	5,903	18.1	3,328	20.4	2,575	15.9
50 - 54	5,219	16.0	2,474	15.1	2,745	16.9
55 - 59	6,493	19.9	3,389	20.8	3,104	19.1
60 - 64	5,819	17.9	2,792	17.1	3,027	18.6
65 - 69	3,852	11.8	1,787	10.9	2,065	12.7
70 - 74	2,688	8.3	1,300	8.0	1,388	8.5
75 - 79	1,746	5.4	841	5.2	905	5.6
80 - 84	860	2.6	425	2.6	435	2.7

Continued on next page

Table 4.1 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)
Region						
East	10,151	31.2	5,089	31.2	5,062	31.2
Middle	9,358	28.7	4,725	28.9	4,633	28.5
Northeast	2,415	7.4	1,181	7.2	1,234	7.6
West	10,656	32.7	5,341	32.7	5,315	32.7

Notes: The results of the Pearson chi-squared test show that the differences in the distribution of sex, urban *hukou*, and region between the 2011 and 2013 waves are not statistically significant, while the differences in the distribution of the other variables between the two waves are statistically significant.
Source: Based on CHARLS 2011 and 2013 survey waves (pooled observations). See Figure 3.3 for the construction of the dataset.

Both waves had more females than males. In terms of health status, the fraction of the respondents who report good health is similar in both waves, at about 23%. However, the 2011 wave has a smaller fraction of respondents who report fair health and a larger fraction of respondents who report poor health. About 49% of respondents reported fair health in the 2011 wave, compared to 52% in the 2013 wave. Twenty-eight percent of respondents report poor health in the 2011 wave, while the percentage in the 2013 wave is about 25%.

The factors determining the self-ranked health state are divided into five groups: health insurance plan, socio-economic, life style, biological age and geographical factors. In the pooled data, about 14% of the respondents have the employee health insurance, while about 81% of the respondents have the resident health insurance. Only about 5% of the residents are not insured. However, the coverage rate slightly varies for each separate wave. The fraction of uninsured decreases from 7% in the 2011 wave to 4% in the 2013 wave. As a result, coverage rates of both the employee health insurance and the resident health insurance increase, by 1 percentage point and 2 percentage points, respectively. The coverage gap between the employee health insurance and the resident health insurance has not been narrowed.

Socio-economic and life style variables are similar in both waves. Thus, I will focus on discussing the pooled data for the summary statistics of these variables. In

terms of education, the illiteracy rate is high (26%), while the fraction of respondents who have a high school degree or above is low (13%). Most of the respondents are married and live with their spouse (80%). Life-style factors include smoking and drinking. About 42% of the respondents have ever smoked and about 26% drink more than once a month. However, the results of the Pearson chi-squared test show that the distributions of these variables in the two waves are statistically different, although the differences are small. The respondents in the 2013 wave have higher education levels, a larger fraction of respondents who are married, more respondents who drink regularly, fewer disabled individuals, and a larger fraction who used to smoke.

Distributions of the geographical variables are similar across the two waves. Only about 22% of the respondents have urban *hukou* in the sample. In terms of region, since the Western region only includes three provinces, the fraction of individuals residing in the Northeastern region is the lowest, at about 7%. The fraction of individuals residing in the other three regions are similar: 31% in the Eastern region, 29% in the Middle region and 33% in the Western region.

In contrast, the distribution of biological age varies across two waves. The 2011 wave is younger than the 2013 wave. In the 2011 wave, the fraction of respondents who are between 45 and 64 years of age is 73%, compared to 71% in the 2013 wave. The differences in the fraction of the youngest group is larger, with the fraction of respondents who are between 45 and 49 years of age being 20% and 16% in the 2011 and 2013 waves, respectively. That of the pooled data is 18%.

I also conduct a model which includes the logarithm of per capita household annual expenditure for a sensitivity check. Section 3.3.5 shows the process of constructing the logarithm of per capita household annual expenditure, and how it is merged with the dataset of the baseline model. Tables 3.9 and 3.10 in section 3.3.5 show the summary statistics of per capita household annual expenditure and the

logarithm of it in each cross-sectional wave.

4.4.2 Pooled data for health state transitions: summary statistics

Table 4.2 reports the summary statistics for the variables used for estimating health state transitions. Compared to the pooled cross-sections, these data have fewer observations, since those who report poor health may die in the next wave, and are excluded from the dataset. As a result, the proportion of respondents reporting poor current health decreases by 0.5 percentage points in both 2011-13 and 2013-15 waves, possibly due to attrition. Moreover, both the 2011-13 and 2013-15 waves show that when comparing the distribution of current health and the subsequent observed health, the fraction of respondents reporting either good or poor health both decreases in the subsequent observed health, while that of the respondents reporting fair health increases. However, death is not the only reason for being excluded from the survey in the subsequent wave.

The pooled data which is used for estimating dynamic health transition has fewer observations compared to the pooled cross-sections, but the distributions of variables only vary slightly across the two datasets (Table 4.2, column 1 and Table 4.1, column 1). In addition, differences in the distribution of variables across the two waves which are observed in the pooled cross-sections are also observed in these data. Results of the Pearson chi-squared test show that the distributions of sex, urban hukou, and region are not statistically different between the 2011-13 and the 2013-15 waves, while those for the other variables are statistically different across the two waves. Similarly, the respondents in the 2013-15 wave have higher education levels, a larger fraction of respondents who are married, more respondents who drink regularly, fewer disabled individuals, and a larger fraction who used to smoke.

However, the distribution of health insurance changes compared to that in the

pooled cross-sections, there are smaller fractions of respondents who have the employee health insurance as well as who do not have any insurance, and the fraction of respondents with the resident health insurance increases by about two percentage points in both the 2011-13 and 2013-15 waves. When comparing with the pooled data, these data also have larger fractions of married and living with their spouse, highly educated, and relatively younger respondents. However, these differences in distributions compared with the cross-sectional dataset are small. Variations in the distributions of other control variables are negligible.

Table 4.2: Pooled Data for Health State Transitions: Summary Statistics

Variables	Pooled data N = 27,229		2011-13 wave N = 13,467		2013-15 wave N = 13,762	
	N	%	N	%	N	%
Sex						
Female	14,163	52.0	6,964	51.7	7,199	52.3
Male	13,066	48.0	6,503	48.3	6,563	47.7
Health state, s						
Good	6,284	23.1	3,092	23.0	3,192	23.2
Fair	13,933	51.2	6,665	49.5	7,268	52.8
Poor	7,012	25.8	3,710	27.6	3,302	24.0
Health state, s'						
Good	6,132	22.5	2,981	22.1	3,151	22.9
Fair	14,366	52.8	7,044	52.3	7,322	53.2
Poor	6,731	24.7	3,442	25.6	3,289	23.9
Health insurance plan						
Employee health insurance	3,314	12.2	1,581	11.7	1,733	12.6
Resident health insurance	22,668	83.3	11,086	82.3	11,582	84.2
No health insurance	1,247	4.6	800	5.9	447	3.3
Educational attainment						
Illiterate	7,180	26.4	3,671	27.3	3,509	25.5
Literate, less than high school	16,868	62.0	8221	61.1	8647	62.8
High school and above	3,181	11.7	1,575	11.7	1,606	11.7
Marital Status						
Married, spouse present	22,641	83.2	11,122	82.6	11,519	83.7
Married, spouse away	1,463	5.4	808	6.0	655	4.8
Unmarried	3,125	11.5	1537	11.4	1,588	11.5
Regular drinker						
Yes	7,200	26.4	3,454	25.7	3,746	27.2
No	20,029	73.6	10,013	74.4	10,016	72.8
Smoking						
Yes	11,214	41.2	5,380	40.0	5,834	42.4
No	16,015	58.8	8,087	60.1	7,928	57.6
Chronic disease						
Yes	19,227	70.6	9,301	69.1	9,926	72.1
No	8,002	29.4	4,166	30.9	3,836	27.9
Disability						
Yes	4,440	16.3	2,292	17.0	2,148	15.6
No	22,789	83.7	11,175	83.0	11,614	84.4
Urban *hukou*						
Yes	5,313	19.5	2,623	19.5	2,690	19.6
No	21,916	80.5	10,844	80.5	11,072	80.5
Age group						
45 - 49	4,832	17.8	2,677	19.9	2,155	15.7
50 - 54	4,358	16.0	2,029	15.1	2,329	16.9
55 - 59	5,579	20.5	2,878	21.4	2,701	19.6
60 - 64	5,061	18.6	2,392	17.8	2,669	19.4

Continued on next page

Table 4.2 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)
65 - 69	3,349	12.3	1,548	11.5	1,801	13.1
70 - 74	2,245	8.2	1,061	7.9	1,184	8.6
75 - 79	1,261	4.6	598	4.4	663	4.8
80 - 84	544	2.0	284	2.1	260	1.9
Region						
East	8,498	31.2	4,173	31.0	4,325	31.4
Middle	7,853	28.8	3,913	29.1	3,940	28.6
Northeast	1,930	7.1	944	7.0	986	7.2
West	8,948	32.9	4,437	33.0	4,511	32.8

Notes: Unless otherwise stated, all variables are measured in the survey year corresponding to self-ranked health state s. s' is the respondent's subsequent observed health state. The results of the Pearson chi-squared test show that the differences in the distribution of sex, urban *hukou*, and region between the 2011-13 and 2013-15 waves are not statistically significant, while the differences in the distribution of the other variables between the two waves are statistically significant.
Source: Based on CHARLS 2011, 2013 and 2015 survey waves (pooled observations). See Figure 3.3 for the construction of the dataset.

4.5 Results

4.5.1 Health state and health insurance plan

I start with the estimation results of the model that links the type of one's health insurance plan to their self-ranked health state. Since I hypothesize that the association between several variables and individuals' health status varies by sex, I report the estimation results for males and females, separately. Tables 4.3 and 4.4 show the estimated ordered probit coefficients based on equation (4.1) for males and females, respectively. My baseline specification for the determinants of current-period self-ranked health state in the cross-sectional model includes health insurance plan, sex, age group (in 5-year intervals), marital status, chronic disease, disability, smoking, drinking, education, *hukou* status and region. For both sexes, I first estimate the baseline model using the pooled dataset (column 1 in Tables 4.3 and 4.4). Section 4.4.1 shows that the distributions of several variables are statistically different in the two separately waves. Hence, I also report the estimates of the baseline model for

each wave (Tables 4.3 and 4.4, columns 2 and 3).

I also test whether the impact of key independent variables is different between the waves by including interaction terms for all independent variables and a dummy variable, which takes a value of 1 if the observation is in the 2013 wave, and 0 otherwise.

From the standpoint of understanding the determinants of self-ranked health state, there are both pros and cons of including the prevalence of chronic disease and disability in the model—both measured as binary, which precludes us from measuring their severity. Having a chronic disease or disability is highly correlated with several explanatory variables, including age. Moreover, in China, people with disabilities are often unable to find employment, and thus are unlikely to have the employee health insurance. I thus report the estimation results from the model without both of these variables (column 4) and when only disability is excluded (column 5). Finally, I estimate the baseline model with the age variable grouped within 10-year intervals (column 6).

Table 4.3: Determinants of Health State: the Ordered Probit Model by Sex (Male)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.161***	−0.180***	−0.148***	−0.169***	−0.146***	−0.155***
	(0.038)	(0.051)	(0.049)	(0.038)	(0.038)	(0.038)
No health insurance	0.016	0.015	−0.028	−0.043	−0.047	0.012
	(0.046)	(0.054)	(0.078)	(0.046)	(0.046)	(0.046)
Chronic disease	0.727***	0.772***	0.687***	-	-	0.730***
	(0.022)	(0.028)	(0.029)	-	-	(0.022)
Disability	0.386***	0.407***	0.353***	-	0.434***	0.389***
	(0.028)	(0.035)	(0.036)	-	(0.028)	(0.028)
Regular drinker	−0.264***	−0.256***	−0.278***	−0.291***	−0.288***	−0.268***
	(0.020)	(0.026)	(0.026)	(0.020)	(0.020)	(0.020)
Past smoker	0.068***	0.058**	0.097***	0.093***	0.083***	0.070***
	(0.024)	(0.030)	(0.032)	(0.024)	(0.024)	(0.024)
Unmarried	0.076**	0.081*	0.070	0.087**	0.057	0.079**
	(0.038)	(0.048)	(0.050)	(0.039)	(0.038)	(0.038)
Married, spouse away	−0.013	−0.004	−0.043	−0.024	−0.026	−0.016
	(0.039)	(0.050)	(0.060)	(0.039)	(0.039)	(0.039)
High school and above	−0.164***	−0.155***	−0.165***	−0.179***	−0.169***	−0.161***
	(0.030)	(0.038)	(0.038)	(0.031)	(0.031)	(0.030)
Illiterate	0.061*	0.118***	−0.010	0.056	0.030	0.073**
	(0.035)	(0.043)	(0.045)	(0.034)	(0.035)	(0.034)
Urban *hukou*	−0.048	−0.067	−0.028	−0.022	−0.018	−0.050
	(0.034)	(0.044)	(0.044)	(0.034)	(0.034)	(0.034)
Middle	0.170***	0.211***	0.127***	0.222***	0.213***	0.169***
	(0.027)	(0.033)	(0.034)	(0.027)	(0.027)	(0.027)
Northeast	0.174***	0.217***	0.131**	0.241***	0.214***	0.174***
	(0.042)	(0.053)	(0.054)	(0.043)	(0.043)	(0.042)
West	0.197***	0.218***	0.177***	0.270***	0.247***	0.195***
	(0.026)	(0.032)	(0.033)	(0.026)	(0.026)	(0.026)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
Threshold μ_1	0.029	0.107**	−0.051	−0.373***	−0.345***	−0.025
	(0.037)	(0.047)	(0.051)	(0.036)	(0.036)	(0.034)
Threshold μ_2	1.566***	1.612***	1.526***	1.066***	1.113***	1.511***
	(0.039)	(0.050)	(0.053)	(0.036)	(0.036)	(0.036)
N	15,816	7,999	7,817	15,816	15,816	15,816

Notes: All models include age effects by 5- or 10-year windows, which are reported in Table B.1. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married male with resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table 4.4: Determinants of Health State: the Ordered Probit Model by Sex (Female)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.111***	−0.209***	−0.024	−0.135***	−0.121***	−0.114***
	(0.040)	(0.054)	(0.052)	(0.040)	(0.040)	(0.040)
No health insurance	−0.029	−0.067	−0.010	−0.106***	−0.106***	−0.031
	(0.041)	(0.052)	(0.065)	(0.041)	(0.041)	(0.041)
Chronic disease	0.732***	0.761***	0.707***	-	-	0.734***
	(0.022)	(0.028)	(0.029)	-	-	(0.022)
Disability	0.333***	0.351***	0.306***	-	0.389***	0.333***
	(0.029)	(0.036)	(0.038)	-	(0.029)	(0.029)
Regular drinker	−0.249***	−0.232***	−0.257***	−0.267***	−0.265***	−0.249***
	(0.035)	(0.048)	(0.045)	(0.035)	(0.035)	(0.035)
Past smoker	0.038	0.044	0.035	0.080**	0.059	0.036
	(0.039)	(0.050)	(0.046)	(0.038)	(0.039)	(0.039)
Unmarried	−0.022	−0.051	0.005	0.002	−0.013	−0.023
	(0.030)	(0.038)	(0.039)	(0.030)	(0.030)	(0.030)
Married, spouse away	−0.048	−0.071	−0.027	−0.044	−0.045	−0.050
	(0.038)	(0.050)	(0.054)	(0.038)	(0.038)	(0.038)
High school and above	−0.133***	−0.121**	−0.143***	−0.162***	−0.154***	−0.126***
	(0.037)	(0.048)	(0.047)	(0.038)	(0.038)	(0.037)
Illiterate	0.049**	0.048*	0.043	0.044*	0.031	0.051**
	(0.023)	(0.029)	(0.029)	(0.023)	(0.023)	(0.023)
Urban *hukou*	−0.196***	−0.137***	−0.252***	−0.171***	−0.163***	−0.196***
	(0.033)	(0.043)	(0.042)	(0.033)	(0.033)	(0.033)
Middle	0.164***	0.175***	0.154***	0.199***	0.193***	0.163***
	(0.026)	(0.033)	(0.033)	(0.026)	(0.026)	(0.026)
Northeast	0.146***	0.144***	0.150***	0.189***	0.189***	0.146***
	(0.041)	(0.053)	(0.050)	(0.042)	(0.042)	(0.041)
West	0.232***	0.215***	0.251***	0.299***	0.284***	0.229***
	(0.025)	(0.032)	(0.031)	(0.025)	(0.025)	(0.025)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
Threshold μ_1	−0.170***	−0.162***	−0.173***	−0.561***	−0.540***	−0.202***
	(0.030)	(0.038)	(0.040)	(0.028)	(0.028)	(0.028)
Threshold μ_2	1.315***	1.293***	1.345***	0.837***	0.870***	1.282***
	(0.032)	(0.040)	(0.042)	(0.029)	(0.029)	(0.029)
N	16,764	8,337	8,427	16,764	16,764	16,764

Notes: All models include age effects by 5- or 10-year windows, which are reported in Table B.2. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married female with resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

In Tables 4.3 and 4.4, a positive coefficient means that the explanatory variable is associated with a larger probability of being in a higher (poorer) self-ranked health state relative to the reference group, and a negative coefficient means a smaller probability of being in a higher self-ranked health state (poor). The baseline estimation results refer to a married individual (male in Table 4.3 and female in Table 4.4) with resident health insurance, between 45 and 49 years of age, living with his/her spouse, and who is not a regular drinker, did not smoke, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region.

The results show that there is a strong association between the self-ranked health state and one's health insurance plan for both males and females, though the size of the association varies by sexes. Self-ranked health state is significantly different for the respondents with the employee health insurance and the resident health insurance. The negative signs of the coefficients for the employee health insurance mean that compared with those who have the resident health insurance, those who are insured by the employee health insurance have a smaller probability of reporting poor health, and larger probability of reporting good health. In the baseline models (Tables 4.3 and 4.4, column 1), the size of the association is larger for males (-0.161) than for females (-0.111). However, the size of this association between types of health insurance and health states varies across waves, especially for females. For both sexes, the size of

the association is larger in the 2011 wave and is smaller in the 2013 wave. For males, the employee health insurance is significant in both waves (-0.180 compared to -0.148 in Table 4.3). For females, while the association between having an employee health insurance plan and reporting poorer health is larger in the 2011 wave (-0.209), it is small and insignificant in the 2013 wave (-0.024).

For both sexes, the coefficient for the no-insurance group is not statistically significant in the baseline model and in the models with separate waves of data (Tables 4.3 and 4.4, columns 1-3). Recall that the number of respondents without any insurance is small. However, for females, the coefficients of "no health insurance" are negative and significant when chronic disease is not included in the model (Table 4.4, columns 4 and 5). This implies that having a chronic disease might be highly correlated with having the resident health insurance for females. Since the resident health insurance is voluntary, individuals with chronic diseases could be more likely to enroll in the resident health insurance plan, so excluding chronic disease from the regression results in significant negative coefficients of "no health insurance" for females.

Regarding the control variables, several variables are significant only for males, such as smoking and being unmarried. In Table 4.3, the coefficients for being unmarried are significant and positive, implying that unmarried males have higher probabilities of reporting poor health compared to married males living with their spouses. In contrast, these coefficients are insignificant and negative for females in Table 4.4. Urban *hukou* is the sole variable that is significant only for females. In Table 4.4, all coefficients for urban *hukou* are significant at the 1% significance level, while they are insignificant in Table 4.3 for males.

Associations between other control variables and health states are similar for both sexes. Having a chronic disease and being disabled each is significantly associated with self-ranked health state, and the estimated coefficients are positive. This indicates that compared to the individuals who do not have any chronic diseases,

those who have a chronic disease have a higher probability of reporting poor health and a lower probability of reporting good health. The explanations are similar for being disabled. In terms of education, the estimated coefficients are negative and significant for having a high school diploma or higher, and they are positive and significant for being illiterate. Moreover, regional disparities exist. Positive estimated coefficients for the Middle, Northeastern and Western region in the baseline model suggest that compared to those living in the East (the richest region), those living in the less economically affluent regions have a higher probability of reporting poor health state, and have a lower probability of reporting good health. These coefficient estimates indicate that the economically most disadvantaged region is also at a disadvantage in terms of self-ranked health state.

Age, not surprisingly, is an important determinant of self-ranked health state. In the baseline model, compared to those between 45 to 49 years of age (the reference group), coefficients for other age groups are all positive and significant at 1% significant level for males, which indicates that aging reduces the likelihood of reporting good health and increases the likelihood of reporting poor health. In addition, the size of the positive association increases then decreases with age for males in groups younger than 75, though in this case, the standard errors are also large. For females, the coefficients for other age groups are also positive, but the size of the association is smaller compared to the corresponding age group of males, and the coefficient of the 75 to 79 age group is not statistically significant (Appendix Tables B.1 and B.2). Figures 4.1 and 4.2 illustrate the estimated probabilities of reporting each health state as age group increases for each health insurance plan. The predicted probability of each health state shows similar age profiles across all health insurance plans. For males younger than 75, the probability of reporting good health decreases with age, while the probabilities of reporting fair and poor health increase. However, for females, the probability of reporting good health decreases as age group increases in the relatively

young groups (45 to 59). Then it fluctuates as age group increases.

However, in ordered probit models, coefficients of variables are less explanatory than the marginal effects, so I also represent the average marginal effect for each variable. For both sexes, the baseline model is used to estimate the average marginal effects. First, chronic disease and disability are significant and correlated with health insurance plans, so they should be included in the model. Second, the association between health insurance plans and health states varies in the 2011 and 2013 waves for both sexes. The reason for this needs to be further explored later, but using pooled data can average out the disturbance here. Additionally, results are consistent while using the 10-year age interval, compared to the baseline model which uses the 5 year age interval, but the five-year age interval provides more precise association between age and health state.

Tables 4.5 and 4.6 show the average marginal effect for each variable in the baseline model for males and females, respectively. Each coefficient indicates the change in probabilities for each health state if the value of explanatory variable changes at the margin while holding all other characteristics the same. For example, for males (Table 4.5), for those in the reference group (who have the resident health insurance, all other variables unchanged), the estimated marginal effect of switching to the employee health insurance corresponds to a 4.8 percentage point higher likelihood of reporting good health, and a 0.6 and 4.2 percentage point lower likelihood of reporting fair and poor health, respectively (row 1). For females (Table 4.6), for those in the reference group (who have the resident health insurance, all other variables unchanged), the estimated marginal effect of switching to the employee health insurance corresponds to a 2.9 and 0.5 percentage point higher likelihood of reporting good and fair health, respectively, as well as a 3.5 percentage point lower likelihood of reporting poor health.

There is a notable difference in the marginal effects of most variables across sexes.

The signs for "fair" are opposite for all variables except for "no health insurance", "unmarried" and "married, spouse away." For example, for both males and females, compared to being in the 45-49 age group, being in the 50-54 age group is associated with a lower probability of reporting good health and a higher probability of reporting fair health. However, the association between age and fair health differs: younger ages are associated with higher probabilities of reporting fair health for males, but are associated with lower probabilities of reporting fair health for females.

Additionally, Tables 4.5 and 4.6 show that the marginal effects of other control variables on health states are larger than the marginal effects of the employee health insurance plan on health states.

Table 4.5: Marginal Effects on Health State by Sex (Male)

Dependent variable: Self-ranked health state

Variables	(1) Good	(2) Fair	(3) Poor
Employee health insurance	0.048***	−0.006***	−0.042***
	(0.012)	(0.002)	(0.010)
No health insurance	0.005	0.000	0.004
	(0.013)	(0.000)	(0.013)
50-54	−0.035***	0.006***	0.029***
	(0.010)	(0.002)	(0.008)
55-59	−0.037***	0.006***	0.029***
	(0.010)	(0.002)	(0.008)
60-64	−0.057***	0.008***	0.050***
	(0.010)	(0.002)	(0.009)
65-69	−0.058***	0.008***	0.050***
	(0.011)	(0.002)	(0.010)
70-74	−0.095***	0.005**	0.090***
	(0.012)	(0.002)	(0.012)
75-79	−0.086***	0.006***	0.079***
	(0.014)	(0.002)	(0.014)
80-84	−0.081***	0.007***	0.075***
	(0.020)	(0.002)	(0.021)
Chronic disease	−0.212***	0.014***	0.199***
	(0.006)	(0.003)	(0.006)
Disability	−0.113***	0.007***	0.105***
	(0.008)	(0.002)	(0.008)
Regular drinker	0.077***	−0.005***	−0.072***
	(0.006)	(0.001)	(0.005)
Past smoker	−0.020***	0.001**	0.019***
	(0.007)	(0.001)	(0.006)
Unmarried	−0.022***	0.001*	0.021*
	(0.011)	(0.000)	(0.011)
Married, spouse away	0.004	0.000	−0.003
	(0.012)	(0.001)	(0.011)
High school and above	0.050***	−0.007***	−0.043***
	(0.009)	(0.002)	(0.008)
Illiterate	−0.017*	0.000	0.017*
	(0.010)	(0.000)	(0.010)
Urban *hukou*	0.014	−0.001	−0.013***
	(0.010)	(0.001)	(0.009)
Middle	−0.051***	0.006***	0.045***
	(0.008)	(0.001)	(0.007)
Northeast	−0.052***	0.006***	0.046***
	(0.012)	(0.001)	(0.012)
West	−0.059***	0.006***	0.053***
	(0.008)	(0.001)	(0.007)

Notes: The average marginal effects of self-ranked health state are based on the estimation results from the baseline ordered probit model by sex for males (Table 4.3, column 1). The reference predicted probabilities refer to a married male between 45 and 49 years of age living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.6: Marginal Effects on Health State by Sex (Female)

Dependent variable: Self-ranked health state

Variables	(1) Good	(2) Fair	(3) Poor
Employee health insurance	0.029***	0.005***	−0.035***
	(0.011)	(0.001)	(0.012)
No health insurance	0.007	0.002	−0.009
	(0.011)	(0.002)	(0.013)
50-54	−0.019**	−0.003**	0.023**
	(0.008)	(0.002)	(0.009)
55-59	−0.027***	−0.005***	0.033***
	(0.008)	(0.002)	(0.010)
60-64	−0.020**	−0.004**	0.024**
	(0.009)	(0.002)	(0.010)
65-69	−0.047***	−0.012***	0.059***
	(0.009)	(0.003)	(0.012)
70-74	−0.030***	−0.006**	0.036**
	(0.011)	(0.003)	(0.014)
75-79	−0.019	−0.003	0.022
	(0.014)	(0.003)	(0.017)
80-84	−0.041**	−0.010*	0.050**
	(0.017)	(0.006)	(0.023)
Chronic disease	−0.188***	−0.046***	0.234***
	(0.006)	(0.003)	(0.007)
Disability	−0.085***	−0.021***	0.106***
	(0.008)	(0.002)	(0.009)
Regular drinker	0.064***	0.016***	−0.080***
	(0.009)	(0.002)	(0.011)
Past smoker	−0.010	−0.002	0.012
	(0.010)	(0.002)	(0.012)
Unmarried	0.006	0.001	−0.007
	(0.008)	(0.002)	(0.010)
Married, spouse away	0.013	0.003	−0.015
	(0.010)	(0.002)	(0.012)
High school and above	0.036***	0.005***	−0.041***
	(0.010)	(0.001)	(0.011)
Illiterate	−0.012**	−0.003**	0.016***
	(0.006)	(0.002)	(0.007)
Urban *hukou*	0.053***	0.008***	−0.061***
	(0.009)	(0.001)	(0.010)
Middle	−0.044***	−0.007***	0.051***
	(0.007)	(0.001)	(0.008)
Northeast	−0.039*****	−0.006**	0.045***
	(0.011)	(0.003)	(0.013)
West	−0.061***	−0.013***	0.074***
	(0.007)	(0.002)	(0.008)

Notes: The average marginal effects of self-ranked health state are based on the estimation results from the baseline ordered probit model by sex for females (Table 4.4, column 1). The reference predicted probabilities refer to a married female between 45 and 49 years of age living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

I also compare the predicted probabilities of self-ranked health state by health insurance plan, keeping all other correlates the same across individuals. The representative respondent is otherwise a married individual, between 45 and 49 years of age, living with his/her spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Table 4.7 shows that for males with the employee health insurance and with such characteristics, the predicted probabilities of being in good, fair, and poor health state are, respectively 59.4%, 36.8%, and 3.8%. For males who have the resident health insurance and otherwise identical characteristics, the probability of being in good health decreases to 53.1%, while probabilities of being in fair and poor health state increase to 41.6% and 5.3%, respectively. Thus, compared to those who have the resident health insurance plan, those who have the employee health insurance plan with otherwise similar personal characteristics are 4.7 percentage points more likely to report good health, which is almost identical to the average marginal effect of the employee health insurance on reporting good health in Table 4.5.

For females, Table 4.8 shows that the marginal effect of having the employee health insurance on reporting good health is smaller than for males, which is consistent with the findings in Table 4.6. Moreover, for females, the probability of reporting good health is lower than that for males among individuals with each health insurance plan. For example, for females with the employee health insurance and with the baseline characteristics, the predicted probabilities are 55.4%, 39.3% and 5.2% (compared to 59.4%, 36.8% and 3.8% for males).

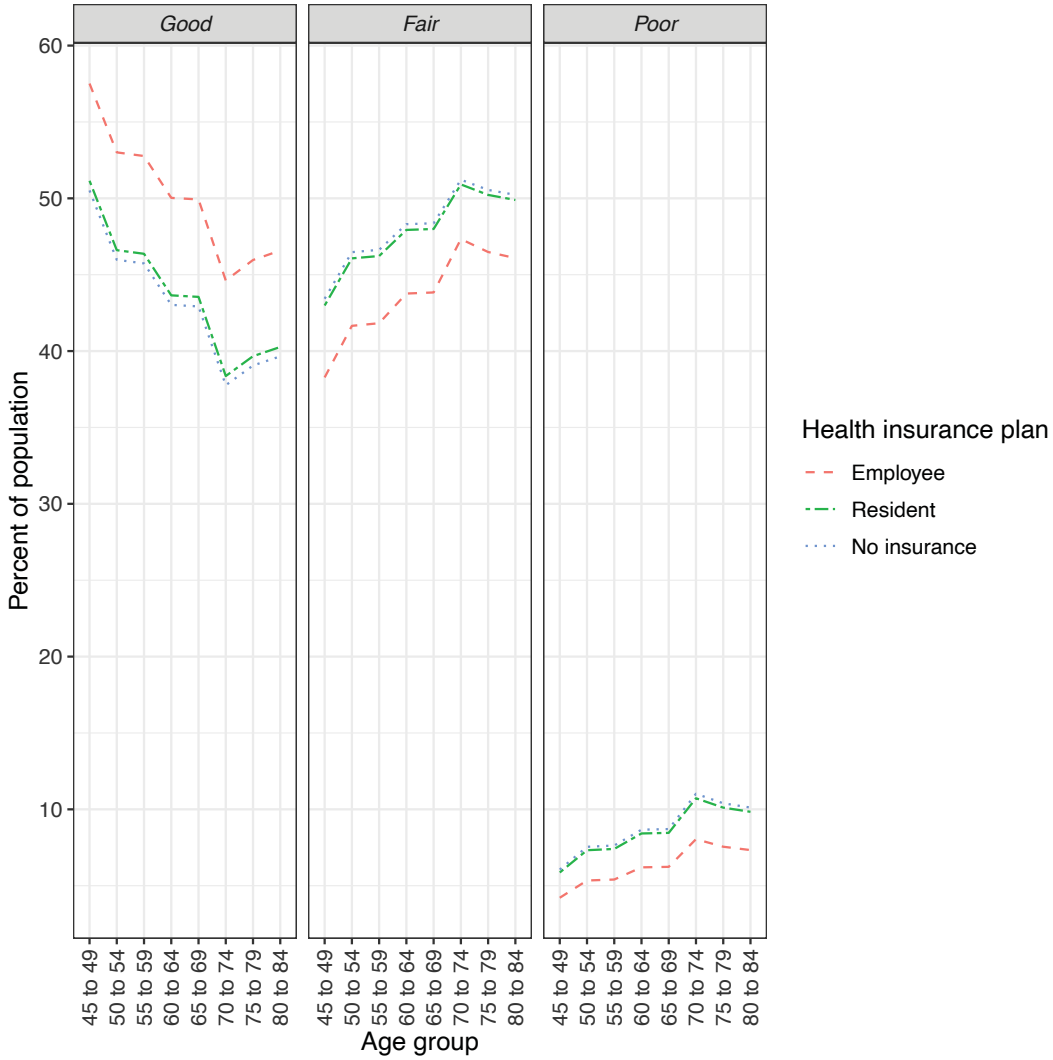


Figure 4.1: Self-ranked Health State by Insurance Plan and Age Group (male)

Notes: Based on CHARLS 2011 and 2013 waves (pooled). Other control variables are evaluated as in the baseline estimation results with reference to a married male living with his spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate and has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region.

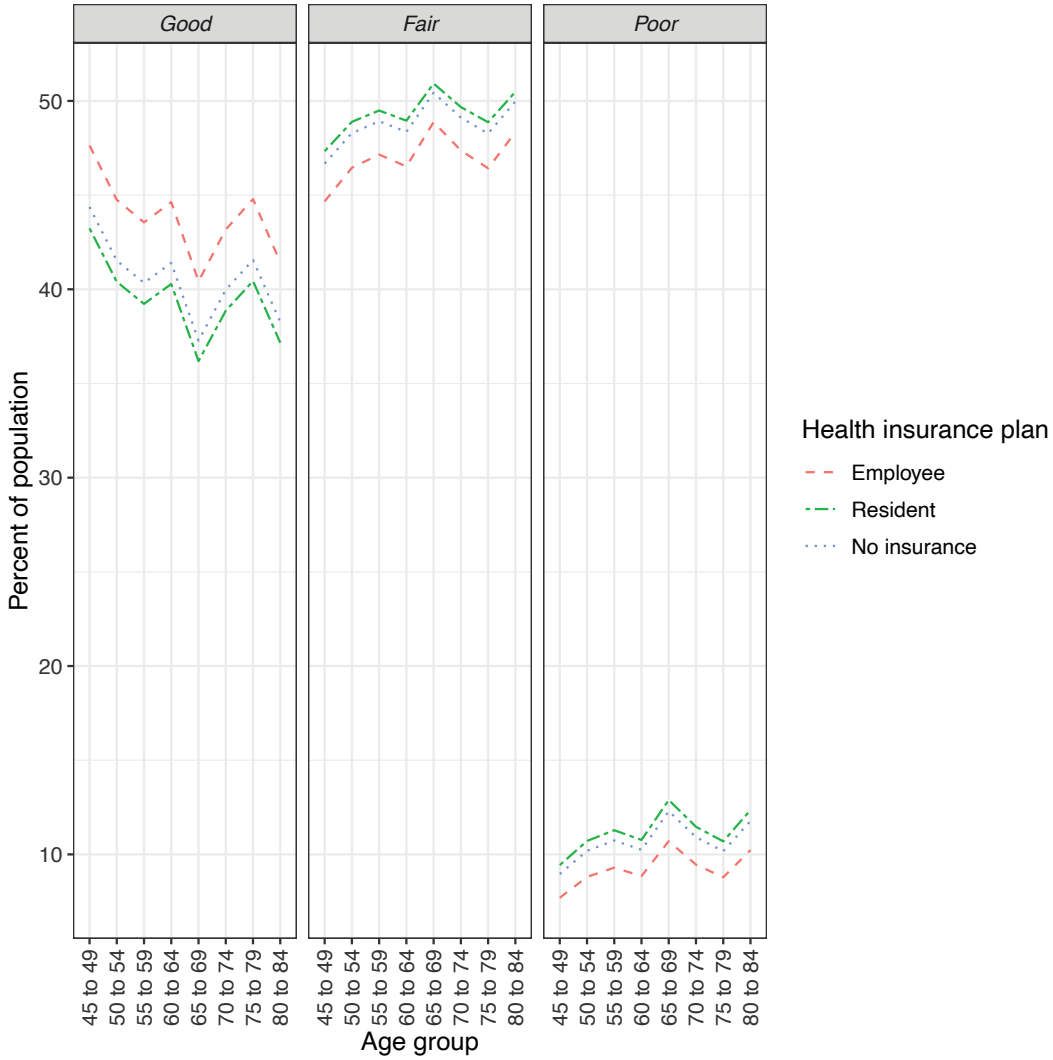


Figure 4.2: Self-ranked Health State by Insurance Plan and Age Group (female)

Notes: Based on CHARLS 2011 and 2013 waves (pooled). Other control variables are evaluated as in the baseline estimation results with reference to a married female living with her spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate and has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region.

Table 4.7: Predicted Probabilities of Health State by Type of Health Insurance Plan (Male)

Health state	Health insurance plan Employee	Resident	No plan
Good	0.594***	0.531***	0.524***
	(0.018)	(0.019)	(0.024)
Fair	0.368***	0.416***	0.421***
	(0.014)	(0.014)	(0.018)
Poor	0.038***	0.053***	0.055***
	(0.004)	(0.005)	(0.007)

Notes: Predicted probabilities of self-ranked health state are based on the estimation results from the baseline ordered probit model by sex for males (Table 4.3, column 1). The reference predicted probabilities refer to a married male between 45 and 49 years of age, living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%. The significance level applies to the estimated probabilities, but they are not applied to the differences across health insurance groups. The estimated coefficients used to generate the probabilities are statistically significant, suggesting the distributions are statistically different.

Table 4.8: Predicted Probabilities of Health State by Type of Health Insurance Plan (Female)

Health state	Health insurance plan Employee	Resident	No plan
Good	0.554***	0.510***	0.522***
	(0.017)	(0.017)	(0.022)
Fair	0.393***	0.424***	0.416***
	(0.012)	(0.012)	(0.015)
Poor	0.052***	0.065***	0.062***
	(0.005)	(0.006)	(0.007)

Notes: Predicted probabilities of self-ranked health state are based on the estimation results from the baseline ordered probit model by sex for females (Table 4.4, column 1). The reference predicted probabilities refer to a married female between 45 and 49 years of age, living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%. The significance level applies to the estimated probabilities, but they are not applied to the differences across health insurance groups. The estimated coefficients used to generate the probabilities are statistically significant, suggesting the distributions are statistically different.

Sensitivity analysis

Per capita household annual consumption expenditures Expenditure, as an indicator of economic status, is also an important variable that affects individuals'

health status, despite being endogenous. Hence, I do a sensitivity check by including the logarithm of per capita household annual consumption expenditure, and the results for males and females are shown in Tables 4.9 and 4.10, respectively. After adding the logarithm of per capita household annual consumption expenditure, the total number of observations decreases from 32,580 to 28,024 (a loss of 14% of the original pooled observations). Compared to columns 1-3 in Tables 4.3 and 4.4, columns 1-3 in Tables 4.9 and 4.10 include the logarithm of per capita household annual expenditure as a control variable. Since the expenditure data are missing for certain observations, including expenditure results in fewer observations in the estimated regressions. To see whether fewer observations will affect the estimated coefficients of the main explanatory variables, columns 4-6 in Tables 4.9 and 4.10 present the results of models in Tables 4.3 and 4.4, but only using observations whose logarithm of per capita household annual expenditure is not missing.

After including the logarithm of per capita household annual consumption expenditure in the regression, for both sexes, all coefficients of the employee health insurance remain significant if they were significant in Table 4.3 and 4.4, but the size of the association between having employee health insurance and health state decreases. "No health insurance" remains insignificant.

For the pooled dataset (column 1), the coefficients of the logarithm of per capita household annual consumption expenditure are negative and significant at the 1% significance level for both sexes; the size of the association is larger for males than for females. This indicates that the higher the household annual expenditure, the lower the likelihood of reporting poor health, and the greater the likelihood of reporting good health.

Using fewer observations is another reason for the change in coefficients. For males, the model specifications are the same in Table 4.3, columns 1-3, and Table 4.9, columns 4-6. However, the size of the association between the employee health

insurance and self-rated health is larger in Table 4.9, which only uses observations for which the expenditure data are usable. Conversely, for females, the association is smaller in Table 4.9, while using observations for which the expenditure data are usable.

For the models which include per capita household annual consumption expenditure, I also show the average marginal effects of the baseline model, which are reported in Tables 4.11 and 4.12 for males and females, respectively. The average marginal effects of the employee health insurance on reporting good health and poor health become smaller. For example, for males (Table 4.11), for those in the reference group (who have the resident health insurance, all other variables are unchanged), the estimated marginal effect of switching to the employee health insurance corresponds to a 4.4 percentage point increase in the likelihood of reporting good health, and a 3.8 percentage point decrease in the likelihood of reporting poor health (compared to 4.8 and 4.2 percentage points in Table 4.5). For females, the average marginal effect of employee health insurance on reporting good health becomes insignificant after including the logarithm of per capita household annual consumption expenditure in the regression, and the size of the association between employee health insurance and reporting fair health and poor health become smaller. Compared to the reference group, having the employee health insurance is associated with a 0.3 percentage point lower likelihood of reporting fair health and a 2.2 percentage point likelihood of reporting poor health. These marginal effects are 0.5 and 3.5 percentage points, respectively, in Table 4.6.

Moreover, for males, a 1% increase in per capita household annual consumption expenditure is associated with a 1.7 percentage point higher likelihood of reporting good health and 0.1 and 1.6 percentage point lower likelihoods of reporting fair health and poor health, respectively. For females, a 1% increase in per capita household annual consumption expenditure is associated with 1.1 and 0.3 percentage point higher

likelihoods of reporting good and fair health, respectively, while it is associated with a 1.4 percentage point lower likelihood of reporting poor health.

Tables 4.13 and 4.14 show the predicted probabilities of self-ranked health state by health insurance plan, for males and females, respectively, while keeping all other correlates the same across individuals. The representative respondent's logarithm of per capita household annual consumption expenditure equals the average of that among all observations, and the other characteristics of the representative respondent are the same as Tables 4.7 and 4.8, when expenditure is not included in the regression. Compared to the representative respondents in Tables 4.7 and 4.8, the representative respondents of each health insurance plan in Tables 4.13 and 4.14 have smaller probabilities of reporting good health and larger probabilities of reporting fair and poor health.

I also compare the predicted probabilities of self-ranked health state by health insurance plan, keeping all other correlates the same across individuals. The representative respondent is otherwise a married individual, between 45 and 49 years of age, living with his/her spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region.

Table 4.9: Determinants of Health State: the Ordered Probit Model (Sensitivity Check: Including Expenditure, Male)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	2011	2013
Expenditure	Y	Y	Y	N	N	N
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.146***	−0.169***	−0.131**	−0.169***	−0.186***	−0.158***
	(0.041)	(0.053)	(0.054)	(0.040)	(0.053)	(0.053)
No health insurance	−0.002	0.001	−0.065	0.007	0.007	−0.059
	(0.049)	(0.057)	(0.088)	(0.049)	(0.057)	(0.087)
Logarithm of per capita	−0.059***	−0.041**	−0.071***	-	-	-
household annual expenditure	(0.013)	(0.016)	(0.019)	-	-	-
Chronic disease	0.743***	0.782***	0.698***	0.738***	0.777***	0.696***
	(0.024)	(0.029)	(0.033)	(0.024)	(0.029)	(0.033)
Disability	0.364***	0.386***	0.323***	0.370***	0.390***	0.328***
	(0.030)	(0.036)	(0.041)	(0.030)	(0.036)	(0.041)
Regular drinker	−0.258***	−0.260***	−0.263***	−0.263***	−0.261***	−0.271***
	(0.021)	(0.027)	(0.030)	(0.021)	(0.027)	(0.029)
Past smoker	0.085***	0.076**	0.115***	0.080***	0.072**	0.114***
	(0.025)	(0.031)	(0.036)	(0.025)	(0.030)	(0.036)
Unmarried	0.068	0.091*	0.038	0.059	0.085*	0.028
	(0.041)	(0.050)	(0.058)	(0.041)	(0.050)	(0.058)
Married, spouse away	0.012	0.015	−0.014	−0.002	0.005	−0.034
	(0.042)	(0.052)	(0.066)	(0.042)	(0.052)	(0.066)
High school and above	−0.150***	−0.140***	−0.155***	−0.159***	−0.146***	−0.167***
	(0.032)	(0.039)	(0.042)	(0.032)	(0.039)	(0.042)
Illiterate	0.054	0.127***	−0.050	0.062*	0.134***	−0.043
	(0.037)	(0.044)	(0.052)	(0.037)	(0.044)	(0.052)
Urban *hukou*	−0.040	−0.057	−0.015	−0.053	−0.068	−0.027
	(0.036)	(0.046)	(0.048)	(0.036)	(0.046)	(0.048)
Middle	0.173***	0.216***	0.126***	0.178***	0.220***	0.129***
	(0.028)	(0.035)	(0.038)	(0.028)	(0.035)	(0.038)
Northeast	0.201***	0.237***	0.164***	0.205***	0.241***	0.167***
	(0.045)	(0.055)	(0.060)	(0.045)	(0.055)	(0.060)
West	0.200***	0.214***	0.186***	0.199***	0.215***	0.181***
	(0.027)	(0.034)	(0.037)	(0.027)	(0.034)	(0.037)
Threshold μ_1	−0.478***	−0.236	−0.681***	0.043	0.119**	−0.047
	(0.118)	(0.150)	(0.181)	(0.040)	(0.049)	(0.058)
Threshold μ_2	1.064***	1.273***	0.909***	1.583***	1.627***	1.540***
	(0.118)	(0.150)	(0.181)	(0.041)	(0.051)	(0.060)
N	13,720	7,490	6,230	13,720	7,490	6,230

Notes: All models include age effects by 5-year windows, which are reported in Appendix Table B.3. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". Compared to the baseline model reported in Table 4.3, the first three columns in this table include expenditure as a control variable, columns 4-6 use the same model specification as those in Table 4.3, columns 1-3, and the same numbers of observations as in the models which include expenditure. The baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age'), living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table 4.10: Determinants of Health State: the Ordered Probit Model (Sensitivity Check: Including Consumption Expenditure, Female)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	2011	2013
Expenditure	Y	Y	Y	N	N	N
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.070*	−0.189***	0.049	−0.087**	−0.187***	0.013
	(0.042)	(0.056)	(0.057)	(0.042)	(0.056)	(0.056)
No health insurance	−0.020	−0.064	0.023	−0.015	−0.065	0.029
	(0.045)	(0.054)	(0.078)	(0.045)	(0.054)	(0.078)
Logarithm of per capita	−0.044***	0.005	−0.094***	-	-	-
household annual expenditure	(0.013)	(0.016)	(0.018)	-	-	-
Chronic disease	0.730***	0.747***	0.710***	0.725***	0.748***	0.706***
	(0.024)	(0.029)	(0.033)	(0.024)	(0.029)	(0.033)
Disability	0.342***	0.356***	0.317***	0.347***	0.355***	0.327***
	(0.031)	(0.038)	(0.043)	(0.031)	(0.038)	(0.043)
Regular drinker	−0.239***	−0.220***	−0.247***	−0.242***	−0.220***	−0.257***
	(0.037)	(0.050)	(0.051)	(0.037)	(0.050)	(0.051)
Past smoker	0.063	0.050	0.075	0.061	0.051	0.074
	(0.042)	(0.051)	(0.054)	(0.042)	(0.051)	(0.054)
Unmarried	−0.030	−0.063	0.001	−0.028	−0.063	0.008
	(0.032)	(0.040)	(0.046)	(0.032)	(0.040)	(0.046)
Married, spouse away	−0.059	−0.078	−0.049	−0.065	−0.077	−0.058
	(0.042)	(0.053)	(0.064)	(0.042)	(0.053)	(0.064)
High school and above	−0.124***	−0.118**	−0.127**	−0.132***	−0.118**	−0.145***
	(0.039)	(0.050)	(0.053)	(0.039)	(0.049)	(0.052)
Illiterate	0.040	0.055*	0.013	0.047*	0.055*	0.027
	(0.025)	(0.030)	(0.034)	(0.025)	(0.030)	(0.033)
Urban *hukou*	−0.190***	−0.144***	−0.245***	−0.201***	−0.142***	−0.259***
	(0.035)	(0.045)	(0.046)	(0.035)	(0.045)	(0.046)
Middle	0.179***	0.187***	0.176***	0.182***	0.186***	0.179***
	(0.028)	(0.034)	(0.037)	(0.028)	(0.034)	(0.037)
Northeast	0.132***	0.136**	0.135**	0.135***	0.135**	0.137**
	(0.044)	(0.055)	(0.056)	(0.044)	(0.055)	(0.056)
West	0.244***	0.229***	0.273***	0.244***	0.229***	0.268***
	(0.027)	(0.033)	(0.036)	(0.027)	(0.033)	(0.036)
Threshold μ_1	−0.556***	−0.111	−1.014***	−0.163***	−0.156***	−0.165***
	(0.115)	(0.149)	(0.169)	(0.033)	(0.040)	(0.045)
Threshold μ_2	0.938***	1.349***	0.528***	1.330***	1.304***	1.373***
	(0.116)	(0.150)	(0.169)	(0.034)	(0.042)	(0.048)
N	14,304	7,749	6,555	14,304	7,749	6,555

Notes: All models include age effects by 5-year windows, which are reported in Appendix Table B.4. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". Compared to the baseline model reported in Table 4.4, columns 1-3, the first three columns in this table include expenditure as a control variable, columns 4-6 use the same model specification as those in Table 4.4, columns 1-3, and the same numbers of observations as in the models which include expenditure. The baseline estimation results refer to a married female with a resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table 4.11: Marginal Effects on Health State: the Baseline Model (Sensitivity Check: Including Consumption Expenditure, Male)

Dependent variable: Self-ranked health state

Variables	(1) Good	(2) Fair	(3) Poor
Employee health insurance	0.044***	−0.006**	−0.038***
	(0.012)	(0.002)	(0.010)
No health insurance	0.000***	0.000***	0.000***
	(0.014)	(0.001)	(0.013)
Logarithm of per capita	0.017***	−0.001***	−0.016***
household annual expenditure	(0.004)	(0.000)	(0.003)
50-54	−0.028***	0.005**	0.024***
	(0.011)	(0.002)	(0.009)
55-59	−0.031***	0.005***	0.026***
	(0.011)	(0.002)	(0.009)
60-64	−0.044***	0.006***	0.038***
	(0.011)	(0.002)	(0.009)
65-69	−0.054***	0.006***	0.047***
	(0.012)	(0.002)	(0.010)
70-74	−0.090***	0.004*	0.086***
	(0.013)	(0.002)	(0.013)
75-79	−0.071***	0.006***	0.065***
	(0.015)	(0.002)	(0.015)
80-84	−0.072***	0.006***	0.066***
	(0.023)	(0.002)	(0.023)
Chronic disease	−0.217***	0.016***	0.201***
	(0.006)	(0.003)	(0.006)
Disability	−0.107***	0.008***	0.099***
	(0.009)	(0.002)	(0.008)
Regular drinker	0.076***	−0.006***	−0.070***
	(0.006)	(0.001)	(0.006)
Past smoker	−0.025***	0.002***	0.023***
	(0.007)	(0.001)	(0.007)
Unmarried	−0.019*	0.001***	0.019
	(0.012)	(0.000)	(0.012)
Married, spouse away	−0.003	0.000	0.003
	(0.012)	(0.001)	(0.011)
High school and above	0.045***	−0.006***	−0.039***
	(0.010)	(0.002)	(0.008)
Illiterate	−0.015	0.000	0.015
	(0.010)	(0.000)	(0.010)
Urban *hukou*	0.012	−0.001	−0.011
	(0.011)	(0.001)	(0.010)
Middle	−0.052***	0.007***	0.046***
	(0.009)	(0.001)	(0.007)
Northeast	−0.060***	0.007***	0.053***
	(0.013)	(0.001)	(0.013)

Continued on next page

Table 4.11 (*Continued*)

	(1)	(2)	(3)
West	−0.060***	0.007***	0.053***
	(0.008)	(0.001)	(0.007)

Notes: The average marginal effects of self-ranked health state are based on the estimation results from the baseline ordered probit model including expenditure (Table 4.9, column 1). The reference predicted probabilities refer to a married male between 45 and 49 years of age, who comes from a household where the logarithm of the per capita annual household expenditure is equal to the average of all observations, is living with his spouse, and who is not a regular drinker, does not have any chronic disease, is literate but has less than a high school diploma, is not disabled, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.12: Marginal Effects on Health State: the Baseline Model (Sensitivity Check: Including Consumption Expenditure, Female)

Dependent variable: Self-ranked health state

Variables	(1) Good	(2) Fair	(3) Poor
Employee health insurance	0.018	0.003**	−0.022*
	(0.011)	(0.002)	(0.013)
No health insurance	0.005	0.001	−0.006
	(0.012)	(0.002)	(0.014)
Logarithm of per capita	0.011***	0.003***	−0.014***
household annual expenditure	(0.003)	(0.001)	(0.004)
50-54	−0.017**	−0.003*	0.020**
	(0.009)	(0.002)	(0.010)
55-59	−0.021**	−0.004**	0.025**
	(0.009)	(0.002)	(0.010)
60-64	−0.014	−0.002	0.017
	(0.009)	(0.002)	(0.011)
65-69	−0.041***	−0.010***	0.051***
	(0.010)	(0.003)	(0.013)
70-74	−0.020	−0.004	0.024
	(0.012)	(0.003)	(0.015)
75-79	−0.024	−0.005	0.029
	(0.015)	(0.004)	(0.018)
80-84	−0.042**	−0.010	0.052**
	(0.019)	(0.007)	(0.026)
Chronic disease	−0.189***	−0.043***	0.232***
	(0.006)	(0.003)	(0.007)
Disability	−0.088***	−0.020***	0.108***
	(0.008)	(0.002)	(0.010)
Regular drinker	0.062***	0.014***	−0.076***
	(0.010)	(0.002)	(0.012)
Past smoker	−0.016	−0.004	0.020
	(0.011)	(0.002)	(0.013)
Unmarried	0.008	0.002	−0.009
	(0.008)	(0.002)	(0.010)
Married, spouse away	0.015	0.003*	−0.018
	(0.011	(0.002	(0.013
High school and above	0.034***	0.004***	−0.038***
	(0.011)	(0.001)	(0.012)
Illiterate	−0.010	−0.003	0.013
	(0.006)	(0.002)	(0.008)
Urban *hukou*	0.051***	0.007***	−0.059***
	(0.010)	(0.001)	(0.010)
Middle	−0.048***	−0.007***	0.055***
	(0.008)	(0.001)	(0.009)
Northeast	−0.036***	−0.004*	0.040***
	(0.012)	(0.002)	(0.014)

Continued on next page

Table 4.12 (*Continued*)

	(1)	(2)	(3)
West	−0.064***	−0.013***	0.077***
	(0.007)	(0.002)	(0.008)

Notes: The average marginal effects of self-ranked health state are based on the estimation results from the baseline ordered probit model including expenditure (Table 4.10, column 1). The reference predicted probabilities refer to a married female between 45 and 49 years of age, who comes from a household where the logarithm of the per capita annual household expenditure is equal to the average of all observations, is living with her spouse, and who is not a regular drinker, does not have any chronic disease, is literate but has less than a high school diploma, is not disabled, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.13: Predicted Probabilities of Health State by Type of Health Insurance Plan (Sensitivity Check: Including Consumption Expenditure, Male)

Health state	(1) Employee health insurance	(2) Resident health insurance	(3) No health insurance
Good	0.511***	0.452***	0.453***
	(0.019)	(0.020)	(0.025)
Fair	0.431***	0.470***	0.470***
	(0.014)	(0.013)	(0.016)
Poor	0.058***	0.077***	0.077***
	(0.006)	(0.008)	(0.009)

Notes: Predicted probabilities of self-ranked health state are based on the estimation results from the ordered probit model including expenditure (Table 4.9, column 1). The reference predicted probabilities refer to a married male, between 45 and 49 years of age, who comes from a household where the logarithm of the per capita annual household expenditure is equal to the average of all observations, is living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Province dummies Since health insurance policies and health care resources also vary at the province level, I conduct a sensitivity check that uses province dummies as the geographical factor. Results are shown in Appendix Tables B.5 and B.6. The estimated coefficients of the employee health insurance in the baseline model vary slightly for both sexes (−0.159 compared to −0.161 for males and −0.102 compared to −0.111 for females). Coefficients of other control variables either do not change or

Table 4.14: Predicted Probabilities of Health State by Type of Health Insurance Plan (Sensitivity Check: Including Consumption Expenditure, Female)

Health state	(1) Employee health insurance	(2) Resident health insurance	(3) No health insurance
Good	0.439***	0.412***	0.420***
	(0.018)	(0.017)	(0.023)
Fair	0.471***	0.486***	0.482***
	(0.011)	(0.010)	(0.013)
Poor	0.090***	0.102***	0.098***
	(0.008)	(0.008)	(0.010)

Notes: Predicted probabilities of self-ranked health state are based on the estimation results from the ordered probit model including expenditure (Table 4.10, column 1). The reference predicted probabilities refer to a married female, between 45 and 49 years of age, who comes from a household where the logarithm of the per capita annual household expenditure is equal to the average of all observations, is living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

change at the third decimal point.

The wave effect In Tables 4.3 and 4.4, certain estimated coefficients are different in the 2011 and 2013 waves. To test whether the difference in coefficients is significant, I consider a model that includes interaction terms of these independent variables and the wave dummy. Appendix Table B.7 shows that the association between employee health insurance and health status is statistically different in the 2011 and 2013 waves for both males and females, and the same pattern holds for chronic disease. For males, the association between age and health state is also statistically different across the two waves.

4.5.2 Health state transitions and health insurance plans

I now test whether an individual's health insurance plan is also a factor that matters for transitions of health states over time. Table 4.15 shows the estimated ordered probit coefficients based on equation (4.6). Column 1 reports the estimation results

for the baseline model with the full set of controls and pooled data (all waves). The reference group in the baseline model is a male person without an urban *hukou* aged 45 to 49 living with his spouse, and who is literate, but has a less than high school diploma, was in fair health in the previous survey wave, has no chronic disease or disability, is not a regular drinker, was not a past smoker, and is living in the Eastern region. Columns 2 and 3 only use the 2011–13 and 2013-15 waves, respectively. Columns 4 and 5 also use the pooled data, but report the results for males and females separately. Table 4.16 shows the estimation results for a parsimonious model, which only includes type of insurance, age, sex, and region as independent variables.

The estimated associations between health insurance plan and current health are stronger in the reduced form models. Nevertheless, there is a statistically significant association between the current-period self-ranked health state and health insurance plan regardless of the previous-period self-ranked health state. In Table 4.15, the estimated coefficient for employee health insurance becomes insignificant while only using the 2011–13 wave, and when estimating the baseline model for males and females separately. In the parsimonious model, the estimated coefficients of the employee health insurance are similar. Hence, I will focus on discussing the model which use the pooled dataset and the female indicator (column 1 in Tables 4.15 and 4.16).

In the baseline model (column 1 in Table 4.15), the sign of the estimated coefficient of the interaction term involving the employee health insurance and self-ranked good health in the previous period is negative. This implies that the association between types of health insurance and health states is stronger for those who were in good health in the previous period compared to those who were in fair health. As well, the estimated coefficients of the employee health insurance are negative and significant regardless of the initial self-ranked health state, indicating that those who have the employee health insurance are less likely to transit to or stay in a poorer health state regardless of their previous health state. The interaction terms of age

dummies and initial health state imply a strong association between age and self-ranked health for those who were in good health in the previous period. However, it is not strong for those who were initially in poor health.

Table 4.15: Determinants of Health State Transitions: Estimation Results of the Ordered Probit Model

Dependent variable: Current wave (t) self-ranked health state

Data	Pooled	2011-13	2013-15	Pooled, Male	Pooled, Female
Variable	(1)	(2)	(3)	(4)	(5)
Employee health insurance	−0.064**	−0.017	−0.108**	−0.063	−0.060
	(0.033)	(0.048)	(0.044)	(0.044)	(0.048)
No health insurance	0.063	0.066	0.059	0.025	0.094
	(0.046)	(0.057)	(0.080)	(0.068)	(0.060)
Good_{t-1}	−0.654***	−0.583***	−0.735***	−0.682***	−0.563***
	(0.045)	(0.059)	(0.069)	(0.059)	(0.056)
Poor_{t-1}	0.844***	0.752***	0.971***	0.792***	0.875***
	(0.054)	(0.067)	(0.082)	(0.081)	(0.059)
Employee health insurance	−0.104**	−0.055	−0.162**	−0.136**	−0.076
$\times \text{Good}_{t-1}$	(0.053)	(0.073)	(0.070)	(0.068)	(0.082)
Employee health insurance	0.046	0.004	0.085	0.141*	−0.095
$\times \text{Poor}_{t-1}$	(0.062)	(0.084)	(0.086)	(0.085)	(0.088)
No health insurance	−0.132	−0.159	−0.118	−0.252**	−0.021
$\times \text{Good}_{t-1}$	(0.083)	(0.104)	(0.136)	(0.120)	(0.116)
No health insurance	−0.026	−0.042	0.040	0.044	−0.079
$\times \text{Poor}_{t-1}$	(0.091)	(0.111)	(0.163)	(0.149)	(0.115)
Female	0.068*	0.010	0.126**	-	-
	(0.037)	(0.050)	(0.051)	-	-
Female $\times \text{Good}_{t-1}$	0.068*	0.010	0.126**	-	-
	(0.037)	(0.050)	(0.051)	-	-
Female $\times \text{Poor}_{t-1}$	0.002	0.018	−0.023	-	-
	(0.038)	(0.050)	(0.053)	-	-
Chronic disease	0.340***	0.345***	0.341***	0.336***	0.344***
	(0.016)	(0.023)	(0.023)	(0.023)	(0.023)
Disability	0.128***	0.197***	0.056*	0.150***	0.110***
	(0.020)	(0.029)	(0.030)	(0.028)	(0.029)
Regular drinker	−0.097***	−0.134***	−0.055**	−0.105***	−0.068*
	(0.018)	(0.026)	(0.025)	(0.021)	(0.036)
Past smoker	0.032	0.020	0.051*	0.067***	−0.038
	(0.020)	(0.028)	(0.030)	(0.024)	(0.036)
Unmarried	0.002	−0.043	0.044	0.031	0.001
	(0.025)	(0.035)	(0.035)	(0.041)	(0.031)
Married, spouse away	0.018	−0.034	0.085*	−0.013	0.040
	(0.031)	(0.042)	(0.048)	(0.047)	(0.041)
High school and above	−0.098***	−0.122***	−0.077**	−0.090***	−0.112***
	(0.024)	(0.034)	(0.034)	(0.030)	(0.039)
Illiterate	−0.016	0.003	−0.039	−0.058*	0.006
	(0.019)	(0.027)	(0.027)	(0.035)	(0.022)
Urban *hukou*	−0.105***	−0.121***	−0.091***	−0.117***	−0.092***
	(0.024)	(0.036)	(0.033)	(0.035)	(0.034)
Middle	0.093***	0.103***	0.084***	0.071***	0.108***
	(0.019)	(0.027)	(0.027)	(0.026)	(0.025)

Continued on next page

Table 4.15 (*Continued*)

	(1)	(2)	(3)	(4)	(5)
Northeast	0.029	0.110**	−0.053	0.002	0.066
	(0.030)	(0.043)	(0.042)	(0.042)	(0.041)
West	0.123***	0.161***	0.083***	0.086***	0.156***
	(0.018)	(0.026)	(0.026)	(0.026)	(0.024)
Threshold μ_1	−0.479***	−0.473***	−0.469***	−0.502***	−0.510***
	(0.033)	(0.046)	(0.047)	(0.044)	(0.034)
Threshold μ_2	1.214***	1.187***	1.265***	1.223***	1.158***
	(0.034)	(0.047)	(0.048)	(0.045)	(0.035)
N	27,229	13,467	13,762	13,066	14,163

Notes: The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married male with the resident insurance plan, between 45 and 49 years of age, living with his spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate, but has less than a high school diploma, does not have an urban *hukou*, and lives in the Eastern region. All models include age variables grouped by 5-year windows, and the interaction terms for each age group and the dummies of self-ranked health state in the previous wave $(t-1)$, which are reported in Appendix Table B.8. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table 4.16: Determinants of Health State Transitions: Estimation Results of the Ordered Probit Model (Parsimonious Model)

Dependent variable: Current wave (t) self-ranked health state

Data	Pooled	2011-13	2013-15	Pooled, Male	Pooled, Female
Variable	(1)	(2)	(3)	(4)	(5)
Employee health insurance	-0.164^{***}	-0.145^{***}	-0.184^{***}	-0.162^{***}	-0.167^{***}
	(0.026)	(0.037)	(0.035)	(0.034)	(0.038)
No health insurance	0.013	0.009	0.011	-0.016	0.035
	(0.046)	(0.055)	(0.079)	(0.068)	(0.058)
Good_{t-1}	-0.737^{***}	-0.673^{***}	-0.811^{***}	-0.764^{***}	-0.639^{***}
	(0.045)	(0.059)	(0.068)	(0.059)	(0.056)
Poor_{t-1}	0.936^{***}	0.858^{***}	1.049^{***}	0.881^{***}	0.953^{***}
	(0.054)	(0.067)	(0.082)	(0.081)	(0.059)
Employee health insurance	-0.127^{**}	-0.084	-0.176^{**}	-0.157^{**}	-0.102
$\times \text{Good}_{t-1}$	(0.053)	(0.072)	(0.070)	(0.068)	(0.082)
Employee health insurance	0.035	0.007	0.066	0.128	-0.103
$\times \text{Poor}_{t-1}$	(0.062)	(0.084)	(0.086)	(0.084)	(0.088)
No health insurance	-0.146^{*}	-0.167	-0.134	-0.272^{**}	-0.032
$\times \text{Good}_{t-1}$	(0.082)	(0.103)	(0.135)	(0.120)	(0.115)
No health insurance	-0.001	-0.011	0.057	0.061	-0.047
$\times \text{Poor}_{t-1}$	(0.091)	(0.110)	(0.162)	(0.149)	(0.114)
Female	0.092^{***}	0.110^{***}	0.076^{***}	-	-
	(0.018)	(0.026)	(0.025)	-	-
Female $\times \text{Good}_{t-1}$	0.078^{**}	0.034	0.123^{**}	-	-
	(0.037)	(0.050)	(0.051)	-	-
Female $\times \text{Poor}_{t-1}$	-0.014	0.000	-0.039	-	-
	(0.038)	(0.049)	(0.052)	-	-
Middle	0.102^{***}	0.109^{***}	0.095^{***}	0.087^{***}	0.111^{***}
	(0.019)	(0.027)	(0.027)	(0.025)	(0.024)
Northeast	0.030	0.109^{**}	-0.048	-0.001	0.058
	(0.030)	(0.043)	(0.042)	(0.041)	(0.039)
West	0.144^{***}	0.183^{***}	0.103^{***}	0.111^{***}	0.173^{***}
	(0.018)	(0.026)	(0.026)	(0.025)	(0.024)
Threshold μ_1	-0.636^{***}	-0.602^{***}	-0.666^{***}	-0.679^{***}	-0.693^{***}
	(0.027)	(0.037)	(0.038)	(0.037)	(0.031)
Threshold μ_2	1.034^{***}	1.031^{***}	1.046^{***}	1.019^{***}	0.954^{***}
	(0.027)	(0.038)	(0.039)	(0.037)	(0.032)
N	27,229	13,467	13,762	13,066	14,163

Notes: The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state. The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married male with resident insurance, aged between 45 and 49 years of age, living with his/her spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate, but has less than a high school diploma , does not have an urban *hukou*, and lives in the Eastern region. All models include age variables grouped by 5-year windows, and the interaction terms for each age group and the dummies of self-ranked health state in the previous wave $(t-1)$, which are reported in Appendix Table B.9. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

In this model, the estimated coefficients for the female indicator are positive and significant regardless of the initial self-ranked health state, indicating that females are more likely to transit to or stay in relatively poorer health at all levels of health state. As well, the association between health insurance plan and health transition varies by sex. Interaction terms of female indicators and initial health state are positive—female is associated with a lower probability of remaining in good health and a higher probability of transitioning into poor health.

I use these estimates from the ordered probit model to estimate the dynamic (3×3) health transition matrices. I focus on the estimates from column 1, Table 4.15, based on pooled observations and report them in Table 4.17. I report the results of the transition matrices for the base line age group, which is 45-49 years old.

Table 4.17: Estimated Health Transition Matrices by Health Insurance Plan: Baseline Model

Employee health insurance				Resident health insurance			
	Good	Fair	Poor		Good	Fair	Poor
Good	0.634	0.345	0.021	Good	0.569	0.400	0.031
Fair	0.339	0.560	0.101	Fair	0.316	0.572	0.112
Poor	0.096	0.555	0.349	Poor	0.093	0.551	0.356

Notes: The reference group refers to a married male, between 45 and 49 years of age, living with his spouse, who does not have any chronic disease or disability, is literate but has less than a high school diploma, does not drink or smoke, does not have an urban *hukou*, and lives in the Eastern region. Based on the estimation results from the ordered probit model reported in Table 4.15, column 1, and are estimated using Equation (4.7).

Focusing on the differences in health transitions between those who have the employee health insurance and those who have the resident health insurance, while keeping other control variables the same, I observe a significant association between self-ranked health state transition probabilities and the health insurance plan (Table 4.17). The employee health insurance is uniformly associated with better health transitions. Regardless of their self-ranked health state in the previous survey year, those who had the employee health insurance in the previous period are less likely to report being in fair or poor health states in the current period, and more likely to report being in good health. For example, for those who have the employee health insurance and who were in good health in the previous survey wave, the likelihood of staying in good health is 63%, whereas for those with the resident health insurance, the same likelihood is only 57%. The respondents with the employee health insurance who are in poor health in the previous survey wave have about 35% likelihood of staying in poor health, while for those with the resident health insurance the likelihood is 36%. Moreover, I obtain similar relative probabilities (Table 4.18, panel a) when I use the estimates from the parsimonious model specification shown in Table 4.16, column 1.

I use the parsimonious model to show the association between sex and health transitions, as well as between region and health transitions. Although the absolute

Table 4.18: Estimated Health Transition Matrices by Health Insurance Plan: the Parsimonious Model

Employee health insurance				Resident health insurance			
a) Male, Eastern region							
	Good	Fair	Poor		Good	Fair	Poor
Good	0.652	0.328	0.020	Good	0.540	0.421	0.038
Fair	0.318	0.566	0.115	Fair	0.262	0.587	0.151
Poor	0.075	0.515	0.410	Poor	0.058	0.481	0.461
b) Female, Eastern region							
	Good	Fair	Poor		Good	Fair	Poor
Good	0.588	0.383	0.029	Good	0.472	0.473	0.055
Fair	0.286	0.579	0.134	Fair	0.210	0.596	0.194
Poor	0.064	0.495	0.441	Poor	0.041	0.431	0.529
c) Male, Western region							
	Good	Fair	Poor		Good	Fair	Poor
Good	0.566	0.401	0.033	Good	0.451	0.488	0.061
Fair	0.243	0.592	0.165	Fair	0.203	0.596	0.200
Poor	0.048	0.453	0.499	Poor	0.039	0.423	0.538

Notes: 'Good', 'Fair', and 'Poor' refer to self-ranked health states. The reference group refers to an individual aged between 45 and 49 years, based on the estimation results from the ordered probit model reported in Table 4.16, column 1 (panel a), and are estimated using Equation (4.7). The reference group refers to a male who lives in the Eastern region. Remaining panels use the coefficient estimates on females (panel b) and the Western region (panel c).

value of the estimated coefficients of employee health insurance is larger in the parsimonious model, compared to the baseline model for the health state transitions, the health transition matrices are similar for the baseline group across these two models. Moreover, the parsimonious model is more frequently adopted in general equilibrium models. A comparison of panels a and b in Table 4.18 indicates that the coefficients of female indicators are significant. For those covered by the employee and resident health insurances alike, females are less likely to report a good health state than males given that their self-ranked health was good in the previous census wave, and they are less likely to transition from fair or poor health to good health. For example, while the females with the employee health insurance have a 59% probability of remaining in good health, males have about 65% probability, and while females with the employee health insurance only have a 6% probability of transitioning from poor to good health, males have a 8% probability. These probabilities are reported for those who are 45 to 49 years old, and live in the Eastern region.

Using the HRS, Jung (2020) finds gender gaps in health transitions and thus this is not the first study to observe that females are more likely to transition over time into relatively poor health using HRS data. However, he also find that in the US, young women have worse health than their male counterparts at younger ages, but as individuals age, being female is associated with higher probabilities of transitioning into better health states than men.

There are also regional differences. A comparison of panels a and c in Table 4.18 indicates that the estimated coefficients of regions are significant. Males in the West with the employee health insurance plan have a lower likelihood of reporting good health compared to their Eastern counterparts (results for females are similar) given that their health was good in the previous survey wave, and they are less likely to transit from fair or poor health to good health over survey years. For example, for the males who are between 45 and 49 years of age, and living in the East, the probability

of remaining in good health is 65%, while it is only 57% for those in the West. Thus, compared to those who live in more economically developed regions, those who live in more deprived regions have lower probabilities of remaining in good health and higher probabilities of remaining in poor health.

The size of the association between health insurance plans and health transitions varies slightly across the survey waves (Table B.10). In particular, a comparison of the estimation results by survey waves suggests that for those with both the employee and resident health insurance plans, the respondents to the 2013–15 waves have a higher probability of remaining in good health. These differences are more marked in the case of the employee health insurance plan. While those with the employee health insurance plan in the 2011–13 waves have a higher probability of moving from poor to fair health, this is not the case for those with the resident insurance. Overall, while the results vary across the survey years, these differences do not appear to be systematic across health insurance plans.

Overall, the differences in self-ranked health state accounted for by differences in the coverage of health insurance plans suggest that the on-going public insurance reforms in China will likely be effective in mitigating the self-ranked health gaps between those who have the more generous employee health insurance plan and those who have the less generous resident health insurance plan—although by themselves, these reforms are unlikely to eliminate the gap between males and females.

4.6 Conclusion

Using the CHARLS dataset and ordered probit models, I found that the currently available two health insurance plans that vary in their benefits are not only systematically associated with self-ranked health, but also affect individual health state transitions over time. Those who have the employee health insurance have higher probabilities of reporting good health, as well as of remaining in good health and

recovering from poor health. Aside from these overall associations, I also find a considerable sex gap and significant regional disparities in self-ranked health state, as well as the health transitions. Females and the residents of the less developed regions are more vulnerable, suggesting that the health benefits afforded by the employee health insurance are even larger for these groups. The findings are relatively robust[5] across models using different independent variables.

My findings support the results in the existing literature, which suggest that despite achieving nearly universal coverage, neither of the two concepts of health equity has been achieved in China. Moreover, my findings indicate that the main goal of the public insurance reform should now shift from improving the coverage rate and integrating the rural resident health insurance and urban resident health insurance to closing the gap between the generosity of the employee health insurance and the resident health insurance plans. Although other studies have found that integrating the rural and urban resident health insurance, as well as increasing the coverage rate and benefits in the resident health insurance plans, are effective in improving the health of the target population, the inequalities in self-ranked health state between the individuals covered by the employee health insurance and the resident health insurance are still large. Thus, integrating the insurance plan pools, and elevating the benefits of the resident health insurance plan to those of the employee health insurance plan would have significant health benefits.

This chapter, however, did not explore the causal relationship or the exact pathways through which health insurance affects health. In China, health insurance plans can influence an individual's health through multiple pathways. For example, a more generous health insurance plan can increase the likelihood of utilizing health services, as the out-of-pocket expenditure is relatively lower. Moreover, by reducing economic

[5] I use three criteria to determine whether the results are consistent: (1) the sign of the coefficients, (2) the significance level, and (3) changes in coefficient size while accounting for corresponding standard errors.

uncertainty, a more generous plan can improve an individual's mental health. Individuals with more generous employee health insurance may also use more preventative care services, as they have a forced savings account, which can only be used to purchase preventative health care, health-promoting products, and hospital visits. Since eligibility for employee health insurance is tied to employment status, it is difficult to identify the causal effects on health when a person switches from resident health insurance to employee health insurance.

In addition to preventative health care through employee health insurance, many workplaces in China that provide employee health insurance also offer regular health check-ups as part of their fringe benefits. These check-ups typically cover a range of preventative services such as blood pressure monitoring, cholesterol checks, and other routine screenings. These preventative check-ups are not part of the employee health insurance plan, but individuals whose workplaces offer health insurance plans also receive these as fringe benefits. It is worth noting that this dissertation does not disentangle the effects of these fringe benefits from the effects directly attributable to the employee health insurance.

This study has several limitations. First, I do not have the information on those who exit the survey potentially due to death. Thus, the likelihood of remaining in poor health or having recovered from poor to fair may somewhat be overestimated. Second, the resident health insurance plan is prone to self-selection. While there is some evidence that those without health insurance choose not to obtain a plan, the number of individuals without any health insurance is very small, in part because resident health insurance premiums are not too onerous and there is a one year waiting period between the enrolment in the plan and eligibility for benefits. Third, the health state used here is self-ranked, and future work may combine these measures with those derived from "objective" measures of health, although such measures have their own difficulties of interpretation. Even though I control for chronic disease and

disability, these are also self reported. The underestimation of these chronic diseases is severe, suggesting that self-ranked health state may overstate the health of Chinese population.

Chapter 5

Health Insurance and Retirement

5.1 Introduction

Prior to the economic reforms that began in 1978, the elderly in China had little social safety net to rely on, resulting in what one commentator described as "ceaseless toil", which meant that the elderly had to work their entire lives without sufficient social security support and possibility to retire (Davis, 1991). Using the data after the economic reform, Benjamin et al. (2003) find that "ceaseless toil" still exists in Chinese's elderly people's retirement pattern. According to Benjamin et al. (2003), after the reforms, this pattern persisted. This raises intriguing questions about the broader benefits of a more generous health insurance plan, particularly its potential to influence an individual's retirement decision[1] through enhancing their financial security in old age. Given that health insurance may act as a crucial component of perceived financial security by alleviating health-related financial burdens during old age, an important question arises: how does an individual's retirement decision differ across health insurance plans? Specially, do individuals with the more generous employee health insurance plan retire early due to the health insurance plan's role in securing financial stability in later life? Using three waves of the CHARLS dataset,

[1]Here, "retirement" is defined economically: a person is considered retired if they have previously worked but are no longer engaged in wage-earning or self-employed activities. It should be distinguished with mandatory (or administrative) retirement, which refers to the requirement that employees of a government institution or a formal sector employer covered under the Employee Pension program must leave their original position, process their retirement, and start collecting their pension at the mandatory retirement age, although they can find another job and stay in the labour force.

this chapter aims to explore how the type of health insurance is associated with work-to-retirement transitions among the middle-aged and elderly in China. Retirement marks a crucial transition in an individual's life and is often associated with improvements in well-being (Zhang et al., 2022; Che and Li, 2018). This study explores the implications of health insurance policies on broader economic well-being, particularly how they may encourage retirement at an older age.

In China, factors affecting an individual's retirement decision have attracted growing attention. Factors that have been examined in existing studies include pension plans, the individual's health status, the spouse's health status, the number of children, and the children's education level (Benjamin et al., 2003; Giles et al., 2015). However, research on the effect of health insurance on retirement patterns in China has been neglected, and little is known about the implications of different health insurance plans on an individual's retirement decision. Part of the reason this field has not been studied in China is that the mandatory retirement ages are the key factors determining retirement for those with the employee pension plan (Giles et al., 2015). In China, individuals with the employee pension plan have to retire at their mandatory retirement ages, which vary by sex and type of work (see chapter 2.3); therefore, it is difficult to identify the effect of other variables on their retirement decisions. However, they can still find another job or be self-employed after mandatory retirement, and remain in the labour force. There are no financial penalties or other restrictions on post-retirement employment. Hence, it is still applicable to identify the effect of the health insurance plan on an individual's retirement decision. Moreover, only about 20% of the middle aged and elderly population in the CHARLS dataset have the employee pension plan, which means about 80% of the target population in the CHARLS dataset do not face mandatory retirement ages. Without the possible influence of mandatory retirement ages on retirement, it is appropriate to examine the effect of the health insurance plan on an individual's retirement decision.

To see whether an individual's retirement decision is associated with the type of health insurance, in this chapter, I aim to answer the following question: among the middle aged and elderly population in China, do those with the more generous health insurance plan have larger probabilities of transitioning from work to retirement in each age group? To answer this question, I use three waves of the CHARLS dataset and a probit model to estimate the probability of transitioning from work to retirement for individuals with the employee health insurance and those with the resident health insurance.

While the link between the type of health insurance and an individual's retirement decision has not been studied in China, health insurance is generally one of the factors that affect one's retirement decision (Madrian et al., 1994; Karoly and Rogowski, 1994; Johnson et al., 2003; French and Jones, 2011; Kapur and Rogowski, 2011). There are at least two channels through which health insurance could affect an individual's work-to-retirement transition.

The first channel through which health insurance could affect an individual's work-to-retirement transition, which is also the focus of this chapter, is by playing a role in preventing the risk of high medical expenses. In China, the employee health insurance has a higher coverage rate and is maintained during retirement. For an individual covered by the employee health insurance, there will be less financial uncertainty in the future resulting from potential medical expenses. The existing literature has shown that individuals with a history of adverse health shocks would have a smaller likelihood of encountering catastrophic health expenditures if they have a more generous health insurance plan (Heeley et al., 2009; Li et al., 2012, 2014; Sun and Lyu, 2020). Since individuals with the more generous health insurance plan face smaller risks of catastrophic health expenditures, they need less precautionary savings. Thus, the first channel implies that having a more generous health insurance plan may increase the likelihood of a work-to-retirement transition.

The second channel is through an individual's health state. Under this channel, there are two effects–an income effect and a substitution effect–which work in opposite directions. Having a more generous health insurance plan is associated with reporting better health (see chapter 4), and as a result, individuals with better health can earn more income. From the income effect point of view, individuals with the more generous employee health insurance accumulate more wealth, and then retire earlier. However, from the substitution effect point of view, individuals who have better health and could earn higher income at old age have incentives to postpone retirement, as the opportunity cost of retirement is higher if they retire. Overall, the effect of health insurance on one's retirement transition through their health state is intermediated through complex income and substitution effects (Sickles and Taubman, 1986; Laitner and Silverman, 2005; Cutler et al., 2006; Jürges et al., 2014).

Pension eligibility is also an important factor in an individual's transition to retirement, not only because, in China, individuals with an employee pension plan have a mandatory retirement age, while others do not. It is also the case that pensions affect the resources available to an individual during retirement (Madrian et al., 1994; Karoly and Rogowski, 1994; Johnson et al., 2003; French and Jones, 2011; Kapur and Rogowski, 2011). However, health insurance remains an important determinant of an individual's retirement decision, even when a pension is available, due to its potential to provide financial protection against significant health shocks during retirement.

What complicates the identification of the more generous health insurance effect on an individual's retirement decision is that, in China, the health insurance system interacts with the pension system. The employee health insurance is employment-based; only certain jobs offer employee health insurance. These jobs are usually in the non-agricultural sector and are long-term permanent positions. Typically, these jobs also offer the employee pension plan. Hence, those who have the employee health insurance plan are more likely to have the employee pension plan. However, about

23% of individuals with employee health insurance do not have the employee pension plan (see chapter 3), which allows me to separately observe the health insurance effect from the pension effect on an individual's retirement decision.

Given the considerable differences in individuals' retirement behaviors that their pension plan status could cause, in terms of both mandatory retirement ages and as a source of post-retirement income, I estimate the work-to-retirement transitions for individuals with different pension plans separately. Since health insurance can affect an individual's retirement decision by affecting their health status, I control for self-ranked health state while estimating an individual's work-to-retirement transition.

In the context of China, and in relation to an individual's pension plan and the two channels through which health insurance affects an individual's work-to-retirement transition, this chapter presents three main findings. First, a more generous health insurance plan is associated with a higher likelihood of transitioning to retirement by more than 10 percentage points for both groups: those with and without an employee pension plan. Second, among individuals with an employee pension plan, this effect is more pronounced for females. When the model is estimated by sex, for females with an employee pension, having employee health insurance is associated with a 16 percentage point higher likelihood of transitioning to retirement, while no significant association is found for males. Among individuals without an employee pension plan, the health insurance effect on the work-to-retirement transition is similar for females (12 percentage points) and males (11 percentage points). Third, reporting poor health is associated with a higher likelihood of transitioning to retirement. The association between health and work-to-retirement transition is less pronounced for individuals without an employee pension plan, both for females and males. For females with an employee pension plan, reporting poor health is associated with a 14 percentage point higher likelihood of transitioning to retirement compared to reporting fair health. For females without an employee pension plan, poor health is associated with only a three

percentage point higher likelihood of transition. A similar pattern is observed in males (9 percentage points versus 3 percentage points).

My results imply that a generous health insurance plan is an important factor in an individual's transition from work to retirement, especially for those without an employee pension plan. My findings also suggest that individuals remain in the labour force even in poor health, as they lack financial protection from a pension plan with a strong safety net. A strong safety net is crucial, as it influences an individual's retirement decision through health-spending related financial security.

5.2 Related Literature

This study contributes to three strands of literature. First, my study contributes to the body of literature which studies the determinants of an individual's retirement decision in China (Benjamin et al., 2003; Giles et al., 2015). Second, my study also contributes to a large international literature examining the effect of health insurance on an individual's retirement decision by providing the evidence in China. My findings that health insurance plays a role in an individual's retirement decision is consistent with the evidence from the United States. Third, my findings support the existing evidence from China that more generous health insurance decreases the risk of catastrophic health expenditures (Heeley et al., 2009; Li et al., 2012, 2014; Sun and Lyu, 2020), while pointing out a neglected link in the Chinese context that health insurance affects individuals' transitions to retirement.

Direct evidence of the effect of health insurance plans on an individual's retirement decision is limited in China. Most existing evidence on how health insurance affects retirement is from the United States (see Madrian et al. (1994); Karoly and Rogowski (1994); Johnson et al. (2003); French and Jones (2011); Kapur and Rogowski (2011)). In the United States, health insurance is often linked to employment, although in 1994, only less than half of the work force was employed by firms offering

health insurance to retirees in 1994 (Madrian et al., 1994).

In the United States, post-retirement health insurance is found to affect an individual's retirement decision. The existing literature can be divided into two groups depending on the explanatory variable, but the results are consistent. One group of studies compares the probability of retirement and the age of retirement for individuals with and without post-retirement health insurance. These studies find that employer-based post-retirement health insurance increases one's probability of retirement before age 65, and those with the post-retirement health insurance tend to retire 5-16 months earlier than those without post-retirement health insurance (Madrian et al., 1994; Karoly and Rogowski, 1994; Kapur and Rogowski, 2011). But the effect of post-retirement health insurance on individual's retirement decision varies across sex and marital status. Kapur and Rogowski (2011) find that in the United States, having a retiree health insurance plan does not increase the probability of retirement for single men, and the role of health insurance on women's retirement decision hinges on their husbands' employment status. Married women are more responsive to the availability of post-retirement health insurance from their own jobs if their spouses have retired. However, studies in the US primarily examine the impact of having any form of health insurance on retirement decisions, rather than the impact of insurance generosity.

The other group of studies examines the policy change in Medicare health insurance. French and Jones (2011) study the effect of changing the age of eligibility of Medicare on an individual's retirement decision. They find that raising the Medicare eligibility age by two years results in individuals aged between 60 and 69 working on average for an additional 27 days. Medicare offers a buy-in option which allows individuals to become eligible at age 62, and the premiums of the buy-in option change. Johnson et al. (2003) find significant but small effects of changing the premiums; a $1,000 increase in the premium reduces the probability of early retirement by 0.17

percentage points for males and 0.24 percentage points for females.

Although there is limited work on China exploring the link between health insurance and retirement, there is evidence indicating that health insurance decreases the risk of catastrophic medical expenses (Heeley et al., 2009; Li et al., 2012, 2014; Sun and Lyu, 2020). Moreover, the more generous an individual's health insurance is, the lower the risk of experiencing catastrophic health expenditures (Sun and Lyu, 2020). In principle, when individuals have a smaller risk of catastrophic health expenditures, they need less precautionary savings and may have a tendency to retire early. Hence, these findings indicate that more generous health insurance could affect an individual's retirement decision through enhanced financial security.

In contrast to the aforementioned studies, my study focuses on whether differences in the generosity of different health insurance plans matter for an individual's probability of transitioning from work to retirement in China. Unlike the US studies, which adopt the probability of retirement, early retirement, and age of retirement as indicators of a retirement decision (Madrian et al., 1994; Karoly and Rogowski, 1994; Kapur and Rogowski, 2011), I use the probability of transitioning from working to retirement, which is the probability of retiring in the next observed period conditional on working in the current period, as an indicator of retirement. Compared to the probability of being retired, the work-to-retirement transition is not necessarily endogenous[2]. Moreover, an advantage of using the probability of transitioning from work to retirement as an indicator is that it offers a dynamic view of retirement decisions. It allows for the observation of retirement as a process, not just an outcome. This dynamic measure can capture how individuals with different types of health

[2]Several control variables can be endogenous if the dependent variable is the probability of being retired. For example, although health affects the probability of being retired, being retired also affects an individual's health. Several studies find that retirement affects an individual's health (Lei et al., 2011; Che and Li, 2018; Feng et al., 2019; Xie et al., 2021; Zhang et al., 2022) and health-related behaviors (Zhang et al., 2018; Feng et al., 2020; Fang and Shi, 2022). Typically, individuals with an employee pension plan are used to study the effect of retirement on individual health (Feng et al., 2019; Zhang et al., 2022; Che and Li, 2018).

insurance plans approach retirement, which could be obscured when only considering the static state of being retired or the age at which one retires (Tang and Burr, 2015; National Academies of Sciences et al., 2022).

Studies in the United States also consider wealth as one of the factors that affect an individual's retirement decision, but they use the present value of pension benefits as the indicator of an individual's wealth at retirement (Madrian et al., 1994; Karoly and Rogowski, 1994; Johnson et al., 2003; French and Jones, 2011; Kapur and Rogowski, 2011). However, in China, rural residents account for 60% of the population, and they do not have an employee pension. A fraction of the rural residents have the rural pension plan, but having the rural pension plan is not a significant determinant of retirement for older adults in rural China (Ning et al., 2016a). Transfers from children are the primary source of old-age income for rural residents, amounting for about 70% of their total economic resources after retirement. For urban residents, pensions are the largest source of income, accounting for about 83% of their total economic resources after retirement (Giles et al., 2015).

Besides using the present value of pension benefits as an individual' wealth (Johnson et al., 2003; French and Jones, 2011; Kapur and Rogowski, 2011), there are other methods of incorporating pensions into the model. Karoly and Rogowski (1994) control for the type of pension plan, while Madrian et al. (1994) divide the entire sample by pension plan to examine the heterogeneous health insurance effects among different pension plans. My approach is similar to that of Madrian et al. (1994). They study the effect of retiree health insurance on retirement age by pension plans, which are the defined benefit pension and defined contribution pension in the US context. However, contrary to their study, which finds that the direction of effect on retirement age varies across type of pension plan, I find that in the Chinese context, the directions of effect of employee health insurance on an individual's work-to-retirement transitions are consistent across pension plans.

5.3 Model specification

To model the work-to-retirement transition, I consider a probit model with lagged independent variables

$$Pr(R_{it} = 1|R_{i,t-1} = 0, X_{i,t-1}) = \Phi(X'_{i,t-1}\beta + Z'_{i,t}\gamma), \tag{5.1}$$

where R_{it} is individual i's retirement status at survey wave t, with a value of 1 indicating "retirement" and 0 indicating "working". $R_{i,t-1}$ is the retirement status at $t-1$, so that this model is estimated conditional on $R_{i,t-1} = 0$ (working at $t-1$).

All independent variables are observed at $t-1$, except for sex and age. Using lagged independent variables allows me to predict the current retirement status, given an individual's characteristics in the last survey wave. It is also consistent with the fact that an individual's consideration and the processing of retirement application often take time.

$X_{i,t-1}$ is a vector of individual characteristics observed in the previous survey wave. Individual characteristics included are health state, education level, marital status, type of health insurance plan, region of residence, total number of children, an indicator for having children in school, and children's highest education level[3]. I also include the logarithm of per capita household annual expenditure for sensitivity analysis. The reason for not including expenditure in the baseline model is that, for individuals facing mandatory retirement ages, income or expenditure does not have a material impact on work-to-retirement transitions. I use expenditure instead

[3]Savings constitute only a small portion of their post-retirement financial resources, for both the urban and rural residents, at 1.8% and 4.2%, respectively (Giles et al., 2015). Other categories of financial resources include children, pension and others. Benjamin et al. (2003) used housing, farm productive assets, and consumer durables to construct household wealth, and found that wealth is not a significant factor associated with rural residents' retirement decisions. For rural residents, the most important source of wealth is housing, which accounts for 78% of household wealth (Benjamin et al., 2003). However, housing equity is not included in retirement studies as households rarely will liquidate it finance their retirement (Johnson et al., 2003). In the CHARLS dataset, wealth will be constructed in the same way as that in the study by Benjamin et al. (2003). Hence, I do not include a wealth variable in my analysis.

of income as the indicator of economic status, since income varies before and after retirement, while expenditure is smoothed over the life-cycle (Li et al., 2016). Incorporating consumption smoothing is necessary, especially in the case that my sample includes a large fraction of observations working in the agricultural sector. Income from agricultural work is volatile, as it is influenced by temporary shocks, for example, the weather. $Z_{i,t}$ is a vector of individual characteristics, age and sex, observed at t.

I divide the whole sample based on pension plan and sex, resulting in four groups: females with the employee pension plan, females without the employee pension plan, males with the employee pension plan and males without the employee pension plan. The rationale for this categorization is rooted in the role that employee pension plans play in retirement decisions. Those covered by the employee pension plan have mandatory retirement ages. Thus, age, health state, and health insurance plan have differential impact on retirement transitions for individuals with and without the employee pension plan. In addition, the mandatory retirement age is different for males and females. Hence, age will have a differential impact on retirement transitions depending on an individual's sex as well. I thus estimate equation (6.1) while grouping individuals by their pension plans and sexes.

However, dividing the entire sample by both sex and pension plan yields small groups, with, for example, only 411 observations for women with the employee pension plan. In contrast, dividing the sample by just one variable increases the number of observations in each group. I select pension plan over sex because I expect a greater variation in how the main explanatory variables are associated with retirement transition probabilities based on pension plan rather than sex. Overall, I estimate equation (6.1) across six sub-samples. I obtain two of them by dividing the entire sample by pension plan, and the other four are obtained by dividing the entire sample by sex and pension plan.

5.4 Data

I use information on enrolment in health insurance plans, self-ranked health state, demographic characteristics, family information, work, retirement and pension, and socio-economic variables from CHARLS. The data used in this chapter is based on the pooled cross-sectional data presented in Figure 3.3. I merge the cross-sectional data in Figure 3.3 with retirement status, sector of employment and pension plan from the work, retirement, and pension section, as well as the total number of children, an indicator of having children in school and children's highest education level from the family information section. Since I use respondents' retirement status and characteristics at time $t-1$ to predict individuals' retirement status at time t, I then merge the dataset with the retirement status variable observed one wave ahead. Thus, only respondents who appeared in two consecutive waves are included. There are 13,039 individual observations from the 2011-13 sample and 14,120 from the 2013-15 sample. Hence, the pooled data has 27,159 observations. In the regression analysis, I only use those who reported working at time $t-1$. The pooled data used for estimation has 19,209 observations who were working at time $t-1$ (either in 2011 or 2013), and of those, 16,045 continued to work and 3,164 of them subsequently retired (either in 2013 or 2015, Table 5.1).

Section 3.3 presents the precise definitions for the variables used for estimation. Table 5.1 compares the distribution of these these variables for the retired group and the working group at time t, conditional on working at $t-1$ in the pooled dataset. Most of the respondents were working only in the agricultural sector. Conditional on working at $t-1$, the proportion of individuals who continue to work only in the agricultural sector at t is 66%, and those who retire at t is also 66% (Table 5.1). However, an individual's sector of employment is highly correlated with health insurance plan and employee pension plan (see Tables 3.6 and 3.8 in section 3.3). Therefore, it is not included in the model.

Within the retired group, 12% have an employee pension plan, compared to only 7% among those who remain employed. The socio-economic variables examined in this analysis include sex, educational attainment, marital status, age, and region of residence. Within the retired group, females represent a higher proportion, approximately 57%, compared to 46% in the group that continues to work. There is a notable discrepancy in literacy rates between the two groups, with a greater percentage of the retired group being illiterate and fewer having attained less than a high school diploma. Conversely, the proportion of individuals with at least a high school education is comparable between the retired and working groups. Additionally, the retired group includes a higher percentage of unmarried individuals, at 13%, in contrast to 6% among their working counterparts. Age differences are also evident, with the retired group being generally older. The geographic distribution by region of residence is similar across both the working and retired groups.

Family information characteristics include the total number of children, whether children are currently enrolled in school, and the highest level of education achieved by children. Approximately 14% of individuals in both the working and retired groups have a single child. However, there is a higher proportion of respondents within the retired group who have four or more children, and a correspondingly lower proportion with children currently attending school. The educational attainment of these children spans from primary schooling to doctoral studies. Predominantly, respondents who have at least one child still in school belong to the 45-49 age group. Overall, children's highest education levels are similar across the working and retired groups.

The summary statistics of per capita household annual expenditure by health insurance plan are in Appendix Table C.2. The data reveal that households with employee health insurance have higher average per capita annual expenditure compared to those with resident health insurance. However, the differences in average expenditure between the resident health insurance group and those without any health

insurance are not statistically significant.

Table 5.1: Summary Statistics for Work-to-Work and Work-to-Retirement Transitions, The Pooled Dataset

Conditional on working at $t-1$	Working, t N = 16,045		Retired, t N = 3,164	
Variables	N (1)	% (2)	N (3)	% (4)
Sex				
Female	7,414	46.2	1,819	57.5
Male	8,631	53.8	1,345	42.5
Pension plan				
Employee pension	1,126	7.0	375	11.9
Resident/no pension	14,919	93.0	2,789	88.2
Sector of employment, $t-1$				
Only in agriculture	10,649	66.4	2,082	65.8
Only in non-agriculture	2,740	17.1	836	26.4
Work in both sectors	2,656	16.6	246	7.8
Sector of employment, t				
Only in agriculture	10,305	64.2	-	-
Only in non-agriculture	3,083	19.2	-	-
Work in both sectors	2,657	16.6	-	-
Health insurance plan				
Employee	872	5.4	317	10.0
Resident	14,547	90.7	2,691	85.1
No health insurance	626	3.9	156	4.9
Health state				
Good	4,107	25.6	674	21.3
Fair	8,565	53.4	1,630	51.5
Poor	3,373	21.0	860	27.2
Educational attainment				
Illiterate	3,806	23.7	974	30.78
Literate, less than high school	10,625	66.2	1,872	59.2
High school and above	1,614	10.1	318	10.1
Marital status				
Married, spouse present	13,959	87.0	2,562	81.0
Married, spouse away	1,086	6.8	188	5.9
Unmarried	1,000	6.2	414	13.1
Age group, t				
45 - 49	1,851	11.5	192	6.1
50 - 54	3,754	23.4	474	15.0
55 - 59	3,391	21.1	543	17.2
60 - 64	3,479	21.7	673	21.3
65 - 69	2,091	13.0	531	16.8
70 - 74	999	6.2	410	13.0
75 - 79	375	2.3	227	7.2
80 - 84	105	0.7	114	3.6
Region				

Continued on next page

Table 5.1 (*Continued*)

	(1)	(2)	(3)	(4)
East	4,894	30.5	979	30.9
Middle	4,679	29.2	969	30.6
Northeast	921	5.7	222	7.0
West	5,551	34.6	994	31.4
Total number of children				
1	2,301	14.3	418	13.2
2	6,707	41.8	1,075	34.0
3	4,168	26.0	775	24.5
4	1,802	11.2	481	15.2
5 or more	1,065	6.6	415	13.1
Having children in school				
Yes	1,714	10.7	190	6.0
No	14,331	89.3	2,974	94.0
Educational attainment children (highest)				
Less than high school	9,190	57.3	1,794	56.7
High school	1,658	10.3	336	10.6
Above high school	5,197	32.4	1,034	32.7

Notes: Unless otherwise stated, all variables are observed at time $t-1$. "-" means not applicable. The results of the Pearson chi-squared test show that the distribution of all variables, except for "educational attainment of children (highest)," is statistically different across the working and retired groups.
Source: Calculations based on pooled CHARLS 2011, 2013 and 2015 survey waves.

I use the pooled data to estimate the retirement transition probabilities, I also present the summary statistics for each wave sample, separately (Appendix Table C.1). For the working and retired groups, distributions of certain variables are statistically different between two wave samples.

The Pearson chi-squared test results indicate that, for the working group, work sector, health insurance coverage, health status, education, marital status, age, and children's highest education level statistically vary between waves. The more recent 2013-15 wave, compared to the 2011-13 wave, features fewer respondents working in agriculture, better reported health, higher education levels, more married respondents living with spouses, an older population, and children with higher education. These changes point to an overall increase in the education level.

Regarding the retired group, employee pension distribution and the number of

children differ statistically across the two samples, while marital status does not. The 2013-15 wave shows more respondents with employee pensions and reporting fewer children compared to the 2011-13 wave. The influence of the one-child policy, initiated in 1980, is more pronounced in the younger generation. The patterns in other variable distributions mirror those I observe in the working group.

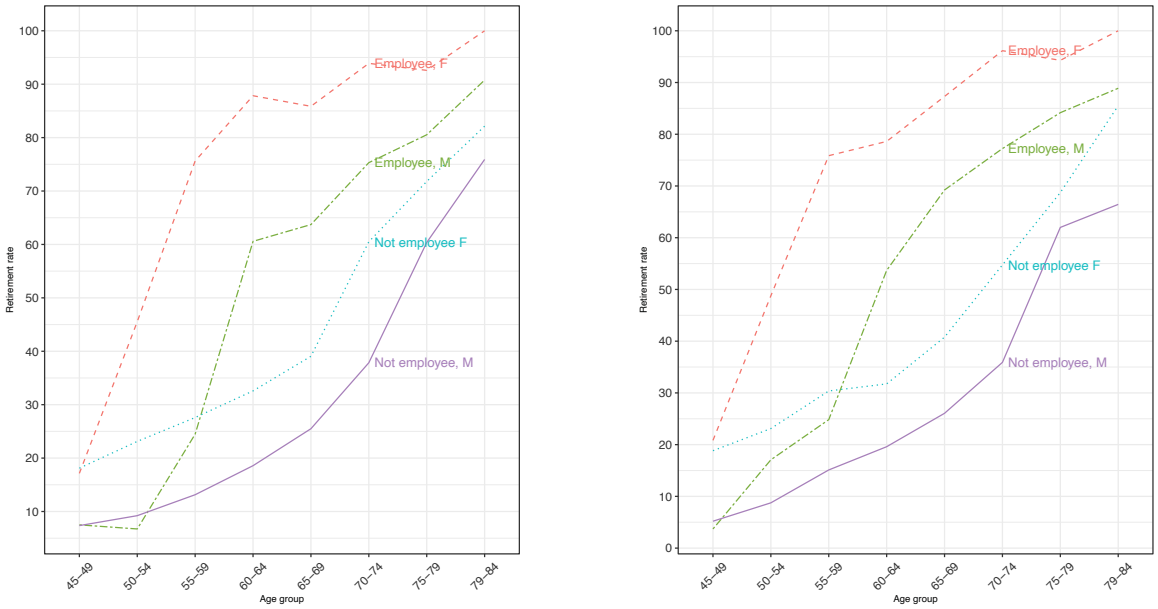
In addition, the survey also collects information on reasons for *early* retirement where early retirement is defined as retirement prior to the mandatory retirement age. Respondents identify health as the primary reason. However, retirement reasons are not asked for other respondents.

Pension plans and retirement rates Although this paper focuses on estimating work-to-retirement transitions, defined as an individual's likelihood to retire in the current period conditional on working in the previous period, I also show the retirement rate by age for individuals with different pension plans. Retirement age profiles are different for individuals with different pension plans.

Figure 5.1 shows that the retirement rate jumps at certain age groups for individuals having employee pension plans, while that for individuals not having employee pensions increases slowly and steadily by age. The retirement rates by sex for individuals with and without employee pensions are similar at the youngest age group (45-49) in the data. However, they begin to diverge as age increases. The divergence emerges earlier for females. The retirement rate of females who have an employee pension plan jumps between 50 and 59, while that of males jumps between 55 and 64. The timing of these jumps coincides with the mandatory retirement ages for females and males. The retirement rate for individuals who are not covered by employee pension plans increases steadily with a lower rate by age.

The retirement rates of individuals with and without employee pension plans do not converge even in the oldest age group. In 2011, for the individuals in the 79-84 age

group and without an employee pension plan, the retirement rates were about 80% for females and 75% for males, increasing in 2013 to 85% for females and decreasing to 65% for males. Retirement rates for individuals with the employee pension plan in this oldest group do not change, remaining at 100% for females and 90% for males.



(a) 2011 and 2013 waves (b) 2013 and 2015 waves

Figure 5.1: Retirement Rate by Pension Plan and Sex

Notes: These two figures show the retirement rate for each age group by individuals' pension plan and sex. "Employee" refers to employee pension plan, "Not employee" refers to resident or no pension plan. An individual's pension plan was based on survey responses at time $t-1$, while sex (F: female; M: male) and retirement rate are based on survey responses at time t.
Source: CHARLS 2011, 2013 and 2015.

5.5 Results

5.5.1 Results of the probit model

My baseline specification for the determinants of work-retirement transition focuses on the lagged health insurance plan, lagged self-ranked health state, sex and region. I report estimates for individuals with different pension plans separately.

Table 5.2 shows the main coefficients from the probit model based on equation

(5.1) and Table 5.3 shows the corresponding marginal effects. (Appendix Table C.3 shows the number of respondents in each cell for health insurance and health states in Table 5.2.) I estimate the baseline model using pooled data. Table 5.2, columns 1-3 report the coefficient estimates for those with an employee pension plan, and for each sex, males and females, respectively. Table 5.2, columns 4-6 report the estimates for the same groups but without an employee pension plan. The baseline estimations in Table 5.2, columns 1 and 4 refer to a married male with a resident insurance plan, who is between 45 and 49 years of age, is living with his spouse, is literate but has less than a high school diploma, resides in the eastern region, has one child, does not have any child currently in school, and whose child's highest education level is lower than high school. The baseline estimations in columns 2, 3, 5 and 6 are the same except for sex. A positive sign indicates that compared to the base category of the explanatory variable, being in that category of the explanatory variable is associated with a higher probability of being retired in the current period, conditional on working in the last survey wave. On the other hand, a negative coefficient represents a lower probability of being retired in the current period, conditional on working during the previous survey wave.

Average marginal effects of the main explanatory variables by sex and pension plan are estimated using coefficients estimated in Table 5.2 (columns 2, 3, 5 and 6). In Table 5.3, each coefficient indicates the change in probabilities for the work-to-retirement transition, which is the probability of retiring at time t, conditional on working at $t-1$, when the value of the explanatory variable changes at the margin. For example, when compared to those in the reference group (who have resident health insurance) in column 1 (who have an employee pension plan and are female), individuals who have employee health insurance have a 16 percentage point higher likelihood of retirement at the last sample wave.

I also estimate a model that includes the logarithm of per capita household annual

expenditure. However, I report only the estimated coefficients of the main explanatory variables (Appendix Table C.4). Inclusion of the logarithm of per capita household annual expenditure in the model results in a decrease in the number of observations for those with the employee pension from 1,501 to 1,308, a loss of 193 observations. Similarly, for those without employee health insurance, the number decreases from 17,706 to 15,188, a loss of 2,518 observations. To ensure that the changes in the estimated coefficients are not solely due to the reduction in observations, I also present the estimation results of the model using the same observations but without including expenditure (Appendix Table C.5).

Health insurance

The results indicate that having employee health insurance is associated with a higher likelihood of retirement compared to having resident health insurance. This association between employee health insurance and the work-to-retirement transition remains consistent across all models except for males with the employee pension plan. Table 5.2 shows that, while controlling for pension plan status, the association between employee health insurance and an individual's probability of transitioning from work to retirement is positive and significant for both groups, those with and without an employee pension plan. The coefficient is smaller for those with the employee pension plan, at 0.242 compared to 0.418 for those who do have the employee pension plan. Furthermore, when the model is estimated by sex for both groups, the effect of having employee health insurance on the work-to-retirement transition becomes insignificant for males among those with the employee pension plan. Hence, although the magnitude is smaller for those with the employee pension plan, this smaller magnitude is driven by the insignificant association between employee health insurance and the work-to-retirement transition of males.

Among individuals who do not have the employee pension plan, when the model

is estimated by sex, the magnitude of the coefficient is almost identical but slightly larger for males at 0.452, compared to 0.408 for females. Estimated coefficients of the ordered probit model are less explanatory than the marginal effects. Therefore, I present the average marginal effects in Table 5.3. Since the effect of the health insurance plan on the work-to-retirement transition varies by sex among individuals with the employee pension plan, I present the average marginal effects for each sex separately.

Table 5.3 shows that, among individuals with an employee pension plan, having employee health insurance, compared to resident health insurance, is associated with a 16 percentage point higher probability of transitioning from working to retirement for females; this association is not significant for males. Among individuals without the employee pension plan, the average marginal effect is 12 percentage points for females and 11 percentage points for males, respectively.

Table 5.2 also indicates that not having any health insurance is not associated with the likelihood of transitioning from working to retirement for individuals with an employee pension plan, but it is associated with a larger likelihood of being retired compared to having resident health insurance. Recall that in section 5.2, some studies show that resident health insurance has a limited effect on reducing a household's likelihood of experiencing catastrophic medical expenditure, it is within expectation that compared to individuals who have the resident health insurance, individuals who do not have any health insurance do not have a smaller likelihood of transitioning from working to retirement. Positive signs of "no health insurance" for individuals without the employee pension plan imply that there might be some unobservable factors which affect the retirement decision for those who do not have any health insurance.

To summarize, having employee health insurance is associated with a higher probability of transitioning from work to retirement, except for males with an employee

pension plan. This association is substantial, as the average marginal effect exceeds 10 percentage points for females with the employee pension plan and for both sexes without the employee pension plan.

Adding the logarithm of per capita household annual expenditure confirms the stability and significance of the estimated coefficients for employee health insurance. These coefficients exhibit slightly larger magnitudes for individuals of both sexes with employee health insurance. For those lacking employee pensions, the coefficients for employee health insurance increase for females but decrease for males. Variations in the number of observations also affect the magnitude of the association between employee health insurance and the probability of transitioning from work to retirement. The non-random nature of the missing expenditure data undermines the reliability of these regression results. The effect of expenditure on an individual's retirement transition probabilities is limited. Overall, after adding the logarithm of per capita household annul expenditure, the estimated coefficients on employee health insurance remain stable and significant.

Table 5.2: Estimation Results for Retirement Status: Probit Model and Pooled Data

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee health insurance	0.242***	0.562***	0.123	0.418***	0.408**	0.452***
	(0.089)	(0.181)	(0.106)	(0.097)	(0.170)	(0.117)
No health insurance	−0.109	−0.281	0.046	0.199***	0.197***	0.205**
	(0.244)	(0.467)	(0.294)	(0.058)	(0.075)	(0.084)
Good health	−0.064	−0.253	0.016	−0.071**	−0.042	−0.095**
	(0.085)	(0.156)	(0.102)	(0.031)	(0.042)	(0.044)
Poor health	0.352***	0.443*	0.318**	0.125***	0.103***	0.160***
	(0.117)	(0.246)	(0.136)	(0.029)	(0.037)	(0.045)
Female	0.531***	-	-	0.358***	-	-
	(0.090)			(0.025)		
Middle	0.063	−0.043	0.143	−0.022	−0.058	0.022
	(0.101)	(0.180)	(0.120)	(0.033)	(0.041)	(0.047)
Northeast	0.274**	0.176	0.328**	−0.099***	−0.142***	−0.042
	(0.112)	(0.193)	(0.135)	(0.036)	(0.045)	(0.053)
West	0.324***	0.143	0.395***	−0.102***	−0.190***	0.013
	(0.112)	(0.226)	(0.129)	(0.037)	(0.046)	(0.053)
Constant	−2.197***	−1.570***	−2.244***	−1.536***	−1.065***	−1.760***
	(0.246)	(0.350)	(0.340)	(0.061)	(0.071)	(0.102)
N	1,501	411	1,089	17,706	8,820	8,886

Notes: Retirement status is observed at t and independent variables are observed at $t-1$, except for sex and age. All models include age effects by a 5-year window, which are reported in Table 5.4. Estimated coefficients of other control variables are reported in Table 5.6. The baseline estimation results in columns 1 and 4 refer to a married male with resident health insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5.3: Average Marginal Effects by Pension Plan and Sex

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)
Employee pension	Y	Y	N	N
Sex	Female	Male	Female	Male
Employee health insurance	0.160***	0.033	0.123**	0.105***
	(0.048)	(0.028)	(0.058)	(0.032)
No health insurance	−0.060	0.012	0.056**	0.042**
	(0.090)	(0.077)	(0.023)	(0.019)
Good health	−0.071*	0.004	−0.011	−0.017**
	(0.043)	(0.027)	(0.010)	(0.008)
Poor health	0.144*	0.091**	0.028***	0.033***
	(0.083)	(0.041)	(0.010)	(0.010)
Middle	−0.012	0.036	−0.016	0.004
	(0.051)	(0.030)	(0.011)	(0.009)
Northeast	0.052	0.087**	−0.038***	−0.008
	(0.058)	(0.037)	(0.012)	(0.010)
West	0.042	0.107***	−0.049***	0.002
	(0.068)	(0.036)	(0.012)	(0.010)
N	411	1,089	8,820	8,886

Notes: Average marginal effects are estimated using the coefficients in Tables 5.2 and 5.4. Columns 1-4 in this table correspond to columns 2, 3, 5 and 6 in Tables 5.2 and 5.4, respectively. In each column, the baseline estimation results refer to a married individual with resident health insurance plan, who is between 45 and 49 years of age, is living with his/her spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Health states

In terms of the effect of health state on retirement, Tables 5.2 and 5.3 show that compared to reporting fair health, reporting good health is associated with a lower likelihood of retirement for males without an employee pension plan, while for others, the association between good health and the probability of transitioning from work to retirement is insignificant. In addition, for males without an employee pension plan, the association between reporting good health and the probability of transitioning from work to retirement is small. Compared to reporting fair health, reporting good health is associated with a lower likelihood of transitioning from work to retirement

by only two percentage points on average for males without an employee pension plan. Overall, the association between reporting good health and work-to-retirement transition probabilities is significant only within one group with a small magnitude.

In contrast, the association between reporting poor health and work-to-retirement transition probabilities is larger and significant within all four groups. Reporting poor health is associated with a higher likelihood of retirement for both sexes and pension plans. For example, for females with an employee pension plan, reporting poor health instead of fair health is associated with a 14 percentage point higher likelihood of transitioning from work to retirement, while reporting good health is associated with a seven percentage point lower likelihood of transitioning from work to retirement. For females with an employee pension plan, the magnitude of the association (in absolute value) of reporting poor health is seven percentage points higher than that of reporting good health.

The association between poor health and work-to-retirement probabilities varies across pension plans for both males and females. The association between poor health and transitioning to retirement is more pronounced for individuals with an employee pension plan than for those without, across both sexes. For females, the average marginal effect of reporting poor health is 14 percentage points if they have an employee pension plan, while it is only three percentage points if they do not have an employee pension plan. For males, the average marginal effects of reporting poor health are nine and three percentage points, respectively, for those with and without an employee pension plan. These estimates suggest that individuals without an employee pension plan who are in poor health are less likely to transition to retirement compared to their counterparts with an employee pension plan. Pension plans are an important source of an individual's retirement income. With the lack of it, or with a poor pension plan, individuals may need to continue working even if they are in poor health at an old age.

Region

Table 5.2 demonstrates that the regional associations with work-to-retirement probabilities are statistically significant for the Northeast and West regions. However, the directions of these associations are opposite across pension plans. Specifically, compared to residing in the Eastern region, residing in the Northeastern and Western regions is associated with a higher likelihood of retirement for those with an employee pension plan. Moreover, these associations are predominantly driven by males. For males with an employee pension plan, compared to residing in the Eastern region, residing in the Northeastern region and the Western region is associated with an approximately nine and 11 percentage point higher likelihood of transitioning from work to retirement, respectively.

Conversely, for individuals without an employee pension plan, residing in the Northeastern and Western regions is associated with a lower likelihood of retirement (Table 5.2, columns 2 and 5). This association is predominantly driven by females. For females without an employee pension plan, residing in the Northeastern region and the Western region is associated with an approximately four and five percentage point lower likelihood of transitioning from work to retirement, respectively.

The opposite direction of the regional association with the work-to-retirement transition for individuals with and without employee pension plans can be explained by differences in relative income. Although the overall income level is the highest in the Eastern region, individuals with an employee pension plan do not rank at the top of the income distribution (Figure 3.5 in section 2.3) in Beijing (Eastern region). State-owned enterprises tend to have more generous pension plans. Yet their average wage ranks in the middle of the wage distribution. Average wages are higher in foreign funded, Hong Kong, Macao and Taiwan funded enterprises as well as shareholding corporations. In contrast, in Liaoning (Northeastern region) and Sichuan (Western region), state-owned enterprises again with most generous pension plans, have the

highest average wages.

Age

Tables 5.4 and 5.5 display the associations between age and the work-to-retirement transition. The results indicate that, compared to the youngest age group (45 to 49), being in an older age group is associated with a higher likelihood of transitioning from work to retirement, except for the 50 to 54 age group, where the coefficient is significant solely for females with an employee pension plan. Moreover, for all age groups and both sexes, the magnitude of the association between age and work-to-retirement transition is greater for those with an employee pension plan. This larger age association observed in individuals with an employee pension plan is partly due to mandatory retirement ages, but it is also attributed to the significant role the pension plays in providing retirement income.

Table 5.4: Probit: Age Coefficients by Pension Plan and Sex Using Pooled Data

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
50-54	0.488**	0.573**	0.405	0.066	0.057	0.101
	(0.203)	(0.239)	(0.312)	(0.050)	(0.060)	(0.094)
55-59	0.711***	0.926***	0.621**	0.206***	0.180***	0.292***
	(0.212)	(0.267)	(0.313)	(0.052)	(0.063)	(0.092)
60-64	1.318***	0.776***	1.472***	0.267***	0.201***	0.401***
	(0.215)	(0.294)	(0.312)	(0.052)	(0.063)	(0.092)
65-69	1.311***	1.254***	1.367***	0.447***	0.331***	0.628***
	(0.227)	(0.337)	(0.323)	(0.057)	(0.070)	(0.096)
70-74	1.690***	2.152***	1.723***	0.700***	0.642***	0.808***
	(0.246)	(0.501)	(0.337)	(0.064)	(0.083)	(0.103)
75-79	1.459***	1.722***	1.420***	0.989***	0.662***	1.260***
	(0.290)	(0.560)	(0.390)	(0.080)	(0.110)	(0.117)
80-84	2.150***	-	2.267***	1.296***	1.270***	1.356***
	(0.409)		(0.467)	(0.108)	(0.146)	(0.154)
N	1,501	411	1,089	17,706	8,820	8,886

Notes: This table shows the age effects in Table 5.2. All models include age effects by a 5-year window. The baseline estimation results in columns 1 and 4 refer to a married male with resident health insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. "-" means that the coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5.5: Average Marginal Effects of Age by Pension Plan and Sex

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)
Employee pension	Y	Y	N	N
Sex	Female	Male	Female	Male
50-54	0.133***	0.045	0.013	0.012
	(0.050)	(0.028)	(0.013)	(0.011)
55-59	0.247***	0.081***	0.043***	0.039***
	(0.067)	(0.031)	(0.015)	(0.011)
60-64	0.196***	0.319***	0.048***	0.059***
	(0.075)	(0.040)	(0.015)	(0.012)
65-69	0.367***	0.282***	0.085***	0.107***
	(0.102)	(0.046)	(0.018)	(0.014)
70-74	0.674***	0.412***	0.186***	0.155***
	(0.130)	(0.059)	(0.025)	(0.019)
75-79	0.538***	0.301***	0.194***	0.303***
	(0.184)	(0.089)	(0.036)	(0.031)
80-84	-	0.616***	0.428***	0.339***
		(0.126)	(0.054)	(0.049)
N	411	1,089	8,820	8,886

Notes: Average marginal effects are estimated using the coefficients in Tables 5.2 and 5.4. Columns 1-4 in this table correspond to columns 2, 3, 5 and 6 in Tables 5.2 and 5.4, respectively. In each column, the baseline estimation results refer to a married individual with resident health insurance, who is between 45 and 49 years of age, is living with his/her spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. "-" means that the coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Other control vairables

Tables 5.6 and 5.7 show coefficients and average marginal effects of other control variables. The influence of these control variables on work-to-retirement transitions differs by pension plan type. Notably, the coefficients of some variables are significant exclusively for individuals lacking an employee pension plan, and the signs of the coefficients for several variables differ between pension plans.

The two tables indicate that the effects of education, marital status and having children in school on the work-to-retirement transition are only significant for those without an employee pension plan. Within this group, compared to individuals who

are literate but have less than a high school diploma, having a high school diploma or above is associated with a five percentage point higher likelihood of transitioning from work to retirement for females, and a three percentage point higher likelihood for males. In terms of marital status, compared to respondents who are married and living with their spouses, being unmarried is associated with a five percentage point higher likelihood of transitioning from work to retirement for females and an eight percentage point higher likelihood for males. Having a child in school also reduces the likelihood of transitioning to retirement. Compared to not having any child in school, having at least one child in school is associated with a four percentage point lower likelihood of transitioning from work to retirement for females without an employee pension.

The association between children's educational attainment and work-to-retirement transition probabilities is significant for both pension plan statuses with consistent directions. Compared to having a child whose highest education level is less than high school, having a child whose highest education level is above high school is associated with a higher likelihood of transitioning from work to retirement. However, among those with an employee pension, this association is observed only in males (eight percentage points), while among those without an employee pension, the association is observed only in females (two percentage points).

The number of children influences the probability of transitioning from work to retirement differently depending on the pension plan. For individuals without an employee pension plan, having more than one child is associated with a higher likelihood of retiring. Specifically, for males, having four children is associated with a 2.5 percentage point higher probability of retiring, while for females, having five or more children is associated with a 3.8 percentage point higher probability of retiring. In contrast, among those with an employee pension, having five or more children is associated with an 8.1 percentage point lower likelihood of males transitioning

from work to retirement. These opposing associations underscore how the impact of children on retirement transitions varies according to pension status. These findings regarding the control variables are consistent with those reported in the study by Giles et al. (2023).

Table 5.6: Probit: Control Variables by Pension Plan and Sex using Pooled Data

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Illiterate	−0.186	−0.210	−0.329	−0.032	−0.030	0.005
	(0.198)	(0.274)	(0.291)	(0.029)	(0.035)	(0.054)
High school or above	−0.046	−0.220	0.034	0.124**	0.178**	0.137**
	(0.089)	(0.169)	(0.102)	(0.048)	(0.084)	(0.061)
Married, spouse away	0.009	0.466	−0.134	0.075	0.004	0.161**
	(0.217)	(0.391)	(0.247)	(0.047)	(0.063)	(0.071)
Unmarried	0.179	0.109	0.164	0.237***	0.194***	0.352***
	(0.174)	(0.245)	(0.270)	(0.042)	(0.053)	(0.065)
Have children in school	0.186	−0.134	0.313	−0.104**	−0.159**	−0.022
	(0.180)	(0.318)	(0.216)	(0.049)	(0.062)	(0.073)
Children's education	0.167	−0.034	0.208	−0.059	−0.059	−0.048
being high school	(0.141)	(0.290)	(0.162)	(0.043)	(0.054)	(0.063)
Children's education	0.216**	−0.018	0.297**	0.055*	0.082**	0.019
being above high school	(0.102)	(0.217)	(0.118)	(0.030)	(0.037)	(0.043)
Having two children	−0.116	0.035	−0.170	0.018	−0.016	0.074
	(0.102)	(0.171)	(0.120)	(0.042)	(0.054)	(0.062)
Having three children	−0.188	0.089	−0.306**	−0.009	−0.027	0.026
	(0.134)	(0.275)	(0.154)	(0.045)	(0.058)	(0.067)
Having four children	−0.111	−0.151	−0.214	0.064	0.032	0.131*
	(0.177)	(0.355)	(0.198)	(0.053)	(0.068)	(0.075)
Having five or more children	0.011	0.523	−0.125	0.114*	0.137*	0.126
	(0.224)	(0.602)	(0.247)	(0.059)	(0.076)	(0.085)
N	1,501	411	1,089	17,706	8,820	8,886

Notes: This table reports the coefficients of several control variables which are not reported in Table 5.2. Retirement status is as observed at t and independent variables are as observed at $t-1$, except for sex and age. All models include age effects by a 5-year window, which are reported in Table 5.4. The baseline estimation results in columns 1 and 4 refer to a married male with resident health insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but with a less than high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5.7: Average Marginal Effects of Other Control Variables by Pension Plan and Sex

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)
Employee pension	Y	Y	N	N
Sex	Female	Male	Female	Male
Illiterate	−0.061	−0.078	−0.008	0.001
	(0.077)	(0.061)	(0.009)	(0.010)
High school or above	−0.064	0.009	0.050**	0.027**
	(0.049)	(0.027)	(0.025)	(0.013)
Married, spouse away	0.147	−0.034	0.001	0.032**
	(0.130)	(0.060)	(0.016)	(0.015)
Unmarried	0.032	0.046	0.054***	0.077***
	(0.074)	(0.079)	(0.016)	(0.016)
Middle	−0.012	0.036	−0.016	0.004
	(0.051)	(0.030)	(0.011)	(0.009)
Have children in school	−0.039	0.084	−0.042**	−0.004
	(0.092)	(0.058)	(0.016)	(0.014)
Children's education	−0.010	0.052	−0.015	−0.009
being high school	(0.084)	(0.041)	(0.013)	(0.011)
Children's education	−0.005	0.076***	0.022**	0.004
above high school	(0.064)	(0.029)	(0.010)	(0.008)
Having two children	0.010	−0.047	−0.004	0.014
	(0.050)	(0.033)	(0.014)	(0.011)
Having three children	0.026	−0.081**	−0.007	0.005
	(0.082)	(0.040)	(0.015)	(0.012)
Having four children	−0.042	−0.058	0.008	0.025*
	(0.095)	(0.052)	(0.018)	(0.014)
Having five or more children	0.166	−0.035	0.038*	0.024
	(0.204)	(0.067)	(0.021)	(0.016)
N	411	1,089	8,820	8,886

Notes: Average marginal effects are estimated using the coefficients in Table 5.2 and Appendix Table 5.4. Columns 1-4 in this table correspond to columns 2,3,5 and 6 in Table 5.2 and Appendix Table 5.4, respectively. In each column, the baseline estimation results refer to a married individual with resident health insurance, who is between 45 and 49 years of age, is living with his/her spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

5.5.2 Predicted conditional retirement transition probabilities

I also compare the predicted probabilities of retirement at time t, conditional on working at the last sample wave and based on health insurance plan, pension plan,

and sex. The representative respondent in terms of other characteristics, is married, is living with their spouse, is literate but has a less than high school diploma, resides in the Eastern region, has one child, does not have any child in school, and has children whose highest education level is below high school. Table 5.8 shows that the work-to-retirement transition rate is higher for individuals with employee health insurance, regardless of the pension plan and sex. For example, for a female who was working at the last sample wave, and is aged between 45 to 49, with an employee pension and employee health insurance, the probability of retiring at the next sample wave is 16%, whereas for a female with resident insurance, it is 6%, which is also not statistically significant. For females who do not have an employee pension plan, the work-to-retirement transition rate is 26% for individuals having an employee health insurance versus 14% for individuals who have a resident health insurance.

Furthermore, among females who are between 45 and 49 years of age, the work-to-retirement transition rate of individuals with an employee pension plan is lower than that of those without an employee pension plan in the 45-49 age group, regardless of employee health insurance. However, it is higher for older age groups.

Table 5.8: Predicted Conditional Retirement Transitions

Dependent variable: probability of transitioning to retirement

	With an employee pension plan			Without an employee pension plan		
	(1)	(2)	(3)	(4)	(5)	(6)
	Employee health insurance	Resident health insurance	No health insurance	Employee health insurance	Resident health insurance	No health insurance
				A) Female		
45-49	0.157**	0.058	0.032	0.256***	0.144***	0.193***
	(0.079)	(0.041)	(0.040)	(0.058)	(0.016)	(0.027)
50-54	0.332***	0.160**	0.101	0.275***	0.157***	0.209***
	(0.104)	(0.070)	(0.096)	(0.059)	(0.016)	(0.028)
55-59	0.468***	0.260***	0.178	0.317***	0.188***	0.246***
	(0.124)	(0.099)	(0.133)	(0.064)	(0.018)	(0.031)
60-64	0.408***	0.214**	0.141	0.324***	0.194***	0.253***
	(0.125)	(0.099)	(0.123)	(0.064)	(0.018)	(0.031)
65-69	0.598***	0.376**	0.275	0.372***	0.232***	0.296***
	(0.140)	(0.146)	(0.192)	(0.068)	(0.022)	(0.035)
70-74	0.874***	0.720***	0.618**	0.494***	0.336***	0.411***
	(0.107)	(0.180)	(0.267)	(0.074)	(0.031)	(0.043)
75-79	0.762***	0.561**	0.449	0.502***	0.344***	0.419***
	(0.170)	(0.232)	(0.288)	(0.080)	(0.041)	(0.052)
80-84	-	-	-	0.730***	0.581***	0.656***
				(0.073)	(0.058)	(0.061)
				B) Male		
45-49	0.017	0.012	0.014	0.095***	0.039***	0.060***
	(0.015)	(0.011)	(0.016)	(0.025)	(0.009)	(0.015)
50-54	0.043**	0.033**	0.036	0.114***	0.049***	0.073***
	(0.018)	(0.014)	(0.027)	(0.025)	(0.008)	(0.015)
55-59	0.067***	0.052***	0.057	0.155***	0.071***	0.103***
	(0.024)	(0.020)	(0.038)	(0.031)	(0.010)	(0.019)
60-64	0.258***	0.220***	0.234**	0.182***	0.087***	0.124***
	(0.056)	(0.054)	(0.101)	(0.035)	(0.012)	(0.022)
65-69	0.225***	0.190***	0.203**	0.248***	0.129***	0.177***
	(0.056)	(0.053)	(0.094)	(0.042)	(0.016)	(0.029)
70-74	0.345***	0.301***	0.317**	0.308***	0.171***	0.228***
	(0.078)	(0.078)	(0.124)	(0.049)	(0.022)	(0.035)
75-79	0.242***	0.205**	0.218*	0.481***	0.309***	0.384***
	(0.092)	(0.086)	(0.118)	(0.059)	(0.036)	(0.049)
80-84	0.558***	0.509***	0.527***	0.519***	0.343***	0.421***
	(0.151)	(0.155)	(0.189)	(0.072)	(0.053)	(0.065)

Notes: A) is estimated using the estimated coefficients in Table 5.2, columns 2 and 5; B) is estimated using the estimated coefficients in Table 5.2, columns 3 and 6. The baseline estimation results refer to a married individual with resident health insurance, who is between 45 and 49 years of age, is living with his/her spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. "-" means that the coefficient cannot be estimated due to the low number of observations. * Significant at 10%; ** significant at 5%; *** significant at 1%.

5.5.3 Sensitivity analysis

Based on the results of estimating equation (5.1), this study conducts two sensitivity analyses by making slight modifications to the model or dataset.

Heterogeneous health insurance effect on work-to-retirement transition by health outcome Does the association between type of health insurance and retirement status vary depending on health outcomes? To investigate this, I include the interaction of health insurance plan and self-ranked health state. The estimated coefficients on employee health insurance remain stable and significant (Appendix Table C.7). Only the interaction term of employee health insurance and poor health for females is statistically significant, regardless of the pension plan. However, the sign of the coefficient for this interaction term is opposite for females with and without an employee pension plan. It is positive for females with an employee pension plan, while being negative for females without an employee pension plan.

The results indicate that the association between health insurance and retirement status varies by health state. Moreover, this heterogeneity differs across types of pension plans. For females with an employee pension plan and in poor health, those with employee health insurance have a higher probability of retiring than those with resident health insurance. Conversely, for females without an employee pension but also in poor health, those with employee health insurance have a lower probability of retiring compared to those with resident health insurance. Overall, females with both an employee pension plan and employee health insurance are more likely to retire if

they are in poor health. These findings suggest that health insurance can mitigate the negative impact of poor health on retirement transitions for females who do not have an employee pension. However, the marginal effect of health state and "no health insurance" cannot be estimated using the coefficients from Appendix C.7 due to the limited number of observations.

Alternative age intervals I conduct a sensitivity check by shifting the age intervals by 1 year so that each age interval starts after mandatory retirement ages. I find that the sign and size of the estimated coefficients and their statistical significance are not sensitive to the alternative specification of age intervals (Appendix Tables C.8, C.9, C.10 and C.11). The only difference is that the effect of the second youngest age (51 to 55 years old) becomes significant for those without an employee pension plan. The estimated coefficient of this age-group becomes smaller for females, as the base group 45-50 now includes the mandatory retirement age for female workers.

5.6 Conclusion

I use the pooled CHARLS dataset to examine how work-to-retirement transitions are associated with the type of health insurance plan in China. I find that individuals enrolled in the relatively more generous employee health insurance plan are more likely to transition from work to retirement.

One mechanism that can account for this finding is that potential medical expenses resulting from a health shock serve as contingent precautionary savings. Individuals with a more generous employee health insurance plan face lower medical expenditures in the event of a negative health shock, requiring less precautionary savings and enabling them to retire earlier. This effect is significant for individuals with and without an employee pension plan, underscoring the importance of health insurance even for those who are financially protected by a more generous pension

plan. Health insurance hence affects their quality of life after retirement. If the social safety net of a generous health insurance plan is lacking, individuals have to rely on their personal savings or other means, like support through their children. If they do not have sufficient savings, the risk of anxiety or depression will increase. Overall, in the Chinese context, retirement transitions respond to the contingent financial protection provided by a more generous health insurance plan.

In addition to these overall associations between type of health insurance plan and work-to-retirement transitions, there are notable sex and regional disparities in work-to-retirement transitions. Overall, females and individuals residing in less developed regions are more likely to retire. However, these associations differ based on the pension plan. For females with an employee pension plan, residing in a less developed region is associated with a higher probability of retirement compared to females without an employee pension. Furthermore, the association between employee health insurance and the work-to-retirement transition is more pronounced for females than for males.

While the primary focus of this study is not on examining how health affects work-to-retirement transitions, I find that reporting poor health is associated with a higher likelihood of retirement. Notably, this association is less pronounced for individuals without an employee pension plan. This suggests that individuals without generous pension coverage continue working even in poor health due to insufficient pension protection.

Overall, my findings suggest that individuals with resident health insurance do not have sufficient health insurance-based financial protection to mitigate the risk of high medical expenses resulting from negative health shocks. Consequently, they remain in the labour force and are less likely to retire, even at an advanced age, which will negatively affect the well-being of elderly people. The estimated coefficient is significant even for individuals with an employee pension. Females in less developed

regions and individuals without an employee pension plan are particularly vulnerable. These finding suggests that integration of the insurance pools and equalization of the benefits, deductibles, and reimbursement rates for employee and resident health insurance will have implications for work-to-retirement transitions and is essential to improve the well-being of elderly people in China.

Chapter 6

Health insurance, Health Status, and Workdays Missed

6.1 Introduction

The number of reported workdays missed (also known as "sick days," sickness absence, or absenteeism in the literature) is one of the indirect economic costs of disease and poor health, which has attracted a large amount of attention (Association, 2013, 2018; Stang and Osterhaus, 1993; Goetzel et al., 2004; Fang et al., 2011; Simon, 2012; Cawley et al., 2007; Kowal et al., 2013; Nurmagambetov et al., 2018; Hammond and Levine, 2010). Given its importance, the determinants of missed workdays have also been extensively studied (Brown and Sessions, 1996; Alexanderson, 1998; Harrison and Martocchio, 1998; Li et al., 2019), with health insurance being one of the widely studied determinants (Vistnes, 1997; Gilleskie, 1998; Lofland and Frick, 2006; Xu and Jensen, 2012; Dizioli and Pinheiro, 2016; Nguyen et al., 2021). Although the literature has extensively studied the association between health insurance and missed workdays, most of the studies are about the United States and little is known about this association for middle-aged and elderly people in China. Given the ongoing health insurance reforms in China, it is important to understand whether the generosity of health insurance plans has an effect on an individual's probability of missing workdays and the number of workdays missed.

This chapter investigates whether an individual's missed workdays are associated with their type of health insurance plan by answering two questions: 1) while controlling for individuals' initial health, are those with the more generous employee health insurance plan less likely to miss workdays?; and 2) while controlling for individuals's

initial health, do those with the more generous employee health insurance plan report fewer missed workdays?

In principle, the type of health insurance plan affects an individual's workdays missed via three channels. First, as I found in chapter 4, individuals with the employee health insurance plan have a lower probability of reporting poor health and transitioning into poor health. From this perspective, better overall health may be associated with fewer workdays missed (Dizioli and Pinheiro, 2016). Second, if in poor health, those with the more generous employee health insurance plan may be more likely to experience successful treatment and recovery (Li and Zhang, 2013; Tan et al., 2019; Lee et al., 2022). From this perspective, it is possible that those with the more generous employee health insurance plan will subsequently have a lower number of workdays missed while controlling for their initial health states (Dizioli and Pinheiro, 2016). Third, individuals with the more generous health insurance plan may visit a hospital or a medical clinic more often, which may result in more workdays missed. The effect of more generous health insurance on an individual's missed workdays through the third context is the opposite of the first two.

In this chapter, I use a logit model to address the first question–the association between the type of health insurance plan and the likelihood of missing workdays. For the second question–the association between the type of health insurance plan and the number of workdays missed–I employ a heaped generalized Poisson model. Individuals who do not miss any workdays (i.e., have zero workdays missed) are also included in the analysis for the second question. Since an individual's health directly affects their likelihood of missing workdays and the number of workdays missed, I control for the individual's current health in these models through self-ranked health states and indicators of chronic disease and disability.

In China, significant differences exist between agricultural and non-agricultural work (Benjamin et al., 2003). For instance, agricultural work is characterized by its

seasonal nature, unlike non-agricultural work. Moreover, agricultural tasks in China are not fully industrialized with a limited use of machinery, resulting in more physical labour demanded compared to non-agricultural jobs (Yu and Wu, 2018). Hence, I estimate these models separately by sector of employment.

The findings indicate that while controlling for an individual's initial health, compared to individuals with the resident health insurance plan, individuals with the employee health insurance plan are less likely to report any workdays missed and they also miss fewer workdays. However, these effects are predominantly driven by those employed solely in the agricultural sector and those employed in both the agricultural and non-agricultural sectors. For example, in the agricultural sector, the likelihood of missing any workdays is 13 percentage points lower for individuals with employee health insurance than for those who have the resident health insurance. Also, in the agricultural sector, individuals who have employee health insurance tend to miss eight fewer workdays annually than those who have the resident health insurance. In the non-agricultural sector, the employee health insurance does not exhibit a statistically significant influence on the probability of missing workdays, nor does it show a statistically significant effect on the number of workdays missed. But individuals without any health insurance have an eight percentage point higher likelihood of missing any work days and also tend to have three more workdays missed annually.

The above results indicate that, for individuals working in the agricultural sector or in both sectors, the employee health insurance plan is associated with a lower likelihood of missing any workdays and with fewer workdays missed. For individuals who work solely in the non-agricultural sector, being insured has a similar association, regardless of the type of health insurance plan.

Self-ranked health states themselves also have a significant association with workdays missed. The association is similar across those who work in the agricultural, non-agricultural, or both sectors in terms of the likelihood of missing workdays. Across

these sectors, the association of good health with workdays missed is smaller than that of poor health. However, the magnitude of this association differs across employment sectors in terms of the number of workdays missed. The effect is most pronounced in the agricultural sector, where individuals reporting poor health on average experience 12 more missed workdays compared to those reporting fair health. It is smallest in the non-agricultural sector, where individuals reporting poor health on average have four more missed workdays.

This study contributes to two strands in the body of literature on the link between health insurance and workdays missed. First, it complements studies that explore the association between health insurance and workdays missed by providing novel evidence from China. My findings are consistent with the findings of most studies from the United States. Second, this study carefully models self-reported count data, wherein respondents tend to round the count data when the recall period is long, and are less reliable compared to the survey data that uses a short recall period (Mattke et al., 2007; Li et al., 2019b). I use a heaping model, which is designed to deal with this type of measurement issue of self-reported data (Wang and Heitjan, 2008; Cummings et al., 2015). This measurement issue has been overlooked in the existing studies on workdays missed.

6.2 Related Literature

The effect of health insurance on the number of workdays missed is intermediated through three channels. Results on the direction of the effect in existing studies are mixed, depending on the variation in country selection, the use of different datasets, and the use of different models.

Most studies find that having health insurance is associated with a smaller likelihood of missing workdays and reporting fewer numbers of workdays missed. Using a self-conducted retrospective study using the 1996-1999 Medical Expenditure Panel

Survey (MEPS) within the United States, Lofland and Frick (2006) discovered a significant association between having health insurance and a reduced likelihood of missing workdays and having fewer missed workdays if they have any. Using the same dataset but with more recent waves, Dizioli and Pinheiro (2016) find similar results, in which workers with health coverage miss, on average, about 77% fewer work days than uninsured workers. Similar effects are observed in Gilleskie (1998), which finds that, on average, full insurance coverage decreases an individual's probability of missing workdays by 9% and reduces the number of workdays missed by 1.7 days per year. Nguyen et al. (2021) provide evidence that improvements in the generosity of illness benefits in health insurance reduces the incidence and number of missed work days using data from Vietnam.

Although the four studies report similar results, there are differences among them. The first three studies use U.S. data, while the fourth uses Vietnamese data. Among the first three studies, Gilleskie (1998) uses a older dataset than Lofland and Frick (2006) and Dizioli and Pinheiro (2016), namely the 1987 National Medical Expenditure Survey (NMES). In addition, Gilleskie (1998) uses a dynamic stochastic model, while Lofland and Frick (2006) and Dizioli and Pinheiro (2016) use reduced form estimation models. These four studies use respondents with a wide age range, for example, employees aged between 18 to 65 years old in Lofland and Frick (2006). The age range may be different in other studies, but the variation is small. Overall, these studies all find that health insurance decreases the likelihood of missing workdays and is associated with fewer missed workdays, despite using different datasets and models.

On the other hand, there are a few studies which arrive at different conclusions. Xu and Jensen (2012) use a smaller age range of respondents, between 52 to 64 years old, in the 2004-2006 Health and Retirement Study (HRS), and find that there is no significant difference in the number of illness-related workdays missed between

insured and uninsured workers in the United States. Vistnes (1997) uses the same dataset as Dizioli and Pinheiro (2016), but draws the opposite conclusion: health insurance increases the probability of missing workdays. Unlike the study of Dizioli and Pinheiro (2016), which does not control for individuals' working sectors, Vistnes (1997) focuses on respondents who are not self-employed and are employed in the non-agricultural sector. Overall, in these studies, health insurance is observed to either increase the probability of missing any workdays, resulting in more workdays missed, or to have no effect on workdays missed.

In China, there is no study linking insurance to workdays missed. However, there is a body of literature that studies missed workdays. For example, using the CHARLS dataset, Li et al. (2019) study the effects of health states on work exit rates (the probability of stopping work) using a probit model and the number of workdays missed due to health problems using a Tobit model. Although they do not consider the health insurance effect, they find that, compared to those who report fair health, females who report good health have on average 11 fewer workdays missed and males in good health have 14 fewer. Moreover, females who report poor health have on average 26 more workdays missed and males have 36 more workdays missed. Guo et al. (2020) study the relationship between commute time and workdays missed using a zero-inflated negative binomial model. They find that a longer commuting time leads to more workdays missed for urban citizens, which is defined as urban residents with urban *hukou*, but not for rural migrants, which is defined as individuals who live in urban area, but with rural *hukou*.

The methodologies employed in the aforementioned studies exhibit both similarities and differences. Except for Gilleskie (1998) and Dizioli and Pinheiro (2016), which use general equilibrium models, reduced form studies share some similarities. First, all of these studies use two step models to investigate the effect of health insurance on missed workdays (see Vistnes (1997); Lofland and Frick (2006); Xu and

Jensen (2012)), where the first step studies the effect on an individual's probability of missing any workdays and the second step studies the effect on the exact number of workdays missed. In the first step, most studies use a logit model (Vistnes, 1997; Lofland and Frick, 2006; Xu and Jensen, 2012; Guo et al., 2020), while only one study uses a probit model (Li et al., 2019). However, variations in model selection are observed in the second step. Models used to identify the effect of health insurance or other factors on an individual's exact number of workdays missed include the OLS with a log-transformed dependent variable (Lofland and Frick, 2006), the Tobit model (Xu and Jensen, 2012; Li et al., 2019), the Poisson regression (Dizioli and Pinheiro, 2016; Kowal et al., 2013), the negative binomial model (Nurmagambetov et al., 2018; Vistnes, 1997) and the zero-inflated negative binomial (ZINB) model (Nguyen et al., 2021; Guo et al., 2020).

One of the reasons for choosing different models in these studies is that they have different assumptions in the second step. While studying the exact number of workdays missed, most studies include the observations with zero missed workdays (Vistnes, 1997; Xu and Jensen, 2012; Li et al., 2019; Nguyen et al., 2021; Li et al., 2019), while others only use observations with a non-zero number of missed workdays, so the estimation is conditional on missing any workday (Lofland and Frick, 2006). But Lofland and Frick (2006) use an OLS model with a log-transformed number of missed workdays, so the reason for not including the zeros could simply be that zeros cannot be used in a log transformation. However, in studies which include zeros, models still vary. The Tobit model (Xu and Jensen, 2012; Li et al., 2019), negative binomial model (Nurmagambetov et al., 2018; Vistnes, 1997), and generalized Poisson regression model (Dizioli and Pinheiro, 2016; Kowal et al., 2013) recognize that the number of workdays missed includes a substantial number of zeros which right skews the distribution of workdays missed. In contrast, the zero-inflated negative binomial model, which is an example of the zero-inflated models, varies slightly with the other

models which are used for estimating over-dispersed[1] count data (Nguyen et al., 2021; Guo et al., 2020).

The zero-inflated model is usually used when there are two sources of zero outcomes: one is from individuals who do not enter into the counting process and the other is from individuals who enter into the counting process and have zero outcomes (Hardin and Hilbe, 2007). Using the number of workdays missed as an example, if individuals who do not work are also used for estimating the number of workdays missed, then their number of workdays is also zero. In this case, the zeros have two sources, one from individuals who participate in the workforce and miss zero workdays and the other from individuals who do not work at all. When only the working population is considered, the zero-inflated model is not suitable.

However, these models do not consider a measurement issue, which is that the respondents have a tendency to round the number of workdays missed to "focal" values (such as 15 days) when their recall period is long. Survey-based information on the "number of workdays missed" is based on a respondent's recall. The recall period varies across surveys. Most surveys measure absenteeism with the number of workdays missed in the past 1 to 3 months (Kessler et al., 2003; Broadhead et al., 1990). Self-reported data like this have been found to be reliable and valid when the recall periods are short (Mattke et al., 2007), while in the HRS and MEPS data, the recall period for the missed workdays is the past year. Given such a long recall period, the respondents have a tendency to round the number of missed workdays, with the concrete implication that the number of workdays missed are count data with multiple inflated values. This is called "heaped data" (Li et al., 2019b). The heaping model, which is used to estimate heaped data, has been adopted in estimating self-reported cigarette consumption (Wang and Heitjan, 2008; Cummings et al., 2015). However, this type of measurement error due to heaping has not been carefully addressed in

[1] Over-dispersion is observed in the dependent variable, which is the number of missed work days. Over-dispersion is defined as the variance of a variable being greater than the mean.

research about missed workdays. According to Harris (2013), not modelling heaping in discrete quantitative data will result in biased estimates. His simulations show that the heaped Poisson model fits the heaped count data better than a regular Poisson regression model.

There are two heaped models for overdispersed count data: the heaped generalized Poisson[2] and the heaped negative binomial models. Famoye (2005) shows where the data contain a substantial proportion of zeros and thus is over-dispersed, the generalized Poisson regression model is more suitable than the negative binomial regression model. Famoye (2005) compares the two models using a test statistic proposed by Vuong (1989) and a simulation study. Both methods suggest that the generalized Poisson regression estimator is more consistent compared to the negative binomial regression estimator when the data have a high proportion of zeros. This evidence and the fact that the self-reported data on the number of workdays missed in the CHARLS dataset are not only heaped but also contain a substantial number of zeros together suggest that the heaped generalized Poisson regression model is the most suitable for my study.

My model is similar to that of Xu and Jensen (2012), who use the HRS dataset, which is a US version of CHARLS. Xu and Jensen (2012) include two models, one for estimating the likelihood of missing workdays, and the other for estimating the number of workdays missed. To estimate the likelihood of missing workdays, I use the same model as Xu and Jensen (2012), a logit regression. The logit model is the most common model estimating the likelihood of missing workdays (Xu and Jensen, 2012; Vistnes, 1997).

However, my analysis differs from that of Xu and Jensen (2012) in two ways. When estimating the effect of health insurance on the number of workdays missed, I

[2]The generalized Poisson model introduces an additional parameter that allows the variance to be either greater than or less than the mean, addressing over-dispersion and under-dispersion. This model is flexible in handling different kinds of deviations from the Poisson assumption, providing a better fit for data sets where these deviations are evident.

use a heaped generalized Poisson model which has an advantage of estimating heaping data. I also use a different age range, which is 45-85 (Benjamin et al., 2003). Using consistent age groups with chapters 4 and 6 makes the implications of these three chapters compatible.

6.3 Model specification

I study the association between health insurance generosity and the number of missed workdays in two steps. The first step investigates the likelihood of missing any workday using a logit model (Vistnes, 1997; Lofland and Frick, 2006; Xu and Jensen, 2012); the second step examines the number of workdays missed.

In the CHARLS dataset, the number of missed workdays is obtained through the question: "How many days did you miss work last year due to health problems?" Given that various factors aside from health can affect the number of workdays missed, my analysis includes indicators and factors such as demographic and socioeconomic characteristics (Xu and Jensen, 2012), behavioral risk factors (such as smoking and drinking) (Farrell and Stamm, 1988; Harrison and Martocchio, 1998; Aldana and Pronk, 2001), job characteristics (determined by the sector of employment) (Lau et al., 2003; Väänänen et al., 2003; Collins et al., 2005; Association, 2013, 2018; Goetzel et al., 2004; Cawley et al., 2007; Hammond and Levine, 2010; Gilleskie, 2010), and a geographical factor based on the region of residence. The control variables are observed ahead of the dependent variable, so the model can be seen as a predictive model of workdays missed. In addition, I control for an individual's initial health, so the estimated association between health insurance and missed workdays accounts for the influence of health insurance on an individual's health state.

6.3.1 The logit model

To investigate the association between a working individual's health insurance and their probability of missing any workday, I use a logit model

$$P_{i,t} = F(X'_{i,t-1}\beta), \tag{6.1}$$

where F is the cumulative standard logistic distribution, $P_{i,t}$ is individual i's probability of missing any workdays due to health problems over a 12 month period before the survey wave t, $X_{i,t-1}$ is a vector of individual characteristics observed during the previous survey wave $t-1$, and a constant. Individual characteristics include health insurance plan, self-ranked health state, sex, age group, indicators of being disabled and having chronic diseases, indicators of drinking and smoking, educational attainment, marital status, and region of residence.

Given that the demand for physical activity and average working time differs for individuals working in different sectors, the model is estimated separately for individuals working only in the agricultural sector, individuals working only in the non-agricultural sector, and individuals working in both sectors.

6.3.2 The heaped generalized Poisson model

To account for data heaping, I consider the heaped generalized Poisson model. There are two common ways to model heaped data: an interval-censored method and a mixture of rescaled distributions method. Both methods use a mixture of likelihood functions for the heaped and nonheaped count data, allowing for two distinct behavioural responses. One group of respondents reports an exact number of requested missed work days, which is treated as the nonheaped count data, and the other group rounds their missed work days to multiples of certain numbers which are treated as heaped count data. The two methods differ in modelling how the second group rounds

and heaps their responses.

The interval-censored method assumes that the exact numbers from the respondents of the second group are not known but are censored over the half-width of the heaping multiple. For example, in the case that the reported count data are heaped at multiples of 5, the probability function of the exact count data at y is $P\{\in (y-[5/2], y+[5/2])\}$. If the count data are not reported at multiples of 5, which are in the first group, they will be treated as exact numbers and the the probability function of the exact numbers is $P(Y = y)$ (Cummings et al., 2015).

In contrast, the mixture of rescaled distributions method assumes that respondents in the second group remember the requested counts over $1/k$ th of a specified period of time and then report k times that amount. For example, if $k = 5$, individuals in the second group report their missed workdays in the past year (365 days) by recalling their missed workdays in the past $1/5 \times 365$ days, and then by multiplying that number by 5. In the mixture of rescaled distributions method, group one is assumed to behave the same as in group one in the interval-censored method. It is not plausible to implement the mixture of rescaled distribution method here. In the CHARLS dataset, the distribution of number of workdays missed is observed to heap at multiples of 5. If the respondents behave as described by the mixture of rescaled distributions method, then they would all have recalled their workdays missed in the past 365/5 days, which is impossible to verify. It is less likely that all respondents were "rescaling" based on an identical k, even if some of them were extrapolating as such. Thus, I employ the interval-censored method in the heaped generalized Poisson model.

In the model description, I adopt the notation and formulas outline by Cummings et al. (2015). Within this framework, y_{Li} and y_{Ri} denote the left and right endpoints of the interval censored count observations,

$$y_{Li} = \max(0, y_i - \Delta_i)$$

$$y_{Ri} = y_i + \Delta_i$$

$$\Delta_i = \max_{j=1,\ldots,H}\{\lfloor h_j/2 \rfloor \times I(y_i \mod h_j = 0)$$

where $I(\ell)$ is an indicator function, which is equal to 0 if ℓ is false and is equal to 1 if ℓ is true, $\lfloor h_j/2 \rfloor$ is the half-width of the heaping interval, $h_1 = 1$ and H is the total number of heaping intervals. Here, h_2 is set to 5 (except for h_1, as the heaped values are observed at multiples of 5 in the number of workdays missed).

In the generalized Poisson (GP) model, the mean and variance are:

$$E(Y_i) = \frac{\mu_i}{1-\alpha},$$

$$\text{Var}(Y_i) = \frac{\mu_i}{(1-\alpha)^3},$$

where α is the dispersion parameter, $\mu_i > 0$, $\max\left(-1, \frac{-\mu_i}{4}\right) < \alpha < 1$, and $\mu_i = \exp(x_i\beta)$. When the dispersion parameter $\alpha \to 0$, the generalized Poisson distribution is reduced to the Poisson model. Then, the probability of being in the interval is calculated using the interval survival probabilities

$$p_{1i} = P(Y > y_{Li} - 1 \mid Y \sim \text{Gen. Poisson}) = \Gamma_I(y_{Li}\alpha, \mu_i),$$
$$p_{2i} = P(Y > y_{Ri} \mid Y \sim \text{Gen. Poisson}) = \Gamma_I\left((y_{Ri}\alpha) + 1, \mu_i\right),$$

where Γ_I is the regularized incomplete gamma function (Cummings et al. (2015)). Therefore, the log-likelihood function for heaped generalized Poisson model is

$$L = \sum_{i=1}^{n} \ln(p_{1i} - p_{2i}).$$

In this model, Y_i is measured as individual $i's$ number of workdays missed over the year prior to the survey year t. All independent variables are observed at $t-1$ (two years ahead of t) and are the same as those in the logit model. I used the lagged independent variable because, although the dependent variable is collected at time t, it contains information from the year preceding the survey. Using control variables measured at t would place the occurrence of the dependent variable prior to that of the controls. In addition, I control for an individual's initial health to ensure that the estimated effect of health insurance on missed workdays is distinct from its impact on the individual's health status. I also conduct two tests for the model estimation: a likelihood ratio test (LRT) to see whether the dispersion parameter α equals zero and a Vuong test to determine whether the heaped model is preferred over the nonheaped model. The null hypothesis of the LRT test is $\alpha = 0$, and the alternative hypothesis is $\alpha > 0$. Rejecting the null hypothesis indicates that the mean is not equal to the variance of the number of workdays missed, which also implies that the variable is over-dispersed. A significant Vuong test statistic, which implies the null hypothesis is rejected, indicates preference for the heaped model and a non-significant Vuong statistic indicates no preference for either model.

6.4 Data

To estimate these models, I merge the pooled cross-sectional dataset described in Section 3.3 with specific work-related variables that are exclusive to this chapter. These variables include an individual's sector of employment, as observed in the current wave and the subsequent wave, along with their annual number of workdays missed, as reported in the subsequent wave. There are two requirements of the sample. One is that respondents stay employed throughout two contiguous waves. The other is that respondents do not change sector of employment. Only a small fraction of respondents switched their sector of employment (Appendix Table D.1). The two

requirements reduce the pooled dataset to 12,190 observations; 8,707 of them are solely employed in the agricultural sector, 2,396 of them are solely employed in the non-agricultural sector, and 1,097 of them are employed in both sectors.

Table 6.1 presents the mean and standard deviation of the number of workdays missed by sector of employment. For individuals who are employed in both sectors and missed workdays while working at each sector, the total number of workdays missed is defined as the greater of the two numbers reported for each sector. I use the greater number instead of the sum of two numbers for the following reasons. In the CHARLS survey, individuals report their number of workdays missed for each sector separately if they work in both sectors. However, the information is not provided on how an individual will split their time on agricultural work and non-agricultural work. Individuals employed in both sectors can spend half of the day on agricultural work and the other half on non-agricultural work. If this is the case, when they miss one workday, that missed workday can be considered as a missed workday under both the agricultural work and the non-agricultural work. Then summing up the number of missed workdays of each working sector will result in an over-counting. But if the smallest splitting unit is larger than a day, for example, certain individuals spend a few months doing agricultural work while spending the other months doing non-agricultural work, my method of calculating total workdays missed will result in an under-counting. Given that both methods of calculating the total number of workdays missed may lead to biased counts, I use the more conservative method that may lead to an under-counting of workdays missed by those who are employed in both sectors.

Individuals employed only in the agricultural sector on average have the highest number of workdays missed, while individuals employed only in the non-agricultural sector have the lowest. Across the two sample waves, the numbers of missed work days are similar for individuals employed in either the agricultural and non-agricultural

sector, while those employed in both sectors on average had two more missed work days reported in the 2015 wave relative to the 2013 wave. Appendix Tables D.2 to D.7 present more detailed of summary statistics regarding the number of missed work days, categorized by sector of employment and survey wave. Across all three groups, and after pooling, over half of the respondents report not having any missed workdays, with the non-agricultural group having the largest fraction at 84%, and the agricultural group having the smallest fraction at 59%. Additionally, smaller fractions of respondents across all three groups report not having any workdays missed in the 2015 wave. Appendix Tables D.2 to D.7 also indicate that the number of workdays missed are heaped at multiples of 5.

Table 6.1: Summary Statistics: Number of Workdays Missed

Number of workdays missed	Agricultural			Non-agricultural			Both sectors		
	Mean	St.D.	N	Mean	St.D.	N	Mean	St.D.	N
2011-13 wave									
Employee health insurance	12.74	42.74	62	2.68	12.06	342	3.73	8.34	26
Resident health insurance	17.54	44.24	4,357	4.39	19.00	719	10.18	28.20	491
No health insurance	12.66	36.42	216	8.69	33.53	97	6.41	17.66	22
All groups	17.21	43.84	4,653	4.17	18.84	1,184	9.64	27.15	543
2013-15 wave									
Employee health insurance	10.07	33.13	94	4.34	28.62	399	2.72	8.97	39
Resident health insurance	18.73	44.99	4,163	4.68	21.41	914	9.78	24.83	550
No health insurance	15.91	46.02	124	8.22	26.25	95	39.70	74.19	10
All groups	18.37	44.67	4,418	4.80	23.85	1,437	9.82	25.88	602

Notes: This table only includes individuals who were working in two consecutive waves, and who did not switch their sector of employment. Most respondents did not switch between different sectors. Appendix Table D.1 shows the incidence of switching working sectors. Distributions of number of workdays missed are shown in Appendix Tables D.2 and to D.7. St.D. means standard deviation.
Source: Based on pooled CHARLS 2011, 2013 and 2015 survey waves.

Table 6.2 reports the summary statistics for all the variables included in the model. In the pooled data, the agricultural group has the highest proportion of females at 55.6%, while those employed in both sectors have the lowest proportion of females at 21.7%. Only around 2% of respondents employed only in the agricultural sector have employee health insurance, and among those employed only in the non-agricultural sector, the percentage is 29%. The uninsured rate is highest among those who are employed only in the non-agricultural sector, at 7%.

Individuals employed only in the non-agricultural sector report the best health, with the largest proportion of respondents reporting good health and the smallest proportion of respondents reporting poor health. This group also has the lowest fraction of respondents who have at least one type of chronic disease or are disabled. However, they have a higher fraction of respondents who drink regularly and who used to smoke, compared to those who are employed only in the agricultural sector. Individuals working in both sectors have the highest rates of drinking and smoking.

Individuals exclusively employed in the non-agricultural sector exhibit the highest

levels of educational attainment, whereas those exclusively employed in the agricultural sector have the lowest educational attainment. Moreover, the agricultural group displays the highest proportion of individuals who are unmarried, whereas those employed in both sectors are the least likely to be married. In terms of age demographics, the agricultural group has a larger percentage of respondents aged 60 or above, with the age distribution being similar for those employed solely in the non-agricultural sector and those employed in both sectors. Geographically, the agricultural sector presents a higher proportion of respondents residing in the West, and fewer respondents residing in the East.

Table 6.2: Summary Statistics: Workdays Missed Model Variables

Variables	Agricultural sector N = 8,707		Non-agricultural sector N = 2,386		Both sectors N = 1,097	
	N	%	N	%	N	%
	(1)	(2)	(3)	(4)	(5)	(6)
Sex						
Female	4,837	55.6	842	35.3	238	21.7
Male	3,870	44.5	1,544	64.7	859	78.3
Health insurance plan						
Employee	156	1.8	693	29.0	63	5.7
Resident	8,221	94.4	1,518	63.6	1,002	91.3
No health insurance	330	3.8	175	7.3	32	2.9
Health state						
Good	1,824	21.0	907	38.0	375	34.2
Fair	4,604	52.9	1,264	53.0	604	55.1
Poor	2,279	26.2	215	9.0	118	10.8
Chronic disease						
Yes	5,961	68.5	1,358	56.9	646	58.9
No	2,746	31.5	1,028	43.1	451	41.1
Disabled						
Yes	1,192	13.7	166	7.0	101	9.2
No	7,515	86.3	2,220	93.0	996	90.8
Drinking regularly						
Yes	2,334	26.8	908	38.1	513	46.8
No	6,373	73.2	1,478	61.9	584	53.2
Smoked						
Yes	3,389	38.9	1,151	48.2	637	58.1
No	5,318	61.1	1,235	51.8	460	41.9
Educational attainment						
Illiterate	2,839	32.6	162	6.8	102	9.3
Literate, less than high school	5,437	62.4	1,467	61.5	807	73.6
High school and above	431	5.0	757	31.7	188	17.1
Marital Status						
Married, spouse present	7,581	87.1	2,046	85.8	974	88.8
Married, spouse away	450	5.2	201	8.4	83	7.6
Unmarried	676	7.8	139	5.8	40	3.7
Age group						
45 - 49	1,468	16.9	952	39.9	413	37.7
50 - 54	1,383	15.9	632	26.5	267	24.3
55 - 59	2,035	23.4	484	20.3	228	20.8
60 - 64	1,888	21.7	207	8.7	134	12.2
65 - 69	1,171	13.5	72	3.0	39	3.6
70 - 74	511	5.9	26	1.1	13	1.2
75 - 79	207	2.4	10	0.4	3	0.3
80 - 84	44	0.5	3	0.1	-	-
Region						
East	2,088	24.0	1,033	43.3	442	40.3

Continued on next page

Table 6.2 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)
Middle	2,464	28.3	599	25.1	373	34.0
Northeast	447	5.1	216	9.1	57	5.2
West	3,708	42.6	538	22.6	225	20.5

Source: Based on pooled CHARLS 2011 and 2013 survey waves.

6.5 Results

6.5.1 The logit model

Table 6.3 shows the estimated coefficients of the logit model (6.1) by sector of employment. Appendix Table D.8 shows the estimated coefficients of the remaining control variables. In both tables, the first three columns show the estimated coefficients for respondents employed only in the agricultural sector, only in the non-agricultural sector and those employed in both sectors, respectively. The baseline estimation results refer to a married male who has a resident health insurance plan, is between 45 and 49 years of age, is living with his spouse, is not a regular drinker, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, and is living in the Eastern region. A positive sign in columns 1 to 3 means that compared to the baseline category, the category of the explanatory variable is associated with a larger probability of having any missed work days due to health problems, and a negative coefficient represents the category of the explanatory is associated with a smaller probability of having any health related missed work days. Columns 4 to 6 report the average marginal effects.[3] These marginal effects represent the average change in the probability of missing any workdays when the value of the discrete control variable changes from the baseline category to other categories.

[3]In a logit model, average marginal effect is informative about the magnitude of the estimated effect, whereas the odds-ratio is not (Perraillon, 2019; Howell-Moroney, 2023).

Table 6.3: Logit Model: Results and Average Marginal Effects

Dependent variable: having missed workdays

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficients			Average marginal effects		
	Agriculture	Non-agriculture	Both	Agriculture	Non-agriculture	Both
Employee health insurance	−0.627***	0.049	−0.689**	−0.131***	0.006	−0.130**
	(0.199)	(0.151)	(0.332)	(0.038)	(0.019)	(0.055)
No health insurance	−0.137	0.545**	−0.203	−0.030	0.080**	−0.042
	(0.133)	(0.214)	(0.446)	(0.029)	(0.035)	(0.089)
Good health	−0.408***	−0.503***	−0.399***	−0.088***	−0.060***	−0.082***
	(0.064)	(0.139)	(0.152)	(0.014)	(0.016)	(0.031)
Poor health	0.624***	0.632***	0.627***	0.148***	0.106***	0.144***
	(0.056)	(0.175)	(0.223)	(0.013)	(0.033)	(0.052)
Being disabled	0.196***	0.004	0.303	0.044***	0.001	0.063
	(0.066)	(0.218)	(0.216)	(0.015)	(0.028)	(0.045)
Having chronic disease	0.447***	0.612***	0.684***	0.100***	0.079***	0.142***
	(0.055)	(0.131)	(0.150)	(0.012)	(0.017)	(0.030)
Female	0.030	0.127	0.564**	0.007	0.016	0.117***
	(0.071)	(0.157)	(0.222)	(0.016)	(0.020)	(0.045)
Constant	−1.196***	−2.127***	−1.504***	-	-	-
	(0.109)	(0.208)	(0.263)	-	-	-
N	8,707	2,383	1,097	8,707	2,383	1,097

Notes: All independent variables are observed at the survey wave preceding the observations on workdays missed. All models include age effects based on a 5-year window. Appendix Table D.8 reports the estimated coefficients of other control variables. The baseline estimation results refer to a married male who has a resident health insurance plan, is between 45 and 49 years of age, is living with their spouse, is literate but with less than a high school diploma, does not have any chronic disease, is not disabled, does not drink regularly and hasn't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means not applicable.

The results show that there is a strong association between the generosity of the health insurance plan and the probability of missing workdays for individuals who are employed in the agricultural sector or in both sectors. In these two groups, the sign of the coefficient for employee health insurance is negative. The average marginal effect of employee health insurance for those employed in the agricultural sector and those employed in both sectors is similar at -0.13. This means that, compared to individuals with resident health insurance, the probability of missing workdays for those with employee health insurance is, on average, 13 percentage points lower.

For individuals employed only in the non-agricultural sector, the association between employee health insurance and missing workdays is statistically insignificant. However, in this sector, not having any health insurance is associated with a higher probability of missing workdays compared to having resident health insurance. For individuals without any health insurance, the probability of missing workdays is eight percentage points higher than for those with resident health insurance.

The associations between reporting good health and poor health on the probability of missing workdays are almost identical among individuals employed in different sectors. The estimated coefficients for good health are negative. In the agricultural sector, compared to individuals who reported fair health, the probability of missing workdays is nine percentage points lower for those who reported good health. This association is six percentage points and eight percentage points among those employed in the non-agricultural sector and both sectors, respectively. The estimated coefficients for poor health are positive, indicating that compared to individuals who report fair health, the probability of missing workdays is higher for those who reported poor health. The magnitude of this association is 15, 11, and 14 percentage points, respectively, for individuals employed in the agricultural, non-agricultural, and both sectors. Sex is only significant among respondents employed in both sectors. Among individuals employed in both sectors, females, on average, have a 12 percentage point higher probability of missing workdays.

Regarding the other control variables (Appendix Table D.8), only living in the West is statistically significant across all three groups. Living in the West is associated with a higher probability of missing workdays. Having a high school diploma or higher is associated with a lower probability of missing workdays, except for those employed in both sectors. Being married but not living with a spouse is associated with a higher probability of missing workdays only for those employed in both sectors. Another variable significant only for those employed in both sectors is smoking, which

is associated with a higher probability of reporting missed workdays.

Most of the age group variables are statistically insignificant. The coefficient on the age group 60 to 64 is significant only for individuals employed in the agricultural sector and is associated with a higher probability of missing workdays; and the coefficient on the age group 70 to 74 is only significant in the non-agricultural sector and is associated with a lower probability of missing workdays. These associations should be interpreted with the understanding that the estimation results are based on individuals who were willing and able to work in two consecutive survey waves, significantly mitigating the impact of age alone on the number of workdays missed.

Table 6.4 reports the predicted probabilities of missing workdays based on their sector of employment and health insurance plan, while controlling for other variables. The probability of missing workdays is the lowest in the non-agricultural sector overall, although the effect of having employee health insurance is insignificant in this group. Among individuals employed only in the agricultural sector and individuals employed in both sectors, those with an employee health insurance plan have lower probabilities of missing workdays compared to individuals with a resident health insurance plan, with rates of 16% compared to 26% in the agricultural sector and 10% compared to 19% in the "both sectors" group.

Table 6.4: Logit Model: Predicted Probabilities

Dependent variable: having missed workdays

	Sector of employment		
Health insurance plan	Agriculture	Non-agriculture	Both sectors
Employee	0.158***	0.089***	0.103**
	(0.029)	(0.024)	(0.040)
Resident	0.260***	0.085***	0.186***
	(0.020)	(0.021)	(0.046)
No health insurance	0.234***	0.138***	0.157**
	(0.030)	(0.040)	(0.072)
N	8,707	2,383	1,097

Notes: All independent variables are observed at the survey wave preceeding the observations on workdays missed. All models include age effects based on a 5-year window. The baseline estimation results refer to a married male, who is between 45 and 49 years of age, is living with their spouse, is literate but with less than a high school diploma, does not have any chronic disease, is not disabled, does not drink regularly and hasn't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

6.5.2 The heaped generalized Poisson model

Table 6.5 presents the estimated coefficients for the main explanatory variables in the heaped generalized Poisson model and Table 6.6 shows the corresponding average marginal effects. Appendix Tables D.9 and D.10 report the estimated coefficients and the average marginal effects of other control variables, respectively.

In Table 6.5, a positive sign indicates that, when compared to the base category, that particular category of the variable is associated with reporting more workdays missed, whereas a negative sign indicates reporting fewer workdays missed. Estimation results show that for individuals employed in the agricultural sector and employed in both sectors, the employee health insurance plan is associated with reporting fewer workdays missed. In particular, for those in the agricultural sector, having the employee health insurance plan is associated with eight fewer workdays missed; for those employed in both sectors, the association is four fewer workdays missed (Table 6.6). In the case of individuals employed solely in the non-agricultural sector, although the association between employee health insurance and the number of workdays missed is

statistically insignificant, the association between having no health insurance and the number of workdays missed is statistically significant. Having no health insurance is associated with missing three more workdays. This association is not statistically significant for the other two groups. Notably, the number of individuals who are employed in both sectors and do not have any health insurance is small, which may contribute to the insignificance of the coefficient.

Reporting good health is associated with fewer missed workdays, while reporting poor health is associated with more missed workdays. The effects of reporting good health and reporting poor health are both the largest in the agricultural group and the smallest in the non-agricultural group. Compared to those reporting fair health, reporting good health is associated with five fewer missed workdays in the agricultural sector, compared to two fewer days in the non-agricultural sector. Conversely, reporting poor health is associated with 12 more missed workdays in the agricultural group, compared to four more days in the non-agricultural group.

Having a chronic disease is associated with more workdays missed in all three groups, while being disabled is associated with more workdays missed only in the agricultural sector, though the number of observations in this category is small. The association of sex is only significant for those employed in both sectors. Females, on average, are associated with missing five more workdays than males.

Table 6.5: Heaped Generalized Poisson: Estimation Results of Main Explanatory Variables

Dependent variable: number of workdays missed

Variable	(1)	(2)	(3)
Sectors of employment	Agriculture	Non-agriculture	Both
Employee	−0.597***	0.061	−0.616**
	(0.170)	(0.134)	(0.297)
No health insurance	−0.155*	0.524***	−0.250
	(0.090)	(0.169)	(0.321)
Good health	−0.377***	−0.479***	−0.397***
	(0.051)	(0.126)	(0.126)
Poor health	0.572***	0.596***	0.492***
	(0.036)	(0.140)	(0.141)
Being disabled	0.182***	−0.018	0.219
	(0.045)	(0.188)	(0.161)
Having chronic disease	0.422***	0.590***	0.612***
	(0.041)	(0.119)	(0.116)
Female	−0.001	0.127	0.465***
	(0.052)	(0.149)	(0.163)
Constant	2.017***	0.940***	1.407***
	(0.101)	(0.289)	(0.259)
α	0.960***	0.955***	0.945***
	(0.001)	(0.004)	(0.004)
N	8,707	2,386	1,097

Notes: All models include age effects based on a 5-year window. The baseline estimation results refer to a married male who has a resident health insurance plan, is between 45 and 49 years of age, is living with their spouse, is literate but with less than a high school diploma, does not have any chronic disease, is not disabled, does not drink regularly and hasn't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. α is the dispersion parameter. A statistically significant likelihood-ratio test (LRT) of $\alpha = 0$ means that the generalized Poisson model is preferred to the Poisson model. The LRT statistic is statistically significant in all three employment sector groups. The p-value of the Vuong test of heap versus non-heap model is 0.000 in all three employment sector groups. Statistically significant Vuong test of the model shows that the heaped Poisson model is better than the nonheaped model. Estimated coefficients of other control variables are in Appendix Table D.9. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6.6: Heaped Generalized Poisson: Marginal Effects of Main Explanatory Variables

Dependent variable: number of workdays missed

Variable	(1)	(2)	(3)
Sectors of employment	Agriculture	Non-agriculture	Both
Employee	−8.128***	0.263	−4.595**
	(1.904)	(0.637)	(1.886)
No health	−2.592	2.862**	−2.211
insurance	(1.843)	(1.376)	(3.009)
Good health	−4.800***	−1.754***	−3.223***
	(0.698)	(0.484)	(1.108)
Poor health	11.790***	3.754***	6.247**
	(1.101)	(1.308)	(2.830)
Being disabled	3.243***	−0.079	2.125
	(1.020)	(0.924)	(1.777)
Having chronic disease	7.532***	2.620***	5.943***
	(0.929)	(0.641)	(1.444)
Female	−0.011	0.564	4.518**
	(1.161)	(0.685)	(1.865)
N	8,707	2,386	1,097

Notes: All models include age effects by a 5-year window. The baseline estimation results refer to a married male who has a resident health insurance plan, is between 45 and 49 years of age, is living with their spouse, is literate but with less than a high school diploma, does not have any chronic disease, is not disabled, does not drink regularly and hasn't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. The marginal effect of other control variables are in Appendix Table D.10. * Significant at 10%; ** significant at 5%; *** significant at 1%.

6.6 Discussion

Why is the association between employee health insurance and fewer workdays missed significant only for those employed in the agricultural sector and those employed in both sectors? Understanding the role of sick leave policies may help to answer this question. Sick leave policy is a factor of determining number of missed workdays, but is not controlled in the study[4]. Specifically, could the significant association in these two groups be attributed to jobs with employee health insurance also having

[4]Controlling for sick leave policies is challenging because, although general regulations exist, the specific rules can vary between firms. Furthermore, these details are not captured in the questionnaire. Additionally, even if several firms share the same sick leave policies, it remains uncertain whether these rules are strictly enforced.

stricter sick leave policies? Furthermore, in the non-agricultural sector, where the association with employee health insurance is statistically insignificant, are these sick leave policies so stringent that workers enrolled in various health insurance plans end up missing a similar number of workdays, thereby obscuring any potential association with employee health insurance?

To exclude the possibility that differences in workdays missed among individuals with different health insurance plans are due to variations in the regulations of sick leave, section D.1 in Appendix D shows the regulations of medical leave and paid leave of employees. It shows that although the maximum duration of medical leave varies depending on an employee's years of experience, for individuals with the least years of experience, it exceeds three months, which is higher than the average number of observed missed workdays in all groups in the sample. Hence, from the evidence of general regulations on sick leave, having fewer missed work days is unlikely due to the restrictions on sick leave among individuals employed solely in the agricultural sector and both sectors.

However, regulations on maximum sick leave days cannot explain the insignificant employee health insurance effect in the non-agricultural sector. It is possible that unobservable factors, such as common work attitudes within non-agricultural jobs, play a role. Regardless of their health insurance plans, workers might perceive that missing too many workdays could negatively impact their chances for promotion. Such social norms toward working could result in a similar number of missed workdays among individuals in the non-agricultural sector. If workers continue to work even when they are unwell, it becomes difficult to identify the potential benefits of employee health insurance on reducing missed workdays.

Missing fewer workdays does not mean having more workdays. Section D.2 shows that, although individuals with the employee health insurance overall work more days, in the agricultural sector, where the impact of employee health insurance on missed

workdays is most pronounced, individuals with the employee health insurance have a lower average number of workdays, only 140 days annually. Their counterparts with the resident health insurance have, on average, 199 workdays. Hence, in the agricultural sector, missing fewer workdays does not mean having more workdays, and there is a possibility that individuals with the employee health insurance in the agricultural sector miss less workdays since they also work fewer days. However, for individuals who are employed in both sectors, where the employee health insurance also reduces workdays missed, those with the employee health insurance work more days. Within the non-agricultural sector, the annual average numbers of workdays for those with the resident health insurance and for those not having any health insurance are not statistically different, but my previous results show that individuals without any health insurance coverage have more workdays missed relative to those with the resident health insurance. This latter two points support the efficacy of the employee health insurance in reducing workdays missed to some extent.

To further substantiate the effectiveness of employee health insurance in reducing workdays missed, Table D.13 shows the hospital visit rates segmented by individuals' health insurance plans across different working sectors. It reveals that across all sectors, those with resident health insurance exhibit the highest hospital visit rates. A higher hospital visit rate not only suggests poorer health conditions but also partially explains the greater number of workdays missed by individuals with resident health insurance.

In summary, within the agricultural sector and for those employed in both sectors, individuals with more generous employee health insurance not only are less likely to miss workdays but also miss fewer workdays overall. However, it is important to note that some individuals who miss fewer workdays also work fewer days annually.

6.7 Conclusion

I use three waves of the pooled CHARLS dataset to investigate the association between different types of health insurance plans and working individuals' likelihood of missing any workdays due to health reasons, as well as the number of annual workdays missed. I find that, within the agricultural sector and for those employed in both sectors, individuals with employee health insurance not only have lower probabilities of missing any workdays due to health reasons but also miss fewer workdays. These associations are substantial, with eight fewer missed workdays in the agricultural sector and five fewer missed workdays for those employed in both sectors. Although employee health insurance is insignificant for those employed only in the non-agricultural sector, individuals without any health insurance in this group are more likely to miss workdays and miss three more workdays annually. These results suggest that the type of health insurance matters for individuals employed in the agricultural sector and those employed in both sectors, whereas simply having insurance matters for those employed in the non-agricultural sector.

One possible mechanism is that individuals with more generous employee health insurance may experience more successful treatment and recovery when they are in poor health. In this study, individuals' health states and explanatory variables, such as health insurance, are controlled for and observed one year prior to any missed workdays. Therefore, the model used in this study can be considered a predictive model for missed workdays.

The association between health insurance and an individual's missed workdays is even larger when considering the effect of health insurance on health, given that the relationship between health and missed workdays is substantial. For example, individuals reporting poor health have 12 more missed workdays in the agricultural group. Moreover, I find that the association between poor health and the number of missed workdays due to health issues is the largest among all explanatory variables

and is significant across all sectors of employment. In previous chapters, I find that individuals with less generous health insurance are also more likely to report poor health or transition into poor health. Hence, the effect of health insurance on an individual's missed workdays is larger when considering the preventative effect through their health condition. However, this study does not rule out the possibility that individuals with employee health insurance may try to avoid missing workdays due to fear of job loss if they take sick days. The specific mechanisms or pathways through which health insurance affects the number of missed workdays could be explored in future studies.

Overall, my findings suggest that the type of health insurance is associated with the likelihood of an individual missing workdays and the number of workdays missed due to health reasons, even after controlling for their self-ranked health states. For individuals working in the agricultural sector, part-time, or self-employed, a higher number of missed workdays is associated with lower income. Since missed workdays lead to lower income and productivity, the findings suggest that improving the generosity of resident health insurance could mitigate substantial economic losses and boost the productivity of the middle-aged and elderly workforce in China.

Chapter 7

Conclusion

This dissertation compares health and labour market outcomes for the middle-aged and elderly population under China's two main public health insurance plans: employee health insurance and resident health insurance. These two health insurance plans vary significantly in terms of the degree of benefits and financial protection, with employee health insurance being notably more generous. Addressing the disparities between these plans is crucial for developing an equitable health insurance system in China, as these inequalities have profound implications for both health and labour market outcomes.

I first study the association between individuals' health insurance plans and their self ranked health states. The analysis shows that individuals with the less generous resident health insurance are disadvantaged in both their current health state and health transitions over time. These health inequalities between individuals with different health plans could further influence their labour market outcomes. Thus, in chapter 5, I study the association between the type of health insurance plan and individuals' retirement transitions. I find that the more generous employee health insurance plan is associated with a higher likelihood of transitioning to retirement, regardless of individuals' pension plans. However, the employee pension plan has an add-on association. The association between the type of health insurance plan and retirement is stronger for females with an employee pension plan and also stronger for males without an employee pension plan. However, working more years and being less likely to retire are not necessarily a "choice" in this context. The extended working

years could be driven by unaffordable retirement, particularly due to less generous health insurance that fails to provide adequate protection against catastrophic health spending.

The inability to retire due to financial constraints can severely impact the wellbeing of the elderly. Prolonged working years, especially when not by choice, can lead to increased physical strain and exacerbate health issues, thereby diminishing the overall quality of life. Moreover, the financial pressure to continue working can contribute to psychological stress and mental health problems, such as anxiety and depression. The constant worry about healthcare costs and insufficient retirement savings creates a pervasive sense of insecurity, undermining the mental and emotional wellbeing of the elderly. This situation implies that the poor social safety net may have significant negative implications for the quality of life of the elderly, which requires further investigation in future research.

In addition, working longer years does not necessarily equate to working efficiently. Although individuals with the less generous resident health insurance are less likely to retire and thus work more years, chapter 6 reveals that they miss more workdays. Compared to individuals with the resident health insurance plan, those with more generous employee health insurance are likely to miss fewer workdays, which may be due to better access to healthcare services and preventive care. This not only enhances their productivity but also reduces the economic burden on employers and the healthcare system. By contrast, those with less generous health insurance coverage are likely to have more missed workdays.

These findings underscore the urgent need to enhance the social safety net to provide sufficient economic security for the middle-aged and elderly people in China. Improving health insurance coverage and making it more equitable across different demographic groups is essential.

This dissertation, however, did not explore the causal relationship or the exact

pathways through which health insurance affects health, retirement and health-related workdays missed. This is because the eligibility for employee health insurance is determined by an individual's job, making it challenging to identify the causal relationship if a person switches from resident health insurance to employee health insurance. Nevertheless, the unique cohort selected for this study allows the introduction of health insurance plans in China to serve as a quasi-natural experiment. Thus, the observed association between health insurance plan and health or labour outcomes may suggest the effects of health insurance on these variables, even though a causal relationship is not directly analysed.

Despite the existing inequalities in health and work outcomes among people with different types of health insurance, mitigating the differences in the generosity of various health insurance plans remains challenging. Compared to the resident insurance plan, the employee health insurance plan requires higher premiums. Although significant health and labour inequalities exist among individuals with different types of health insurance, closing these gaps requires substantial financial input from the government. Understanding the fiscal and general equilibrium implications of a health insurance plan reform with elevated benefits is left for future research.

To conclude, this dissertation provides compelling evidence that in China, individuals with resident health insurance are at a significant disadvantage in terms of both health and labour market outcomes. The disparities in health insurance benefits between employee health insurance and resident health insurance result in unequal access to healthcare, which in turn affects the ability of individuals to work efficiently and maintain good health. Addressing these disparities is crucial for fostering a more equitable and effective health insurance system. By ensuring that all individuals, regardless of their employment status, have access to comprehensive health insurance, policymakers can enhance the overall wellbeing of the population, reduce health-related missed workdays, and improve labour market outcomes. This would

ultimately contribute to a more productive and healthier society.

Bibliography

51 Shebao (2023). Effective from november 1st, the duration for medical insurance payments is extended. Can't enjoy lifetime medical insurance benefits with only 15 years of payment? http://www.51shebao.com/article/detail/7162. Accessed: 2023-11-16.

Administration, N.H.S (2021). Statistical report on the development of national basic medical insurance. http://www.nhsa.gov.cn/art/2021/6/8/art_7_5232.html. Accessed: 2023-09-24.

Aldana, S. G. and Pronk, N. P. (2001). Health promotion programs, modifiable health risks, and employee absenteeism. *Journal of Occupational and Environmental Medicine*, pages 36–46.

Alexanderson, K. (1998). Sickness absence: A review of performed studies with focused on levels of exposures and theories utilized. *Scandinavian journal of social medicine*, 26(4):241–249.

Association, A. D. (2013). Economic costs of diabetes in the US in 2012. *Diabetes Care*, 36(4):1033–1046.

Association, A. D. (2018). Economic costs of diabetes in the US in 2017. *Diabetes Care*, 41(5):917–928.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., and O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 17(3):351–375.

Baidin, V., Gerry, C. J., and Kaneva, M. (2021). How self-rated is self-rated health? Exploring the role of individual and institutional factors in reporting heterogeneity in russia. *Social Indicators Research*, 155(2):675–696.

Bairoliya, N., Canning, D., Miller, R., and Saxena, A. (2018). The macroeconomic and welfare implications of rural health insurance and pension reforms in China. *The Journal of the Economics of Ageing*, 11:71–92.

Beijing Municipal Medical Insurance Bureau (2023). Medical insurance benefits for urban employees in Beijing. https://ybj.beijing.gov.cn/2020_zwfw/2020_bmcx/202004/t20200409_1798870.html. Accessed: 2023-11-16.

Benitez-Silva, H. and Ni, H. (2008). Health status and health dynamics in an empirical model of expected longevity. *Journal of health economics*, 27(3):564–584.

Benjamin, D., Brandt, L., and Fan, J.-Z. (2003). Ceaseless toil? Health and labor supply of the elderly in rural China. *Health and Labor Supply of the Elderly in Rural China (June 2003)*.

Bhattacharyya, O., Delu, Y., Wong, S. T., and Bowen, C. (2011). Evolution of primary care in China 1997–2009. *Health policy*, 100(2-3):174–180.

Bian, Y. and Zhang, Z. (2002). Marketization and income distribution in urban China, 1988 and 1995. *Research in Social Stratification and Mobility*, 19:377–415.

Boyreau-Debray, G. and Wei, S.-J. (2004). Can China grow faster? A diagnosis of the fragmentation of its domestic capital market.

Bramall, C. (2001). The quality of China's household income surveys. *The China Quarterly*, 167:689–705.

Briggs, A. and Sculpher, M. (1998). An Introduction to Markov modelling for economic evaluation. *PharmacoEconomics*, 13(4):397–409.

Broadhead, W. E., Blazer, D. G., George, L. K., and Tse, C. K. (1990). Depression, disability days, and days lost from work in a prospective epidemiologic survey. *Jama*, 264(19):2524–2528.

Brown, S. and Sessions, J. G. (1996). The economics of absence: Theory and evidence. *Journal of economic surveys*, 10(1):23–53.

Cao, Q., Shi, L., Wang, H., and Dong, K. (2012). Report from China: Health insurance in China—evolution, current status, and challenges. *International Journal of Health Services*, 42(2):177–195.

Carro, J. M. and Traferri, A. (2009). Correcting the bias in the estimation of a dynamic ordered probit with fixed effects of self-assessed health status.

Cawley, J., Rizzo, J. A., and Haas, K. (2007). Occupation-specific absenteeism costs associated with obesity and morbid obesity. *Journal of Occupational and Environmental Medicine*, pages 1317–1324.

Che, Y. and Li, X. (2018). Retirement and health: Evidence from China. *China Economic Review*, 49:84–95.

Chen, Y. and Jin, G. Z. (2012). Does health insurance coverage lead to better health and educational outcomes? Evidence from rural China. *Journal of Health Economics*, 31(1):1–14.

Chen, Y., Shi, J., and Zhuang, C. C. (2019). Income-dependent impacts of health insurance on medical expenditures: Theory and evidence from China. *China Economic Review*, 53:290–310.

Cheng, L., Liu, H., Zhang, Y., Shen, K., and Zeng, Y. (2015). The impact of health insurance on health outcomes and spending of the elderly: Evidence from China's new cooperative medical scheme. *Health Economics*, 24(6):672–691.

Collins, J. J., Baase, C. M., Sharda, C. E., Ozminkowski, R. J., Nicholson, S., Billotti, G. M., Turpin, R. S., Olson, M., and Berger, M. L. (2005). The assessment of chronic health conditions on work performance, absence, and total economic impact for employers. *Journal of Occupational and Environmental Medicine*, pages 547–557.

Craig, B. A. and Sendi, P. P. (2002). Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics*, 11(1):33–42.

Culyer, A. J., Lavers, R., and Williams, A. (1971). Social indicators: Health. *Social Trends*, 2:31–42.

Cummings, T. H., Hardin, J. W., McLain, A. C., Hussey, J. R., Bennett, K. J., and Wingood, G. M. (2015). Modeling heaped count data. *The Stata Journal*, 15(2):457–479.

Cutler, D. M., Liebman, J. B., and Smyth, S. (2006). How fast should the social security eligibility age rise? *Cambridge, MA: Harvard University*.

Dalstra, J. A., Kunst, A. E., Mackenbach, J. P., on Socioeconomic Inequalities in Health, E. W. G., et al. (2006). A comparative appraisal of the relationship of education, income and housing tenure with less than good health among the elderly in Europe. *Social science & medicine*, 62(8):2046–2060.

Davis, D. (1991). Long lives: Chinese elderly and the communist revolution, volume 100. *Stanford University Press.*

Deaton, A. (1999). Inequalities in income and inequalities in health.

Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health affairs*, 21(2):13–30.

Deng, C., Li, X., and Pan, J. (2018). Private hospital expansion in China: A global perspective. *Global Health Journal*, 2(2):33–46.

DeSalvo, K. B., Fisher, W. P., Tran, K., Bloser, N., Merrill, W., and Peabody, J. (2006). Assessing measurement properties of two single-item general health measures. *Quality of Life Research*, 15:191–201.

Dizioli, A. and Pinheiro, R. (2016). Health insurance as a productive factor. *Labour Economics*, 40:1–24.

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

Etilé, F. and Milcent, C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from france. *Health economics*, 15(9):965–981.

Famoye, F. (2005). Count data modeling: Choice between generalized poisson model and negative binomial model. *Journal of Applied Statistical Science*, 14(1-2):99–106.

Fan, G., Deng, Z., Wu, X., and Wang, Y. (2020). Medical insurance and health equity in health service utilization among the middle-aged and older adults in China: A quantile regression approach. *BMC Health Services Research*, 20:1–12.

Fan, H., Yan, Q., Coyte, P. C., and Yu, W. (2019). Does public health insurance coverage lead to better health outcomes? Evidence from Chinese adults. *The Journal of Health Care Organization, Provision, and Financing*, 56:1–10.

Fan, X., Su, M., Si, Y., Zhao, Y., and Zhou, Z. (2021). The benefits of an integrated social medical insurance for health services utilization in rural China: Evidence from the China Health and Retirement Longitudinal Study. *International Journal for Equity in Health*, 20(1):126.

Fang, H. and Feng, J. (2018). The Chinese pension system. *Technical report, National Bureau of Economic Research.*

Fang, L. and Shi, R. (2022). Comparative analysis of the effects of retirement on health status of older adulthood. *International Journal of Environmental Research and Public Health*, 19(16):9957.

Fang, P., Han, X., Chen, J., Feng, L., Tang, S., and Yu, H. (2011). How many working days would be missed due to moderate or severe influenza pandemic in China? *Scandinavian Journal of Public Health*, 39(6):656–664.

Farrell, D. and Stamm, C. L. (1988). Meta-analysis of the correlates of employee absence. *Human Relations*, 41(3):211–227.

Feng, J., Li, Q., and Smith, J. P. (2020). Retirement effect on health status and health behaviors in urban China. *World Development*, 126:104702.

Feng, Q., Yeung, W.-J. J., Wang, Z., and Zeng, Y. (2019). Age of retirement and human capital in an aging China, 2015–2050. *European Journal of Population*, 35:29–62.

Frankovic, I. and Kuhn, M. (2023). Health insurance, endogenous medical progress, health expenditure growth, and welfare. *Journal of Health Economics*, 87:102717.

Franks, P., Gold, M. R., and Fiscella, K. (2003). Sociodemographics, self-rated health, and mortality in the US. *Social science & medicine*, 56(12):2505–2514.

French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.

Frijters, P., Haisken-DeNew, J. P., and Shields, M. A. (2005). The causal effect of income on health: Evidence from german reunification. *Journal of health economics*, 24(5):997–1017.

Fuchs, V. R. (2004). Reflections on the socio-economic correlates of health. *Journal of health economics*, 23(4):653–661.

General Office of the State Council, PRC (2007). Regulations on paid annual leave for employees. https://www.gov.cn/ziliao/flfg/2007-12/16/content_835527.htm. Accessed: 2023-09-24.

Giles, J., Lei, X., Wang, G., Wang, Y., and Zhao, Y. (2023). One country, two systems: Evidence on retirement patterns in China. *Journal of Pension Economics & Finance*, 22(2):188–210.

Giles, J., Lei, X., Wang, G., Wang, Y., and Zhao, Y. (2023). One country, two systems: Evidence on retirement patterns in China. *Journal of Pension Economics and Finance*, 22(2):188–210.

Giles, J., Lei, X., Wang, Y., and Zhao, Y. (2015). One country, two systems: Evidence on retirement patterns in China. *Asian Bureau of Finance and Economic Research Working Paper, February*.

Gilleskie, D. (2010). Work absences and doctor visits during an illness episode: The differential role of preferences, production, and policies among men and women. *Journal of Econometrics*, 156(1):148–163.

Gilleskie, D. B. (1998). A dynamic stochastic model of medical care use and work absence. *Econometrica*, pages 1–45.

Goetzel, R. Z., Long, S. R., Ozminkowski, R. J., Hawkins, K., Wang, S., and Lynch, W. (2004). Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting US employers. *Journal of Occupational and Environmental Medicine*, pages 398–412.

Gong, C. H., Kendig, H., and He, X. (2016). Factors predicting health services use among older people in China: An analysis of the China Health and Retirement Longitudinal Study 2013. *BMC Health Services Research*, 16(1):63.

Greene, W. H. (2000). Econometric analysis 4th edition. *International edition, New Jersey: Prentice Hall*, pages 201–215.

Guo, M., Tang, K., and Wang, Z. (2020). Commuting time and sickness absence in China: Rural/urban variations and *hukou* impacts. *The Economic and Labour Relations Review*, 31(1):76–95.

Hammond, R. A. and Levine, R. (2010). The economic impact of obesity in the United States. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, pages 285–295.

Hardin, J. W. and Hilbe, J. M. (2007). Generalized linear models and extensions. *Stata press*.

Harris, T. (2013). Heaped data in count models.

Harrison, D. A. and Martocchio, J. J. (1998). Time for absenteeism: A 20-year review of origins, offshoots, and outcomes. *Journal of management*, 24(3):305–350.

Healthcare Security Administration of Guangdong Province (2022). Interim measures for the transfer and continuation of basic medical insurance relations within guangdong province (guangdong sheng jiben yiliao baoxian guanxi shengnei zhuanyi jiexu zanxing banfa). http://hsa.gd.gov.cn/zwgk/zcfg/ylbz/content/post_3953358.html. Accessed: 2023-11-16.

Heeley, E., Anderson, C. S., Huang, Y., Jan, S., Li, Y., Liu, M., Sun, J., Xu, E., Wu, Y., Yang, Q., et al. (2009). Role of health insurance in averting economic hardship in families after acute stroke in China. *Stroke*, 40(6):2149–2156.

Hernández-Quevedo, C., Jones, A. M., Rice, N., et al. (2004). Reporting bias and heterogeneity in self-assessed health. Evidence from the British household panel survey. *Health, Econometrics and Data Group (HEDG) Working paper 05*, 4.

Holstein, A. D. et al. (2004). Health outcomes and the cost-quality trade-off in health care: Empirical study of OECD countries. *International Business & Economics Research Journal (IBER)*, 3(7).

Holtz-Eakin, D. (1993). Health insurance provision and labor market efficiency in the United States. *Technical report, National Bureau of Economic Research Working Paper No.4388*.

Hou, X. and Zhang, J. (2017). The effects of public health insurance expansion on private health insurance in urban China. *International Journal of Health Economics and Management*, 17(3):359–375.

House, J. S., Lepkowski, J. M., Kinney, A. M., Mero, R. P., Kessler, R. C., and Herzog, A. R. (1994). The social stratification of aging and health. *Journal of health and social behavior*, pages 213–234.

Howell-Moroney, M. (2023). Inconvenient truths about logistic regression and the remedy of marginal effects. *Public Administration Review*.

Hu, S., Tang, S., Liu, Y., Zhao, Y., Escobar, M.-L., and De Ferranti, D. (2008). Reform of how health care is paid for in China: Challenges and opportunities. *Lancet*, 372(9652):1846–1853.

Huang, X. and Wu, B. (2020). Impact of urban-rural health insurance integration on health care: Evidence from rural China. *China Economic Review*, 64:101543.

Huguet, N., Kaplan, M. S., and Feeny, D. (2008). Socioeconomic status and health-related quality of life among elderly people: Results from the joint canada/united states survey of health. *Social science & medicine*, 66(4):803–810.

Idler, E. and Cartwright, K. (2018). What do we rate when we rate our health? Decomposing age-related contributions to self-rated health. *Journal of health and social behavior*, 59(1):74–93.

Idler, E. L. and Kasl, S. V. (1995). Self-ratings of health: Do they also predict change in functional ability? *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 50(6):S344–S353.

Jaelani, A. K., Nuryanto, A. D., Fenitra, R. M., Mujib, M. M., and Luthviati, R. D. (2023). Legal protection of employee wage rights in bankrupt companies: Evidence from China. *Legality: Jurnal Ilmiah Hukum*, 31(2):202–223.

Jiang, J., Huang, W., Wang, Z., and Zhang, G. (2019). The effect of health on labour supply of rural elderly people in China—an empirical analysis using CHARLS data. *International Journal of Environmental Research and Public Health*, 16(7):1195.

Jin, Y., Hou, Z., and Zhang, D. (2016). Determinants of Health Insurance Coverage among People Aged 45 and over in China: Who buys public private and multiple insurance. *PLOS ONE*, 11(8):e0161774.

JIN, Z., FENG, X., ZHAO, Y.-d., et al. (2014). Cognition and utilization of individual account of medical insurance among employees in Shenyang city. *Chinese Journal of Public Health*, 30(7):923–925.

Johansson, P.-O. (2000). Properties of actuarially fair and pay-as-you-go health insurance schemes for the elderly. an olg model approach. *Journal of health economics*, 19(4):477–498.

Johnson, R. W., Davidoff, A. J., and Perese, K. (2003). Health insurance costs and early retirement decisions. *ILR Review*, 56(4):716–729.

Johnstone, D. B. and Shroff-Mehta, P. (2003). Higher education finance and accessibility: An international comparative examination of tuition and financial assistance policies. *Globalization and reform in higher education*, 11.

Jung, J. (2006). Estimating Markov transition probabilities between health states in the HRS dataset. *Indiana University*, pages 1–42.

Jung, J. (2020). Estimating Markov transition probabilities between health states using U.S. longitudinal survey data. *SSRN Electronic Journal*.

Jung, J. and Tran, C. (2016). Market inefficiency, insurance mandate and welfare: US health care reform 2010. *Review of Economic Dynamics*, 20:132–159.

Jürges, H., Thiel, L., Bucher-Koenen, T., Rausch, J., Schuth, M., and Börsch-Supan, A. (2014). Health, financial incentives, and early retirement: microsimulation Evidence for Germany. *Social Security Programs and Retirement around the World: Disability Insurance Programs and Retirement*, pages 285–330. University of Chicago Press.

Kapur, K. and Rogowski, J. (2011). How does health insurance affect the retirement behavior of women? *Journal of Health Care Organization, Provision, and Financing*, 48(1):51–67.

Karoly, L. A. and Rogowski, J. A. (1994). The effect of access to post-retirement health insurance on the decision to retire early. *ILR Review*, 48(1):103–123.

Kessler, R. C., Barber, C., Beck, A., Berglund, P., Cleary, P. D., McKenas, D., Pronk, N., Simon, G., Stang, P., Ustun, T. B., et al. (2003). The World Health Organization health and work performance questionnaire (HPQ). *Journal of Occupational and Environmental Medicine*, pages 156–174.

Kowal, S. L., Dall, T. M., Chakrabarti, R., Storm, M. V., and Jain, A. (2013). The current and projected economic burden of Parkinson's disease in the United States. *Movement Disorders*, 28(3):311–318.

Laitner, J. and Silverman, D. (2005). Estimating life-cycle parameters from consumption behavior at retirement.

Lardy, N. R. (2016). China: Toward a consumption-driven growth path. In *Seeking changes: The economic development in contemporary China*, pages 85–111. World Scientific.

Lau, V. C., Au, W. T., and Ho, J. M. (2003). A qualitative and quantitative review of antecedents of counterproductive behavior in organizations. *Journal of Business and Psychology*, 18:73–99.

Lee, D.-C., Wang, J., Shi, L., Wu, C., and Sun, G. (2022). Health insurance coverage and access to care in China. *BMC Health Services Research*, 22(1):140.

Lee, S., Lee, J., and Chang, Y. (2014). Is dual income costly for married couples? An analysis of household expenditures. *Journal of Family and Economic Issues*, 35:161–177.

Lei, X. and Lin, W. (2009). The New Cooperative Medical Scheme in rural China: Does more coverage mean more service and better health? *Health Economics*, 18(S2).

Lei, X., Sun, X., Strauss, J., Zhao, Y., Yang, G., Hu, P., Hu, Y., and Yin, X. (2014). Health outcomes and socio-economic status among the mid-aged and elderly in China: Evidence from the CHARLS national baseline data. *The Journal of the Economics of Ageing*, 3:29–43.

Lei, X., Tan, L., and Zhao, Y. (2011). The impact of retirement on health: Evidence from China. *Unpublished Manuscript, China Center for Economic Research, Peking University, Peking*.

Li, C., Tang, C., and Wang, H. (2019). Effects of health insurance integration on health care utilization and its equity among the mid-aged and elderly: Evidence from China. *International Journal for Equity in Health*, 18(1):1–12.

Li, H., Shi, X., and Wu, B. (2016). The retirement consumption puzzle revisited: Evidence from the mandatory retirement policy in China. *Journal of Comparative Economics*, 44(3):623–637.

Li, Q., Tso, G. K., Qin, Y., Lovejoy, T. I., Heckman, T. G., and Li, Y. (2019b). Penalized multiple inflated values selection method with application to safer data. *Statistical Methods in Medical Research*, 28(10-11):3205–3225.

Li, X., Huang, J., and Zhang, H. (2008). An analysis of hospital preparedness capacity for public health emergency in four regions of China: Beijing, Shandong, Guangxi, and Hainan. *BMC Public Health*, 8(1):319.

Li, X. and Zhang, W. (2013). The impacts of health insurance on health care utilization among the older people in China. *Social Science & Medicine*, 85:59–65.

Li, X., Zhang, W., Sun, H., and Anis, A. H. (2019). Effects of health status on work exit and absenteeism among the older working population in China: A secondary analysis of a cohort sample. *BMJ open*, 9(9):e024115.

Li, Y., Wu, Q., Liu, C., Kang, Z., Xie, X., Yin, H., Jiao, M., Liu, G., Hao, Y., and Ning, N. (2014). Catastrophic health expenditure and rural household impoverishment in China: What role does the new cooperative health insurance scheme play? *Plos one*, 9(4):e93253.

Li, Y., Wu, Q., Xu, L., Legge, D., Hao, Y., Gao, L., Ning, N., and Wan, G. (2012). Factors affecting catastrophic health expenditure and impoverishment from medical expenses in China: Policy implications of universal health insurance. *Bulletin of the World Health Organization*, 90:664–671.

Lim, K. M. (2016). Public provision of health insurance and welfare. *The BE Journal of Macroeconomics*, 16(2):439–483.

Liu, G. G., Zhao, Z., Cai, R., Yamada, T., and Yamada, T. (2002). Equity in health care access to: Assessing the urban health insurance reform in China. *Social Science & Medicine*, 55(10):1779–1794.

Liu, Q., Wang, B., Kong, Y., and Cheng, K. (2011). China's primary health-care reform. *The Lancet*, 377(9783):2064–2066.

Liu, X., Wong, H., and Liu, K. (2015). Outcome-based health equity across different social health insurance schemes for the elderly in China. *BMC Health Services Research*, 16(1):1–12.

Liu, Z. (1998). Earnings, education, and economic reforms in urban China. *Economic development and cultural change*, 46(4):697–725.

Lixin Cai (2009). Be wealthy to stay healthy: An analysis of older Australians using the HILDA survey. *Journal of Sociology*, 45(1):55–70.

Lofland, J. H. and Frick, K. D. (2006). Effect of health insurance on workplace absenteeism in the US workforce. *Journal of Occupational and Environmental Medicine*, pages 13–21.

Lundberg, O. and Manderbacka, K. (1996). Assessing reliability of a measure of self-rated health. *Scandinavian journal of social medicine*, 24(3):218–224.

Madrian, B. C., Burtless, G., and Gruber, J. (1994). The effect of health insurance on retirement. *Brookings Papers on Economic Activity*, 1994(1):181–252.

Mattke, S., Balakrishnan, A., Bergamo, G., and Newberry, S. J. (2007). A review of methods to measure health-related productivity loss. *American Journal of Managed Care*, 13(4):211.

McCallum, J., Shadbolt, B., and Wang, D. (1994). Self-rated health and survival: A 7-year follow-up study of Australian elderly. *American Journal of Public Health*, 84(7):1100–1105.

Meer, J., Miller, D. L., and Rosen, H. S. (2003). Exploring the health–wealth nexus. *Journal of health economics*, 22(5):713–730.

Mei, J. and Kirkpatrick, I. (2019). Public hospital reforms in China: Towards a model of new public management? *International Journal of Public Sector Management*, 32(4):352–366.

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

Meng, X. (2003). Unemployment, consumption smoothing, and precautionary saving in urban China. *Journal of Comparative Economics*, 31(3):465–485.

Meng, X. (2004). Economic restructuring and income inequality in urban China. *Review of Income and Wealth*, 50(3):357–379.

Ministry of Human Resources and Social Security of the People's Republic of China (2012). Interpretation of the social insurance law of the people's republic of China (xii). http://www.mohrss.gov.cn/SYrlzyhshbzb/rdzt/syshehuibaoxianfa/bxffaguijijiedu/201208/t20120807_28573.html#:~:

text=%E5%9F%8E%E9%95%87%E5%B1%85%E6%B0%91%E5%9F%BA%E6%9C%AC%E5%8C%BB%E7%96%97%E4%BF%9D%E9%99%A9,%E6%A0%87%E5%87%86%E5%92%8C%E8%B4%A2%E6%94%BF%E8%A1%A5%E5%8A%A9%E6%A0%87%E5%87%86%E3%80%82. Accessed: 2023-11-16.

Ministry of Human Resources and Social Security, PRC (1953). Draft amendment to the implementation rules of the people's republic of China labor insurance regulations. http://www.mohrss.gov.cn/xxgk2020/fdzdgknr/zcfg/gfxwj/shbx/201308/t20130808_109736.html. Accessed: 2023-09-24.

Ministry of Human Resources and Social Security, PRC (1994). Regulations for the medical leave of enterprise employees due to illness or non-work-related injuries. http://www.mohrss.gov.cn/xxgk2020/gzk/gz/202112/t20211228_431556.html. Accessed: 2023-09-24.

Ministry of Human Resources and Soical Security, PRC (1995). Notice regarding the issuance of opinions on the implementation of the labor law of the people's republic of China. http://www.mohrss.gov.cn/xxgk2020/fdzdgknr/zcfg/gfxwj/zh/202103/t20210330_412011.html. Accessed: 2023-09-24.

Modigliani, F. and Cao, S. L. (2004). The Chinese saving puzzle and the life-cycle hypothesis. *Journal of economic literature*, 42(1):145–170.

Molina, T. (2016). Reporting heterogeneity and health disparities across gender and education levels: Evidence from four countries. *Demography*, 53(2):295–323.

Molla, M. T. and Madans, J. H. (2008). Estimating healthy life expectancies using longitudinal survey data: Methods and techniques in population health measures. *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research*, (146):1–24.

Mossialos, E., Ge, Y., Hu, J., Wang, L., Organization, W. H., et al. (2016). Pharmaceutical policy in China: Challenges and opportunities for reform. *World Health Organization. Regional Office for Europe.*

Mu, R. (2014). Regional disparities in self-reported health: Evidence from chinese older adults. *Health economics*, 23(5):529–549.

Muldoon, K. A., Galway, L. P., Nakajima, M., Kanters, S., Hogg, R. S., Bendavid, E., and Mills, E. J. (2011). Health system determinants of infant, child and maternal mortality: A cross-sectional study of un member countries. *Globalization and health*, 7:1–10.

Mullins, J. T. and White, C. (2019). Temperature and mental health: Evidence from the spectrum of mental health outcomes. *Journal of health economics*, 68:102240.

National Academies of Sciences, E., Medicine, et al. (2022). Understanding the aging workforce: Defining a research agenda.

National Bureau of Statistics (2023). What are poverty standards and poverty incidence rate. http://www.stats.gov.cn/zs/tjws/tjzb/202301/t20230101_1903716.html. Accessed: 2023-04-28.

National Health and Family Planning Commision (2000-2018). Yearbook of health and family planning in the people's republic of China. http://60.16.24.131/CSYDMirror/area/Yearbook/Single/N2014010012?z=D13. Accessed: 2021-02-16.

National Health and Family Planning Commision (2013). Yearbook of health and family planning in the people's republic of China. http://60.16.24.131/CSYDMirror/area/Yearbook/Single/N2014010012?z=D13. Accessed: 2021-02-16.

National Healthcare Security Administration (2020). National medical security development statistical bulletin in 2020. http://www.nhsa.gov.cn/art/2021/6/8/art_7_5232.html. Accessed: 2023-10-01.

Nee, V. and Cao, Y. (2005). Market transition and the firm: Institutional change and income inequality in urban China. *Management and Organization Review*, 1(1):23–56.

Nguyen, K.-D., Tran, V.-A., and Nguyen, D.-T. (2021). Social insurance reform and absenteeism in Vietnam. *International Journal of Social Welfare*, 30(2):193–207.

Ning, M., Gong, J., Zheng, X., and Zhuang, J. (2016a). Does new rural pension scheme decrease elderly labor supply? Evidence from CHARLS. *China Economic Review*, 41:315–330.

Ning, M., Zhang, Q., and Yang, M. (2016b). Comparison of self-reported and biomedical data on hypertension and diabetes: Findings from the China Health and Retirement Longitudinal Study (CHARLS). *BMJ open*, 6(1).

Nurmagambetov, T., Kuwahara, R., and Garbe, P. (2018). The economic burden of asthma in the United States, 2008–2013. *Annals of the American Thoracic Society*, 15(3):348–356.

O'Doherty, M., French, D., Steptoe, A., and Kee, F. (2017). Social capital, deprivation and self-rated health: Does reporting heterogeneity play a role? Results from the english longitudinal study of ageing. *Social Science & Medicine*, 179:191–200.

Okun, M. A., Stock, W. A., Haring, M. J., and Witter, R. A. (1984). Health and subjective well-being: A meta-analyis. *The International journal of aging and human development*, 19(2):111–132.

Pan, J., Lei, X., and Liu, G. G. (2016). Health insurance and health status: exploring the causal effect from a policy intervention: Health Insurance and health status: Exploring the causal effect. *Health Economics*, 25(11):1389–1402.

Pavalko, E. K. and Caputo, J. (2013). Social inequality and health across the life course. *American Behavioral Scientist*, 57(8):1040–1056.

Perraillon, M. C. (2019). Interpreting model estimates: Marginal effects. *University of Colorado Anschutz Medical Campus, Nova School of Business and Economics Lisbon, Portugal*.

Pfarr, C., Schmid, A., and Schneider, U. (2012). Reporting heterogeneity in self-assessed health among elderly Europeans. *Health economics review*, 2:1–14.

Pijls, B. G., Jolani, S., Atherley, A., Derckx, R. T., Dijkstra, J. I., Franssen, G. H., Hendriks, S., Richters, A., Venemans-Jellema, A., Zalpuri, S., et al. (2021). Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ open*, 11(1):e044640.

Pradhan, A. and Olsson, P.-E. (2020). Sex differences in severity and mortality from covid-19: are males more vulnerable? *Biology of sex Differences*, 11(1):53.

Preston, S. H., Stokes, A., Mehta, N. K., and Cao, B. (2014). Projecting the effect of changes in smoking and obesity on future life expectancy in the United States. *Demography*, 51(1):27–49.

Qingyue, M. and Shenglan, T. (2013). Universal health care coverage in China: Challenges and opportunities. *Procedia-Social and Behavioral Sciences*, 77:330–340.

Ranabhat, C. L., Atkinson, J., Park, M.-B., Kim, C.-B., and Jakovljevic, M. (2018). The influence of universal health coverage on life expectancy at birth (leab) and healthy life expectancy (hale): A multi-country cross-sectional study. *Frontiers in pharmacology*, 9:399630.

Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human resource management*, 57(3):739–750.

Rözer, J. J. and Volker, B. (2016). Does income inequality have lasting effects on health and trust? *Social science & medicine*, 149:37–45.

Rust, J. P. (1990). Behavior of male workers at the end of the life cycle: An empirical analysis of states and controls. In *Issues in the Economics of Aging*, pages 317–382. University of Chicago Press, 1990.

Schneider, U., Pfarr, C., Schneider, B. S., and Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *The European Journal of Health Economics*, 13:251–265.

Schwartzman, S., Pinheiro, R., and Pillay, P. (2015). *Higher education in the BRICS countries: Investigating the pact between higher education and society*, volume 44. Springer.

Shmueli, A. (2003). Socio-economic and demographic variation in health and in its measures: The issue of reporting heterogeneity. *Social Science & Medicine*, 57(1):125–134.

Sickles, R. C. and Taubman, P. (1986). An analysis of the health and retirement status of the elderly. *Econometrica: Journal of the Econometric Society*, pages 1339–1356.

Simon, L. S. (2012). Relieving pain in America: A blueprint for transforming prevention, care, education, and research. *Journal of Pain & Palliative Care Pharmacotherapy*, 26(2):197–198.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic perspectives*, 13(2):145–166.

Smith, J. P. (2004). Unraveling the sex: Health connection. *Population and development review*, 30:108–132.

Sokoya, T., Zhou, Y., Diaz, S., Law, T., Himawan, L., Lekey, F., Shi, L., Gimbel, R. W., and Jing, X. (2022). Health indicators as measures of individual health status and their public perspectives: Cross-sectional survey study. *Journal of Medical Internet Research*, 24(6):e38099.

Stang, P. E. and Osterhaus, J. T. (1993). Impact of migraine in the United States: data from the National Health Interview Survey. *Headache: The Journal of Head and Face Pain*, 33(1):29–35.

State Council of the People's Republic of China (1998). Decision of the state council on establishing the basic medical insurance system for urban employees. https://www.gov.cn/banshi/2005-08/04/content_20256.htm. Accessed: 2023-11-16.

Su, D., Chen, Y.-c., Gao, H.-x., Li, H.-m., Chang, J.-j., Jiang, D., Hu, X.-m., Lei, S.-h., Tan, M., and Chen, Z.-f. (2019). Effect of integrated urban and rural residents medical insurance on the utilisation of medical services by residents in China: A propensity score matching with difference-in-differences regression approach. *BMJ Open*, 9(2):e026408.

Sun, J. and Lyu, S. (2020). Does health insurance lead to improvement of health status among Chinese Rural Adults? Evidence from the China Family Panel Studies. *International Journal of Health Services*, 50(3):350–359.

Sun, J. and Lyu, S. (2020). The effect of medical insurance on catastrophic health expenditure: Evidence from China. *Cost Effectiveness and Resource Allocation*, 18(1):1–11.

Tan, S. Y., Wu, X., and Yang, W. (2019). Impacts of the type of social health insurance on health service utilisation and expenditures: Implications for a unified system in China. *Health Economics, Policy and Law*, 14(4):468–486.

Tang, F. and Burr, J. A. (2015). Revisiting the pathways to retirement: A latent structure model of the dynamics of transition from work to retirement. *Ageing & Society*, 35(8):1739–1770.

Tang, S., Meng, Q., Chen, L., Bekedam, H., Evans, T., and Whitehead, M. (2008). Tackling the challenges to health equity in China. *The Lancet*, 372(9648):1493–1501.

The People's Government of Beijing Municipality (2023). How is the equivalent payment period for basic medical insurance calculated for insured persons (including those in flexible employment) who participate in the basic medical insurance for urban employees in this city? http://banshi.beijing.gov.cn/pubtask/task/1/110113000000/53193448-1617-4fb5-94f0-09d133fce9d0_cjwt.html. Accessed: 2023-11-16.

The Regents of the University of Michigan (2024). Health and Retirement Study. https://hrs.isr.umich.edu/about. Accessed: 2024-11-09.

The World Bank (2014-2018). Consumer price index (2010 = 100) - China. https://data.worldbank.org/indicator/FP.CPI.TOTL?end=2018&locations=CN&start=2014. Accessed: 2024-10-31.

Tubeuf, S., Jusot, F., Devaux, M., Sermet, C., et al. (2008). Social heterogeneity in self-reported health status and measurement of inequalities in health. *Paris: Irdes*, 24.

Väänänen, A., Toppinen-Tanner, S., Kalimo, R., Mutanen, P., Vahtera, J., and Peiró, J. M. (2003). Job characteristics, physical and psychological symptoms, and social support as antecedents of sickness absence among men and women in the private industrial sector. *Social Science & Medicine*, 57(5):807–824.

Van Den Hout, A., Jagger, C., and Matthews, F. E. (2009). Estimating life expectancy in health and ill health using a hidden Markov model. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(4):449–465.

Vistnes, J. P. (1997). Gender differences in days lost from work due to illness. *ILR Review*, 50(2):304–323.

Wagstaff, A. and Van Doorslaer, E. (2000). Equity in health care finance and delivery. *Handbook of Health Economics*, 1:1803–1862.

Wagstaff, A., Yip, W., Lindelow, M., and Hsiao, W. C. (2009). China's health system and its reform: A review of recent studies. *Health economics*, 18(S2):S7–S23.

Wang, H., Gusmano, M. K., and Cao, Q. (2011). An evaluation of the policy on community health organizations in China: Will the priority of new healthcare reform in China be a success? *Health policy*, 99(1):37–43.

Wang, H. and Heitjan, D. F. (2008). Modeling heaping in self-reported cigarette counts. *Statistics in Medicine*, 27(19):3789–3804.

Ware Jr, J. E., Brook, R. H., Davies, A. R., and Lohr, K. N. (1981). Choosing measures of health status for individuals in general populations. *American Journal of public health*, 71(6):620–625.

Xie, L., Shen, Y., Wu, Y., and Yang, H. (2021). The impact of retirement on mental health. *The International journal of health planning and management*, 36(5):1697–1713.

Xing-Yuan, G. and Sheng-Lan, T. (1995). Reform of the Chinese health care financing system. *Health Policy*, 32(1-3):181–191.

Xu, X. and Jensen, G. A. (2012). Does health insurance reduce illness-related worker absenteeism? *Applied Economics*, 44(35):4591–4603.

Yang, R. (2015). Cost sharing in China's higher education: Analyses of major stakeholders. *Higher education in the BRICS countries: Investigating the pact between higher education and society*, pages 237–251. Springer.

Yang, X., Chen, M., Du, J., and Wang, Z. (2018). The inequality of inpatient care net benefit under integration of urban-rural medical insurance systems in China. *International Journal for Equity in Health*, 17(1):1–12.

Ye, L., Shia, B.-C., Fang, Y., and Lee, T.-S. (2019). Heterogeneous health profiles and healthcare utilization of the middle-aged and elderly with multiple health insurance schemes in China. *Public Health*, 170:61–69.

Ying, M., Wang, S., Bai, C., and Li, Y. (2020). Rural-urban differences in health outcomes, healthcare use, and expenditures among older adults under universal health insurance in China. *PloS One*, 15(10):e0240194.

Yip, W. and Hsiao, W. C. (2015). What drove the cycles of chinese health system reforms? *Health Systems & Reform*, 1(1):52–61.

Yip, W. C.-M., Hsiao, W. C., Chen, W., Hu, S., Ma, J., and Maynard, A. (2012). Early appraisal of China's huge and complex health-care reforms. *The Lancet*, 379(9818):833–842.

Yu, H. (2015). Universal health insurance coverage for 1.3 billion people: What accounts for China's success? *Health Policy*, 119(9):1145–1152.

Yu, J. and Wu, J. (2018). The sustainability of agricultural development in China: The agriculture–environment nexus. *Sustainability*, 10(6):1776.

Zhang, A., Nikoloski, Z., and Mossialos, E. (2017a). Does health insurance reduce out-of-pocket expenditure? Heterogeneity among China's middle-aged and elderly. *Social Science & Medicine*, 190:11–19.

Zhang, A., Zhang, Y., and Tao, Y. (2022). Does retirement make people happier? Evidence from China. *Frontiers in Public Health*, 10:874500.

Zhang, C., Lei, X., Strauss, J., and Zhao, Y. (2017b). Health insurance and health care among the mid-aged and older Chinese: Evidence from the national baseline survey of CHARLS. *Health Economics*, 26(4):431–449.

Zhang, L., Chen, R., and Fang, Y. (2023). Effects of Urban and Rural Resident Basic Medical Insurance on Healthcare Utilization Inequality in China. *International Journal of Public Health*, 68:1605521.

Zhang, X. and Kanbur, R. (2009). Spatial inequality in education and health care in China. In *Regional Inequality in China*, pages 92–110. Routledge.

Zhang, Y., Salm, M., and van Soest, A. (2018). The effect of retirement on healthcare utilization: Evidence from China. *Journal of health economics*, 62:165–177.

Zhao, M., Liu, B., Shan, L., Li, C., Wu, Q., Hao, Y., Chen, Z., Lan, L., Kang, Z., Liang, L., et al. (2019). Can integration reduce inequity in healthcare utilization? Evidence and hurdles in China. *BMC Health Services Research*, 19(1):1–11.

Zhao, Y., Hu, Y., Smith, J. P., Strauss, J., and Yang, G. (2014). Cohort profile: the China health and retirement longitudinal study (CHARLS). *International Journal of Epidemiology*, 43(1):61–68.

Zhao, Y., Strauss, J., Yang, G., Giles, J., Hu, P., Hu, Y., Lei, X., Park, A., Smith, J. P., and Wang, Y. (2013). China health and retirement longitudinal study–2011–2012 national baseline users' guide. *Beijing: National School of Development, Peking University*, 2.

Zhou, Q., He, Q., Eggleston, K., and Liu, G. G. (2022). Urban-rural health insurance integration in China: impact on health care utilization, financial risk protection, and health status. *Applied Economics*, 54(22):2491–2509.

Ziebarth, N. (2010). Measurement of health, health inequality, and reporting heterogeneity. *Social Science & Medicine*, 71(1):116–124.

Appendix A

Supplementary Material for Chapter 2

In China, the reimbursement rates for specific health insurance plans vary depending on the level of the hospital. Accordingly, I will first present institutional details about Chinese hospitals in Section A.1. Subsequently, in Section A.2, I provide a comprehensive overview of the employee health insurance plan and the resident health insurance plan. This section will include tables of deductibles and reimbursement rates, exemplified using data from Beijing.

A.1 Hospitals in China

In China, hospitals can be public or private. Public hospitals, which are funded and operated by the government, dominate the health care sector in China, despite that the number of private hospitals is growing. The number of private hospitals has exceeded the number of public hospitals in 2015, with percentage of private hospitals in all hospitals increased from 17.22% in 2005 to 56.39% (Deng et al., 2018). Although the number of private hospital is larger, the share of hospital beds, total visits and bed occupancy rate (BOR)[1] are lower. The share of hospital beds in private hospitals was 21.69% among all hospitals and the total private hospital visits was only 12.90% of all hospitals visits. Moreover, in 2015, 83.2% of the workforce in health care sector work in public hospitals. Public hospitals also dominant the inpatient services, with 90% of the inpatient services are provided by the public hospitals (Yip and Hsiao, 2015; Mei and Kirkpatrick, 2019).

[1] $\text{BOR } (\%) = \frac{Culmulative Inpatient days \times 100}{Number of beds \times days}$

Hospitals in China are categorized into three levels: primary, secondary, and tertiary. Primary hospitals include community and township hospitals, which provide basic and preventative medical care as well as rehabilitation services. They serve as the basic healthcare institutions in rural areas. Secondary hospitals usually refer to county hospitals, offering comprehensive medical and health services to multiple communities, with bed capacities ranging from 100 to 499. Tertiary hospitals encompass city, provincial, and national hospitals, delivering high-level healthcare services and typically affiliated with universities or research institutions. These hospitals are considered the best in terms of medical technology and expertise in China and are required to have at least 500 beds.

There is no referral system in China, patients can freely choose from hospitals of different levels. However, the primary hospitals are week. Although there are reforms which aimed at increasing government funding for building community health center in cities, township hospitals and village clinics in rural areas (Bhattacharyya et al., 2011; Yip et al., 2012), the primary hospitals hospitals still lack well-trained doctors and they also face public trust issues (Liu et al., 2011; Mossialos et al., 2016; Wang et al., 2011). Moreover, the health care system is designed to encourage the usage of lower level hospitals. For example, the lower level of the hospital is, the higher the price and reimbursement rate are for the same type of service. Nevertheless, Zhang et al. (2017a) found the effect of this incentive is limited. Since the primary hospitals are too week, most people still get access to health care from the secondary and tertiary hospitals, even though they are more expensive.

A.2 Health insurance plans

In this section, I compare the benefits of different types of public health insurance. I compare employee health insurance and integrated resident health insurance in sections A.2.1 and A.2.2, respectively.

In China, public health insurance is pooled and administered by the local government. Employee health insurance and urban resident health insurance are pooled at the city level, while rural resident health insurance (before its integration with urban resident health insurance) is pooled at the county level.

A.2.1 Employee health insurance

According to the State Council of the People's Republic of China, all urban employers, including enterprises (state-owned enterprise, collective enterprises, foreign-invested enterprises, private enterprises, etc.), government and public institutions, social organizations, private non-enterprise units, and their employees, must participate in employee health insurance. In terms of the township enterprise and their employees, as well as owners and employees of urban private household organizations, whether they should participate in the employee health insurance is determined by the local government[2]. Moreover, self-employed individuals, part-time workers who do not participate in the employee health insurance, and other flexibly employed individuals can choose to participate in the employee health insurance, with individuals paying all the premiums. They can also choose to participate in the urban resident health insurance.

The employee health insurance is primarily pooled at the city level, but it can also be pooled at the county level. The three municipalities - Beijing, Tianjin, and Shanghai are pooled at the city level. All employees must participate the employee health insurance in their city or county, depending on the pooling level. Policies are unified within a pooling unit. Enterprises with cross-regional operations and high mobility in production, such as railways, electricity, and ocean transportation, and their employees, may participate in the basic medical insurance of the overall planning area in a relatively concentrated manner from different locations.

[2]It refers to the government of the provinces, autonomous regions, and municipalities.

The premiums of the employee health insurance are co-paid by employers and employees. The amount paid by the employer should be 6% of employers' wage and the amount paid by the employee should be 2% of their wage. However, with the economic development, these rates for both employers and employees can be adjusted accordingly (State Council of the People's Republic of China, 1998). Moreover, these rates also vary, depending on the financial situation of the local government. In 2012, the national average is 7.43%, with the lowest at 3%. Higher rates in cities like Shanghai and Beijing reaches 10% and 9%, respectively. The national average for individual premium contribution is 2% of their wage (Ministry of Human Resources and Social Security of the People's Republic of China, 2012).

The employee health insurance has two accounts. One is a personal saving account and one is the social risk pooling account. All individuals' premium contribution goes to their personal saving account. A proportion of employer's contribution also goes to employees' personal account, which could vary across province or city and may also vary depending on employees' age, but usually, it is 30%. The larger proportion goes to the social risk pooling account (State Council of the People's Republic of China, 1998).

The personal saving account and social risk pooling account have different payment scopes. Reimbursement is paid by the social risk pooling account. However, the reimbursement rate is not 100%, individuals still need to pay a proportion of expenditure. The reimbursement varies depending on individuals' employment status, types of hospitals, the type of services and is progressive by the amount of expenditure (see Tables A.1 and A.2 for an example of Beijing). In general, the reimbursement rates are higher for the outpatient services compared to the inpatient services, and those for the retired employee are higher for the working employees.

Moreover, employee health insurance also has deductibles and ceilings, which vary depending on an individual's employment status and the type of services. Thus, the

personal savings account is used to pay for expenditures beyond the reimbursement, which includes expenditures below the deductibles and the proportion of expenses borne by individuals between the deductibles and the ceilings. Expenses exceeding the ceiling are also be paid by individuals, either through out-of-pocket expenses or private health insurance.

Overall, the reimbursement rates, deductibles, and ceilings are set by local governments and may vary across provinces or cities. However, there are principles established by the central government that local authorities should adhere to. According to State Council of the People's Republic of China (1998), reimbursements are determined based on the principle of balancing the income and expenses of employee health insurance. Deductibles should be set at approximately 10% of the average annual salary of local workers, and the ceiling should be around four times the average annual salary of local workers.

Table A.1: Employee Health Insurance Payment and Reimbursement Benefits of Outpatient Services in Beijing

Personnel Category	Deductibles		Reimbursement ratio	
	Age group	Amount	Below 20,000	Above 20,000
Working employee		1,800	Secondary & Tertiary 70%	60%
			Primary 90%	
Retired employee	Below 70	1,300	Secondary & Tertiary 85%	80%
			Primary90%	
	Above 70	1,300	Secondary & Tertiary 90%	
			Primary 90%	

Notes: 1. All numbers are in RMB (yuan). 2. Primary to tertiary refers to the level of hospitals.
Source: Beijing Municipal Medical Insurance Bureau (2023)

A.2.2 Integrated urban and rural resident health insurance

Initially, the urban and rural resident health insurance has two main forms. Most areas implement the "one system, two standards" form, where two standards of payment and reimbursement are set, and individuals are free to choose from (Su et al.,

Table A.2: Employee Health Insurance Payment and Reimbursement Benefits of Inpatient Services in Beijing

Personnel Category	Deductibles			
	Range of cost	Primary	Secondary	Tertiary
Working employee	1300-30,000	90.0%	87.0%	85.0%
	30,000-40,000	95.0%	92.0%	90.0%
	40,000-100,000	97.0%	97.0%	95.0%
	100,000-500,000	85.0%		
Retired employee	1,300-30,000	97.0%	96.1%	95.5%
	30,000-40,000	98.5%	97.6%	97.0%
	40,000-100,000	99.1%	99.1%	98.5%
	100,000-500,000	90.0%		

Notes: 1. All numbers are in RMB (yuan). 2. The deductible is 1300 yuan for the first use of outpatient service in a year. That for the second or use is 650 yuan.
Source: Beijing Municipal Medical Insurance Bureau (2023)

2019). The second form, which is "one system, one standard", was implemented in economically developed areas, which the payment and reimbursement are unified. However, in 2016, the Chinese central government issued a guideline which required that the resident health insurance have to be unified in terms of coverage, premiums and reimbursement rate.

The insurance is voluntary, and the premiums are funded by both the central and local governments. However, individuals still need to pay a large proportion. Premiums also vary depending on age, with that for minors being the lowest, follows by the elderly people, then the working age population.

Taking Beijing as an example, the premiums paid by individual in 2023 for children/students, elderly people and working age population are 375, 400 and 695 yuan, respectively. Although individuals also needs to pay a fraction of premiums, the portion of personal payments required by individuals can also be subsidized by the government for those who are live on government's subsistence allowance, who are disabled, elderly people who are over sixty years of age and minors from low income families[3] (Ministry of Human Resources and Social Security of the People's Republic

[3]The cut-off for low household income is determined by local governments (city, municipality,

of China, 2012).

Similar to the employee health insurance, the reimbursement rate, premiums and ceilings are set by the local government, which could vary across provinces/cities. The deductibles and reimbursement rates vary depending on the level of hospital, type of services and age. In general, the better the hospital is, the higher the deductible is and lower the reimbursement rate is. Compared to the inpatient services, deductibles of the outpatient services are higher, but the reimbursement rates are also higher. Deductibles of the outpatient services are slightly lower for children/students (The People's Government of Beijing Municipality, 2023) (For more detail, see Tables A.4 and A.3).

Table A.3: Urban and Rural Residents' Medical Insurance Payment and Reimbursement Benefits of Outpatient Services in Beijing

Type	Premium	Deductible		Reimbursement Rate	
		Level 1	Level 2 & 3	Level 1	Level 2 & 3
Children/students	375	100	550	55%	50%
Elderly people	400	100	550	55%	50%
Working age population	695	100	550	55%	50%

Notes: 1. All numbers are in RMB (yuan). 2. Premiums are yearly. 3. The ceiling of reimbursement is 4,500 yuan. 4. The levels represent the ranking of hospitals; the higher the level, the better the hospital.
Source: The People's Government of Beijing Municipality (2023)

Compared to the urban employee health insurance, the deductibles for both the inpatient and outpatient services are lower for urban and rural resident health insurance (1800 yuan vs 695 yuan for the working age population in Beijing). However, the reimbursement rate and ceiling of the employee health insurance are also lower (90% vs 55% of the hospital for the working age population in Beijing; 500,000 yuan vs 250,000 yuan for the working age population in Beijing).

county), influenced by the local economic situation and the amount of money required to meet the basic living needs of urban residents. Two main indicators are considered when categorizing a household as low-income: family income and family assets.

Table A.4: Urban and Rural Residents' Medical Insurance Payment and Reimbursement Benefits of Inpatient Services in Beijing

Type	Premium	Deductible			Reimbursement Rate		
		Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Children /students	375	150	400	650	80%	78%	75%-78%
Elderly people	400	300	800	1300	80%	78%	75%-78%
Working age population	695	300	800	1300	80%	78%	75%-78%

Notes: 1. All numbers are in RMB (yuan). 2. Premiums are yearly. 3. The ceiling of reimbursement is 250,000 yuan. 4. The levels represent the ranking of hospitals; the higher the level, the better the hospital. 5. Deductibles in this table refer to those for the first use of inpatient services in a year. For all populations, the deductibles are halved for the second and subsequent uses of inpatient services within the same year.

Source: The People's Government of Beijing Municipality (2023)

Appendix B

Supplementary Material for Chapter 4

B.1 Additional tables

Table B.1: Age Effects in the Ordered Probit Model by Sex (Male)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
50-54	0.114***	0.165***	0.055	0.142***	0.132***	
	(0.033)	(0.044)	(0.046)	(0.033)	(0.032)	-
55-59	0.120***	0.172***	0.055	0.184***	0.170***	
	(0.032)	(0.041)	(0.044)	(0.032)	(0.032)	-
60-64	0.189***	0.242***	0.129***	0.287***	0.263***	
	(0.033)	(0.043)	(0.045)	(0.033)	(0.033)	-
65-69	0.191***	0.254***	0.128***	0.323***	0.286***	
	(0.037)	(0.048)	(0.049)	(0.036)	(0.037)	-
70-74	0.325***	0.394***	0.249***	0.473***	0.422***	
	(0.043)	(0.055)	(0.057)	(0.042)	(0.043)	-
75-79	0.291***	0.408***	0.185***	0.443***	0.371***	
	(0.049)	(0.066)	(0.066)	(0.049)	(0.049)	-
80-84	0.275***	0.402***	0.155	0.462***	0.352***	
	(0.071)	(0.098)	(0.095)	(0.067)	(0.069)	-
55-64	-	-	-	-	-	0.095***
	-	-	-	-	-	(0.024)
65-74	-	-	-	-	-	0.187***
	-	-	-	-	-	(0.029)
75-84	-	-	-	-	-	0.223***
	-	-	-	-	-	(0.032)
N	15,816	7,999	7,817	15,816	15,816	15,816

Notes: This table contains the age effects in Table 4.3. The baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. - means that the variable is not used in regression.

Table B.2: Age Effects in the Ordered Probit Model by Sex (Female)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
50-54	0.073**	0.078*	0.079*	0.138***	0.129***	
	(0.030)	(0.041)	(0.041)	(0.030)	(0.030)	-
55-59	0.103***	0.118***	0.093**	0.211***	0.199***	
	(0.031)	(0.039)	(0.041)	(0.030)	(0.030)	-
60-64	0.076**	0.101**	0.064	0.218***	0.199***	
	(0.032)	(0.042)	(0.042)	(0.032)	(0.032)	-
65-69	0.183***	0.246***	0.143***	0.354***	0.321***	
	(0.036)	(0.048)	(0.048)	(0.036)	(0.036)	-
70-74	0.113***	0.136**	0.106*	0.281***	0.237***	
	(0.043)	(0.058)	(0.056)	(0.043)	(0.043)	-
75-79	0.072	0.067	0.088	0.271***	0.198***	
	(0.053)	(0.071)	(0.070)	(0.052)	(0.052)	-
80-84	0.157**	0.245***	0.083	0.370***	0.272***	
	(0.069)	(0.088)	(0.096)	(0.068)	(0.069)	-
55-64	-	-	-	-	-	0.057**
	-	-	-	-	-	(0.023)
65-74	-	-	-	-	-	0.121***
	-	-	-	-	-	(0.030)
75-84	-	-	-	-	-	0.069
	-	-	-	-	-	(0.045)
N	16,764	8,337	8,427	16,764	16,764	16,764

Notes: This table contains the age effects in Table 4.4. The baseline estimation results refer to a married female who lives with her spouse, with a resident insurance, between 45 and 49 years of age (except when we use 10-year windows, in which case it is 'between 45 and 54 years of age') and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. - means that the variable is not used in regression.

Table B.3: Age Effects in the Ordered Probit Model (Sensitivity Check: including Expenditure, Male)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	2011	2013
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Chronic disease	Y	Y	Y	Y	Y	Y
Disability	Y	Y	Y	Y	Y	Y
Expenditure	Y	Y	Y	N	N	N
Age window	5y	5y	5y	5y	5y	5y
50-54	0.092***	0.155***	0.006	0.100***	0.161***	0.018
	(0.035)	(0.045)	(0.051)	(0.035)	(0.045)	(0.051)
55-59	0.100***	0.164***	0.009	0.117***	0.176***	0.031
	(0.034)	(0.043)	(0.050)	(0.034)	(0.042)	(0.049)
60-64	0.146***	0.211***	0.062	0.165***	0.224***	0.089*
	(0.036)	(0.045)	(0.051)	(0.035)	(0.045)	(0.050)
65-69	0.179***	0.240***	0.106*	0.203***	0.257***	0.137**
	(0.039)	(0.050)	(0.055)	(0.039)	(0.049)	(0.055)
70-74	0.311***	0.373***	0.222***	0.341***	0.392***	0.264***
	(0.046)	(0.058)	(0.066)	(0.046)	(0.057)	(0.065)
75-79	0.242***	0.379***	0.090	0.275***	0.399***	0.137*
	(0.054)	(0.070)	(0.077)	(0.053)	(0.069)	(0.076)
80-84	0.242***	0.379***	0.090	0.275***	0.399***	0.137*
	(0.054)	(0.070)	(0.077)	(0.053)	(0.069)	(0.076)
N	13,720	7,490	6,230	13,720	7,490	6,230

Notes: This table contains the age effects in Table 4.9. The baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age, living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.4: Age Effects in The Ordered Probit Model (Sensitivity Check: including Expenditure, Female)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	2011	2013
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Chronic disease	Y	Y	Y	Y	Y	Y
Disability	Y	Y	Y	Y	Y	Y
Expenditure	Y	Y	Y	N	N	N
Age window	5y	5y	5y	5y	5y	5y
50-54	0.065**	0.088**	0.052	0.071**	0.087**	0.067
	(0.032)	(0.042)	(0.046)	(0.032)	(0.042)	(0.046)
55-59	0.080**	0.115***	0.042	0.092***	0.114***	0.069
	(0.033)	(0.041)	(0.046)	(0.033)	(0.040)	(0.046)
60-64	0.054	0.106**	0.010	0.067**	0.104**	0.040
	(0.034)	(0.043)	(0.047)	(0.034)	(0.043)	(0.047)
65-69	0.159***	0.245***	0.083	0.176***	0.243***	0.123**
	(0.039)	(0.050)	(0.055)	(0.039)	(0.050)	(0.055)
70-74	0.075	0.139**	0.016	0.094**	0.137**	0.060
	(0.047)	(0.060)	(0.066)	(0.047)	(0.060)	(0.065)
75-79	0.091	0.096	0.101	0.110*	0.094	0.143*
	(0.057)	(0.074)	(0.083)	(0.057)	(0.073)	(0.082)
80-84	0.162**	0.283***	−0.012	0.182**	0.281***	0.039
	(0.079)	(0.093)	(0.125)	(0.079)	(0.092)	(0.125)
N	14,304	7,749	6,555	14,304	7,749	6,555

Notes: This table contains the age effects in Table 4.10. The baseline estimation results refer to a married female with a resident insurance, between 45 and 49 years of age, living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.5: Determinants of Health State: the Ordered Probit Model by Sex (Sensitivity Check: Using Provincial Dummies, Male)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.159***	−0.179***	−0.143***	−0.162***	−0.142***	−0.154***
	(0.038)	(0.052)	(0.050)	(0.039)	(0.039)	(0.038)
No health insurance	0.005	0.008	−0.050	−0.048	−0.053	0.001
	(0.046)	(0.055)	(0.078)	(0.046)	(0.046)	(0.046)
Chronic disease	0.727***	0.774***	0.685***	-	-	0.729***
	(0.022)	(0.029)	(0.029)	-	-	(0.022)
Disability	0.385***	0.403***	0.357***	-	0.430***	0.389***
	(0.028)	(0.035)	(0.036)	-	(0.028)	(0.028)
Regular drinker	−0.263***	−0.252***	−0.281***	−0.297***	−0.292***	−0.267***
	(0.020)	(0.027)	(0.027)	(0.020)	(0.020)	(0.020)
Past smoker	0.067***	0.057*	0.097***	0.092***	0.083***	0.069***
	(0.024)	(0.030)	(0.033)	(0.024)	(0.024)	(0.024)
Unmarried	0.076**	0.077	0.075	0.084**	0.056	0.080**
	(0.038)	(0.048)	(0.050)	(0.039)	(0.038)	(0.038)
Married, spouse away	−0.017	−0.004	−0.050	−0.027	−0.029	−0.021
	(0.039)	(0.050)	(0.060)	(0.039)	(0.039)	(0.039)
High school and above	−0.164***	−0.159***	−0.163***	−0.181***	−0.169***	−0.162***
	(0.031)	(0.038)	(0.039)	(0.031)	(0.031)	(0.030)
Illiterate	0.083**	0.147***	0.006	0.073**	0.051	0.096***
	(0.035)	(0.043)	(0.046)	(0.034)	(0.035)	(0.035)
Urban *hukou*	−0.055	−0.070	−0.041	−0.037	−0.031	−0.056
	(0.034)	(0.045)	(0.045)	(0.035)	(0.035)	(0.034)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
Province dummies	Y	Y	Y	Y	Y	Y
Threshold μ_1	−0.221***	−0.088	−0.359***	−0.666***	−0.617***	−0.267***
	(0.053)	(0.068)	(0.071)	(0.053)	(0.052)	(0.051)
Threshold μ_2	1.325***	1.427***	1.228***	0.782***	0.850***	1.277***
	(0.054)	(0.070)	(0.072)	(0.053)	(0.053)	(0.052)
N	15,816	7,999	7,817	15,816	15,816	15,816

Notes: All models include age effects by 5- or 10-year windows. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state.The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with his spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table B.6: Determinants of Health State: the Ordered Probit Model by Sex (Sensitivity Check: Using Provincial Dummies, Female)

Dependent variable: Self-ranked health state

Data	Pooled	2011	2013	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee health insurance	−0.102**	−0.207***	−0.013	−0.119***	−0.108***	−0.106***
	(0.041)	(0.056)	(0.053)	(0.041)	(0.041)	(0.041)
No health insurance	−0.022	−0.063	−0.004	−0.100**	−0.101**	−0.024
	(0.041)	(0.053)	(0.065)	(0.041)	(0.041)	(0.041)
Chronic disease	0.733***	0.755***	0.718***	-	-	0.736***
	(0.023)	(0.028)	(0.030)	-	-	(0.023)
Disability	0.328***	0.346***	0.303***	-	0.382***	0.329***
	(0.030)	(0.037)	(0.038)	-	(0.030)	(0.030)
Regular drinker	−0.243***	−0.229***	−0.253***	−0.269***	−0.267***	−0.242***
	(0.035)	(0.049)	(0.046)	(0.035)	(0.035)	(0.035)
Past smoker	0.060	0.061	0.065	0.101**	0.081**	0.058
	(0.040)	(0.051)	(0.048)	(0.039)	(0.039)	(0.040)
Unmarried	−0.029	−0.056	−0.005	−0.002	−0.017	−0.031
	(0.030)	(0.038)	(0.039)	(0.030)	(0.030)	(0.030)
Married, spouse away	−0.053	−0.073	−0.039	−0.050	−0.050	−0.055
	(0.038)	(0.051)	(0.054)	(0.038)	(0.038)	(0.038)
High school and above	−0.133***	−0.123**	−0.139***	−0.166***	−0.157***	−0.126***
	(0.037)	(0.048)	(0.047)	(0.038)	(0.038)	(0.037)
Illiterate	0.049**	0.052*	0.038	0.044*	0.033	0.051**
	(0.024)	(0.030)	(0.030)	(0.024)	(0.024)	(0.024)
Urban *hukou*	−0.199***	−0.142***	−0.251***	−0.185***	−0.175***	−0.199***
	(0.034)	(0.044)	(0.043)	(0.034)	(0.034)	(0.034)
Chronic disease	Y	Y	Y	N	N	Y
Disability	Y	Y	Y	N	Y	Y
Age window	5y	5y	5y	5y	5y	10y
Province dummies	Y	Y	Y	Y	Y	Y
Threshold μ_1	−0.505***	−0.450***	−0.564***	−0.929***	−0.888***	−0.533***
	(0.048)	(0.061)	(0.062)	(0.046)	(0.046)	(0.046)
Threshold μ_2	0.988***	1.014***	0.964***	0.478***	0.530***	0.959***
	(0.048)	(0.062)	(0.062)	(0.046)	(0.046)	(0.047)
N	16,764	8,337	8,427	16,764	16,764	16,764

Notes: All models include age effects by 5- or 10-year windows. The health states are ordered as follows: "good = 1", "fair = 2", and "poor = 3". Thresholds μ_1 and μ_2 correspond to the categories of health state.The value of the latent health state variable for each observation determines the category of health state which it falls into. If the value of the latent health state is below μ_1, the observation is categorized as "good", if the value of the latent health state is between μ_1 and μ_2, it is categorized as "fair", and if the value of the latent health state is above μ_2, it is categorized as "poor". The baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age (except when I use 10-year windows, in which case it is 'between 45 and 54 years of age') living with her spouse, and who is not a regular drinker, does not smoke, does not have any chronic disease, is not disabled, literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. "-" means that the variable is not included in the regression model.

Table B.7: Ordered Probit Model: the Wave Effect

Dependent variable: Self-ranked health status

Data	Pooled	Pooled	Pooled	Pooled
Variable	(1)	(2)	(3)	(4)
Sex	Male	Female	Male	Female
Employee insurance	−0.182***	−0.211***	−0.207***	−0.160***
	(0.051)	(0.055)	(0.044)	(0.048)
No insurance	0.016	−0.068	0.015	−0.057
	(0.055)	(0.053)	(0.054)	(0.052)
2013 wave	0.143**	−0.007	0.103**	−0.014
	(0.063)	(0.048)	(0.047)	(0.041)
Employee ×2013 wave	0.037	0.187***	0.087**	0.089*
	(0.064)	(0.068)	(0.042)	(0.047)
No insurance ×2013 wave	−0.043	0.057	−0.044	0.039
	(0.091)	(0.081)	(0.089)	(0.080)
50-54	0.167***	0.079*	0.168***	0.081**
	(0.045)	(0.041)	(0.045)	(0.041)
55-59	0.174***	0.119***	0.178***	0.120***
	(0.041)	(0.039)	(0.041)	(0.039)
60-64	0.245***	0.103**	0.253***	0.103**
	(0.043)	(0.042)	(0.043)	(0.041)
65-69	0.257***	0.249***	0.265***	0.249***
	(0.049)	(0.048)	(0.048)	(0.047)
70-74	0.399***	0.137**	0.418***	0.135**
	(0.056)	(0.058)	(0.055)	(0.057)
75-79	0.413***	0.067	0.436***	0.064
	(0.066)	(0.071)	(0.066)	(0.070)
80-84	0.407***	0.247***	0.441***	0.243***
	(0.099)	(0.089)	(0.098)	(0.087)
50-54 × 2013 wave	−0.112*	0.000	−0.114*	−0.006
	(0.061)	(0.054)	(0.061)	(0.054)
55-59 × 2013 wave	−0.120**	−0.028	−0.125**	−0.032
	(0.055)	(0.051)	(0.055)	(0.050)
60-64 × 2013 wave	−0.118**	−0.040	−0.133**	−0.041
	(0.057)	(0.054)	(0.056)	(0.052)
65-69 × 2013 wave	−0.130**	−0.108*	−0.147**	−0.110*
	(0.063)	(0.063)	(0.062)	(0.061)
70-74 × 2013 wave	−0.153**	−0.033	−0.186***	−0.030
	(0.072)	(0.074)	(0.071)	(0.070)
75-79 × 2013 wave	−0.230***	0.020	−0.274***	0.026
	(0.088)	(0.092)	(0.086)	(0.086)
80-84 × 2013 wave	−0.254**	−0.165	−0.317**	−0.158
	(0.129)	(0.121)	(0.125)	(0.115)
Chronic disease	0.782***	0.770***	0.783***	0.771***
	(0.028)	(0.028)	(0.028)	(0.028)
Chronic disease × 2013 wave	−0.104***	−0.071**	−0.108***	−0.074**
	(0.036)	(0.035)	(0.035)	(0.034)

Continued on next page

Table B.7 (*Continued*)

	(1)	(2)	(3)	(4)
Disability	0.412***	0.355***	0.381***	0.329***
	(0.036)	(0.037)	(0.028)	(0.029)
Disability× 2013 wave	−0.063	−0.052	-	-
	(0.043)	(0.045)	-	-
Regular drinker	−0.259***	−0.234***	−0.267***	−0.246***
	(0.026)	(0.049)	(0.020)	(0.035)
Regular drinker × 2013 wave	−0.016	−0.020	-	-
	(0.034)	(0.062)	-	-
Past smoker	0.059**	0.044	0.076***	0.038
	(0.030)	(0.050)	(0.024)	(0.039)
Past smoker × 2013 wave	0.038	−0.010	-	-
	(0.039)	(0.057)	-	-
Unmarried	0.082*	−0.051	0.075**	−0.023
	(0.049)	(0.038)	(0.038)	(0.030)
Married, spouse away	−0.004	−0.072	−0.018	−0.051
	(0.051)	(0.051)	(0.040)	(0.038)
Unmarried × 2013 wave	−0.013	0.057	-	-
	(0.061)	(0.047)	-	-
Married, spouse away × 2013 wave	−0.038	0.045	-	-
	(0.076)	(0.071)	-	-
High school and above	−0.157***	−0.123**	−0.160***	−0.133***
	(0.038)	(0.048)	(0.031)	(0.037)
Illiterate	0.120***	0.049*	0.057*	0.046**
	(0.043)	(0.029)	(0.035)	(0.023)
High school and above × 2013 wave	−0.006	−0.018	-	-
	(0.045)	(0.059)	-	-
Illiterate × 2013 wave	−0.130**	−0.006	-	-
	(0.054)	(0.035)	-	-
Urban *hukou*	−0.068	−0.048	−0.138***	−0.195***
	(0.044)	(0.034)	(0.043)	(0.033)
Urban *hukou* × 2013 wave	0.040	−0.111**	-	-
	(0.056)	(0.053)	-	-
Middle	0.214***	0.177***	0.170***	0.164***
	(0.034)	(0.033)	(0.027)	(0.026)
Northeast	0.220***	0.145***	0.175***	0.147***
	(0.054)	(0.053)	(0.042)	(0.042)
West	0.221***	0.218***	0.198***	0.232***
	(0.033)	(0.032)	(0.026)	(0.025)
Middle × 2013 wave	−0.089**	−0.025	-	-
	(0.040)	(0.039)	-	-
Northeast × 2013 wave	−0.091	0.003	-	-
	(0.065)	(0.060)	-	-
West×2013 wave	−0.046	0.031	-	-
	(0.040)	(0.038)	-	-
Threshold μ_1	0.100**	−0.172***	0.079*	−0.176***

Continued on next page

Table B.7 (*Continued*)

	(1)	(2)	(3)	(4)
	(0.048)	(0.038)	(0.043)	(0.036)
Threshold μ_2	1.641***	1.316***	1.620***	1.311***
	(0.049)	(0.040)	(0.044)	(0.037)
N	15,816	16,764	15,816	16,764

Notes: All models include age effects by 5-year windows. The baseline estimation results refer to a married individual with a resident insurance, between 45 and 49 years of age, living with his/her spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%. - means that the variable is not included in the regression model.

Table B.8: Age Effects in the Ordered Probit Model of Health Transitions

Dependent variable: Current wave (t) self-ranked health state

Data	Pooled	2011-13	2013-15	Pooled, Male	Pooled, Female
Variable	(1)	(2)	(3)	(4)	(5)
50 - 54	0.034	0.073*	0.012	0.012	0.052
	(0.031)	(0.045)	(0.042)	(0.045)	(0.041)
50 - 54	−0.158***	−0.110	−0.209**	−0.122	−0.190**
×Good at $t-1$	(0.058)	(0.081)	(0.085)	(0.083)	(0.083)
50 - 54	0.004	−0.008	−0.021	0.015	−0.003
×Poor at $t-1$	(0.066)	(0.089)	(0.097)	(0.106)	(0.083)
55 - 59	0.017	−0.029	0.071*	0.022	0.008
	(0.029)	(0.041)	(0.042)	(0.043)	(0.039)
55 - 59	0.014	0.088	−0.064	0.017	0.018
×Good at $t-1$	(0.056)	(0.074)	(0.083)	(0.080)	(0.079)
55 - 59	−0.099	−0.051	−0.163*	−0.125	−0.074
×Poor at $t-1$	(0.062)	(0.079)	(0.093)	(0.102)	(0.078)
60 - 64	0.042	0.044	0.054	0.045	0.035
	(0.030)	(0.043)	(0.042)	(0.043)	(0.041)
60 - 64	0.040	0.034	0.054	0.046	0.044
×Good at $t-1$	(0.060)	(0.081)	(0.086)	(0.085)	(0.084)
60 - 64	−0.045	−0.047	−0.075	−0.066	−0.017
×Poor at $t-1$	(0.063)	(0.081)	(0.094)	(0.102)	(0.081)
65 - 69	0.075**	0.038	0.121**	0.049	0.101**
	(0.035)	(0.050)	(0.048)	(0.049)	(0.048)
65 - 69	0.028	0.060	0.019	0.122	−0.075
×Good at $t-1$	(0.068)	(0.098)	(0.094)	(0.094)	(0.101)
65 - 69	−0.097	−0.089	−0.132	0.074	−0.218**
×Poor at $t-1$	(0.070)	(0.093)	(0.100)	(0.110)	(0.089)
70 - 74	0.117***	0.085	0.153***	0.098*	0.140**
	(0.041)	(0.060)	(0.056)	(0.057)	(0.058)
70 - 74	0.056	−0.006	0.129	0.172	−0.065
×Good at $t-1$	(0.082)	(0.111)	(0.114)	(0.115)	(0.115)
70 - 74	−0.065	0.005	−0.157	0.154	−0.250**
×Poor at $t-1$	(0.078)	(0.105)	(0.112)	(0.118)	(0.105)
75 - 79	0.067	0.016	0.120	0.083	0.043
	(0.054)	(0.075)	(0.075)	(0.071)	(0.078)
75 - 79	0.225**	0.345**	0.159	0.267*	0.193
×Good at $t-1$	(0.107)	(0.151)	(0.143)	(0.140)	(0.152)
75 - 79	−0.189*	−0.141	−0.261*	−0.117	−0.223*
×Poor at $t-1$	(0.097)	(0.127)	(0.139)	(0.146)	(0.131)
80 - 84	0.154*	0.062	0.266**	0.210*	0.093
	(0.084)	(0.122)	(0.114)	(0.121)	(0.119)
80 - 84	0.065	0.153	−0.016	0.068	0.082
×Good at $t-1$	(0.149)	(0.220)	(0.207)	(0.205)	(0.223)
80 - 84	−0.342**	−0.227	−0.504**	−0.344	−0.314
×Poor at $t-1$	(0.145)	(0.201)	(0.201)	(0.219)	(0.193)
N	27,229	13,467	13,762	13,066	14,163

Notes: This table contains the age effects in Table 4.15. In columns 1-3, the baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age, living with his spouse, and who is not a regular drinker, does not have any chronic disease, is not disabled, is literate but has a less than high school diploma, does not have an urban *hukou*, and lives in the Eastern region. In columns 4 and 5, characteristics of the reference group are the same as those in columns 1-3, except sex. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.9: Age Effects in the Ordered Probit Model of Health Transitions (the Parsimonious Model)

Dependent variable: Current wave (t) self-ranked health state

Data	Pooled	2011-13	2013-15	Pooled, Male	Pooled, Female
Variable	(1)	(2)	(3)	(4)	(5)
50 - 54	0.042	0.075*	0.025	0.011	0.068*
	(0.030)	(0.044)	(0.042)	(0.045)	(0.040)
50 - 54	−0.145**	−0.092	−0.198**	−0.111	−0.171**
×Good at $t-1$	(0.058)	(0.080)	(0.085)	(0.082)	(0.083)
50 - 54	0.008	0.000	−0.021	0.024	−0.001
×Poor at $t-1$	(0.066)	(0.090)	(0.096)	(0.107)	(0.083)
55 - 59	0.054*	0.012	0.101**	0.051	0.053
	(0.029)	(0.040)	(0.041)	(0.043)	(0.038)
55 - 59	0.022	0.090	−0.052	0.023	0.032
×Good at $t-1$	(0.056)	(0.074)	(0.083)	(0.080)	(0.079)
55 - 59	−0.114*	−0.065	−0.178*	−0.136	−0.092
×Poor at $t-1$	(0.063)	(0.079)	(0.093)	(0.102)	(0.078)
60 - 64	0.105***	0.116***	0.105**	0.100**	0.107***
	(0.030)	(0.043)	(0.041)	(0.043)	(0.039)
60 - 64	0.041	0.028	0.061	0.039	0.049
×Good at $t-1$	(0.060)	(0.081)	(0.086)	(0.085)	(0.084)
60 - 64	−0.068	−0.072	−0.092	−0.079	−0.046
×Poor at $t-1$	(0.063)	(0.082)	(0.093)	(0.102)	(0.081)
65 - 69	0.144***	0.107**	0.186***	0.114**	0.173***
	(0.034)	(0.049)	(0.047)	(0.048)	(0.046)
65 - 69	0.051	0.075	0.046	0.141	−0.050
×Good at $t-1$	(0.068)	(0.097)	(0.094)	(0.093)	(0.100)
65 - 69	−0.116*	−0.103	−0.155	0.061	−0.240***
×Poor at $t-1$	(0.070)	(0.093)	(0.100)	(0.111)	(0.089)
70 - 74	0.194***	0.173***	0.219***	0.169***	0.217***
	(0.040)	(0.059)	(0.055)	(0.056)	(0.056)
70 - 74	0.074	0.011	0.146	0.202*	−0.060
×Good at $t-1$	(0.082)	(0.111)	(0.114)	(0.115)	(0.116)
70 - 74	−0.097	−0.029	−0.195*	0.124	−0.286***
×Poor at $t-1$	(0.078)	(0.105)	(0.111)	(0.118)	(0.105)
75 - 79	0.138***	0.105	0.177**	0.137**	0.133*
	(0.052)	(0.072)	(0.073)	(0.070)	(0.074)
75 - 79	0.250**	0.356**	0.183	0.281**	0.225
×Good at $t-1$	(0.107)	(0.150)	(0.143)	(0.141)	(0.152)
75 - 79	−0.206**	−0.160	−0.276**	−0.120	−0.258**
×Poor at $t-1$	(0.097)	(0.126)	(0.140)	(0.146)	(0.131)
80 - 84	0.225***	0.140	0.325***	0.274**	0.170
	(0.081)	(0.116)	(0.111)	(0.117)	(0.114)
80 - 84	0.133	0.203	0.066	0.150	0.133
×Good at $t-1$	(0.148)	(0.217)	(0.206)	(0.204)	(0.221)
80 - 84	−0.359**	−0.229	−0.526***	−0.365*	−0.331*
×Poor at $t-1$	(0.142)	(0.197)	(0.200)	(0.215)	(0.190)

Continued on next page

Table B.9 (*Continued*)

	(1)	(2)	(3)	(4)	(5)
N	27,229	13,467	13,762	13,066	14,163

Notes: This table contains the age effects in Table 4.16. In columns 1-3, the baseline estimation results refer to a married male with a resident insurance, between 45 and 49 years of age, and lives in the Eastern region. In columns 4 and 5, the characteristics of the reference group are the same as those in columns 1-3, except for sex. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.10: Estimated Health Transition Matrices by Insurance Plan: the Parsimonious Model by Wave

Employee health insurance				Resident health insurance			
a) 2011–13 waves							
	Good	Fair	Poor		Good	Fair	Poor
Good	0.618	0.355	0.027	Good	0.528	0.428	0.044
Fair	0.324	0.556	0.120	Fair	0.274	0.575	0.151
Poor	0.093	0.529	0.378	Poor	0.072	0.497	0.431
b) 2013–15 waves							
	Good	Fair	Poor		Good	Fair	Poor
Good	0.693	0.293	0.013	Good	0.558	0.411	0.032
Fair	0.315	0.576	0.109	Fair	0.253	0.600	0.148
Poor	0.055	0.491	0.454	Poor	0.043	0.456	0.501

Notes: 'Good', 'Fair', and 'Poor' refer to self-ranked health states. All the results use as the reference group those 45 to 49 years old based on the estimation results from the ordered probit model reported in Table 4.16, column 2 (panel a) and column 3 (panel b).

Appendix C

Supplementary Material for Chapter 5

Table C.1: Summary Statistics for Work-to-Work and Work-to-Retirement Transitions

Variables	2011-13 sample N = 9,280				2013-15 sample N = 9,927			
	Working, 2013 N = 7,829		Retired, 2013 N = 1,451		Working, 2015 N = 8,214		Retired, 2015 N = 1,713	
	N	%	N	%	N	%	N	%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sex								
Female	3,625	46.3	826	56.9	3,789	46.1	993	58.0
Male	4,206	53.7	625	43.1	4,425	53.9	720	42.0
Pension plan								
Employee pension	546	7.0	148	10.2	580	7.1	227	13.3
Resident/no pension	7,285	93.0	1,303	89.8	7,634	92.9	1,486	86.8
Sector of employment		(2011)				(2013)		
Agricultural	5,289	67.5	1,010	69.6	5,360	65.3	1,072	62.6
Non-agricultural	1,269	16.2	343	23.6	1,471	17.9	493	28.8
Both sectors	1,273	16.3	98	6.8	1,383	16.8	148	8.6
		(2013)				(2015)		
Agricultural	5,267	67.3	-	-	5,038	61.3	-	-
Non-agricultural	1,366	17.4	-	-	1,717	20.9	-	-
Both sectors	1,198	15.3	-	-	1,459	17.8	-	-
Health insurance plan								
Employee	376	4.8	117	8.1	496	6.0	200	11.7
Resident	7,093	90.6	1,233	85.0	7,454	90.8	1,458	85.1
No health insurance	362	4.6	101	7.0	264	3.2	55	3.2
Health state								
Good	2,016	25.7	289	19.9	2,091	25.5	385	22.5
Fair	4,049	51.7	723	49.8	4,516	55.0	907	53.0
Poor	1,766	22.6	439	30.3	1,607	19.6	421	24.6
Educational attainment								
Illiterate	1,972	25.2	490	33.8	1,834	22.3	484	28.3
Less than high school	5,075	64.8	831	57.3	5,550	67.6	1,041	60.8
High school and above	784	10.0	130	9.0	830	10.1	188	11.0
Marital status								
Married, spouse present	6,760	86.3	1,189	81.9	7,199	87.6	1,373	80.2
Married, spouse away	592	7.6	84	5.8	494	6.0	104	6.1
Unmarried	479	6.1	178	12.3	521	6.3	236	13.8
Age group		(2013)				(2015)		
45 - 49	1,038	13.3	99	6.8	813	9.9	93	5.4
50 - 54	1,666	21.3	195	13.4	2,088	25.4	279	16.3
55 - 59	1,846	23.6	274	18.9	1,545	18.8	269	15.7
60 - 64	1,657	21.2	310	21.4	1,822	22.2	363	21.2
65 - 69	952	12.2	241	16.6	1,139	13.9	290	16.9
70 - 74	446	5.7	178	12.3	553	6.7	232	13.5
75 - 79	180	2.3	105	7.2	195	2.4	122	7.1

Continued on next page

Table C.1 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
80 - 84	46	0.6	49	3.4	59	0.7	65	3.8
Region								
East	2,379	30.4	470	32.4	2,515	30.6	509	29.7
Middle	2,328	29.7	437	30.1	2,351	28.6	532	31.1
Northeast	446	5.7	88	6.1	475	5.8	134	7.8
West	2,678	34.2	456	31.4	2,873	35.0	538	31.4
Total number of children								
1	1,127	14.4	181	12.5	1,174	14.3	237	13.8
2	3,202	40.9	468	32.3	3,505	42.7	607	35.4
3	2,073	26.5	354	24.4	2,095	25.5	421	24.6
4	897	11.5	252	17.4	905	11.0	229	13.4
5 or more	530	6.8	196	13.5	535	6.5	219	12.8
Having children in school								
Yes	834	10.7	89	6.1	880	10.7	101	5.9
No	6,997	89.4	1,362	93.9	7,334	89.3	1,612	94.1
Educational attainment children (highest)								
Less than high school	4,662	59.5	865	59.6	4,528	55.1	929	54.2
High school	808	10.3	154	10.6	850	10.4	182	10.6
Above high school	2,361	30.2	432	29.8	2,836	34.5	602	35.1

Notes: Unless otherwise stated, all 2011-13 sample variables correspond to 2011, and 2013-15 sample variables correspond to 2013. "-" means not applicable. The results of the Pearson chi-squared test show that for the working group, the distribution of sector of work, health insurance plan, health state, education, marital status, age group, and children's highest educational attainment are statistically different between the two samples. For the retired group, the distribution of employee pension and the number of children are also statistically different between the two samples, but marital status is not.

Source: Calculations based on pooled CHARLS 2011, 2013 and 2015 survey waves.

Table C.2: Summary Statistics of Per Capita Household Annual Expenditure (Yuan) by Health Insurance Plan (Pooled data)

Health insurance plan	N	Mean	Std.dev.	Min	Max
Employee	1,048	14,045	17,347.73	740	249,985.50
Resident	14,767	7,946	10,209.57	0	307,935.00
No health insurance	688	8,116	10,411.04	284	128,347.50

Notes: Results of a t-test show that the mean of per capita household annual expenditure is statistically different between individuals with employee health insurance and those with resident health insurance (p-vaule is 0.00). Results of a t-test show that the means are not statistically different between the resident health insurance group and the no health insurance group (p-value is 0.67).

Table C.3: Number of Respondents in Regression

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee	952	252	700	237	66	171
No health insurance	44	18	26	737	368	369
Good health	518	252	379	4,263	1,855	369
Poor health	163	18	124	4,068	2,377	1,691
N	1,501	411	1,089	17,706	8,820	8,886

Notes: This table shows the number of respondents in each cell for health insurance and health states in Table 5.2.

Table C.4: Estimation Results for the Retirement Status: Probit Model and Pooled Data (with Expenditure)

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee health insurance	0.277***	0.581***	0.172	0.428***	0.521***	0.395***
	(0.096)	(0.200)	(0.113)	(0.102)	(0.174)	(0.124)
No health insurance	−0.037	−0.263	0.151	0.192***	0.195**	0.193**
	(0.253)	(0.510)	(0.299)	(0.063)	(0.081)	(0.090)
Good health	−0.048	−0.311*	0.047	−0.065**	−0.038	−0.092*
	(0.092)	(0.174)	(0.109)	(0.033)	(0.045)	(0.047)
Poor health	0.355***	0.483*	0.332**	0.131***	0.125***	0.139***
	(0.129)	(0.260)	(0.151)	(0.031)	(0.040)	(0.049)
Logarithm of per capita	0.065	0.103	0.066	0.057***	0.060***	0.047*
household annual expenditure	(0.059)	(0.106)	(0.066)	(0.017)	(0.021)	(0.024)
Female	0.533***	-	-	0.376***	-	-
	(0.097)			(0.027)		
Middle	0.139	0.124	0.187	−0.025	−0.071	0.030
	(0.108)	(0.197)	(0.126)	(0.035)	(0.045)	(0.051)
Northeast	0.262**	0.123	0.334**	−0.096**	−0.146***	−0.028
	(0.122)	(0.216)	(0.142)	(0.038)	(0.048)	(0.055)
West	0.362***	0.095	0.461***	−0.116***	−0.194***	−0.014
	(0.122)	(0.246)	(0.138)	(0.041)	(0.052)	(0.058)
Constant	−2.855***	−2.484**	−2.992***	−2.024***	−1.574***	−2.141***
	(0.544)	(0.974)	(0.644)	(0.161)	(0.201)	(0.238)
N	1,308	348	960	15,188	7,531	7,657

Notes: Retirement status is as observed at t and independent variables are as observed at $t-1$, except for sex and age. All models include age effects by a 5-year window. This table only shows the coefficients of key explanatory variables. The baseline estimation results in columns 1, 4 and 5 refer to a married male with a resident insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but with less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school. whose children's highest education level is below high school. The baseline estimations in columns 2 and 3 are the same, except for sex. Standard errors are clustered at the household level. "-" means not applicable. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.5: Estimation Results for The Retirement Status: Probit Model and Pooled Data (Original Model, Less Observations)

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee health insurance	0.292***	0.609***	0.187*	0.445***	0.551***	0.404***
	(0.095)	(0.197)	(0.112)	(0.101)	(0.174)	(0.123)
No health insurance	−0.051	−0.280	0.135	0.188***	0.190**	0.189**
	(0.251)	(0.504)	(0.296)	(0.063)	(0.081)	(0.090)
Good health	−0.044	−0.300*	0.051	−0.061*	−0.033	−0.090*
	(0.092)	(0.171)	(0.108)	(0.033)	(0.045)	(0.047)
Poor health	0.356***	0.476*	0.334**	0.130***	0.125***	0.138***
	(0.129)	(0.261)	(0.152)	(0.031)	(0.040)	(0.049)
Female	0.540***	-	-	0.375***	-	-
	(0.096)			(0.027)		
Middle	0.130	0.107	0.177	−0.026	−0.071	0.028
	(0.107)	(0.197)	(0.126)	(0.035)	(0.045)	(0.051)
Northeast	0.265**	0.104	0.341**	−0.088**	−0.138***	−0.022
	(0.121)	(0.215)	(0.141)	(0.038)	(0.048)	(0.055)
West	0.362***	0.088	0.464***	−0.117***	−0.196***	−0.014
	(0.121)	(0.244)	(0.137)	(0.041)	(0.052)	(0.058)
Constant	−2.279***	−1.570***	−2.397***	−1.517***	−1.039***	−1.725***
	(0.279)	(0.384)	(0.404)	(0.066)	(0.076)	(0.110)
N	1,308	348	960	15,188	7,531	7,657

Notes: Retirement status is as observed at t and independent variables are as observed at $t-1$, except for sex and age. All models include age effects by a 5-year window. This table only shows the coefficients of key explanatory variables. The baseline estimation results in columns 1, 4 and 5 refer to a married male with a resident insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but with less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school. whose children's highest education level is below high school. The baseline estimations in columns 2 and 3 are the same, except for sex. Standard errors are clustered at the household level. "-" means not applicable. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.6: Average Marginal Effects by Pension Plan and Sex (Including Expenditure)

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)
Employee pension	Y	Y	N	N
Sex	Female	Male	Female	Male
Employee health insurance	0.163***	0.046	0.164***	0.090***
	(0.052)	(0.029)	(0.063)	(0.033)
No health insurance	−0.056	0.040	0.055**	0.040**
	(0.098)	(0.082)	(0.024)	(0.020)
Good health	−0.086*	0.013	−0.010	−0.017**
	(0.047)	(0.029)	(0.011)	(0.008)
Poor health	0.157*	0.095**	0.034***	0.028***
	(0.087)	(0.046)	(0.011)	(0.010)
Logarithm of per capita	0.030	0.018	0.016***	0.009*
household annul expenditure	(0.030)	(0.018)	(0.006)	(0.005)
Middle	0.036	0.047	−0.020	0.006
	(0.057)	(0.032)	(0.012)	(0.010)
Northeast	0.035	0.089**	−0.039***	−0.005
	(0.063)	(0.039)	(0.013)	(0.010)
West	0.027	0.127***	−0.051***	−0.003
	(0.071)	(0.039)	(0.013)	(0.011)
N	348	960	7,531	7,657

Notes: Average marginal effects are estimated using the coefficients in Appendix Table C.4. Columns 1-4 in this table correspond to columns 2, 3, 5 and 6 in Appendix Table C.4. In each column, the baseline estimation results refer to a married individual with a resident insurance, who is between 45 and 49 years of age, is living with his/her spouse, is literate but has less than a high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.7: Probit Model: Estimation Results by Pension Plan and Sex Using Pooled Data, with Interaction Terms

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee health insurance	0.240**	0.431**	0.176	0.495***	0.712***	0.452***
	(0.113)	(0.219)	(0.140)	(0.124)	(0.228)	(0.146)
No health insurance	0.064	0.645	−0.118	0.243***	0.260**	0.232**
	(0.301)	(0.637)	(0.343)	(0.077)	(0.101)	(0.115)
Good health	0.021	−0.224	0.118	−0.059*	−0.028	−0.085*
	(0.153)	(0.285)	(0.183)	(0.032)	(0.043)	(0.046)
Poor health	0.189	−0.206	0.294	0.129***	0.113***	0.156***
	(0.200)	(0.403)	(0.240)	(0.030)	(0.038)	(0.047)
Employee health insurance	−0.100	0.060	−0.166	−0.084	−0.464	0.053
× Good health	(0.182)	(0.345)	(0.218)	(0.211)	(0.369)	(0.256)
Employee health insurance	0.253	1.126**	0.021	−0.575	−1.195**	−0.249
× Poor health	(0.247)	(0.533)	(0.290)	(0.358)	(0.569)	(0.451)
No health insurance	−0.559	-	0.563	−0.205	−0.175	−0.251
× Good health	(0.563)		(0.782)	(0.145)	(0.188)	(0.210)
No health insurance	0.079	-	0.456	0.016	−0.105	0.127
× Poor health	(0.759)		(0.873)	(0.143)	(0.185)	(0.211)
Constant	−2.224***	−1.580***	−2.272***	−1.541***	−1.076***	−1.759***
	(0.257)	(0.386)	(0.348)	(0.061)	(0.071)	(0.103)
N	1,501	398	1,089	17,706	8,820	8,886

Notes: This table is the revision of Table 5.2, with including the interaction terms of health insurance status and health states. Retirement status is as observed at t and independent variables are as observed at $t-1$, except for sex and age. Coefficients of other control variables are not reported here. The baseline estimation results in columns 1 and 4 refer to a married male with a resident insurance, between 45 and 49 years of age, living with his spouse, literate but has a less than high school diploma, who lives in the Eastern region, has one child, does not have any child in school. whose children's highest education level is below high school. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. "-" means that the variable is included in the regression model. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.8: Probit Model: Estimation Results by Pension Plan and Sex, Pooled Data (Sensitivity Check)

Dependent variable: retirement status

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
Employee health insurance	0.249***	0.537***	0.161	0.417***	0.409**	0.447***
	(0.089)	(0.173)	(0.105)	(0.098)	(0.171)	(0.117)
No health insurance	−0.055	−0.224	0.110	0.198***	0.196***	0.203**
	(0.244)	(0.446)	(0.307)	(0.058)	(0.075)	(0.084)
Good health	−0.067	−0.217	−0.014	−0.071**	−0.045	−0.096**
	(0.085)	(0.155)	(0.101)	(0.031)	(0.042)	(0.044)
Poor health	0.323***	0.465*	0.275**	0.126***	0.104***	0.162***
	(0.117)	(0.245)	(0.134)	(0.029)	(0.037)	(0.045)
Illiterate	−0.208	−0.137	−0.381	−0.031	−0.027	0.004
	(0.200)	(0.282)	(0.285)	(0.029)	(0.035)	(0.054)
High school or above	−0.024	−0.215	0.055	0.124**	0.175**	0.140**
	(0.089)	(0.168)	(0.102)	(0.048)	(0.084)	(0.061)
Married, spouse away	−0.013	0.433	−0.159	0.076	0.003	0.165**
	(0.220)	(0.401)	(0.248)	(0.047)	(0.063)	(0.071)
Unmarried	0.166	0.169	0.133	0.238***	0.195***	0.354***
	(0.172)	(0.244)	(0.264)	(0.042)	(0.053)	(0.065)
Female	0.521***	-	-	0.354***	-	-
	(0.090)			(0.025)		
Middle	0.074	−0.071	0.161	−0.022	−0.057	0.019
	(0.101)	(0.178)	(0.119)	(0.033)	(0.041)	(0.047)
Northeast	0.273**	0.149	0.366***	−0.101***	−0.142***	−0.045
	(0.112)	(0.194)	(0.134)	(0.036)	(0.045)	(0.053)
West	0.320***	0.159	0.403***	−0.101***	−0.188***	0.013
	(0.112)	(0.220)	(0.129)	(0.037)	(0.046)	(0.053)
Have children in school	0.140	−0.170	0.247	−0.116**	−0.176***	−0.028
	(0.179)	(0.320)	(0.212)	(0.049)	(0.062)	(0.072)
Children's education	0.149	−0.014	0.173	−0.055	−0.056	−0.044
being high school	(0.139)	(0.286)	(0.160)	(0.042)	(0.054)	(0.063)
Children's education	0.196*	0.001	0.253**	0.054*	0.079**	0.019
being above high school	(0.102)	(0.218)	(0.119)	(0.030)	(0.037)	(0.043)
Having two children	−0.129	0.050	−0.191	0.020	−0.015	0.078
	(0.101)	(0.171)	(0.119)	(0.042)	(0.054)	(0.062)
Having three children	−0.225*	0.046	−0.336**	−0.004	−0.022	0.032
	(0.135)	(0.276)	(0.155)	(0.045)	(0.058)	(0.067)
Having four children	−0.133	−0.063	−0.242	0.071	0.043	0.134*
	(0.178)	(0.377)	(0.198)	(0.053)	(0.068)	(0.076)
Having five or more children	−0.038	0.639	−0.211	0.121**	0.145*	0.129
	(0.225)	(0.591)	(0.248)	(0.060)	(0.076)	(0.085)
Constant	−1.918***	−1.371***	−1.886***	−1.513***	−1.047***	−1.727***
	(0.194)	(0.310)	(0.241)	(0.056)	(0.066)	(0.088)

Continued on next page

Table C.8 (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)
N	1,501	411	1,089	17,706	8,820	8,886

Notes: This table reports the coefficients of the other control variables which are not reported in Table 5.2. Retirement status is as observed at t and independent variables are as observed at $t-1$, except for sex and age. All models include age effects by a 5-year window, which are reported in Appendix Table C.9. The baseline estimation results in columns 1 and 4 refer to a married male with a resident insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but with a less than high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.9: Probit: Age Effects by Pension Plan and Sex (Sensitivity Check)

Dependent variable: retirement at t

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Employee pension	Y	Y	Y	N	N	N
Sex	Both	Female	Male	Both	Female	Male
51-55	0.278*	0.480**	0.056	0.083*	0.091*	0.084
	(0.151)	(0.198)	(0.220)	(0.045)	(0.054)	(0.078)
56-60	0.460***	0.514**	0.391*	0.177***	0.128**	0.283***
	(0.159)	(0.225)	(0.213)	(0.045)	(0.056)	(0.075)
61-65	1.188***	0.822***	1.215***	0.284***	0.217***	0.408***
	(0.166)	(0.266)	(0.214)	(0.047)	(0.057)	(0.076)
66-70	1.019***	1.149***	0.967***	0.451***	0.327***	0.628***
	(0.183)	(0.350)	(0.231)	(0.052)	(0.066)	(0.082)
71-75	1.519***	1.692***	1.519***	0.705***	0.612***	0.833***
	(0.214)	(0.485)	(0.259)	(0.060)	(0.081)	(0.090)
76-80	1.374***	1.460***	1.234***	1.036***	0.770***	1.263***
	(0.255)	(0.535)	(0.319)	(0.079)	(0.112)	(0.110)
81-84	1.696***	-	1.795***	1.279***	1.248***	1.328***
	(0.431)		(0.449)	(0.119)	(0.165)	(0.162)
N	1,501	411	1,089	17,706	8,820	8,886

Notes: This table shows the age effects in Table C.8 All models include age effects by a 5-year window. The baseline estimation results in columns 1 and 4 refer to a married male with a resident insurance, who is between 45 and 49 years of age, is living with his spouse, is literate but has a less than high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. The baseline estimations in columns 2, 3, 5, and 6 are the same as those in columns 1 and 4, except for sex. Standard errors are clustered at the household level. "-" means the coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.10: Average Marginal Effects by Pension Plan and Sex (Sensitivity Check)

Dependent variable: retirement at t

Variable at $t-1$	(1)	(2)	(3)	(4)
Employee pension	Y	Y	N	N
Sex	Female	Male	Female	Male
Employee health insurance	0.156***	0.042	0.124**	0.104***
	(0.047)	(0.027)	(0.058)	(0.032)
No health insurance	−0.050	0.029	0.055**	0.042**
	(0.093)	(0.082)	(0.023)	(0.019)
Good health	−0.062	−0.004	−0.011	−0.017**
	(0.044)	(0.026)	(0.010)	(0.008)
Poor health	0.154*	0.079**	0.028***	0.033***
	(0.084)	(0.040)	(0.010)	(0.010)
Illiterate	−0.041	−0.088	−0.007	0.001
	(0.083)	(0.057)	(0.009)	(0.010)
High school or above	0.064	0.015	0.049**	0.028**
	(0.049)	(0.028)	(0.025)	(0.013)
Married, spouse away	0.139	−0.040	0.001	0.033**
	(0.136)	(0.060)	(0.016)	(0.015)
Unmarried	0.052	0.037	0.055***	0.078***
	(0.077)	(0.076)	(0.016)	(0.017)
Middle	−0.020	0.040	−0.016	0.004
	(0.051)	(0.030)	(0.011)	(0.009)
Northeast	0.045	0.098***	−0.038***	−0.008
	(0.059)	(0.037)	(0.012)	(0.010)
West	0.048	0.109***	−0.049***	0.002
	(0.068)	(0.036)	(0.012)	(0.010)
51-55	0.129**	0.008	0.021*	0.010
	(0.052)	(0.030)	(0.013)	(0.009)
56-60	0.140**	0.067**	0.030**	0.040***
	(0.062)	(0.033)	(0.013)	(0.010)
61-65	0.246***	0.311***	0.054***	0.063***
	(0.084)	(0.042)	(0.014)	(0.011)
66-70	0.368***	0.224***	0.085***	0.112***
	(0.121)	(0.047)	(0.018)	(0.014)
71-75	0.561***	0.426***	0.179***	0.168***
	(0.152)	(0.066)	(0.026)	(0.019)
76-80	0.483***	0.318***	0.237***	0.314***
	(0.185)	(0.094)	(0.040)	(0.033)
81-84	-	0.530***	0.423***	0.338***
		(0.150)	(0.063)	(0.056)
Have children in school	−0.050	0.066	−0.046***	−0.005
	(0.085)	(0.041)	(0.013)	(0.011)
Children's education	0.000	0.065**	0.021**	0.004
being high school	(0.089)	(0.041)	(0.013)	(0.011)
Children's education	0.009	0.061**	0.021**	0.004
being above high school	(0.065)	(0.029)	(0.010)	(0.008)

Continued on next page

Table C.10 (*Continued*)

	(1)	(2)	(3)	(4)
Having two children	0.015	−0.053	−0.004	0.014
	(0.051)	(0.033)	(0.014)	(0.011)
Having three children	0.014	−0.089**	−0.006	0.006
	(0.083)	(0.040)	(0.015)	(0.012)
Having four children	−0.018	−0.066	0.012	0.025*
	(0.107)	(0.052)	(0.018)	(0.014)
Having five or more children	0.210	−0.058	0.040*	0.024
	(0.205)	(0.065)	(0.021)	(0.016)
N	411	1,089	8,820	8,886

Notes: This table is estimated using the coefficients in Appendix Tables C.8 and C.9. Columns 1-4 in this table correspond to columns 2,3,5 and 6 in Appendix Tables C.8 and C.9, respectively. In each column, the baseline estimation results refer to a married individual with a resident insurance, who is between 45 and 50 years of age, is living with their spouse, is literate but has a less than high school diploma, lives in the Eastern region, has one child, does not have any child in school, whose children's highest education level is below high school. Standard errors are clustered at the household level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table C.11: Predicted Conditional Retirement Rate by Sex (Sensitivity Check)

Dependent variable: probability of being retired, conditional on working in the last observed period

	Having an employee pension			Not having employee pension		
	(1)	(2)	(3)	(4)	(5)	(6)
	Employee health insurance	Resident health insurance	No health insurance	Employee health insurance	Resident health insurance	No health insurance
				A) Female		
45-50	0.202**	0.085*	0.055	0.262***	0.148***	0.198***
	(0.085)	(0.048)	(0.058)	(0.058)	(0.015)	(0.027)
51-55	0.362***	0.187**	0.133	0.292***	0.170***	0.224***
	(0.107)	(0.077)	(0.113)	(0.061)	(0.017)	(0.029)
56-60	0.374***	0.196**	0.140	0.305***	0.179***	0.235***
	(0.119)	(0.086)	(0.112)	(0.063)	(0.017)	(0.030)
61-65	0.495***	0.291**	0.220	0.337***	0.203***	0.263***
	(0.131)	(0.120)	(0.164)	(0.066)	(0.019)	(0.032)
66-70	0.623***	0.412**	0.328	0.378***	0.236***	0.300***
	(0.150)	(0.164)	(0.210)	(0.069)	(0.023)	(0.036)
71-75	0.805***	0.626***	0.539**	0.490***	0.332***	0.406***
	(0.144)	(0.206)	(0.274)	(0.075)	(0.032)	(0.044)
76-80	0.734***	0.536**	0.447	0.553***	0.391***	0.468***
	(0.181)	(0.230)	(0.284)	(0.081)	(0.045)	(0.055)
81-84	-	-	-	0.729***	0.580***	0.655***
				(0.079)	(0.066)	(0.068)
				B) Male		
45-50	0.042*	0.030*	0.038	0.100***	0.042***	0.064***
	(0.022)	(0.016)	(0.032)	(0.025)	(0.008)	(0.015)
51-55	0.048**	0.034**	0.043	0.116***	0.050***	0.075***
	(0.020)	(0.015)	(0.033)	(0.026)	(0.008)	(0.016)
56-60	0.091***	0.068***	0.083	0.159***	0.074***	0.107***
	(0.030)	(0.024)	(0.054)	(0.032)	(0.010)	(0.020)
61-65	0.305***	0.251***	0.288**	0.192***	0.094***	0.132***
	(0.062)	(0.058)	(0.118)	(0.036)	(0.012)	(0.023)
66-70	0.224***	0.179***	0.209**	0.257***	0.136***	0.185***
	(0.057)	(0.053)	(0.101)	(0.043)	(0.017)	(0.030)
71-75	0.418***	0.357***	0.399***	0.327***	0.185***	0.245***
	(0.088)	(0.088)	(0.144)	(0.050)	(0.024)	(0.037)
76-80	0.312***	0.257***	0.294**	0.493***	0.321***	0.397***
	(0.106)	(0.097)	(0.139)	(0.061)	(0.039)	(0.052)
81-84	0.528***	0.464***	0.508**	0.519***	0.345***	0.422***
	(0.170)	(0.172)	(0.212)	(0.078)	(0.060)	(0.071)

Notes: A) is estimated using the estimated coefficients in Appendix Table C.8, columns 2 and 5; B) is estimated using the estimated coefficients in Appendix Table C.8, columns 3 and 6. The baseline estimation results refer to a married individual, is living with their spouse, is literate but has a less than high school diploma, lives in the Eastern region, has one child, does not have any child in school and whose children's highest education level is below high school. Standard errors are clustered at the household level."-" means that the coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix D

Supplementary Material for Chapter 6

Table D.1: Incidence of Change in Working Sectors

(a) 2011-13 wave

Working sector in 2011	Working sector in 2013			
	Agriculture	Non-agriculture	Both	Total
Agriculture	4,799	261	589	5,649
Non-agriculture	166	1,215	175	1,556
Both	650	183	566	1,399
Total	5,615	1,659	1,330	8,604

(b) 2013-15 wave

Working sector in 2013	Working sector in 2015			
	Agriculture	Non-agriculture	Both	Total
Agriculture	4,845	408	776	6,029
Non-agriculture	232	1,479	225	1,936
Both	579	314	690	1,583
Total	5,656	2,201	1,691	9,548

Notes: The numbers on the diagonal are the number of participated who remained within the same working sector throughout the survey period.
Source: CHARLS 2011, 2013 and 2015 survey waves.

Table D.2: Number of Workdays Missed: Individuals Who Work Only in the Agricultural Sector, Pooled

Only in Agriculture

daysmissed	Freq.	Percent	Cum.	daysmissed	Freq.	Percent	Cum.	daysmissed	Freq.	Percent	Cum.
0	**5,132**	**58.94**	58.94	39	1	0.01	87.65	156	3	0.03	97.07
1	43	0.49	59.43	**40**	**47**	**0.54**	88.19	159	1	0.01	97.08
2	154	1.77	61.20	45	21	0.24	88.43	160	2	0.02	97.11
3	183	2.10	63.31	46	1	0.01	88.45	170	1	0.01	97.12
4	96	1.10	64.41	48	7	0.08	88.53	175	3	0.03	97.15
5	**199**	**2.29**	66.69	**50**	**41**	**0.47**	89.00	178	1	0.01	97.16
6	57	0.65	67.35	52	1	0.01	89.01	**180**	**123**	**1.41**	98.58
7	154	1.77	69.12	54	1	0.01	89.02	182	4	0.05	98.62
8	37	0.42	69.54	**55**	**3**	**0.03**	89.05	183	2	0.02	98.64
9	7	0.08	69.62	56	1	0.01	89.07	184	1	0.01	98.66
10	**398**	**4.57**	74.19	**60**	**330**	**3.79**	92.86	185	1	0.01	98.67
12	16	0.18	74.38	62	2	0.02	92.88	**200**	**26**	**0.30**	98.97
13	15	0.17	74.55	65	3	0.03	92.91	**210**	**7**	**0.08**	99.05
14	23	0.26	74.81	**70**	**15**	**0.17**	93.09	219	1	0.01	99.06
15	**266**	**3.06**	77.87	72	5	0.06	93.14	**240**	**8**	**0.09**	99.15
16	7	0.08	77.95	73	1	0.01	93.15	242	1	0.01	99.16
17	3	0.03	77.98	**75**	**6**	**0.07**	93.22	243	1	0.01	99.17
18	12	0.14	78.12	76	1	0.01	93.24	245	1	0.01	99.18
20	**184**	**2.11**	80.23	**80**	**17**	**0.20**	93.43	250	1	0.01	99.20
21	3	0.03	80.27	84	2	0.02	93.45	**270**	**7**	**0.08**	99.28
22	3	0.03	80.30	**90**	**130**	**1.49**	94.95	280	1	0.01	99.29
23	4	0.05	80.35	92	1	0.01	94.96	290	1	0.01	99.30
24	10	0.11	80.46	95	2	0.02	94.98	292	1	0.01	99.31
25	**21**	**0.24**	80.71	96	2	0.02	95.00	**300**	**30**	**0.34**	99.66
26	6	0.07	80.77	**100**	**59**	**0.68**	95.68	312	1	0.01	99.67
27	2	0.02	80.80	104	3	0.03	95.72	325	1	0.01	99.68
28	2	0.02	80.82	**110**	**1**	**0.01**	95.73	**330**	**3**	**0.03**	99.71
30	**566**	**6.50**	87.32	**120**	**70**	**0.80**	96.53	**335**	**1**	**0.01**	99.72
31	2	0.02	87.34	126	1	0.01	96.54	**340**	**2**	**0.02**	99.75
32	2	0.02	87.37	128	1	0.01	96.55	**350**	**4**	**0.05**	99.79
35	8	0.09	87.46	**130**	**4**	**0.05**	96.60	**360**	**3**	**0.03**	99.83
36	14	0.16	87.62	**140**	**3**	**0.03**	96.63	**365**	**15**	**0.17**	100.00
38	2	0.02	87.64	**150**	**35**	**0.40**	97.04	Total	8,707	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of workdays missed in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero missed workdays, and they tend to round the number of workdays missed to multiples of 5.
Source: Based on pooled CHARLS 2011, 2013 and 2015 survey waves.

Table D.3: Number of Workdays Missed: Individuals Who Worked Only in the Non-agricultural Sector and Individuals Who Worked in Both Sectors, Pooled

Only in non-agriculture

daysmissed	Freq.	Percent	Cum.
0	**1,998**	**83.74**	83.74
1	13	0.54	84.28
2	44	1.84	86.13
3	27	1.13	87.26
4	12	0.50	87.76
5	**19**	**0.80**	88.56
6	8	0.34	88.89
7	28	1.17	90.07
8	5	0.21	90.28
10	**42**	**1.76**	92.04
11	3	0.13	92.16
12	5	0.21	92.37
13	1	0.04	92.41
14	4	0.17	92.58
15	**23**	**0.96**	93.55
16	3	0.13	93.67
18	2	0.08	93.76
20	**23**	**0.96**	94.72
21	1	0.04	94.76
24	3	0.13	94.89
25	1	0.04	94.93
27	1	0.04	94.47
30	53	2.22	97.19
31	1	0.04	97.23
36	2	0.08	97.32
40	**3**	**0.13**	97.44
45	**5**	**0.21**	97.65
48	1	0.04	97.69
50	**2**	**0.08**	97.78
60	**23**	**0.96**	98.74
69	1	0.04	98.78
90	**3**	**0.13**	98.91
100	1	0.04	98.95
120	**6**	**0.25**	99.20
150	**3**	**0.13**	99.33
153	1	0.04	99.37
178	1	0.04	99.41
180	**10**	**0.42**	99.83
200	**2**	**0.08**	99.92
360	**1**	**0.04**	99.96
365	**1**	**0.04**	100.00
Total	2,386	100.00	

Both sectors

daysmissed	Freq.	Percent	Cum.
0	**717**	**65.36**	65.36
1	8	0.73	66.09
2	28	2.55	68.64
3	22	2.01	70.65
4	7	0.64	71.29
5	**26**	**2.37**	73.66
6	7	0.64	74.29
7	35	3.19	77.48
8	2	0.18	77.67
10	**43**	**3.92**	81.59
12	3	0.27	81.86
14	2	0.18	82.04
15	**30**	**2.73**	84.78
17	1	0.09	84.87
20	**27**	**2.46**	87.33
21	1	0.09	87.42
24	3	0.27	87.69
25	1	0.09	87.78
30	**49**	**4.47**	92.25
35	1	0.09	92.34
40	**7**	**0.64**	92.98
45	**5**	**0.46**	93.44
50	**7**	**0.64**	94.07
55	1	0.09	94.17
60	**29**	**2.64**	96.81
70	1	0.09	96.90
80	1	0.09	96.99
90	**11**	**1.00**	97.99
100	**6**	**0.55**	98.54
120	**8**	**0.73**	99.27
180	**6**	**0.55**	99.82
210	**1**	**0.09**	99.91
365	**1**	**0.09**	100.00
Total	1,097	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of missed work days in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero workdays missed, and they tend to round the number of workdays missed to multiples of 5.

Source: Based on pooled CHARLS 2011, 2013 and 2015 survey waves.

Table D.4: Number of Workdays Missed: Individuals Who Work Only in the Agricultural Sector, 2013 Wave

Only in Agriculture

daysmissed	Freq.	Percent	Cum.
0	**2,959**	**62.82**	62.82
1	15	0.32	63.14
2	56	1.19	64.33
3	67	1.42	65.75
4	36	0.76	66.52
5	**80**	**1.70**	68.22
6	39	0.83	69.04
7	99	2.10	71.15
8	17	0.36	71.51
9	5	0.11	71.61
10	**219**	**4.65**	76.26
12	14	0.30	76.56
13	5	0.11	76.67
14	10	0.21	76.88
15	**106**	**2.25**	79.13
16	7	0.15	79.28
17	3	0.06	79.34
18	9	0.19	79.53
19	1	0.02	79.55
20	**123**	**2.61**	82.17
21	1	0.02	82.19
24	2	0.04	82.23
25	**5**	**0.11**	82.34
26	1	0.02	82.36
28	1	0.02	82.38
30	**292**	**6.20**	88.58
33	1	0.02	88.60
35	**8**	**0.17**	88.77
36	4	0.08	88.85
37	1	0.02	88.87
38	2	0.04	88.92
39	1	0.02	88.94
40	**30**	**0.64**	89.58
45	**13**	**0.28**	89.85
46	1	0.02	89.87
48	1	0.02	89.89
50	**23**	**0.49**	90.38
58	1	0.02	90.40
60	**158**	**3.35**	93.76
63	1	0.02	93.78
65	1	0.02	93.80
70	**10**	**0.21**	94.01
72	1	0.02	94.03
74	1	0.02	94.06
75	1	0.02	94.08
80	**7**	**0.15**	94.23
84	1	0.02	94.25
90	**62**	**1.32**	95.56
98	1	0.02	95.58
100	**36**	**0.76**	96.35
120	**38**	**0.81**	97.15
135	1	0.02	97.18
140	2	0.04	97.22
150	**16**	**0.34**	97.56
160	2	0.04	97.60
170	1	0.02	97.62
180	**46**	**0.98**	98.60
200	**9**	**0.19**	98.79
210	2	0.04	98.83
240	**8**	**0.17**	99.00
250	2	0.04	99.04
270	**4**	**0.08**	99.13
279	1	0.02	99.15
300	**21**	**0.45**	99.60
330	2	0.04	99.64
340	1	0.02	99.66
350	1	0.02	99.68
360	**6**	**0.13**	99.81
365	**9**	**0.19**	100.00
Total	4,710	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of missed work days in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero workdays missed, and they tend to round the number of workdays missed to multiples of 5.

Source: Based on pooled CHARLS 2011 and 2013 survey waves.

Table D.5: Number of Workdays Missed: Individuals Who Worked Only in the Non-agricultural Sector and Individuals Who Worked in Both Sectors, 2013 Wave

Only in non-agriculture			
daysmissed	Freq.	Percent	Cum.
0	**1,015**	**86.31**	86.31
1	4	0.34	86.65
2	21	1.79	88.44
3	18	1.53	89.97
4	8	0.68	90.65
5	**8**	**0.68**	91.33
6	1	0.09	91.41
7	15	1.28	92.69
9	1	0.09	92.77
10	**21**	**1.79**	94.56
15	**10**	**0.85**	95.41
16	2	0.17	95.58
17	1	0.09	95.66
20	**4**	**0.34**	96.00
25	1	0.09	96.09
28	2	0.17	96.26
30	**13**	**1.11**	97.36
36	1	0.09	97.45
40	**4**	**0.34**	97.79
50	2	0.17	97.96
60	**11**	**0.94**	98.89
75	1	0.09	98.98
90	**5**	**0.43**	99.40
100	**3**	**0.26**	99.66
140	**1**	**0.09**	99.74
150	**1**	**0.09**	99.83
180	**1**	**0.09**	99.91
270	**1**	**0.09**	100.00
Total	1,176	100.00	

Both sectors			
daysmissed	Freq.	Percent	Cum.
0	**399**	**72.55**	72.55
1	1	0.18	72.73
2	7	1.27	74.00
3	12	2.18	76.18
4	8	1.45	77.64
5	**11**	**2.00**	79.64
6	3	0.55	80.18
7	8	1.45	81.64
9	1	0.18	81.82
10	**24**	**4.36**	86.18
11	1	0.18	86.36
12	3	0.55	86.91
14	1	0.18	87.09
15	**12**	**2.18**	89.27
18	1	0.18	89.45
20	**8**	**1.45**	90.91
30	**21**	**3.82**	94.73
40	2	0.36	95.09
45	1	0.18	95.27
48	1	0.18	95.45
50	1	0.18	95.64
60	**9**	**1.64**	97.27
75	**1**	**0.18**	97.45
80	**1**	**0.18**	97.64
90	**2**	**0.36**	98.00
100	**2**	**0.36**	98.36
120	**3**	**0.55**	98.91
180	**3**	**0.55**	99.45
200	**1**	**0.18**	99.64
245	**1**	**0.18**	99.82
300	**1**	**0.18**	100.00
Total	550	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of missed work days in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero workdays missed, and they tend to round the number of workdays missed to multiples of 5.
Source: Based on pooled CHARLS 2011 and 2013 survey waves.

Table D.6: Number of Workdays Missed: Individuals Who Work Only in the Agricultural Sector, 2015 Wave

Only in Agriculture

daysmissed	Freq.	Percent	Cum.
0	**2,850**	**60.70**	60.70
1	20	0.43	61.13
2	68	1.45	62.58
3	102	2.17	64.75
4	50	1.06	65.81
5	**102**	**2.17**	67.99
6	33	0.70	68.69
7	82	1.75	70.44
8	16	0.34	70.78
9	5	0.11	70.88
10	**216**	**4.60**	75.48
12	9	0.19	75.68
13	6	0.13	75.80
14	11	0.23	76.04
15	**142**	**3.02**	79.06
16	8	0.17	79.23
17	2	0.04	79.28
18	5	0.11	79.38
20	**110**	**2.34**	81.73
21	1	0.02	81.75
22	2	0.04	81.79
23	1	0.02	81.81
24	9	0.19	82.00
25	**10**	**0.21**	82.22
26	1	0.02	82.24
27	1	0.02	82.26
28	2	0.04	82.30
30	**292**	**6.22**	88.52
31	1	0.02	88.54
32	1	0.02	88.56
35	**5**	**0.11**	88.67
36	2	0.04	88.71
38	2	0.04	88.75
39	2	0.04	88.80
40	**30**	**0.64**	89.44
45	**6**	**0.13**	89.56
46	1	0.02	89.58
48	4	0.09	89.67
50	**26**	**0.55**	90.22
55	1	0.02	90.24
60	**152**	**3.24**	93.48
65	1	0.02	93.50
70	**10**	**0.21**	93.72
72	1	0.02	93.74
73	1	0.02	93.76
75	**2**	**0.04**	93.80
76	1	0.02	93.82
80	**5**	**0.11**	93.93
84	2	0.04	93.97
90	**57**	1.21	95.19
96	1	0.02	95.21
100	**36**	0.77	95.97
104	1	0.02	96.00
120	**28**	0.60	96.59
130	2	0.04	96.63
150	**14**	0.30	96.93
156	1	0.02	96.95
160	**2**	0.04	97.00
175	1	0.02	97.02
180	**65**	1.38	98.40
183	1	0.02	98.42
200	**22**	0.47	98.89
210	4	0.09	98.98
240	**8**	0.17	99.15
245	1	0.02	99.17
270	**4**	**0.09**	99.25
280	**1**	**0.02**	99.28
290	**1**	**0.02**	99.30
300	**12**	**0.26**	99.55
330	**3**	**0.06**	99.62
340	**1**	**0.02**	99.64
350	**1**	**0.02**	99.66
360	**7**	**0.15**	99.81
365	**9**	**0.19**	100.00
Total	4,695	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of missed work days in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero workdays missed, and they tend to round the number of workdays missed to multiples of 5.

Source: Based on pooled CHARLS 2013 and 2015 survey waves.

Table D.7: Number of Workdays Missed: Individuals Who Worked Only in the Non-agricultural Sector and Individuals Who Worked in Both Sectors, 2015 Wave

Only in non-agriculture

daysmissed	Freq.	Percent	Cum.
0	**1,207**	83.82	83.82
1	12	0.83	84.65
2	31	2.15	86.81
3	22	1.53	88.33
4	8	0.56	88.89
5	**11**	**0.76**	89.65
6	6	0.42	90.07
7	13	0.90	90.97
8	2	0.14	91.11
10	**21**	**1.46**	92.57
11	2	0.14	92.71
12	5	0.35	93.06
14	1	0.07	93.12
15	**15**	**1.04**	94.17
16	1	0.07	94.24
20	**11**	**0.76**	95.00
24	2	0.14	95.14
25	1	0.07	95.21
26	1	0.07	95.28
30	**36**	**2.50**	97.78
35	1	0.07	97.85
40	**3**	**0.21**	98.06
45	**2**	**0.14**	98.19
52	1	0.07	98.26
55	1	0.07	98.33
56	1	0.07	98.40
60	**6**	**0.42**	98.82
80	**1**	**0.07**	98.89
90	**5**	**0.35**	99.24
93	1	0.07	99.31
100	**2**	**0.14**	99.44
120	**1**	**0.07**	99.51
178	1	0.07	99.58
180	4	0.28	99.86
200	**1**	**0.07**	99.93
365	**1**	**0.07**	100.00
Total	1,440	100.00	

Both sectors

daysmissed	Freq.	Percent	Cum.
0	**443**	**66.42**	66.42
1	5	0.75	67.17
2	8	1.20	68.37
3	15	2.25	70.61
4	8	1.20	71.81
5	**12**	**1.80**	73.61
6	5	0.75	74.36
7	22	3.30	77.66
8	2	0.30	77.96
9	1	0.15	78.11
10	**39**	**5.85**	83.96
12	1	0.15	84.11
14	2	0.30	84.41
15	**17**	**2.55**	86.96
20	**15**	**2.25**	89.21
24	1	0.15	89.36
25	1	0.15	89.51
30	**30**	**4.50**	94.00
37	1	0.15	94.15
40	2	0.30	94.45
45	1	0.15	94.60
50	2	0.30	94.90
60	**17**	**2.55**	97.45
80	**2**	**0.30**	97.75
90	**3**	**0.45**	98.20
100	**1**	**0.15**	98.35
120	**2**	**0.30**	98.65
180	**3**	**0.45**	99.10
200	**1**	**0.15**	99.25
300	**1**	**0.15**	99.40
365	**4**	**0.60**	100.00
Total	667	100.00	

Notes: This table only includes responses to the question on workdays missed by those who were working in two consecutive waves, and who did not switch between working sectors. "Daysmissed" is the number of missed workdays in the year before the survey year. Numbers in bold indicate that a substantial fraction of respondents report zero workdays missed, and they tend to round the number of workdays missed to multiples of 5.

Source: Based on pooled CHARLS 2013 and 2015 survey waves.

Table D.8: Logit Model: Results and Marginal Effects

Dependent variable: having missed workdays

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Coefficients			Average marginal effects		
	Agricultural	Non-agricultural	Both	Agricultural	Non-agricultural	Both
Regular drinker	−0.016	0.058	0.092	−0.003	0.007	0.019
	(0.060)	(0.131)	(0.150)	(0.013)	(0.017)	(0.031)
Smoked	−0.021	0.107	0.324*	−0.005	0.014	0.067*
	(0.069)	(0.153)	(0.186)	(0.016)	(0.020)	(0.038)
Illiterate	0.061	0.323	0.253	0.014	0.048	0.054
	(0.057)	(0.233)	(0.249)	(0.013)	(0.037)	(0.054)
High school or above	−0.344***	−0.280*	−0.055	−0.074***	−0.035**	−0.011
	(0.121)	(0.146)	(0.199)	(0.025)	(0.017)	(0.041)
Married, spouse away	0.093	−0.083	0.459*	0.021	−0.010	0.099*
	(0.105)	(0.212)	(0.250)	(0.024)	(0.026)	(0.056)
Unmarried	0.134	0.346	0.237	0.030	0.049	0.050
	(0.091)	(0.239)	(0.354)	(0.021)	(0.037)	(0.077)
Middle	0.133*	0.048	0.080	0.029*	0.006	0.016
	(0.070)	(0.159)	(0.164)	(0.015)	(0.020)	(0.034)
Northeast	−0.005	0.052	−0.021	−0.001	0.007	−0.004
	(0.130)	(0.227)	(0.358)	(0.028)	(0.029)	(0.072)
West	0.617***	0.283*	0.390**	0.140***	0.038*	0.083**
	(0.065)	(0.151)	(0.190)	(0.014)	(0.021)	(0.041)
50-54	0.101	0.022	0.109	0.022	0.003	0.023
	(0.084)	(0.148)	(0.176)	(0.019)	(0.020)	(0.037)
55-59	0.063	0.025	−0.261	0.014	0.003	−0.053
	(0.079)	(0.156)	(0.195)	(0.018)	(0.021)	(0.039)
60-64	0.149*	−0.247	0.029	0.033*	−0.030	0.006
	(0.080)	(0.226)	(0.225)	(0.018)	(0.026)	(0.047)
65-69	0.133	−0.514	−0.277	0.030	−0.058	−0.056
	(0.092)	(0.444)	(0.386)	(0.021)	(0.043)	(0.075)
70-74	−0.011	−1.312*	0.763	−0.002	−0.114***	0.170
	(0.120)	(0.770)	(0.735)	(0.026)	(0.040)	(0.167)
75-79	0.151	−0.959	0.035	0.034	−0.094	0.007
	(0.176)	(1.088)	(0.938)	(0.039)	(0.074)	(0.198)
80-84	0.188	-	-	0.042	-	-
	(0.335)	-	-	(0.076)	-	-
Constant	−1.196***	−2.127***	−1.504***	-	-	-
	(0.109)	(0.208)	(0.263)	-	-	-
N	8,707	2,383	1,097	8,707	2,383	1,097

Notes: All models include age effects based on a 5-year window. The baseline estimation results refer to a married male with a resident insurance, but without an employee pension plan, between 45 and 49 years of age, living with their spouse, literate but has a less than high school diploma, who do not have any chronic disease, are not disabled, who do not drink regularly and haven't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. "-" means that coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table D.9: Heaped Generalized Poisson: Estimation Results

Dependent variable: number of missedworkdays

Variable	(1)	(2)	(3)
Sectors of employment	Agricultural	Non-agricultural	Both
Regular drinker	−0.025	0.052	0.091
	(0.042)	(0.118)	(0.117)
Smoked	−0.044	0.126	0.322**
	(0.050)	(0.137)	(0.139)
Literate	0.038	0.342*	0.179
	(0.038)	(0.190)	(0.171)
High school or above	−0.316***	−0.275**	−0.027
	(0.095)	(0.132)	(0.150)
Married, spouse away	0.128*	−0.071	0.335*
	(0.074)	(0.188)	(0.177)
Unmarried	0.095	0.325	0.213
	(0.059)	(0.199)	(0.244)
Middle	0.139***	0.060	0.092
	(0.050)	(0.134)	(0.125)
Northeast	−0.003	0.033	0.032
	(0.092)	(0.192)	(0.257)
West	0.556***	0.286**	0.334**
	(0.045)	(0.129)	(0.134)
50-54	0.090	0.030	0.063
	(0.059)	(0.129)	(0.130)
55-59	0.054	0.020	−0.206
	(0.055)	(0.138)	(0.149)
60-64	0.144***	−0.261	0.006
	(0.055)	(0.203)	(0.169)
65-69	0.144**	−0.542	−0.189
	(0.062)	(0.348)	(0.296)
70-74	0.020	−1.281*	0.621
	(0.084)	(0.718)	(0.384)
75-79	0.114	−0.902	−0.416
	(0.115)	(1.013)	(1.027)
80-84	0.144	−18.280	-
	(0.231)	(922.400)	-
Constant	2.017***	0.940***	1.407***
	(0.101)	(0.289)	(0.259)
N	8,707	2,386	1,097

Notes: All models include age effects based on a 5-year window. The baseline estimation results refer to a married male with a resident insurance, but without an employee pension plan, between 45 and 49 years of age, living with their spouse, literate but has a less than high school diploma, who do not have any chronic disease, are not disabled, who do not drink regularly and haven't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. A statistically significant LRT of $\delta = 0$ shows that the generalized Poisson model is preferred to the Poisson model. The p-value of the Vuong test of heap versus non-heap model is 0.0000. Statistically significant Vuong test of the model shows that the heaped Poisson model is better than the nonheaped model. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table D.10: Heaped Generalized Poisson: Marginal Effect of Control Variables
Dependent variable: number of missed workdays

Variable	(1)	(2)	(3)
Sectors of employment	Agriculture	Non-agriculture	Both
Regular drinker	−0.448	0.233	0.881
	(0.964)	(0.552)	(1.371)
Smoked	−0.780	0.558	3.127*
	(1.134)	(0.656)	(1.694)
Literate	0.694	1.892	1.870
	(0.909)	(1.507)	(2.304)
High school or above	−4.820***	−1.116**	−0.258
	(1.511)	(0.556)	(1.773)
Married, spouse away	2.399	−0.301	3.723
	(1.912)	(0.845)	(2.774)
Unmarried	1.748	1.680	2.223
	(1.541)	(1.404)	(3.052)
Middle	1.942**	0.249	0.833
	(0.894)	(0.651)	(1.356)
Northeast	−0.045	0.133	0.281
	(1.577)	(0.898)	(2.979)
West	9.661***	1.335*	3.427*
	(0.950)	(0.729)	(1.789)
50-54	1.542	0.139	0.644
	(1.272)	(0.679)	(1.553)
55-59	0.916	0.092	−1.839
	(1.177)	(0.712)	(1.571)
60-64	2.552**	−1.061	0.062
	(1.240)	(0.810)	(1.985)
65-69	2.536*	−1.936	−1.709
	(1.457)	(1.203)	(3.074)
70-74	0.339	−3.341***	8.513
	(1.814)	(1.091)	(10.160)
75-79	1.977	−2.748	−3.370
	(2.821)	(2.148)	(3.227)
80-84	2.541	−4.625***	-
	(5.440)	(0.545)	-
N	8,707	2,386	1,097

Notes: All models include age effects based on a 5-year window. The baseline estimation results refer to a married male with a resident insurance, but without an employee pension plan, between 45 and 49 years of age, living with their spouse, literate but has a less than high school diploma, who do not have any chronic disease, are not disabled, who do not drink regularly and haven't smoked before, who lives in the Eastern region. Standard errors are clustered at the household level. "-" means that the coefficient cannot be estimated. * Significant at 10%; ** significant at 5%; *** significant at 1%.

D.1 Regulations of medical leave and paid leave of employees

In China, the maximum duration of medical leave[1] ranges from three months to twenty-four months, depending on an employee's years of work experience and years of service in the current company (Ministry of Human Resources and Social Security, PRC, 1994). For those with actual working experience of fewer than ten years and fewer than five years of service in the current enterprise, the maximum medical leave period is three months; for those whose years of service are more than five years, it extends to six months. The longer the years of experience and service in the current enterprise individuals have, the longer their medical leave can be. Those whose maximum medical leave is twenty-four months should have served in the current enterprise for more than twenty years.

Wages during medical leave are determined by their enterprises. Usually, these wages are determined by the duration of the employee's medical leave and the employee's years of service in the current enterprise. In general, if employee's medical leave is less than six months, their wage during medical leave ranges from 60% to 100% of their own wage, following a progressive scale based on their years of service in the current enterprise[2] (Ministry of Human Resources and Social Security, PRC, 1953). Wages for employees during medical leave also vary by city and by the type of employee's job. The medical leave wage should not be less than 80% of the local minimum wage (Ministry of Human Resources and Soical Security, PRC, 1995).

In addition to medical leave, employees also have annual paid leave. The length of employees' annual leave is determined by their years of experience. Those with years of experience between one and 10 years are entitled to five days of annual leave.

[1]The maximum duration of medical leave is the time limit during which enterprise employees, due to illness or non-work-related injuries, stop working for treatment and rest without the labor contract being terminated.

[2]When employee's medical leave is less than 6 months, for those with less than two years of service, the medical leave wage is 60% of their own wage. For those with two to four years of service, the ratio of medical leave to the original wage increases to 70%. This ratio increases to 100% when employees have eight years or more of service in the current enterprise.

The longest annual leave is 15 days, available to employees with more than 20 years of working experience (General Office of the State Council, PRC, 2007).

In summary, although the maximum duration of medical leave varies depending on an employee's years of experience, for individuals with the least years of experience, it exceeds three months, which is higher than the average number of observed missed workdays in all groups in the sample. Hence, the observed effect of health insurance is unlikely to be attributed to sick leave rules.

D.2 Working days and access to health care

In the context of labour economics, the relation between workdays missed and workdays is complex. A reduction in the number of workdays missed does not linearly translate to an increase in the number of workdays attended. Conversely, the number of workdays may in turn influence the number of workdays missed. In principle, individuals who work more days are more likely to miss workdays due to health reasons and are also expected miss more workdays. Table D.11 compare the average number of workdays by respondents based on their health insurance plan. This data indicates that overall, individuals with employee health insurance have the most of working days, while those with resident health insurance have the least, but within each working sector, this relationship varies. Within the agricultural sector, individuals with the employee health insurance have the fewest workdays, at only 140 days annually. In the non-agricultural sector, the number of workdays are similar among individuals with different health insurance plans, but those with the employee health insurance still report the fewest days worked. In contrast, for individuals employed in both sectors, individuals with the employee health insurance report most workdays.

Table D.11: Summary Statistics: Number of Workdays

Sector of employment	Health insurance plan								
	Employee			Resident			No plan		
	Mean	St.D.	N	Mean	St.D.	N	Mean	St.D.	N
Overall	248.48	100.11	1729	216.94	122.20	17,322	233.06	124.88	957
Agricultural	138.81	127.59	278	199.38	124.55	12,002	195.33	128.11	495
Non-agricultural	272.38	75.16	1325	282.87	102.25	3140	287.84	102.78	370
Both	239.19	98.71	126	218.67	101.23	2180	215.72	107.97	92

Notes: Results of t-test show that, in the overall sample, the mean of workdays for the resident health insurance group is statistically different with those for the employee health insurance and no health insurance groups. In both the agricultural and non-agricultural samples, the mean number of workdays for the resident health insurance group is statistically different from that of the employee health insurance group (p-values are 0.000 and 0.001, respectively), but it shows no statistical difference when compared to the no health insurance group. In the sample of those employed by both sectors, the average missed workdays for the resident health insurance group is statistically different from that of both the employee health insurance group and no health insurance group (p-value is 0.000).
Source: CHARLS, 2011, 2013, and 2015 waves.

Individuals with the employee health insurance also work fewer hours per week. Table D.12 shows the average hours worked per week based on their health insurance plan and type of job. In Table D.12, those who worked on other's farm and own farm are employed in the agricultural sector and those who are employed and self-employed are employed in the non-agricultural sector.

Table D.12: Hours Worked Per Week by Health Insurance Plan and Job Type

	Other's farm		Own farm		Employed		Self-employed	
	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Employee health insurance	36.43	26.70	23.30	20.13	43.02	16.05	50.53	31.79
	(N = 14)		(N = 420)		(N = 1,397)		(N = 199)	
Resident health insurance	44.63	24.48	41.31	22.71	52.78	22.15	51.24	30.72
	(N = 1,174)		(N = 14,925)		(N = 4,015)		(N = 2,577)	
No health insurance	41.11	28.37	42.97	23.50	55.55	24.53	57.46	28.69
	(N = 53)		(N = 605)		(N = 284)		(N = 225)	

Source: CHARLS, 2011, 2013 waves.

Unfortunately, the survey did not provide information on whether the number of workdays missed includes days spent visiting hospitals. In Table D.13, I present the percentage of respondents who visited a hospital within one month before the survey was conducted. Agricultural sector has the largest fraction of people who visited a hospital in the past month, while individuals who are employed only in the non-agricultural sector had the smallest fraction of people who visited a hospital. In all three sectors, the hospital visit rate is the lowest for individuals who do not have any insurance, and it is the highest for those who have a resident health insurance. It is worth noting that in the non-agricultural sector, although those without any health insurance have the lowest rate of hospital visit, they have the largest probability of experiencing missed work days.

Table D.13: Proportion of Respondents Who Visited A Hospital in the Past Month

	Sector of employment		
Health insurance plan	Agriculture	Non-agriculture	Both
Employee	20.00	14.04	15.88
Resident	21.23	15.92	17.94
No health insurance	15.18	10.43	13.22

Notes: In agriculture, the proportion of respondents who visited a hospital in the past month in the no health insurance group is statistically different from that in the resident health insurance group (p-value of a t-test is 0.002). In non-agriculture, the proportion in the no health insurance group is different from that in the resident health insurance group (p-value of a t-test is 0.002). For those employed in both sectors, the proportion of respondents who visited a hospital in the past month is not statistically different in all three health insurance groups.
Source: Author's calculation based on CHARLS, 2011, 2013 and 2015 waves.