THE IMPACT OF INCOME ON THE HEALTH OF CHINESE RESIDENTS

by

Wudi Yang

Bachelor of Economics, Henan University, 2012

A Report Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in the Graduate Academic Unit of Economics

Supervisor: Philip Leonard, PhD, Dept. of Economics

Examining Board: Weiqiu Yu, PhD, Dept. of Economics

Michael G Farnworth, PhD, Dept. of Economics

This report is accepted by the Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

September, 2016

Abstract

Since the implementation of China's reform and opening up policy in 1978, and with the rapid development of China's economy, individual income has increased significantly, and personal health is becoming a growing issue. This paper, through an empirical study, estimates the impact of income on the following measures of health: Self-reported health (SRH) and Health of last four weeks. This study uses a longitudinal dataset, the China Health and Nutrition Survey (CHNS). Based on the Grossman (1972) health demand model, logit regression estimates and fixed effect regression estimates, both before and after controlling for different variables and treating age differently, the results indicate that individual income is positively correlated with personal health. More specifically, each 1% increase in an individual's income for last year is associated with an increase in the odds of health as measured by self-reported health and health for the last four weeks of around 0.045%.

Acknowledgement

First of all, I would like to extend my sincere gratitude to my supervisor, professor Philip Leonard, for his constant encouragement and guidance. He has walked me through all the stages of the writing of this report. Without his consistent and illuminating instruction, this report could not have reached its present form.

High tribute shall be paid to professor Weiqiu Yu and professor Michael G Farnworth, for their instructive advice and useful suggestions on my report.

Finally, I am indebted to my parents, Yaner Lee and my friends for their continuous support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

I. Introduction

With the increase of industrialization and urbanization, and the improvement of national economics, Chinese residents' income has increased year by year. According to *The national economic and social development statistical bulletin* from China's National Bureau of Statistics, in 2015, the disposable income of urban residents was 31195 yuan per capita, up 8.2% over the previous year. After deducting price factors, the real growth rate is 6.6%. The disposable income of rural residents was 11422 yuan per capita, up 8.9% over the previous year; after deducting price factors, the real growth rate is 7.5%. (China, 2015) At the same time, China residents' medical health care service demand increased significantly: in particular, chronic disease continues to rise, and the disease burden is rising over time. (Liu, (2004) The relationship between income and health is receiving more and more attention.

From 1989 to 2011 was the golden time of China's economic development, and was also the most rapid and intense period of economic transformation. In this period, the international economic environment was unstable; China experienced the East Asian financial crisis in 1998 and the international financial crisis in 2008. In the aspect of system reform, China has implemented the tax system reform in 1994 and investment system reform in 2003. Those reforms reduced tax rates and the cost of investment, which can increase people's income. (Naughton, 2007) An increase of income could have a further influence on people's health. From the Chinese residents' perspective, those events and reforms have a significant influence on their lives. This paper will study the relationship between health and income in China based on the

characteristics of the rapid economic development during this period. This paper will be based on the China Health and Nutrition Survey (CHNS) data, which is jointly conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety and the Chinese Center for Disease Control and Prevention, to analyze the relationship between income and health of the people of China during the past 20 years, 1991-2011. From the social point of view, studying the relationship between health and income has a certain practical significance. On the one hand, China put forward the concept that a well-off society with a comprehensive policy to improve the quality of life is an important goal in the future. (Jang, 2003) In the past decades, the Chinese people have had poor living conditions, and the main goal was to solve the problem of food and clothing. With the implementation of the reform and opening-up policy in 1978, the Chinese economy has been increasing rapidly for 30 years. In this process, the majority of Chinese people's living condition have enhanced significantly. (Gan and Zheng, 2009) People started to pay attention to health, environment and other issues. So health issues have increasingly become an important part of life. Health is the foundation of increasing income and building a well-off society. On the other hand, health is the main component of human capital investment; investment in health can directly improve social productivity, and thus increase personal wages and the quality of life.

Based on the characteristics of China's economic development from 1991-2011, this paper analyzes the influence of income on the health of Chinese adults from

1991-2011.From the perspective of health demand, Grossman (1972) and many scholars have done a lot of work; the most important achievement is the Grossman health demand function. At the same time, health also belongs to the category of human capital, so both sides should be fully considered in the modeling process. This study is based on the empirical analysis of the specific health and income of adults in China during the 20 years between 1991 and 2011, through general logit and fixed-effect models which can reduce the problem of endogeneity, and comparing two models not only using different dependent variables, but also controlling for different independent variables and treating age differently. Finally, we find that individual income has a positive relationship with health in China, with each 1% increase in income, the odds of health will increase around 0.045%.

II. Literature Review

Grossman（1972）is a famous scholar for study of individual demand for health; he built a demand model of health from the perspective of human capital, and he also took the concept of health human capital and made it model-based. Before this, some studies treated health as part of human capital. Otherwise, his model also clearly distinguishes the difference between health demand and health investment demand. Next, I will outline Grossman's health demand theory from four aspects:

1. Health can be seen as a kind of lasting capital stock, this capital stock can produce health time, which means the total amount of time to be engaged in market and non-market activities. So health capital is different from other forms

of human capital, because they mainly determine the individual's labour productivity; thereby it leads to changes in yield. Health not only determines labour productivity, but it also determines non-market activities.

2. Everyone inherits an initial capital stock of health from their parents at birth, and the health capital depreciates with the increase of age. People can maintain health through health investment. Health demand led to the demand for health investment, and the medical service is one of the basic forms of investment in health; therefore, health service demand is a derived demand.
3. Health has a "shadow price," that is, the cost to maintain a certain health stock. The "shadow price" of health depends on many factors; in addition to the price of medical services, age, income, nutritional intake and sanitation are non-medical factors that also will affect the stock of health.
4. The Grossman model provides three aspects of forecasting: first of all is the effect of age on health demand. The health capital depreciation rate increases as age increases; in time the cost of getting one unit of health capital increases. Therefore, the demand for health capital becomes smaller until health stock is zero. The second is the wage rate effect on health demand. When the wage rate rises, the value of health time increases, which will greatly motivate individuals to invest more time to work. So individual income has an impact on health. Therefore, consumer's demand for health and medical services are positively correlated with the wage rate. Finally, the other human capital (mainly education) effects health demand. With an increase of education, a variety of

input-output efficiencies will improve, including health output, which implies education reduces the marginal cost of health investment. So when wage rates and health capital-output are unchanged, education has a positive effect on health demand.

The advantage of the Grossman health demand model is that it analyzes some key variables, such as education, income, age, influences on the marginal cost of health capital or marginal benefit. The downside of the model is some assumptions are slightly simplistic. For instance, the rate of depreciation of health capital is seen as an exogenous variable, and the depreciation of health is foreseeable, which assumes that individuals have perfect information of their date of death. Moreover, causality-likely runs in both directions: higher income can increase living conditions and environment and therefore increase health; good health also can increase ability to work and therefore increase income. In the Grossman health demand model, health is considered to be the dependent variable: age, gender, level of education, health care and income are considered to be independent variables. The basic equation this paper used in the empirical study is according to the Grossman health production function, which is Health=β_0+ $\beta_1$logincome+ β_2X+u, and this function will be explained more in the methodology section.

Grossman's ground-breaking health demand model has become the standard model for study the health demand; many experts and scholars have used it in their empirical research. There are many factors which can affect health demand, many western experts and scholars use income or wages as influence factors to discuss the

relationship between them. At the same time, there are also many Chinese experts and scholars who use the existing database or questionnaire to study health. The study of health problems cannot be separated from the data collection. Western scholars' studies of health have been conducted for a long time, because the database is more durative and effective than China's health database. On the other side, China's health database is insufficient because the duration of health research is shorter. The existing databases includes China Health and Nutrition Survey (CHNS), Chinese Family Panel Studies (CFPS), and Chinese Health and Retirement Longitudinal Study (CHARLS).

For the research study of the relationship between health and income, Chinese and Western scholars have different points of view. Chinese scholars focus more on specific groups and income inequality. Pan (2007) used the 1997 and 2006 panel data from CHNS, which indicates that income's influence on health has a lagged effect. The influence of the income gap on health is negative correlated, particularly for low-income individuals. Yuan and Zhou (2009) used the nine provinces of China, 2200 over 55-year-old elder people's micro survey data to prove that income level, health status and health insurance have positive influences on subjective happiness. Wang, Jin and Xu (2009) found that migrant workers' income levels have a positive correlation with nutrition and health, which means the higher trophic, health and health consciousness level of migrant workers, the higher income they will have. Yi, Wang and Wang (2011), based on the 2000, 2004 and 2006 CHNS data, used income function and health functions with seemingly unrelated regression (SUR) and

three-stage least squares (3SLS) methods to prove that rural residents' health and income have positive influences on each other. This suggests there is a causal relationship between income and health running in both directions.

The research of western scholars is more general compared with Chinese scholars. Ettner (1996) used the National Survey of Families and Households (NSFH), the Survey of Income and Program Participation (SIPP), and the National Health Interview Survey (NHIS) data, through ordinary and two-stage least squares methods not only to prove that higher income has a significant influence on health, but to show there is some evidence that increases in disposable income reduce alcohol abuse. Ecob and Smith (1999) chose panel data from the Health and Lifestyle Survey, a survey in in England, Wales and Scotland. They mainly consider the relationship between health and income before and after controlling for socio-economic variables. The result is that, with increasing income, people have a better health condition. However, there are diminishing returns at higher levels of income. Paul Frijters, John Haisken-DeNew and Michael A. Shields (2005) used the panel data from German Socio-Economic Panel (GSOEP) between 1984 and 2002. The authors provided new evidence on the causal effect of income and health satisfaction, which indicates higher income improved health satisfaction in both East and West Germans. Howitt (2005) divided the factors affecting personal health level into two aspects: physical and health education. They hold that with the increase of personal health education, people will have a deeper understanding of physiological knowledge which can help people more easily to find their own health problems and choose a suitable healthy

lifestyle in order to improve their health level.

III. Methodology and Data

3.1 Methodology

The main purpose of this study is to analyze the influence of income on health of Chinese people from 1991-2011. Grossman (1972) introduced the concept of the health production function, and also introduced some variables including age, education, income, etc. Therefore, health is not only closely related with medical service, but also influenced by other factors, for instance, living environment, life behaviors, etc. According to the Grossman model, we can study the impact of income on the health of residents by the following production function and it can be associated with both health production and people's demand for health.

$$Health=\beta_0+ \beta_1 logincome+ \beta_2 X+u$$

where health represents the health status of people; logincome represents the individual income in the natural logarithm; X is the observable factors influencing the residents' health status, including education level, chronic disease prevalence, lifestyle, environment, age. u is the error term.

In the study of the relationship between health and income, endogeneity is a problem that must be faced. Estimates may also be biased by missing variables or reverse causality. If not treated, we may get an inconsistent estimate of the results. One of the common ways to solve and avoid the endogeneity problem is to find an instrumental

variable (IV) for income. At the beginning of preparing this paper, we expected to be able to make some breakthroughs in the search for this instrumental variable. However, the effective health instrumental variables are mainly situated at the community level, such as the provision of community health services, the price of daily commodities in the community. But CHNS community level data is confidential, so community data cannot be obtained. Therefore, we mainly use the fixed effect model to address the endogeneity problem. By applying the fixed effect model, some of the unobserved factors that are related to the explanatory variables and which do not change over time are eliminated, which can solve the endogeneity problem to some extent. There are also some disadvantages of controlling endogeneity using fixed effect. The main disadvantages are: this increase the variance of each parameter estimate; another one is we are unable to examine time constant covariates. Because health is the dummy variable, the fixed effect logit model can be used in this paper as in Wooldridge (2002), Jones and Wildman (2008).

With respect to panel data, if the dependent variable is a dummy variable, then we call it panel binary choice model. People usually use a value of 0 or 1; the binary choice variable y_{it} represents the dependent variable. $y_{it}=x_{it}\boldsymbol{\beta}+u_i+\epsilon_{it}$ $(i=1,\cdots,n;\ t=1,\cdots, T)$

$y_{it}=1$ if $y_{it}>0$; $y_{it}=0$ if $y_{it}\leq 0$;

Given $x_{it,}$ $\boldsymbol{\beta}$, u_i, we have:

$$P(y_{it}=1 \mid x_{it,}\ \boldsymbol{\beta},\ u_i) = P(y_{it}>0 \mid x_{it}\ ,\ \boldsymbol{\beta},\ u_i)$$

$$=P(x_{it}\boldsymbol{\beta}+u_i+\in_{it}>0) \mid x_{it}\ ,\ \boldsymbol{\beta},\ u_i)$$

$$=P(\in_{it} > -u_i - x_{it}\boldsymbol{\beta} \mid x_{it}\ ,\ \boldsymbol{\beta},\ u_i)$$

$$= P(\in_{it} < u_i + x_{it}\boldsymbol{\beta} \mid x_{it}\ ,\ \boldsymbol{\beta},\ u_i)$$

$$=F(u_i + x_{it}\boldsymbol{\beta})$$

Whereby F is the cumulative distribution function of $\in_{it,}$ assume the density function of $\in_{it}$ is symmetrical about the origin.

If $\in_{it} \sim N(0, 1)$, it is Probit model

$$P(y_{it}=1 \mid x_{it,}\ \boldsymbol{\beta},\ u_{i)} = \boldsymbol{\Phi}(u_i + x_{it}\boldsymbol{\beta})$$

If $\in_{it}$ obeys the logistic distribution, it is a Logit model:

$$P(y_{it}=1 \mid x_{it,}\ \boldsymbol{\beta},\ u_{i)} = \boldsymbol{\Lambda}(u_i + x_{it}\boldsymbol{\beta}) = e^{ui + xit\boldsymbol{\beta}} / 1+e^{ui + xit\boldsymbol{\beta}}$$

The main estimate methods of the binary choice model include pooled regression, a random effect model and a fixed effect model. If u_i is unrelated with the explanatory variables x_{it}, then the model is a random effects model; if u_i is unassociated with one or more explanatory variables x_{it}, then the model is a fixed effects model. Whether the random effect or fixed effect model should be chosen can be checked by the Hausman test.

The Hausman test is mainly used to check whether the non-observed factors should be dealt with by a fixed effects model or a random effects model. The null hypothesis of the Hausman test is that there are no systematic differences in the coefficients of random effects and fixed effects. When the null hypothesis is true, the random effects model is more effective than the fixed effect model. If the null hypothesis is not true, the random effects model is inconsistent. Therefore, if the null hypothesis is true, the estimated value of FE and RE will converge to the true value of the parameter, but if the gap of the estimated value between FE and RE is too large, it is more inclined to reject the null hypothesis. In this paper, after the Hausman test, the result shows that the null hypothesis has been rejected, and random effect mode is inconsistent. Therefore, the fixed effect model will be used in the empirical study part. The results of the Hausman test are shown in the Appendix.

The Link test can be used to detect whether a model has specification error. The idea of Link test is that if the model is properly specified, one should not be able to find any additional predictors that are statistically significant except by chance. The results of the Link test for two general logit model are shown in the appendix, and the Link test results will be interpreted in the result section.

In addition, each model controls for different variables including: with and without chronic conditions, and smoking and drinking status. Moreover, the independent variable age has been treated differently in across models, separately using a continuous age variable and dummy variables for 10-year age categories.

3.2 Data and Variables

This paper utilizes the longitudinal data from the China Health and Nutrition Survey (CHNS), which is widely used to study health and related issues in China. This survey is jointly conducted by Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety and the Chinese Center for Disease Control and Prevention. The purpose of the survey is to assess the effect of the implementation of Chinese government policy at all levels of health, nutrition and family planning. Using information on family and personal economic conditions, demographic and social factors are used to comprehensively consider the effects on the nutrition and health of the Chinese social and economic transformation.

Up to 2016, the survey was conducted a total of 9 times, in 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011 respectively. The scope of the survey involved nine provinces: Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi and Guizhou. These provinces have large differences in geographical position, economic development level, public resources and health indicators. CHNS adopts the multistage random sampling method to determine the samples. Cities and counties in the nine provinces were stratified by income (low, middle, and high), and a weighted sampling scheme was used to randomly select for cities or counties in each province. From the survey of regional distribution and the method of sampling perspective, the database is representative in the study of people's health in China.

The CHNS data can be divided into three categories: personal survey data, household survey data and community survey data. In personal survey data, adults aged 18 and over were treated differently from children under 18 years of age; questionnaires were set up accordingly. The adult survey included demographic data, work, housework and child care, personal income and consumption status, health status and health service use status, etc. Household survey data include the head of household demographic data, family time allocation, labor force participation, family income and consumption expenditure, family property and sanitation facilities, health care and medical services information, etc. Community survey data mainly include community infrastructure (electricity, water, transport facilities, communication facilities, etc.), community services (family planning, health facilities, retail stores), the population of the community, community general wage level information. In the official website of the CHNS data, only the community survey questionnaires were shown; specific community survey data were not available to the public, which needed to apply for survey institutions to acquire it. Therefore, our study does not include community level data.

In this paper, a total of 59,255 observations and 20653 individuals, around 7500 observations in each year are chosen from CHNS. The sample is unbalanced panel data, so all individuals do not remain in the sample for all years. The age range is 18 years old to 65 years old. 50.49% of them are male, and 49.51% of them are female. In this sample, people who lived in urban area account for 29.69%, people who lived in rural area account for 70.31%. More details will be shown in the summary

statistics.

How to define health, the answers in different fields is not uniform. People usually hold that health means physically there is no disease, but this is only a narrow sense of health. World Health Organization (WHO, 1946) claim that “Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.” Health not only means no defects in physical aspect, but also includes no mental illness.

From the perspective of economics to define health, health is a kind of human capital. Genetics determine the amount of health stock when everyone is born; as age increases, health stock depreciates gradually. However, people may also pursue health through health investments. Therefore, health not only is seen as a consumer product, but it also is seen as an investment product. In short, health is important human capital and one of the important factor of production. To a certain extent, health determines the labor and capital inputs (Fushs, 1982), which play an important role in the development of social economy.

People have difficulty judging which definition is more accurate; the fundamental reason is that we lack a complete indicator system to measure a person's overall health. Health is different from other forms of human capital, which is very difficult to quantify. Therefore, considering the research topic of this paper, how to measure health is the first important problem.

A typical subjective health measure is the self - reported health status (SRH), which

also is one of the most widely used indicators to measure health status. For instance, Grossman (1972), Acton (1975), Benham and Benham (1982), Wolfe and Hill (1995) used this indicator in their studies. Self-reported health does not take an actual measurement of people's health; it is the psychological orientation of respondents themselves, which is the subjective judgment of comparing their health status with others. This SRH is usually derived from the question on "What do you think of your health?" or similar questions; respondents can typically choose from four options: excellent, good, fair or poor.

There is a lot of controversy about using self-reported health status as an indicator of health. The advantages of this indicator are that it is easy to get the data, data quality is high, because the respondents can very easily make a comprehensive evaluation of their health. In addition, the research of Deaton and Paxson (1998) indicates that this index is highly correlated with the mortality and other objective health indicators, which can reflect the individual's health very well. This indicator has also been more questioned, mainly in two aspects: first, the index is a discrete variable, using it as a continuous variable may affect the validity of estimate results. Second, this indicator is highly subjective, the choice made by the respondents will be affected by many factors, for instance, cognitive ability (especially the cognition of medical knowledge). The people who have extensive medical knowledge may have a different standard of health from those who lack medical knowledge. Compared with people who are in poor health, respondents tend to report better health conditions.

While the subjective health measurement method receives more and more challenges and criticism, objective health measurement methods have been widely used in the health economics field.

Those objective indicators mainly include mortality rate, morbidity rate, nutrient intake, body measurement index, activities of daily living (ADL), etc.

The mortality index is one of the most widely used health indicators in the early studies; the advantage is that the data can be acquired easily. But this index is not suitable for the application for survey data, since dead people do not answer surveys, but it is fine in administrative data.

The morbidity rate is usually recorded as a measure of health two weeks or four weeks before the survey. The advantage of a morbidity rate is that the data are more objective and easy to obtain. However, the limitations of this type of health indicator are also evident. This indicator uses sick or not as a basis for the definition of health which is too narrow; it cannot fully reflect the health status of individuals. Schultz and Tansel (1993) use this indicator to study how morbidity affects wage rates in Cote d'Ivoire and Ghana.

Nutrient intake is closely related to health, so it is widely used as an indicator to reflect the short-term health status of individuals. The intake of calories is related to the health of the heavy manual workers, and the protein intake is related to the health of the middle and light manual workers. Many economists use nutritional intake indicators in their research. For instance, Sahn and Alderman (1988), Hass and

Brownlie (2001) studied the relationship between nutrient intake and personal health, and the effect of individual income and wage rate was estimated.

Body measurement is mainly referring to the body mass index (BMI). The body mass index (BMI) is obtained with kilograms of body weight divided by height in meters squared. People generally use this indicator to measure whether a person's weight is excessive. BMI can be used as a measure of a person's medium-term health indicators. Overweight or obesity status cannot fully reflect a person's health condition. Therefore, in practical studies, there are few scholars: Zhang (2003) who use BMI as the only indicator to measure the health status.

Activity of daily living (ADL) is a commonly used measure of dysfunction, which is used in the early research. Due to ADL measuring respondents' specific activity function, the ADL is not only objective, but also effective. However, the ADL also has obvious limitations. Because ADL generally measures the patient and the elder people, ADL has no statistical significance for the healthy individual and young people. Mete and Schultz (2002) used ADL as health indicator did a research about the labor force participation rate of the elderly population in Taiwan. The result indicates that the people who have a better ADL indicator show a higher labor participation rate.

To sum up, the measurement indicators of health are multidimensional because health is multidimensional. A subjective measurement index is easy to obtain, but it is subjective. Objective measurement is more accurate, but it is not comprehensive

enough for health assessment.

Therefore, in the specific research process, the selection of indicators for health measurement depends on the use of the data and the theme of the study.

Various health indicators have their own advantages and disadvantages; the use of a single index as a measure of health standards is not desirable. In this paper, according to the experience of the previous research literature, two indicators will be used to measure the health of individuals: Self-reported health (SRH), and the Morbidity rate of last four weeks.

In the questionnaire of the individual level of CHNS, there are special questions that ask the respondents to self-evaluate their health status. The options were divided into four grades, of which 1 means excellent, 2 means good, 3 means fair, and 4 means poor. This is the source of our data on self-reported health (SRH) variables. In the CHNS questionnaire, the relevant question about self-reported health (SRH) has been removed after the year 2006. Therefore, the study scope of all models with outcomes of self-reported health (SRH) is set to 1991-2006. In this paper 1 means health status is good as indicated by answering excellent or good on the questionnaire, and 0 means health status is not good as indicated by answering fair or poor on the questionnaire. Ettner (1996); van Doorslaer (1997); Adams (2003); Jones and Wildman (2008); Fang and Rizzo (2011) used the same method for measure of individual health status when studying the effects of income on health. Using the same dichotomous health measure facilitates comparison of their study with previous

work. Moreover, this method is also suitable for measurement of another health indicator: health for the last four weeks. Therefore, a dichotomous measure of individual health is adopted in this paper, and assigning a value of one to respondents in good health or better and zero for respondents in fair or worse which same as previous literature did.

Figure 1 indicates that the under fair rate based on SRH had a slight decrease from 23.36% to 21.73% from 1991 to 1993, but from 1993 to 2006, it kept increasing from bottom point 21.73% to the peak point 39.55%; comparing to 1993 the unhealthy rate almost doubled in 2006. Because CHNS is longitudinal, all respondents are older in each survey, which likely explains at least part of the increase.

Figure 2 shows that how people's health changes in different income groups, and income is divided to four groups: bottom level 0-25%, 25%-50% level, 50%-75% level and the top 25% level. We can clearly see that compared to the same year and from lowest income level to highest income level, people's health status is getting better. This suggests that income has a positive relationship with health, which is in line with the expectation.

Figure 1

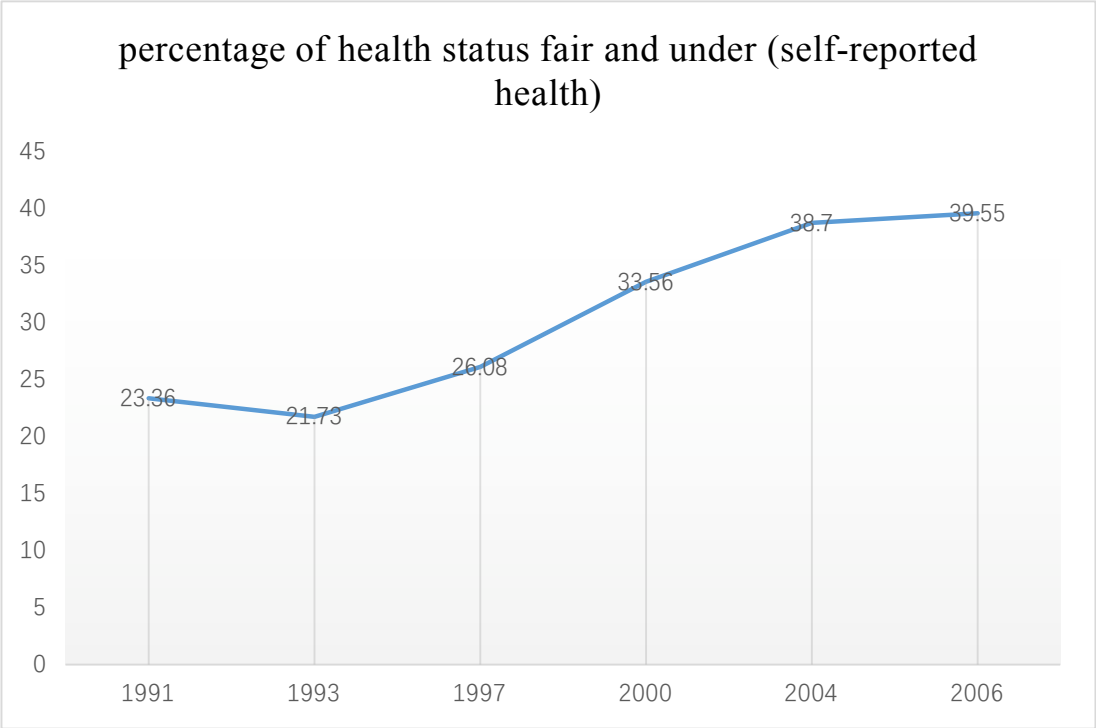
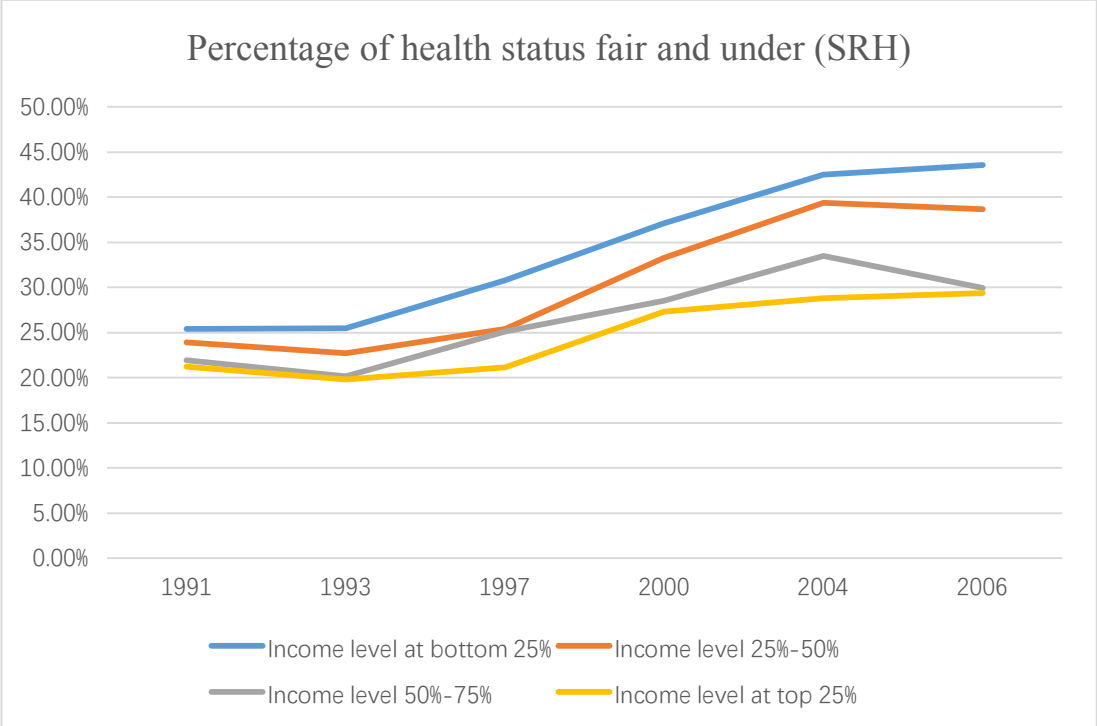


Figure 2



The morbidity rate of the last four weeks is based on the question: “Have you been sick or injured in last four weeks?” The morbidity rate of the last four weeks is more objective than self-reported health and thus is suitable to reflect a person’s health status in a period of time. We set health for the last four weeks as the second health indicator, 1 represents has not been sick or injured in last four weeks; 0 represents has been sick or injured in last four weeks.

Figure 3 clearly shows that the morbidity rate fluctuated from 1991 to 2011. More specifically, the morbidity rate declined from 9.26% to 5.3% and from 15.53% to 13.15% in period of 1991 to 1993 and 2004 to 2006 respectively. It had a sharp increase from 7.18% to 15.53% between 2000 and 2004. From 1993 to 2000 and from 2006 to 2011 morbidity rate had a slight increase. For the overall trend, the morbidity rate increased from 1991 to 2011 in China.

Figure 4 indicates the morbidity rate changes of different income groups. Although in 1991 and 2004 the morbidity rate of the middle income groups, 25%-50% and 50%-75%, are higher than the bottom income class 0-25%, but the main trend sill shows the higher income level people have, the lower morbidity rate people will be, which also same as what we expected.

Figure 3

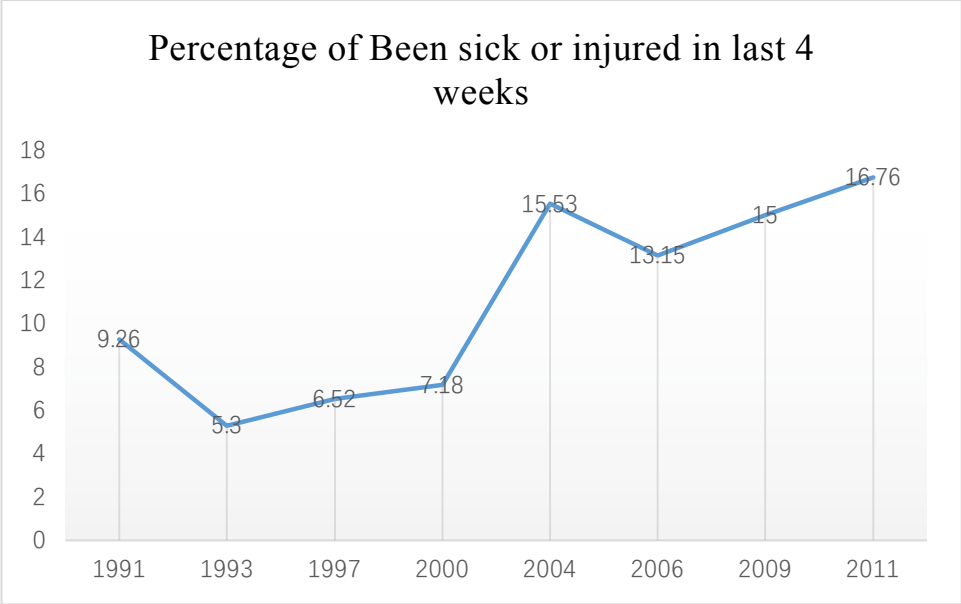
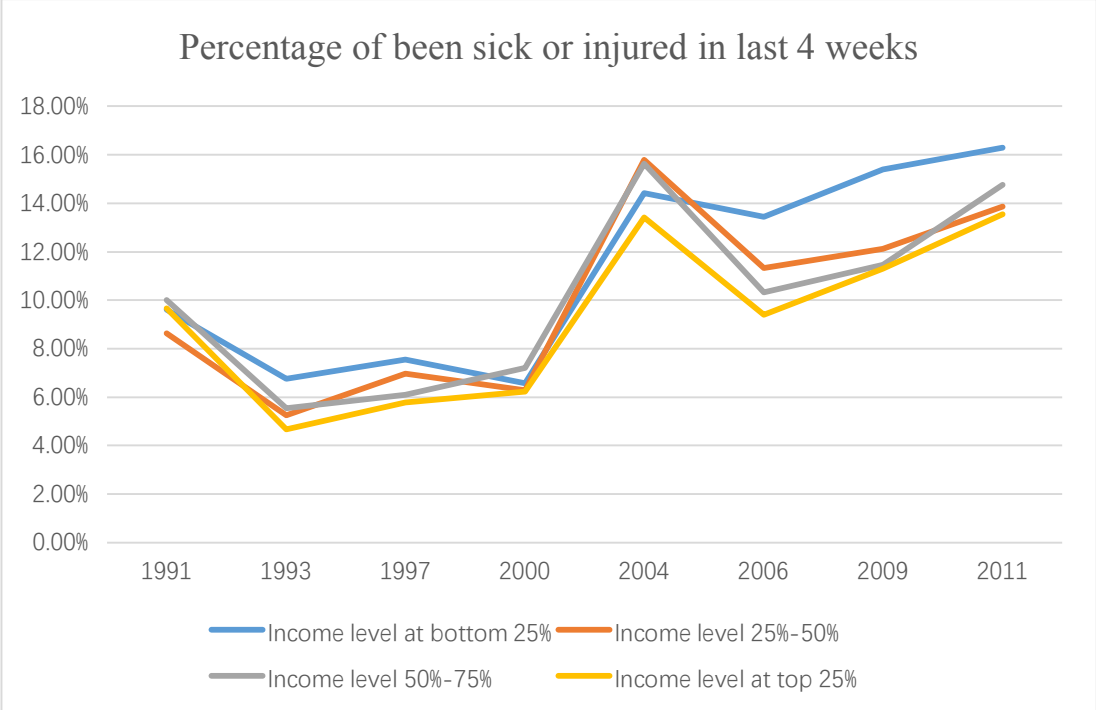
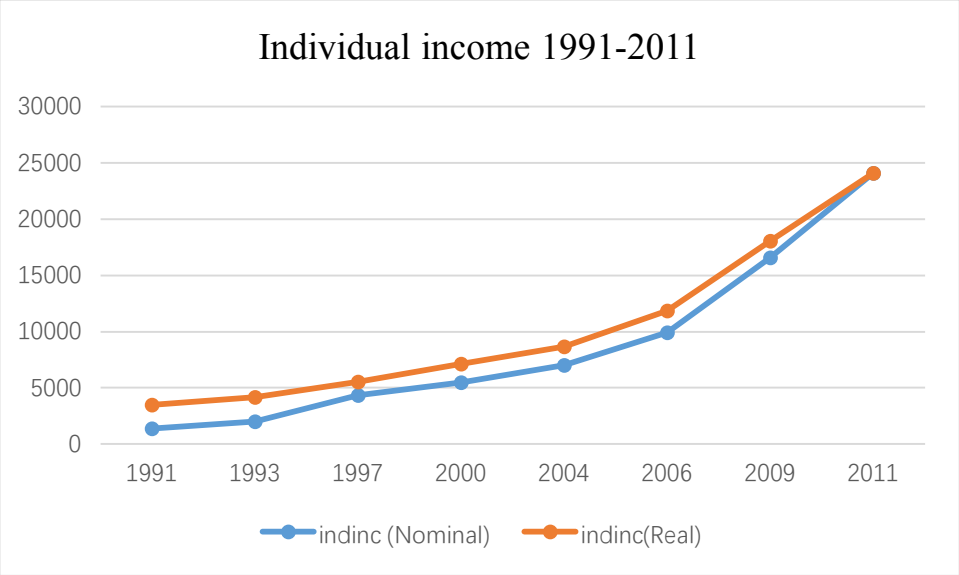


Figure 4



The net individual income is chosen from the CHNS which include wages, bonuses, subsidies, farm income and business income. And in this paper, the net individual income is adjusted according to the inflation rate in different years, so the real annual income of the individual is obtained. In Table 5, from 1991 to 2011, the real annual income of Chinese residents has seen an almost fivefold increase. Importantly, in the CHNS questionnaire, the question about individual income is "How much is your total income last year?" Therefore, this paper actually already studies previous one-year income impact on health this year. So, the lag variable of income has not been added in the regression model.

Figure 5



In this paper, 18-65 years old residents were selected as the study object. This is mainly based on the following two points: firstly, the questionnaire for adults over 18 years of age and children under the age of 18 have large differences, and most of the

people under the age of 18 are still students, which means they do not have personal income. Therefore, the residents who under 18 years old are not suitable for our study. Secondly, we set the maximum age at 65 years. Although the legal retirement age of urban male employees is 60 years old, in rural areas there is no strict retirement age, and many rural residents over 60 years of age are still engaged in agricultural or other production activities. Considering that aging has a negative impact on health, the maximum age of the research subjects was set at 65 years. In order to treat age differently than continue age, we divided age into five categories, which are 18 to 25, 26 to 35, 36 to 45, 46 to 55 and 56 to 65.

In this paper, we also selected the following explanatory variables: gender, education, living habits (smoking, drinking), the chronic disease index, living environment (whether use tap water).

Male is given the value 1, female given the value 0. Years of education means the completed years of formal education in regular school. Smoking is given the value 1, non-smoking is given the value 0. Drinking is given the value 1, non-drinking is given the value 0. This paper uses whether diagnosed with any of hypertension, diabetes, myocardial infarction, or stroke as a chronic disease index. Having being diagnosed with one or more of these diseases is represented by 1. Uses tap water is represented by 1, use other water source is represented by 0. In China, tap water is considered worse than bottled water but better than well water and river water, and is therefore most likely to be drunk by middle-income individuals. Occupation is based

on skill-level divided into three categories: occupation1 includes senior professional workers; junior professional workers; administrators, executives and managers; army police officers. Occupation2 includes office staff; skilled workers; ordinary soldiers, policemen; drivers; service workers; athletes, actors and musicians. Occupation3 means farmers, fishermen, hunters and other labourers. And all variables will be shown in the Table 1 below.

Table 1 Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
health for the last four weeks	58800	0.900272	0.299639	0	1
self-reported health	42979	0.716908	0.450506	0	1
age	59255	41.881060	12.30707	18	65
age18to25	59255	0.107180	0.309345	0	1
age26to35	59255	0.210345	0.407557	0	1
age36to45	59255	0.266340	0.442048	0	1
age46to55	59255	0.238883	0.426405	0	1
age55to65	59255	0.177251	0.381884	0	1
ln real income	59255	8.387481	1.738472	0	13.4805
male	59255	0.504886	0.499980	0	1
years of education	57021	7.820891	4.372689	0	19
urban	59255	0.296853	0.456875	0	1
chronic	59255	0.064737	0.246064	0	1
tapwater	59007	0.699426	0.458512	0	1
smoking	55609	0.339226	0.473451	0	1
drinking	55370	0.382138	0.485915	0	1
occupation1	59255	0.109105	0.311773	0	1
occupation2	59255	0.336140	0.472392	0	1
occupation3	59255	0.422378	0.493942	0	1

Table 2. General logit Regression results.

Dependent Variable=health for the last four weeks						
	Base model		Controlling for different variables		Treating age differently	
Variables	**OR**	**p-values**	**OR**	**p-values**	**OR**	**p-values**
male	1.200949***	0.000	1.243313***	0.000	1.19184***	0.000
age	0.96282***	0.000	0.971198***	0.000		
age 26-35					0.5564669***	0.000
age 36-45					0.4206779***	0.000
age 46-55					0.3059773***	0.000
age 56-65					0.1957756***	0.000
ln real income	1.044444***	0.000	1.050091***	0.000	1.044615***	0.000
smoking			1.011815	0.781		
drinking			0.912917**	0.015		
chronic			0.3410859***	0.000		
tapwater	0.9641605	0.313	0.9876672	0.739	0.9570431	0.224
years of education	1.001776	0.676	1.005827	0.182	1.003886	0.359
urban	0.8374905***	0.000	0.8820343***	0.001	0.8303776***	0.000
occupation1	1.008488	0.879	1.016885	0.770	1.010877	0.847
occupation2	1.029384	0.441	1.000121	0.997	1.044353	0.248
y1993	1.826024***	0.000	1.822009***	0.000	1.819211***	0.000
y1997	1.657297***	0.000	1.694505***	0.000	1.636772***	0.000
y2000	1.623193***	0.000	1.613103***	0.000	1.607367***	0.000
y2004	0.7368123***	0.000	0.7607882***	0.000	0.7253763***	0.000
y2006	1.058029	0.342	1.107653	0.094	1.031759	0.598
y2009	0.9207915	0.157	0.991199	0.883	0.9049026	0.086
y2011	0.8016368***	0.000	0.8806398**	0.027	0.7872889***	0.000

Notes: *Statistically significant at the 10% level; **statistically significant at the 5% level; *** statistically significant at the 1% level.

Table 3. General logit Regression results.

Dependent Variable=self reported health						
	Base model		**Controlling for different variables**		**Treating age differently**	
Variables	**OR**	**p-values**	**OR**	**p-values**	**OR**	**p-values**
male	1.300234***	0.000	1.162893***	0.000	1.290196***	0.000
age	0.9599472***	0.000	0.9631165***	0.000		
age 26-35					0.6688842***	0.000
age 36-45					0.4752502***	0.000
age 46-55					0.323571***	0.000
age 56-65					0.2052285***	0.000
ln real income	1.047906***	0.000	1.054042***	0.000	1.046649***	0.000
smoking			1.022182	0.518		
drinking			1.167204***	0.000		
chronic			0.4188731***	0.000		
tapwater	0.9572383	0.104	0.9566365	0.109	0.9490729	0.052
years of education	1.014592***	0.000	1.015266***	0.000	1.017866***	0.000
urban	0.9121472***	0.004	0.9496992	0.113	0.9006529***	0.001
occupation1	1.218449***	0.000	1.207362***	0.000	1.205143***	0.000
occupation2	1.234387***	0.000	1.217065***	0.000	1.248228***	0.000
y1993	1.068097*	0.095	1.086168*	0.049	1.06239	0.125
y1997	0.914075**	0.022	0.9360916	0.103	0.89876***	0.006
y2000	0.7222864***	0.000	0.7510429***	0.000	0.7118028***	0.000

y2004	0.6476749***	0.000	0.6763014***	0.000	0.632805***	0.000
y2006	0.6770902***	0.000	0.7095735***	0.000	0.653863***	0.000

Notes: *Statistically significant at the 10% level; **statistically significant at the 5% level; *** statistically significant at the 1% level.

Table 4. Fixed-effect Logit Regression results.

Dependent Variable=health for the last four weeks						
	Base model		**Controlling for different variables**		**Treating age differently**	
Variables	**OR**	**p-values**	**OR**	**p-values**	**OR**	**p-values**
age	0.9367964***	0.000	0.9466068***	0.000		
age 26-35					0.6031213***	0.000
age 36-45					0.4006881***	0.000
age 46-55					0.2397223***	0.000
age 56-65					0.1220465***	0.000
ln real income	1.064672***	0.000	1.068257***	0.000	1.050743***	0.000
smoking			0.8937619	0.188		
drinking			0.8098681***	0.000		
chronic			0.4103813***	0.000		
tapwater	1.011514	0.854	0.9895049	0.870	0.9392991	0.306
occupation1	0.869041	0.176	0.8370853	0.101	0.8887484	0.259
occupation2	0.913677	0.143	0.9094379	0.140	0.9219382	0.192

Notes: *Statistically significant at the 10% level; **statistically significant at the 5% level; *** statistically significant at the 1% level.

Table 5. Fixed-effect Logit Regression results.

Dependent Variable=self reported health						
	Base model		**Controlling for different variables**		**Treating age differently**	
Variables	**OR**	**p-values**	**OR**	**p-values**	**OR**	**p-values**
Age	0.9121764***	0.000	0.9184238***	0.000		
age 26-35					0.432746***	0.000
age 36-45					0.2357395***	0.000
age 46-55					0.1189049***	0.000
age 56-65					0.0568238***	0.000
ln real income	1.035509***	0.000	1.040464***	0.000	1.025***	0.010
smoking			0.9690017	0.634		
drinking			0.9823237	0.709		
Chronic			0.4053173***	0.000		
Tapwater	1.090197	0.086	1.072197	0.181	0.9604351	0.411
occupation1	1.183067	0.056	1.156864	0.114	1.269046	0.007
occupation2	1.099546	0.069	1.075945	0.182	1.134907	0.015

Notes: *Statistically significant at the 10% level; **statistically significant at the 5% level; *** statistically significant at the 1% level.

IV. Results

First of all, the Link test results (in appendix) indicate whether health for the last four weeks or self-reported health as the outcome, the variable _hat is a statistically significant predictor and the variable _hatsq is not a statistically significant predictor. More specifically, the fact that the variable _hat is statistically significant means the general logit model is not misspecified; variable _hatsqnot being statistically significant indicates we have not omitted relevant variable(s) or our link function is correctly specified. In short, the Link test results suggest that the logit model used in this paper has no specification error.

Table 1 shows the summary statistics of all samples for this paper. Following the examples in the literature, self-reported health (SRH) and health rate for the last four weeks will be used in two different outcomes in our model. Table 2 indicates the results with health rate for the last four weeks as the outcome with the general logit estimation model from 1991 to 2011. Table 3 shows the result with self-reported health (SRH) as the outcome with general logit estimation model from 1991 to 2006. Considering the endogeneity problem in the logit model and after the Hausman test of the random effect logit model and fixed effect model, the fixed effect model has been chosen to help adjust the endogeneity problem. The results of the Hausman test are shown in the appendix. Table 4 and Table 5 indicate the result of the health rate for the last four weeks and self-reported health (SRH) as the outcome with the fixed-effect model from 1991 to 2011 and 1991 to 2006 respectively. In addition, in

each table except Table 1, we also run the other two regressions which are controlling for different variables and treating age differently for comparison with the base model.

Table 1 indicates that the average age is 41.88 for the individuals in the sample; 50.49% of the individuals are male; the average years of education is 7.8; 29.6% of individuals live in urban areas and 69.9% of them drink tap water.

Table 2 shows the logit estimation; variation comes from differences across individuals, and health for the last four weeks as dependent variable is used in the model.

The base model of Table 2 shows that the coefficient for males is statistically significant at the 1% level and positively correlated with health. This indicates that the odds of health for the last four weeks for males are 20.1% higher than for females. This may be because females not only have a lot of pressure from their work, but also the majority of female have to take care of the children and families. The coefficient on age is statistically significant at the 1% level and negatively correlated with health. This shows that with each year an individual age increases, the odds of health for the last four weeks decrease 3.7%. This is because with the increase in age, all aspects of personal function will be decreased, which can decrease the odds of an individual being healthy. The coefficient of urban is statistically significant at the 1% level. This indicates that the odds of health for the last four weeks of people who live

in urban areas 16.3% lower than people who live in rural areas. A possible reason for this result is that urban pollution is much more serious than that of the rural areas in China, for instance, haze and water pollution. On the other hand, there may be more stress associated with living in urban China. Excessive stress also is harmful for health.

For our most important variable, income, the log of real income is statistically significant at the 1% level and positively correlated with health. It means that if an individual's income for last year increases by 1%, the odds of health for the last four weeks will increase by 0.043% (the tables contains coefficients based on a 1 unit change in *natural log* of real income; in the text, we convert these to a 1% change in real income using the process shown in the Appendix).

In order to observe the influence of human behavior on health, three more independent variables which are chronic, drinking and smoking are added into the base model of Table 2. The results of male, age, urban and log of real income have no significant difference with the base model. For added variables, chronic is statistically significant at the 1%, which means the odds of health for the last four weeks of people who have chronic diseases are 65.9% lower than people who have no chronic diseases. Drinking is statistically significant at the 5% level and has a negatively correlated relationship with health. It shows that odds of health for the last four weeks for drinking people are 8.7% lower than non-drinking people. Smoking

has no statistically significant effect on the health.

If we treat age as 5 separate categories, the results of male, urban and log of real income almost same as the what base model shows. The results of age categories indicate that compare the age group 18-25, the health for the last four weeks is getting worse from age group 26-35 to 56-65.

The logit model and self-reported health as outcome were used in Table 3. For the base model in Table 3, it shows that the coefficient of males is statistically significant at the 1% level. This indicates that the odds of excellent/ good health for males are 30% higher than for females. The coefficient on age is statistically significant at the 1% level and negatively correlated with health. This shows that with each year individual age increases, the odds of excellent/ good health decrease 4%. The coefficient for the year of education is positively correlated with health and is statistically significant at the 1% level. This shows that the odds of excellent/ good health increase 1.4% for each additional year of schooling. The coefficient of urban is statistically significant at the 1% level. This indicates that the odds of excellent/ good health of people who live in urban areas are 8.8% lower than for people who live in rural areas.

The log of real income is statistically significant at the 1% level and positively correlated with health. The odds of excellent/ good health will increase by 0.047% for each 1% increase in an individual's income for last year.

For the occupation categories, occupation1, which represents senior professional

workers; junior professional workers; administrators, executives and managers is statistically significant at the 1% level. It shows that the odds of excellent/ good health for occupation1 are 21.8% higher than for occupation3 which includes hunter, fisher, farmer and other labourers. Occupation2 is statistically significant at the 1%. It indicates that the odds of excellent/ good health for occupation2 is 23.4% higher than for occupation3.

If we control for different variables from the base model in Table 3, the results for all variables (except for the added variables): chronic drinking and smoking are almost same as base model. The variable chronic is statistically significant at the 1% level, which means the odds of excellent/ good health about people with chronic diseases are 58.1% lower than people who have no chronic diseases. The indices of living behavior: smoking, drinking, smoking has no significant effect on self-reported health across individuals. Drinking is statistically significant at the 1% level and is positively correlated with self-reported health. It shows that odds of excellent/ good health for drinking people 16.7% higher than non-drinking people. This is probably a selection effect, because people who are in good health may be more likely to drink; and people who earn more are may be able to afford more for drinking.

If we treat age as a categorical variable, the results on other variables are about the same as before changing the age variable. The results of age categories indicate that compared to the age group 18-25, self-reported health is getting worse from age

group 26-35 to 56-65.

When we compare the base model of Table 2 and Table 3, it clearly indicates that the log of real income in last year for two regressions is both statistically significant at the 1% level, and positively correlated with two different outcomes. With the increase of 1% in individual income in last year, the odds of health for the last four weeks increase by 0.043% and the odds of excellent/ good health increase by 0.047%. Except for the health indicators, there also have the similar results in male, age, and urban aspects. But there are big differences in the variable for years of education and occupation.

More specifically, the results of the base model in Table 2 (morbidity/health rate in last 4 weeks) indicates that the year of education, occupation1 and occupation2 are not significant. However, the results of the base model in Table 3 show that the years of education variable is statistically significant at the 1% level and it has a positively correlated relationship with self-reported health; occupation1 and occupation2 are statistically significant at the 1% level and have a better health status than for occupation3.

The likely reason is that health rate for the last four weeks is the short-run objective evaluation of health, and self-reported health (SRH) is the long-run subjective evaluation of health.

The correlation between years of education and health for the last four weeks is not

significant because the evaluation time is too short, and not many people in the sample are still gaining education. It's time-invariant for most people who have finished school. For instance, an individual may suddenly catch a cold or SARS which may have no relationship with people's education level. Occupation1 and occupation2, which represent skilled workers, are significant because long term high intensity physical work is not beneficial to health.

Table 4 and Table 5 show the results of fixed-effects logit estimates for the base model, the base model controlling for additional variables and base model treating age as separate categories. The variables of male, urban and the years of education are eliminated because they are non-time varying. Table 2 and Table 4 use health rate for the last four weeks as the outcome, Table 3 and Table 5 use self-reported health as the outcome.

The results of the base model in Table 4 clearly show that the coefficient on age is statistically significant at the 1% level and negatively correlated with health. This shows that with each year an individual age, the odds of health for the last four weeks decrease by 6.3%. The coefficient for log of real income is statistically significant at the 1% level and positively correlated with health. The odds of health for the last four weeks will increase by 0.063% for each 1% increase in an individual's income for last year.

When controlling for additional variables (Column 2), the results on other variables

are about the same as before adding the variables. Drinking is statistically significant at the 1% level and has a negatively correlated relationship with health. It shows that the odds of health for the last four weeks for drinking people are 19.0% lower than for non-drinking people. Smoking has no significant effect on health. Chronic is statistically significant at the 1% level, which means that the odds of health for the last four weeks for people with chronic diseases are lower by 59.0% than for the people without chronic diseases.

When age categories replace continuous age (Column 3), the results on other variables are about the same as before changing the age variable. The results of age categories indicate that compared to the age group 18-25, health for the last four weeks is getting worse from age group 26-35 to 56-65.

In Table 5, self-reported health is the outcome; the results of the base model indicate that the coefficient on age is statistically significant at the 1% level and negatively correlated with health. This shows that with each year individual age increases, the odds of excellent/ good health decrease by 8.7%. The coefficient for log of real income is statistically significant at the 1% level and positively correlated with health. The odds of excellent/ good health will increase by 0.035% for each 1% increase in an individual's income for last year.

When controlling for additional variables (Column 2), the results on other variables are about the same as before adding the variables. Drinking and smoking have no significant effect on health within the fixed effect model. The one reason may be just

few individuals changed their drinking and smoking status over time in the sample. Chronic is statistically significant at the 1% level, which means the odds of excellent/ good health for people with chronic diseases are lower by 59.5%.

If age categories replace age in the base model, the results indicate that compared to the age group 18-25, self-reported health is getting worse from age group 26-35 to 56-65. The coefficient of the log of real income is statistically significant at the 1% level and positively correlated with health. The odds of excellent/ good health will increase by 0.025% for each 1% increase in an individual's income for last year.

V. Conclusions

This paper makes an empirical study of the influence of individual income on the health of Chinese residents by using data from the CHNS database; health for the last four weeks and self-reported health were selected as health measures. Logit regression and fixed-effect logit regression were used to estimate the effect of income on health.

The main finding is that whether health for the last four weeks or self-report health is used as the health indicator, individual income is positively associated with health. More specifically, in contrast, the coefficients of the base model for Table 2 (logit for health in last 4 weeks) and Table 4 (logit adding individual fixed effects) are similar, and the coefficients of Table 3 (logit for self-reported health) and Table 5 (logit with

fixed effects) are similar. For the most important income variable of the base model in Table 4, we find that the log of real income is statistically significant at the 1% level and positively correlated with health, which means the odds of health for the last four weeks will increase by 0.063% for each 1% increase in an individual's income for last year. In Table 5, for base model the log of real income is statistically significant at the 1% level and positively correlated with health, which means the odds of excellent/ good health will increase by 0.035% for each 1% increase in an individual's income for last year.

After we use fixed-effects logit models to help adjust for the endogeneity problem in estimating these relationships, for health for the last four weeks as outcome we found that the coefficient of the log of real income for base model in Table 2 and Table 4 with logit regression and fixed-effect logit regression, are both statistically significant at the 1% level and positively correlated with health. The odds of health for the last four weeks will increase by 0.043% and 0.063% respectively for each 1% increase in an individual's income for last year. With self-reported health as the outcome, we found that the coefficient of the log of real income of the base model in Table 3 and Table 5 with logit regression and fixed-effect logit regression, they both are statistically significant at the 1% level and positively correlated with health. The odds of excellent/ good health will increase by 0.047% and 0.035%, respectively, for each 1% increase in an individual's income for last year.

The results of logit regression and fixed-effect logit regression are very close, there may have two following reasons: firstly, the limitation of data. The unbalanced panel data paper used has a total of eight survey years, but because of attrition, individuals only appear in an average of three survey years. This could be one reason fixed-effect logit model can not adjust endogeneity problem efficiently. Secondly, the reason for the endogeneity in this paper may not be mainly because omitted variables are associated with other explanatory variables, but mainly owing to explained variable and explanatory variable have a causal relationship. If so, instrumental variable estimation can adjust the problem of the endogeneity more effectively than the panel fixed effect model.

Moreover, after fixed effect logit estimations for the base model of Table 4 and Table 5, although they use different outcomes to represent health, the results are also very close, which indicates the odds of health will increase by 0.063% and 0.035% respectively for each 1% increase in an individual's income for last year.

The results have no significant difference whether we use different regression method or use different outcomes or control for different variables. Thus, to synthesize the above results, individual income has a positive relationship with the health in China. The odds of health, which are measured by self-reported health and health for the last four weeks, will increase around 0.045% for each 1% increase in an individual's income for last year.

There is a very close and complex relationship between a country's more profound economic growth and the health of its citizens. According to the conclusions of this study, we can offer the following suggestion regarding policy implications.

The government should keep developing national economy and raising the income of residents. From the results, it is found that the individual annual income is statistically significant and has a positive relationship with the health status. Assuming this to be a causal relationship, the government should continue to maintain steady economic growth, develop a variety of incentives on income, to guide the increase in individual income. At the same time, the government should also pay attention to adjusting the income gap between residents and improve the income distribution system, because the results of Pan (2007) indicates that income gap have negative relationship with people's health.

Although this paper has obtained main conclusions and suggestions, it still has some shortcomings, which mainly appear in the following aspects:

First, the limitations of the research method. As we showed in the methodology part, in the study of the relationship between health and income, the endogeneity is a problem that must be faced. An efficient way to help adjust the endogeneity problem is to find an instrumental variable (IV) for health. Because the effective health instrumental variables are mainly concentrated at the community level, such as the provision of community health services and the price of daily commodity in the

community. But CHNS community level data is confidential, so community data cannot be obtained. Therefore, this paper uses fixed-effect logit models to eliminate some non-time varying unobserved factors related to explanatory variables, which can help us get more consistent estimate results.

Second, the limitations of research data. The data used in the empirical analysis are based on China Health and Nutrition Survey (CHNS), the panel data from 1991-2011 have been selected. However, since the year of 2009, the CHNS questionnaire no longer includes the question about self-reported health, so in the model which uses self-reported health as the outcome, data is narrowed down to 1991-2006. And the data are used in this paper is unbalanced panel data, total are 59,255 observations and 20653 individuals. Because of the attrition, every individual only appears in three survey years on average.

Thus, the conclusion of this paper may be affected by the data to some extent, which may make the model have some degree of inaccuracy. These are the deficiencies of the study, which additional time and resources can ameliorate.

VI. References

Acton, J. P. (1975). Nonmonetary factors in the demand for medical services: Some empirical evidence. *Journal of Political Economy, 83*(3), 595-614.

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of econometrics, 112*(1), 3-56.

Benham, L., & Benham, A. (1982). Employment, earnings, and psychiatric diagnosis. In *Economic aspects of health* (pp. 203-220). University of Chicago Press.

China National Bureau of Statistics (2015). The national economic and social development statistical bulletin. http://www.stats.gov.cn/tjsj/zxfb/201602/t20160229_1323991.html

Deaton, A. S., & Paxson, C. H. (1998). Aging and inequality in income and health. *The American Economic Review*, *88*(2), 248-253.

Ecob, R., & Davey Smith, G. (1999). Income and health: What is the nature of the relationship? *Social Science & Medicine, 48*(5), 693–705.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics, 15*(1), 67–85.

Frijters, P., Haisken-DeNew, J. P., & 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. (1982). *Economic aspects of health*. University of Chicago Press.

GAN, C. H., & ZHENG, R. G. (2009). An Empirical Study on Change of Industrial Structure and Productivity Growth since the Reform and Opening-up——A Test for the Structure-bonus Hypotheses from 1978 to 2007 in China [J]. *China Industrial Economics*, *2*, 009.

Grossman, M. (1972). The demand for health: a theoretical and empirical investigation. *NBER Books*.

Hass J E, Brownlie IV T (2001). Iron deficiency anemia (IDA): A review (2016). *International Journal of Science and Research (IJSR), 5*(4), 1999–2003.

Howitt (2005). *Health and economic growth: Findings and policy implications.* Cambridge, MA: MIT Press.

Jones, A. M., & Wildman, J. (2008). Health, income and relative deprivation: Evidence from the BHPS. *Journal of Health Economics, 27*(2), 308–324.

Liu, Y. (2004). China's public health-care system: facing the challenges. *Bulletin of the World Health Organization*, *82*(7), 532-538.

Mete, C., & Schultz, T. P. (2002). Health and labor force participation of the elderly in Taiwan. *Yale University Economic Growth Center Discussion Paper*, (846).

Naughton, B. (2007). *The Chinese economy: Transitions and growth*. MIT press.

Pan sisi (2007). *Health human capital's impact on the income of urban and rural residents in China.* Zhejiang University Press

Sahn, D. E., & Alderman, H. (1988). The effects of human capital on wages, and the determinants of labor supply in a developing country. *Journal of Development Economics, 29*(2), 157–183.

Schultz, P. T., & Tansel, A. (1993). *Measurement of Returns to Adult Health; Morbidity Effects on Wage Rates in Cote d'Ivoire and Ghana.* Washington, D.C.: World Bank Publications.

van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U.-G., Gerfin, M., Winkelhake, O. (1997). Income-related inequalities in health: Some international comparisons. *Journal of Health Economics, 16*(1), 93–112.

Wang zhigang, Jin jingshu & Xu xiao (2009). Research on the influence of the level of nutrition and health on the wage income of migrant workers. *Demographic Journal*, *6*(5), 24-30.

Wolfe, B. L., & Hill, S. C. (1995). The effect of health on the work effort of single mothers. *The Journal of Human Resources, 30*(1), 42.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

World Health Organization. (1975). *Official records of the World Health Organization* (No. 227-228). United Nations, World Health Organization, Interim Commission.

Yi qingshuang, Wang wei & Wang peng (2011). *Analysis on the cycle effects of the income of rural residents and their health in China.* Southwestern University Press.

Yuan shouwei & Zhou shaofu (2009). The influence of income health and medical insurance on the happiness of the elderly. *Journal of public management, 15*(4), 100-127.

Zemin, J. (2003). Build a Well-off Society in an All-round Way and Create a New Situation in Building Socialism with Chinese Characteristics Report to the 16th National Congress of the Communist Party of China (Excerpted)[J]. Chinese Translators Journal, 1, 026.

Zhang chewei (2003). Nutrition, health and efficiency - The evidence from rural of China. *Economic Research, 13*(8), 3-12

Appendix

The result of the Hausman test.

1. Health for the last four weeks as the outcome of the logit fixed effect regression.

```
. hausman fe re
```

	Coefficients			
	(b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
age	-.0652894	-.0487824	-.0165069	.0031004
lnreal_inc	.0626671	.0149361	.047731	.0075572
tapwater	.0114485	-.1185948	.1300433	.0466016
occupation1	-.140365	.0296463	-.1700113	.0847554
occupation2	-.0902782	.0438785	-.1341567	.0451486

b = consistent under Ho and Ha; obtained from xtlogit
B = inconsistent under Ha, efficient under Ho; obtained from xtlogit

Test: Ho: difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 59.83
Prob>chi2 = 0.0000

2. Self reported health as the outcome of the logit fixed effect regression.

```
. hausman fe re
```

	Coefficients			
	(b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
age	-.0919219	-.0538255	-.0380964	.0028482
lnreal_inc	.0348936	.0321844	.0027092	.0059264
tapwater	.0863589	-.0962867	.1826456	.0394608
occupation1	.1681106	.3188319	-.1507212	.0719249
occupation2	.0948972	.2436475	-.1487503	.0390446

b = consistent under Ho and Ha; obtained from xtlogit
B = inconsistent under Ha, efficient under Ho; obtained from xtlogit

Test: Ho: difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 200.61
Prob>chi2 = 0.0000

The result of the Link test.

1. Health for the last four weeks as the outcome of the general logit model.

```
. linktest, nolog

Logistic regression                                     Number of obs   =      56390
                                                        LR chi2(2)      =    1852.05
                                                        Prob > chi2     =     0.0000
Log likelihood = -17552.175                             Pseudo R2       =     0.0501
```

health4	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.7939867	.1540737	5.15	0.000	.4920077	1.095966
_hatsq	.0477669	.0353077	1.35	0.176	-.0214349	.1169687
_cons	.2053505	.1602319	1.28	0.200	-.1086984	.5193993

2. Self reported health as the outcome of the general logit model.

```
.
. linktest, nolog

Logistic regression                                     Number of obs   =      40731
                                                        LR chi2(2)      =    3139.49
                                                        Prob > chi2     =     0.0000
Log likelihood = -22825.417                             Pseudo R2       =     0.0643
```

healthSRH	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	1.018636	.0490264	20.78	0.000	.9225457	1.114726
_hatsq	-.0108411	.0263514	-0.41	0.681	-.062489	.0408067
_cons	-.0039643	.0216517	-0.18	0.855	-.0464008	.0384722

The process of we explain odds ratio for variable log of real income.

Logit functional form is: Ln(odds)=$\boldsymbol{\beta}_0$+ $\boldsymbol{\beta}_1$ln(income)+$\boldsymbol{\beta}_2$X

exp($\boldsymbol{\beta}\hat{}_{\ln(income)}$=1.04444. (The odds ratio of ln real income in Table 2)

$\boldsymbol{\beta}\hat{}_{\ln(income)}$=ln (1.04444) =0.043

$\frac{\%\Delta odds}{\%\Delta income}$=0.043

CURRICULUM VITAE

Candidate's Full Name: Wudi Yang

University Attended: Bachelor of Economics,
Henan University, 2012.7.

Publications: N.A.

Conference Presentations: N.A.