The Effect of Health-Related Behaviors on Income in China: A Panel Data Analysis

Muyan Xie (1756314803@qq.com)
N/A

Research Article

Keywords: Income factors, health status, health-related activities, gender

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-936625/v1

The Effect of Health-Related Behaviors on Income in China: A Panel Data Analysis

Muyan Xie

Abstract

Understanding the interaction between habits and income can be an efficient way to economically improve income and this paper serves to figure out the effectiveness of health-related income factors and gender differences in these factors using regressions, interaction effects and fixed effect model estimations. I used China Health and Nutrition Survey (CHNS) for 2000, 2004, 2006, 2009, 2011 and 2015 and expected the study to be representative of the general population in China.

Results show clear effect of these factors, with that smoking can negatively associate with income and tea, coffee, alcohol, and exercising be in positive association. Also, gender differences are tested and interpreted. Extensions include discuss pathway through which these factors influence income, urban residence's interaction effect with these factors and yearly differences.

Main findings are concluded for future possible uses, as policy makers for women's economical empowerment programs may refer to the statistics and find the area that return the most such as drinking tea and quitting smoking.

Keywords: Income factors, health status, health-related activities, gender

Contents

1 Introduction **2**

2 Literature Review **3**

3 Research Questions **4**

4 Methodology **5**
- 4.1 Data . . . 5
- 4.2 Measures . . . 8
- 4.3 Methods . . . 10
  - 4.3.1 OLS Model . . . 10
  - 4.3.2 Fixed Effects Model . . . 10
  - 4.3.3 Interaction Effects . . . 11

5 Results **11**

6 Extension **17**

7 Conclusion **23**
- 7.1 Advantages and Limitations . . . 23
- 7.2 Main Findings . . . 23

8 Acknowledgement **27**

1 Introduction

Throughout history there is evidence for income inequalities for females.[1] Some people may count the difference on the fact that women tend to work in different industries and have less experience than men, while others may sum that the difference might be caused by discrimination against women and fewer hours working at work.[2] [3]

Also, what is worth noticing is that thorough researches have been conducted on factors that might influence income, including education and health-related activities.[4] [5] Naturally, people with living habits like refusing to smoke are found to earn more as a social award from the society.[6] In addition to catering to social expectation, people with better habits can create themselves a stronger body and thus the superior health status may lead them to be better qualified to work more efficiently for longer time.

Quantitatively analyse the extend to which some factors like health, schooling and habits can have on influencing income is useful. Also, intuitively, inspired by concluded sex differences in immune responses, which lead females to respond more violently to several types of illness, I designed to investigate deeper into different effectiveness of income factors across gender, trying to figure out the sex's degree of influence on income when we divide the determinant of income into several smaller categories.[7] It is of great value if I can figure out if there is differences across genders in the earning effects of improvement of habits, education or occupations, if we are interested in economical empowerment for women and are trying to find a way to most efficiently address the problem by investing in factors that are most crucial in influencing income.[8]

Thus, this paper focus on analysing effect of each income factor both for the general working population and each gender. I hope this paper can help to reveal gender income difference in its most fundamental ways.

[1] Gender and Income Inequality. International Monetary Fund. Accessed August 12, 2021.
$https://www.imf.org/external/pubs/ft/sdn/2015/sdn1520_info.pdf$

[2] Tyson, Laura D'Andrea, and Ceri Parker. "An Economist Explains Why Women Are Paid Less." World Economic Forum, March 2019, accessed August 12, 2021.
$https://www.weforum.org/agenda/2019/03/an-economist-explains-why-women-get-paid-less/$

[3] Bleiweis, Robin. "Quick Facts about the Gender Wage Gap." Center for American Progress, 2019, accessed August 12, 2021.
$https://www.americanprogress.org/issues/women/reports/2020/03/24/482141/quick-facts-gender-wage-gap/$

[4] Schultz, T.Paul. "Human Capital, Schooling and Health." Economics & Human Biology 1, no. 2 (June 2003): 207–21, accessed August 13, 2021. $https://doi.org/10.1016/s1570-677x(03)00035-2.$

[5] Xiao, Yuxi, Haizheng Li, and Belton M. Fleisher. "The Earnings Effects of Health and Health-Related Activities: A Panel Data Approach." Applied Economics 47, no. 14 (2015): 1407–23, accessed August 13, 2021.
$https://doi.org/10.1080/00036846.2014.1000521$

[6] Fox, Maggie. "Smokers Less Likely to Get Hired and Earn Less: Study." NBCNews.com. NBC Universal News Group, April 12, 2016, accessed August 13, 2021.
$https://www.nbcnews.com/health/health-news/smokers-less-likely-get-hired-earn-less-study-n554321.$

[7] Takahashi, Takehiro, and Akiko Iwasaki. "Sex Differences in Immune Responses." Science 371, no. 6527 (2021): 347–48, accessed August 13, 2021. $https://doi.org/10.1126/science.abe7199.$

[8] Rubery, Jill, and Aristea Koukiadaki. "Closing the Gender Pay Gap: A Review of the Issues, Policy" International Labor Office, 2016, accessed August 13, 2021.
$https://www.ilo.org/wcmsp5/groups/public/---dgreports/---gender/documents/publication/wcms_540889.pdf$

2 Literature Review

I found that previous studies tried to correlate income and health-related activities. For instance, previous study has shown that smoking and drinking are correlated with income level.[9] [10] Interestingly, several prior studies have investigated the correlation between drinking and income with no agreed conclusions and experts are still aruging about the effect of alcohol drinking on income in psychological and empirical aspects.[11] [12] Whatever, currently, no paper directly shows the correlation between coffee or tea drinking and income level. Studies, however, show positive effects of coffee and tea have on productivity, an indicator of income level.[13] [14] For exercise, studies show a positive correlation between it and health [15] Moreover, good health status has been established to be able to positively influence a person's income. A longitudinal data set starting from 1750, focusing on generations in US, showing that casual effects exist between health situation and yearly income.[16]

Next, I focused on education and years of experiences, which have long been argued to be in a casual relationship with income. In the famous Mincer Model, education and experiences are found to be determinants of income.[17] However, the Mincer Model only did the statistical proof that these variables are of casual relationships. It was pointed out that the model did not clarify the channels through which education and experiences might affect income in a certain direction.[18] Fortunately, previous research has shown that the model can be improved by counting the effect of education on health improvement within schooling.[19] This was achieved

[9] M. Christopher Auld, "Smoking, Drinking, and Income," Journal of Human Resources XL, no. 2 (2005): pp. 505-518, accessed August 14, 2021. *https://doi.org/10.3368/jhr.xl.2.505.*

[10] In "Smoking, Drinking, and Income", Christopher Auld pointed out that drinking is associated with 10-12 percent higher income than drinking abstention, while smokers are associated with 24 percent lower income than non-smokers after correcting for endogeneity

[11] Jae-Hyung Lee, "The Influence of Alcohol Consumption on Income and Health: Empirical Evidence from a Panel of OECD Countries," Seoul Journal of Economics 26, no. 2 (2013). This article use panel data from 2001 to 2005 to do regression and find statistically significant results that states moderate drinking is positively correlated with income level.

[12] Magdalena Cerdá, Vicki D. Johnson-Lawrence, and Sandro Galea, "Lifetime Income Patterns and Alcohol Consumption: Investigating the Association between Long- and Short-Term Income Trajectories and Drinking," Social Science & Medicine 73, no. 8 (2011): pp. 1178-1185, *https://doi.org/10.1016/j.socscimed.2011.07.025.* This article suggests that alcohol drinking is correlated with lower income levels, as opposed to the results in Lee's paper.

[13] Tom M. McLellan, John A. Caldwell, and Harris R. Lieberman, "A Review of Caffeine's Effects on Cognitive, Physical and Occupational Performance," Neuroscience & Biobehavioral Reviews 71 (2016): pp. 294-312, accessed August 15, 2021. *https://doi.org/10.1016/j.neubiorev.2016.09.001.*

[14] Farrukh Afaq et al., "Health Benefits of Tea Consumption," Beverages in Nutrition and Health, 2004, pp. 143-156, accessed August 15, 2021. *https://doi.org/10.1007/978-1-59259-415-3_10.*

[15] Jaana T. Kari et al., "Income and Physical Activity AMONG Adults: Evidence from Self-Reported AND Pedometer-Based Physical Activity Measurements," PLOS ONE 10, no. 8 (2015), accessed August 16, 2021. *https://doi.org/10.1371/journal.pone.0135651.*

[16] Dora L. Costa, "Health and the Economy in the United States from 1750 to the Present," Journal of Economic Literature 53, no. 3 (January 2015): pp. 503-570, accessed August 17, 2021. *https://doi.org/10.1257/jel.53.3.503.*

[17] Jacob Mincer, Schooling, Experience, and Earnings (New York, New York: National Bureau of Economic Research, 1974).

[18] Hall, Robert E. Journal of Political Economy 83, no. 2 (1975): 444-46. Accessed August 15, 2021. *http://www.jstor.org/stable/1830936.*

[19] Damon Clark and Heather Royer, "The Effect of Education on Adult Mortality and Health: Evidence from Britain," American Economic Review 103, no. 6 (January 2013): pp. 2087-2120, accessed August 16, 2021. *https://doi.org/10.1257/aer.103.6.2087.*

in 2003 by Schultz, who conducted quantified the economic return to schooling in terms of human capital.[20]

When it came to the situation where I am examining gender differences in the income factors stated above, I can find several relevant papers concerning gender and productivity. Education is shown to be more beneficial to women than to men in terms of income growth by the channel of improving one's health.[21] Biologically, health-related habits can impose a deeper influence on females. Smoking is shown to be more detrimental to women because of the way in which nicotine interact with sex hormones.[22] [23] Similarly, metabolism mechanism determines that females are more vulnerable to the effect of heavily drinking alcohol and benefit more from exercise.[24] However, none of the papers about gender and habits uses longitudinal empirical data to support their conclusion. Instead, they did biological experiments only. It is of high value if we can use observed panel data to figure out the gender differences in several income factors quantitatively.

To sum up, there is no perfect fit to the goal of correlating health-related activities to income using unbiased statistics. The paper that almost did that job should be Y. Xiao et al. 's research in 2015.[25]. Still, it fails by using data set from urban families only. Also, in the data choice, they drop all the observations with missing variables, a redundant step that lower the number of observations the model could use to generate precise results. Finally, the paper did the OLS regression for only two years, too short to observe notable patterns.[26] For gender differences, most of the researches are in an angle of biological and physiological differences between the two genders, lacking the support of empirical evidence. In a nutshell, a new research should be conducted to address all the above concerns.

3 Research Questions

Upon reviewing the articles, I found that a comprehensive assessment on income factors and gender differences in these factors is missing. There are two missions for us to accomplish.

[20] Guangjie Ning, "Can Educational Expansion Improve Income Inequality? Evidences from the CHNS 1997 and 2006 Data," Economic Systems 34, no. 4 (April 2010): pp. 397-412, accessed August 21. 2021. *https : //doi.org/10.1016/j.ecosys.2010.04.001.*

[21] Catherine E. Ross and John Mirowsky, "Gender and the Health Benefits of Education," National Institute of Health 51, no. 1 (January 2010): pp. 1-19, accessed August 21, 2021. *https : //doi.org/10.1111/j.1533 − 8525.2009.01164.x.*

[22] Allen, Alicia M., Cheryl Oncken, and Dorothy Hatsukami. "Women and Smoking: The Effect of Gender on The Epidemiology, Health Effects, and Cessation of Smoking." Current Addiction Reports 1, no. 1 (2014): 53–60, accessed August 22, 2021. *https : //doi.org/10.1007/s40429 − 013 − 0003 − 6.*

[23] Joni Hersch, "Gender, Income Levels, and the Demand for Cigarettes," Journal of Risk and Uncertainty 21, no. 2 (2000), accessed August 21, 2021. *https : //doi.org/10.2139/ssrn.247822.*

[24] Richard Wilsnack, "Are Women More Vulnerable To ALCOHOL'S Effects?-Alcohol Alert No. 46-1999," National Institute on Alcohol Abuse and Alcoholism (U.S. Department of Health and Human Services, 2010), accessed August 21, 2021. *https : //pubs.niaaa.nih.gov/publications/aa46.htm.*

[25] Yuxi Xiao, Haizheng Li, and Belton M. Fleisher, "The Earnings Effects of Health and Health-Related Activities: A Panel Data Approach," Applied Economics 47, no. 14 (July 2015): pp. 1407-1423, accessed August 22, 2021. *https : //doi.org/10.1080/00036846.2014.1000521.*

[26] Jim Frost, "7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression," Statistics By Jim, May 19, 2021, accessed August 23, 2021. *https : //statisticsbyjim.com/regression/ols − linear − regression − assumptions/.*

First, although paper had shown correlation between health-related activities and income, I need to avoid the statistical bias in the model and establish causal relations. The can serve to fully point out the monetary benefits of keeping some healthy habits and provide guidelines for future social researchers to follow to make a powerful and persuasive advertisement for healthy lives.

Secondly, previous studies, when came to the situation where females are paid less than men, tend to assess the conditions in lights of discrimination and social expectations. It should be useful if I can try to find gender differences in effectiveness of health-related income factors. Getting the results, policy makers can decide in which, relating to education or alcohol abstention, for example, can they most effectively invest in using their limited resources. Different policies or incentives might be appropriate or effective for women compared to men.

Thus, considering the concerns, I am going to try to solve the following questions in my paper:

1. What's the correlation between health-related activities and income levels?
2. Does the effectiveness of health-related income factors differ by gender?

Much more refined than previous papers in this area, we discuss methods to mitigate biases in the regression, include missing variables to reduce value of standard error, extend the data set to 15 years and do interaction terms to explore gender differences. However, there is still possible omitted variable bias from time-varying omitted factors.

4 Methodology

4.1 Data

We demand that the selected database to have health-related observations and income. Also, the database need to be panel to come closer to establish casual effects. Thus, I choose to acquire data from China Health and Nutrition Survey (CHNS). Conducted by UNC Carolina Population Center China Project, the survey collect data from Mainland China from over 30,000 individuals for over 25 years, focusing on economics, nutrition and demographics features of the country. To the latest update in 2015, the survey includes education, income, occupations, health status and health related activities from 7200 households, covering population aging from 2 to 110, which can be of much use to our present study. Randomly and anonymously picking respondents, the survey can be reasonably used and be treated as representative of Chinese households. [27]

In table 1, descriptive statistics of the qualified individuals are provided.[28] Over the years, income of Chinese people significantly increases by about 20% to 80% each year.

[27] Carolina Population Center University of North Carolina at Chapel Hill, "China Health and Nutrition Survey," China Health and Nutrition Survey - China Health and Nutrition Survey (CHNS) (UNC China Project, 2015), *https://www.cpc.unc.edu/projects/china.*

[28] "–" was used to indicate variables that are not observed for that year.

Table 1. Descriptive Statistics by Year

		2000					2004			
Variable	Observation	Mean	Std. Dev.	Min	Max	Observation	Mean	Std. Dev.	Min	Max
Annual Income	7611	8788.993	10003.362	5.332	157094.59	5426	10855.437	12948.493	2.581	225000
Education and Experience										
Schooling(years)	6536	7.962	3.792	0	18	5415	8.592	3.857	0	18
Experience(age - years schooling - 6)	6518	24.781	12.273	0	54	5405	26.883	11.664	0	52
Current health status(% sample reporting)										
Excellent or good	7611	0.57	0.495	0	1	5426	0.664	0.472	0	1
Fair	6300	0.25	0.433	0	1	5426	0.293	0.455	0	1
Poor	6300	0.033	0.179	0	1	5426	0.041	0.198	0	1
Habits(% sample reporting)										
Smoking or not	6219	0.332	0.471	0	1	5252	0.333	0.471	0	1
Heavily drinking alcohol	7611	0.151	0.358	0	1	5426	0.171	0.377	0	1
Drinking tea	6300	0.406	0.491	0	1	5426	0.389	0.487	0	1
Adequately exercising	6300	0.095	0.294	0	1	5426	0.087	0.282	0	1
Drinking Coffee	6300	0.014	0.118	0	1	5426	0.024	0.152	0	1
Primary Occupation(% sample reporting)										
Technical worker	7583	0.073	0.26	0	1	5421	0.076	0.266	0	1
Manager	7583	0.043	0.203	0	1	5421	0.044	0.206	0	1
Office staff	7583	0.044	0.206	0	1	5421	0.041	0.198	0	1
Farmer	7583	0.47	0.499	0	1	5421	0.344	0.475	0	1
Worker	7583	0.17	0.376	0	1	5421	0.139	0.345	0	1
Other job	7583	0.144	0.351	0	1	5421	0.137	0.344	0	1
Additional Information(% sample reporting)										
Female	7611	0.463	0.499	0	1	5426	0.478	0.5	0	1
Urban residents	6300	0.32	0.466	0	1	5426	0.306	0.461	0	1

		2006					2009			
Variable	Observation	Mean	Std. Dev.	Min	Max	Observation	Mean	Std. Dev.	Min	Max
Annual Income	5098	14541.746	20317.184	1.653	367000	5260	21628.179	34164.28	6.917	730000
Education and Experience										
Schooling(years)	5090	8.933	4.102	0	18	5258	9.014	3.904	0	18
Experience(age - years schooling - 6)	5076	27.497	11.489	0	54	5243	27.727	11.578	0	54
Current health status(% sample reporting)										
Excellent or good	5098	0.675	0.469	0	1	--	--	--	--	--
Fair	5098	0.277	0.448	0	1	--	--	--	--	--
Poor	5098	0.043	0.202	0	1	--	--	--	--	--
Habits(% sample reporting)										
Smoking or not	4940	0.329	0.47	0	1	5124	0.333	0.471	0	1
Heavily drinking alcohol	5098	0.18	0.385	0	1	5260	0.166	0.372	0	1
Drinking tea	5098	0.366	0.482	0	1	5259	0.362	0.481	0	1
Adequately exercising	5098	0.099	0.298	0	1	5259	0.098	0.298	0	1
Drinking Coffee	5098	0.024	0.153	0	1	5259	0.035	0.184	0	1
Primary Occupation(% sample reporting)										
Technical worker	5098	0.083	0.276	0	1	5260	0.086	0.28	0	1
Manager	5098	0.042	0.2	0	1	5260	0.043	0.203	0	1
Office staff	5098	0.051	0.221	0	1	5260	0.051	0.22	0	1
Farmer	5098	0.349	0.477	0	1	5260	0.344	0.475	0	1
Worker	5098	0.157	0.364	0	1	5260	0.162	0.368	0	1
Other job	5098	0.171	0.377	0	1	5260	0.188	0.39	0	1
Additional Information(% sample reporting)										
Female	5098	0.472	0.499	0	1	5260	0.458	0.498	0	1
Urban residents	5098	0.315	0.464	0	1	5259	0.31	0.462	0	1

		2011					2015			
Variable	Observation	Mean	Std. Dev.	Min	Max	Observation	Mean	Std. Dev.	Min	Max
Annual Income	6728	27390.852	33012.501	22.826	714000	6212	41657.325	92709.238	11.612	4528302
Education and Experience										
Schooling(years)	6710	9.982	3.983	0	18	6205	10.719	3.87	0	18
Experience(age - years schooling - 6)	6696	27.039	11.733	0	54	6187	26.257	11.934	0	54
Current health status(% sample reporting)										
Excellent or good	--	--	--	--	--	6212	0.592	0.492	0	1
Fair	--	--	--	--	--	6211	0.361	0.48	0	1
Poor	--	--	--	--	--	6211	0.038	0.19	0	1
Habits(% sample reporting)										
Smoking or not	6513	0.321	0.467	0	1	6092	0.27	0.444	0	1
Heavily drinking alcohol	6728	0.168	0.374	0	1	6212	0.121	0.326	0	1
Drinking tea	6726	0.416	0.493	0	1	--	--	--	--	--
Adequately exercising	6726	0.152	0.359	0	1	5696	0.091	0.288	0	1
Drinking Coffee	6726	0.083	0.276	0	1	--	--	--	--	--
Primary Occupation(% sample reporting)										
Technical worker	6726	0.111	0.314	0	1	6211	0.122	0.328	0	1
Manager	6726	0.058	0.234	0	1	6211	0.05	0.219	0	1
Office staff	6726	0.072	0.258	0	1	6211	0.097	0.296	0	1
Farmer	6726	0.24	0.427	0	1	6211	0.12	0.325	0	1
Worker	6726	0.168	0.374	0	1	6211	0.185	0.388	0	1
Other job	6726	0.231	0.422	0	1	6211	0.217	0.412	0	1
Additional Information(% sample reporting)										
Female	6728	0.466	0.499	0	1	6212	0.453	0.498	0	1
Urban residents	6726	0.391	0.488	0	1	5696	0.389	0.487	0	1

From year 2000 to 2015, years of education increase steadily, rising by about 34%, which can be attribute to the emphasis to schooling in modern China and the fact that the government is encouraging additional education of high school or college level during work.[29]

Health status is self-reported in the database. Through the years, the percentage of population reporting themselves as in excellent of good health status is around 60%. The percentage reporting fair health status varies from 25% to 30%. Very little people report their health as poor over the years.

Observation rate of smoking decreases from 33% to 27%, probably as awareness of the negative effects of smoking is spreading across the nation. Also, relative stable percentage, about 40% and 3%, of the population report drinking tea or coffee.

Also, about 9% of the population report adequately exercising over the years. Percentage of people who frequently drink alcohol increases at first but decreases then, varying from 15% to 18% and finally to 12%.

For sure, all variable summary reported here are processed first. I dropped income observations that are negative in the dataset because we are going to use percent change of income in our future analysis and negative independent variable can pose a problem, which will not statistically affect our results for less than 2% of the income observations in the original database in negative. Also, I include only individual at working ages to exclude the confounding variables of degrees of aging and status of being adolescents. For male, I only investigate individuals with an age between 18 to 60. For females, I only investigate those with an age between 18 and 55.[30] Next, the survey provide discrete dataset for schooling, listed in years of elementary school, years of middle school and years of high school, etc. I calculate the schooling variable in a continuous form by adding the years from each level of education. Chinese students start elementary school from the age of 6, so I decide to calculate years of potential work experience by deducting living years(ages) by years of education and 6. Next, I made dummy variables for health status and get the percent of observation reporting corresponding health situation in a chosen year. Smoking, drinking tea and drinking coffee are all in dummy variable form, with value of 1 indicating "yes" and value of 0 indicating "no".

Dummy variables for heavily drinking alcohol and adequate exercising have been specially considered. The variable of heavily drinking alcohol is generated by the reported frequency of drinking alcohol including wine and beer. It is defined as 1, or "yes", if the individual is female and drinking one or more drink for more than five times a week. For male, heavily drinking is set to the criteria for which the individual take two or more drinks for more than five times a week.[31] For exercise, an observation's adequately exercise dummy variable is set to be 1, or "yes", if the individual is active in any set of sports combination summed in the survey. [32]

[29] Usually, people at work in China can get working benefits and may increase their education levels by attending state directed Night Universities, which is credited with increasing the overall years of education by such a large extent in the dataset.

[30] OCED Publication, "China: Pension System in 2018," OCED for China, 2018, accessed August 25, 2021. *https://www.oecd.org/els/public-pensions/PAG2019-country-profile-China.pdf.*

[31] Defined under the standard set by United States Centers of Disease Control and Prevention *https://www.cdc.gov/alcohol/faqs.htm#moderateDrinking*

[32] In the CHNS dataset, multiple variables are created to reflect whether the individual is active in any sports. The researchers

Overall, rummaging through the survey's dataset, we can see that occupations are extensively reported, which may largely influence income. Due to the fact that we may need to use OLS regression in our model, I decide to include the occupation variables and report them in dummy variables.[33]

Holistically, rate of observing beneficial factors like education, experiences and good habits increases over the years with observations of bad habits decreasing, leading us to have preliminary proposition that health-related activities and health may play a role in determining income level.[34]

4.2 Measures

I use 12 variables in all in our models in the next section. Here I will provide details of the variables.

Independent Variables of Interest

The following variables are the health related activities we are interested in associating with the income of the individual.

- **Smoking:** This dummy variable is generated by individual's response to the question of whether they smoke or not in the investigation year. Positive answer is included as 1 and negative response is included as 0.
- **Heavily Drinking Alcohol:** This dummy variable is generated by self-reported frequency of drinking alcohol and the standard mentioned in the previous section, *Data*.
- **Drinking Tea:** This dummy variable is generated by individual's response to the question of whether they drinking tea or not in the investigation year. Positive answer is included as 1 and negative response is included as 0.
- **Adequately Exercising:** This dummy variable is generated by self-reported frequency of doing various kinds of sports and the standard mentioned in the previous section, *Data*.
- **Drinking Coffee:** This dummy variable is generated by individual's response to the question of whether they drinking coffee or not in the investigation year. Positive answer is included as 1 and negative response is included as 0.

Independent Control Variables

Some variables will affect income while have some relationship with health related activities. As explained in the section of next section, I will include them in the OLS regression model.

report the results in combination of sports, commonly dance and gymnastics, field sports and walking, ball games or martial arts and other unlisted sports.

[33] For OLS, set variables' coefficients will be biased if we exclude variables that have correlation with both the independent variable and dependent variables. More information at Jim Frost, "Confounding Variables Can Bias Your Results," Statistics By Jim, April 26, 2021, *https://statisticsbyjim.com/regression/confounding−variables−bias/.* and section 4.3, methods

[34] As previously explored in relationship between investment on health and income, by Cropper, M. L. "Health, Investment in Health, and Occupational Choice." Journal of Political Economy 85, no. 6 (1977): 1273–94, accessed August 26, 2021. *https://doi.org/10.1086/260637*

- **Health Status:** This variable is self-reported. Individual are asked to choose their current health status during the investigation year to three different levels: Excellent or good, fair, or poor. To better process the information, I made a dummy variable for each health status, with 1 being in that status and 0 meaning not.
- **Occupation:** This variable is self-reported. Individual are asked to choose their current occupation during the investigation year in six different options: Technical Worker, Manager, Office Staff, Farmer, Worker or Other Job. To better process the information, I made a dummy variable for each occupation option, with 1 being in that job and 0 meaning not.
- **Schooling:** This is a self-reported continuous variable indicating years of education the individual has received up until the investigation year.
- **Experience:** This is a continuous variable measuring years of experience on the society, calculated from the age and the years of education(Age−Schooling−6).
- **Experience, square of:** As argued in the Mincer Model, which studied income factors, it is better to include the square of years of experience when studying any models in income factors for the proposition that in standard human capital models, investment in training at work decreases over time .[35] This variable is generated by taking square of the variable *Experience*.
- **Gender:** This dummy variable is included because income differ for gender for various reasons.. This variable is 1 when the individual is a female while being 0 otherwise.
- **Urban Residence:** This dummy variable is included because income is different across locations. This variable is 1 when the individual is an urban resident while being 0 otherwise.

Dependent Variable

- **Annual Income, natural log of:** This variable is generated by taking the natural log of the reported real annual income, which used 2006 as a base year. The annual income was included by adding the following components of income together: wage, insurance, investment, sales of household goods and governmental aids. In OLS regression models, taking the natural log of income enables us to interpret the results in yearly percent changes.

Additional Variables

Checking our variables, we find that some variables such as smoking, has missing observation, being that the individual refuses to respond to the question. Thus, during the processing, I combine each dummy variable with a counterpart, being the situation in which the variable's status is missing.

For example, for smoking, we have three separate recordings, with $Smoking\_Yes$ being the indicator of whether the individual is smoking or not, $Smoking\_no$ being the 1 when the individual does not smoke, $Smoking\_Missing$ being the indicator of whether the individual responded to the question and generate effective answers.

[35] Thomas Lemieux, "The 'Mincer Equation' Thirty Years after Schooling, Experience, and Earnings," Jacob Mincer A Pioneer of Modern Labor Economics, 2000, pp. 127-145, accessed August 27, 2021. $https://doi.org/10.1007/0-387-29175-x_11.$

In comparison to prior research, this method enable us to maintain greater number of observations, which will help these standard error and make it smaller and obtain a sample size two times larger than the previous studies.

4.3 Methods

4.3.1 OLS Model

I estimated the model using the Ordinary Least Squares regression, or OLS regression:

$$Y_i = \beta_0 + \beta_1 Smoking_i + \beta_2 Alcohol_i + \beta_3 Coffee_i + \beta_4 Exercising_i + \beta_5 Tea_i + \beta_K X_{Ki} + \varepsilon_i \quad (1)$$

where Y_i is the outcome variable, being the natural log of real annual income in this case. X_{ni} stands for the independent variables, indicating the five health-related activities of interest like smoking, drinking alcohol, etc, and X_{Ki} being the pool of the other control variables, which includes education, experience, health status and occupation.

I will do OLS regression for six separate times for each year and see the trending of the coefficients in response to research question 1 of the effects of health related activities in earnings.

However, bias problem must be considered. Omitted variable bias occurs when unobserved factors, such as qualities of planfulness or the degree to which the person is oriented to the future, that in fact have influence on income while having correlation with health-related activities like smoking, are not included in the model.[36]

Still, methods exist for us to utilize to mitigate the bias. Fixed effect panel data estimation may help to mitigate this bias, however, differences by gender cannot be analyzed using fixed effect model because gender is constant.

4.3.2 Fixed Effects Model

The unbalanced database is, fortunately, longitudinal. Thus, I can apply fixed effect models to the panel database in order to try to lessen the omitted variable bias.[37]

I used the following equation to write the fixed effect model:

$$Y_{it} = \beta_1 Smoking_i + \beta_2 Alcohol_i + \beta_3 Coffee_i + \beta_4 Exercising_i + \beta_5 Tea_i + \beta_K X_{Ki} + \alpha_i + \varepsilon_{it} \quad (2)$$

where Y_{it} is the real income of person i in year t, α_i being the fixed effect of being the target, which will not vary over the years but is different across people. X_{it} is independent variables

[36] referred from Barreto; Howland (2006). "Omitted Variable Bias". Introductory Econometrics: Using Monte Carlo Simulation with Microsoft Excel. Cambridge University Press. If something in ε, the error term, correlates with the dependent variable, which in this case is the log of annual income, and one independent variable, then the coefficient of that particular independent variable in the OLS model will be biased for that the model will count the effect on the omitted variable on that correlated independent variable

[37] Wikipedia Society, "Fixed Effects Model," Wikipedia (Wikimedia Foundation, June 22, 2021), accessed August 28, 2021. *https://en.wikipedia.org/wiki/Fixed_effects_model.*

at the year t and ε_{it} is the idiosyncratic error term

For convenience, I applied entity-demeaned OLS regression. From equation(2), we can deduce the following equation:

$$Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \beta_1 \left(X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it} \right) + \ldots + \left(\varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it} \right) \tag{3}$$

or simply:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \ldots + \tilde{u}_{it} \tag{4}$$

where $\tilde{Y}_{it}$ stands for the difference between the log of real annual income of the year t and the average of the log of annual income each year in the range of the panel database for the designated individual i. $\tilde{X}_{it}$, on the other hand, means similarly as $\tilde{Y}_{it}$. The α_i coefficients are not all individually estimated in this entity demeaned estimation. However, they are controlled for and they can represent things like a person's innate ability, patience and furture orientedness. The only difference is that $\tilde{X}_{it}$ measures the independent variables.

Using the Entity-demeaned OLS regression, we are able to eliminate the effect of unobserved variables that are constant over time, affect the income, and correlate with health oriented activities. Applying this model, we can improve our regression significantly in evaluation of question of effectiveness of health-related income factors.

4.3.3 Interaction Effects

In research of the effectiveness of each income factors across genders, I utilized the interaction effect model in the OLS regression, calculating the interaction term values of the income factors when they are tested across genders.[38] [39] This coefficient represents the differences in the effects of the health-related variables between females and males and I estimate regressions for each year in the data set.

5 Results

Table 2 reports the OLS cross sectional regression results from year 2000 to year 2015, analyzing the effect of health-related activities and other independent income factors on earnings.

In regard of our main area of investigation, several health-related activities can have effects on income. Most notably, tea drinking is suggested to have positive effects on income, varying from 5% in 2009 to 10.1% in 2000. coefficients on smoking, on the other hand, imply that smoking has negative effects on income and the results are statistically significant.[40] Drinking

[38] Wikipedia Society, "Interaction (Statistics)," Wikipedia (Wikimedia Foundation, July 9, 2021), accessed August 29, 2021. *https://en.wikipedia.org/wiki/Interaction_(statistics).*

[39] the interaction terms stands for the differences in the extent of the income factors for the two genders, with β_1 stands for the unique effect of the variable it represent when the other interacted variable in in β_3, or β_2 is 0 and vice versa

[40] $\beta = -0.065, P < 0.1$ in 2000, $\beta = -0.117, P < 0.01$ in 2004, $\beta = -0.064, P < 0.05$ in 2009 for smokers as relative to nonsmokers

Table 2: Regression Results by Year

	2000		2004		2006		2009		2011		2015	
VARIABLES	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Education and Experience												
Schooling	0.032***	0.005	0.034***	0.006	0.045***	0.006	0.046***	0.005	0.039***	0.005	0.042***	0.006
Experience	0.055***	0.005	0.031***	0.006	0.027***	0.005	0.037***	0.005	0.019***	0.004	0.019***	0.005
Experience, square	-0.001***	0.000	0.000***	0.000	0.000***	0.000	-0.001***	0.000	0.000***	0.000	0.000***	0.000
Health Status												
Excellent or good	0.283***	0.095	0.121	0.082	0.319***	0.079	---	---	---	---	0.492***	0.095
Fair	0.145	0.097	0.004	0.084	0.256***	0.081	---	---	---	---	0.364***	0.096
Missing(Health Status)	0.325***	0.125	0.412	0.371	0.052	0.167	---	---	---	---	0.071	0.185
Habits												
Smoking												
Yes	-0.065*	0.035	-0.117***	0.040	-0.043	0.039	-0.041	0.037	-0.064**	0.031	0.038	0.040
Missing(Smoking)	0.163	0.176	0.956	0.616	-0.978**	0.386	-0.966	0.637	0.182**	0.081	0.334*	0.191
Heavily Drinking Alcohol												
Yes	0.038	0.035	0.061	0.041	0.088**	0.041	0.130***	0.040	0.027	0.033	-0.071	0.049
Missing(Alcohol)	0.022	0.073	-0.123	0.113	0.189	0.123	0.294**	0.141	0.029	0.096	0.018	0.276
Tea Drinking												
Yes	0.101***	0.028	0.074**	0.032	0.116***	0.030	0.050*	0.030	0.092***	0.023	---	---
Missing(Tea)	0.143	0.199	0.525*	0.302	0.547***	0.183	0.191	0.295	0.403	0.346	---	---
Adequately Exercise												
Yes	0.043	0.042	-0.012	0.048	0.130***	0.039	0.05	0.041	0.106***	0.028	0.080*	0.044
Missing(Exercise)	-0.081*	0.049	-0.076	0.180	---	---	---	---	---	---	0.587**	0.286
Coffee Drinking												
Yes	0.023	0.108	0.256***	0.064	0.209***	0.066	0.119*	0.064	0.204***	0.035	---	---
Missing(Coffee)	-0.034	0.159	0.267	0.261	-0.320	0.266	-0.286	0.249	0.162	0.261	---	---
Occupation												
Technical Worker	0.189***	0.052	0.324***	0.050	0.364***	0.050	0.340***	0.047	0.338***	0.036	0.293***	0.041
Manager	0.208***	0.055	0.174***	0.063	0.373***	0.065	0.335***	0.068	0.333***	0.046	0.259***	0.060
Office Staff	0.135**	0.061	0.285***	0.055	0.275***	0.053	0.202***	0.060	0.201***	0.041	0.222***	0.043
Farmer	-0.759***	0.040	-0.918***	0.047	-0.594***	0.047	-0.494***	0.043	-0.465***	0.040	-0.491***	0.055
Worker	0.123***	0.040	-0.003	0.046	0.217***	0.044	0.101**	0.043	0.151***	0.032	0.135***	0.042
Missing(Occupation)	-0.644***	0.078	-0.767***	0.053	-0.816***	0.060	-0.615***	0.059	-0.550***	0.049	-0.845***	0.049
Additional Information												
Being Female	-0.096***	0.035	-0.222***	0.040	-0.184***	0.037	-0.092**	0.036	-0.149***	0.030	-0.114***	0.035
Living in Urban	0.107***	0.031	0.189***	0.032	0.228***	0.031	0.104***	0.030	0.093***	0.025	0.150***	0.030
Constant	7.730***	0.132	8.395***	0.127	8.122***	0.127	8.662***	0.099	9.153***	0.081	9.024***	0.137
Observations	5,279		5,227		4,918		5,107		6,484		5,559	
R-squared	0.251		0.26		0.286		0.207		0.218		0.236	
Adjusted R^2	0.251		0.26		0.286		0.207		0.218		0.236	

Note: *** p<0.01, ** p<0.05, * p<0.1

coffee can have positive influence in income. Coefficients on drinking coffee suggests as relative to non coffee drinkers, coffee drinkers earn 11.9% to 25.6% more. Exercising is also positively associated with earnings, increasing income by around 10% over the years. Interestingly enough, heavily drinking alcohol is associated with higher income, from 2.7% to 13.0%, though the statistical evidence is not substantiated over years and coefficients are not constantly positive.

Highly similar to results generated in the classical Mincer Model, one of experience have almost same effect on increasing income as one year of education. As time goes, however, the effect of schooling ($\beta = 0.042, P < 0.01$) began to outweigh that of experience ($\beta = 0.019, P < 0.01$) by 2.3 % as indicated in the regression results for 2015.

As for self-reported health status, the coefficients are always positive. Holding other independent variables constant, excellent or good health status is correlated with 28.3% to 49.2% higher income than poor health, as there is some evidence in year 2000, 2006 and 2015, all being statistically significant. This sufficiently shows that good health is important to earning highly.

Also, occupation plays a key role in determining income, as testified in the statistically significant coefficients along the years. Though occupation can be largely associated with education, differences other than years of education can be manifested in the occupation the individual finally take as a result of class ranking and curriculum quality. Comparing with holding "other jobs" like being a driver, being a farmer indicates much lower income. However, this phenomenon fades in recent years, as the effect changed from -91.8% in 2004 to -49.1% in 2015. Besides, all other jobs like being technical worker or office staff, will correlate with 13.5% to 29.3% higher than "other jobs".

Finally, gender and location of residence can be earning factors too. Being a female while holding all other independent variables constant is associated with around 11% lower income and living in cities is associated with 9.3% to 22.8% more income than rural residents.

Table 3 shows the results from the fixed effect model, which eliminated the bais generated by individual differences that does not vary over the years.

The identifying variations for the effect of health-related activities on income are the changes in behaviors within a person over time. For example, the effect of smoking comes from people who change their smoking behavior. Answering the research question of effects of health-related activities, the results show that smoking and alcohol drinking do have association with income, but is statistically insignificant ($\beta = 0.020, P = 0.15$ and $\beta = 0.037, P = 0.21$). Tea drinking can have a positive effect, 4.6%, on income ($\beta = 0.046, P < 0.05$). Adequate exercisers can have 3.0% more income than those who exercise insufficiently ($\beta = 0.030, P < 0.05$). Finally, coffee drinking associates with 3.3% more income than nondrinkers, with p-value equals 0.10, while containing 0 in its standard error boundaries.[41]

Notably, the effects of gender and urban residence can't be investigated using the new model due to the fact that no observations in the survey ever changed their gender and no people move

[41] Thus, the results should be viewed with caution

Table 3: Fixed Effect Regression Results

Variables	Year 2000 - 2015 Coef.	Std. Err.	Variables	Continue Coef.	Std. Err.
Education and Experience			Coffee Drinking		
Schooling	0.078 ***	0.006	Yes	0.033 *	0.033
Experience	0.104 ***	0.006	Missing(Coffee)	-0.031	0.126
Experience, square	-0.001 ***	0.000	Occupation		
Health Status			Technical Worker	0.078 **	0.039
Excellent or good	0.254 ***	0.056	Manager	0.090 **	0.041
Fair	0.178 ***	0.056	Office Staff	0.053	0.035
Missing(Health Status)	0.564 ***	0.057	Farmer	-0.417 ***	0.033
Habits			Worker	0.111 ***	0.031
Smoking			Missing(Occupation)	-0.609 ***	0.035
Yes	0.020	0.031	Additional Information		
Missing(Smoking)	-0.181	0.151	Being Female	---	---
Heavily Drinking Alcohol			Living in Urban	---	---
Yes	0.037	0.024	Constant	6.127 ***	0.142
Missing(Alcohol)	-0.004	0.059			
Tea Drinking					
Yes	0.046 **	0.019	Observations	32574	
Missing(Tea)	0.369 ***	0.127	R-squared	0.218	
Adequately Exercise			Adjusted R^2	0.218	
Yes	0.030 **	0.021			
Missing(Exercise)	-0.041	0.062			

Note: *** p<0.01, ** p<0.05, * p<0.1

between towns and rural villages.[42] Thus, I use "—" as placeholders in the table.

Education here is shown to have statistically significant effect on earnings, which is definitely expected. Also, as predicted in the Mincer Model, years of experience have a effect similar to (in case large than) that of education. One years of experience can even generate 2.6% more income than one year of education.

Compared to those in poor health status, holding all other independent variables constant, better health statuses are associated with higher earnings ($\beta = 0.254, P < 0.01$ and $\beta = 0.178, P < 0.01$). Occupation, similarly, can largely influence one's income, as in accordance with the results in the cross-sectional data analysis.

Table 4 is designed to answer the question of gender differences in the earning effects of health-related activities. While applying the interaction effects principles, I used cross-sectional OLS regression instead of fixed effect estimation, preserving the significance of gender in the model at the expense of keep constant individual unobserved variable bias. Gender, as indicated above, cannot be included in the fixed effect model due to the fact that no one in the sample changed their sexual identity and thus gender will be omitted in the panel data model. Thus we have to apply OLS regression by year. Also, I include other independent variables in concern of omitted variable bias.

In the table, only coefficients on the interaction terms of interest are reported for conciseness and clarity.[43] In 2015, the interaction term of tea drinking and coffee drinking are nonexistent for sake of no observations of these two activities in that year.

Females are more sensitive to the negative effects of smoking on wages. At year 2004, the interaction term of smoking and sex shows females are more vulnerable to smoking in affecting income than to men ($\beta = -0.209, P < 0.05$) and similarly in 2015 ($\beta = -0.344, P < 0.01$). The interaction term of smoking and sex is -0.209 in 2004, showing that females are more vulerable to smoking in affecting income than to men. In particular, for women, being a smoker relative to a nonsmoker, is associated with 27.1%(-6.2% + -20.9% = 27.1%) decrease in wages, whereas for men, being a smoker relative to a nonsmoker is associated with 6.2% lower wages

In terms of drinking alcohol, the interaction is only significant in 2006 ($\beta = 0.231, P < 0.01$), indicating that for females, being a alcohol drinker relative to a non-drinker, is associated with 30.4% increase in wages. Throughout the cross sectional regression, the interaction term coefficients of alcohol and gender do change sign, indicating the effect is not stable over the years.

For tea drinking, females are shown to benefit more than men. In 2000, 2006, 2009 and 2011, the interaction term is positive. The interaction is statistically significant in 2000 ($\beta = 0.039, P < 0.1$) and 2011 ($\beta = 0.136, P < 0.01$).

[42] This can be caused by the fact that whenever the target change their location during the surveys years, the researchers will lose track of them and cannot include them in the reported database.

[43] As explained before, the interaction term is the difference between the effect of the independent variable on females and the effect generally when I pool the two genders together for analysis.

Table 4. Interaction Terms for Gender

	2000		2004		2006		2009		2011		2015	
VARIABLES	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Health related activities												
Smoking * Female	-0.151	0.110	-0.209 **	0.173	-0.011	0.149	0.111	0.133	-0.025	0.150	-0.344 ***	0.148
Heavily Drinking Alcohol * Fem	-0.035	0.092	-0.035	0.190	0.231 ***	0.110	-0.022	0.115	-0.004	0.095	0.020	0.162
Tea Drinking * Female	0.039 *	0.056	-0.016	0.063	0.016	0.061	0.011	0.059	0.136 ***	0.046	---	---
Adequately Exercise * Female	-0.013	0.097	-0.056	0.098	0.193 ***	0.074	-0.051	0.079	-0.059 **	0.055	-0.273 ***	0.254
Coffee Drinking * Female	0.119	0.218	0.072	0.128	0.073	0.132	-0.168 ***	0.123	-0.122 *	0.068	---	---
Constant	7.726 ***	0.133	8.388 ***	0.128	8.145 ***	0.128	8.667 ***	0.100	9.162 ***	0.083	8.996 ***	0.137
Observations	5,279		5,227		4,918		5,107		6,484		5,559	
R-squared	0.253		0.261		0.288		0.208		0.219		0.237	
Adjusted R^2	0.253		0.261		0.288		0.208		0.219		0.237	

Note: *** p<0.01, ** p<0.05, * p<0.1. The model was run using other independent variables in concerns of omitted variable bias. This table only choose to not include the other variables for clarity. The omitted independent variables are available in previous cross-sectional OLS result list.

It is complicate for effect of adequately exercising. The term is positive in 2006 ($\beta = 0.193, P < 0.01$) and negative in 2011 ($\beta = -0.059, P < 0.05$) and 2015 ($\beta = -0.273, P < 0.01$). And for drinking coffee, females tend to benefit less, as the interaction term in 2009 ($\beta = -0.168, P < 0.01$) and 2011 ($\beta = -0.122, P < 0.1$) is negative and statistically significant.

6 Extension

In goal of furthering the understanding of the research questions, I conducted the interaction effects of place of residence, fixed effect model of levels of the different health-related activities with occupation and/or health status removed, and include dummy variables of each year in an independent fixed effect model.

Table 5 shows the interaction results of the urban residence and health-related activities. For smoking, drinking alcohol or tea or coffee, the interaction effect with the urban residence is not so noteworthy for the fact that the coefficients generated can vary in sign and is not constantly statistically significant around one particular value.

For adequately exercising, the interaction effect with urban residence is notable for the coefficients are all positive and statistically significant in 2009 ($\beta = 0.159, P < 0.05$) and 2011 ($\beta = 0.116, P < 0.01$), which can be possibly caused by the fact that urban residents experience less violent physical activities during work than rural villagers and thus active exercise can make a huge difference in physical fitness.[44]

[44] Guo, Xinyan. "Research on Perceived Profiles and Stages of Exercise Behavior Change in Urban Residents." Proceedings of the 2015 International Conference on Education Technology, Management and Humanities Science, 2015, accessed September 1, 2021. *https://doi.org/10.2991/etmhs − 15.2015.86*

Table 5. Interaction Term for Urban Residence

	2000		2004		2006		2009		2011		2015	
VARIABLES	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Health related activities												
Smoking * Urban	-0.034	0.062	0.022	0.068	0.048	0.064	-0.030	0.062	-0.049 *	0.051	-0.040	0.070
Heavily Drinking Alcohol * Urba	-0.093	0.072	-0.223 ***	0.085	-0.044	0.077	0.081 *	0.078	-0.020	0.065	-0.129 **	0.100
Tea Drinking * Urban	0.148 **	0.058	0.209 ***	0.063	-0.036	0.059	-0.029	0.059	-0.092 **	0.045	---	---
Adequately Exercise * Urban	0.041	0.083	0.136	0.093	0.065	0.077	0.159 **	0.080	0.116 **	0.058	0.113	0.088
Coffee Drinking * Urban	-0.002	0.217	0.197 *	0.150	0.112	0.145	-0.308 **	0.131	-0.101 *	0.077	---	---
Constant	7.760 ***	0.133	8.434 ***	0.126	8.112 ***	0.128	8.658 ***	0.099	9.128 ***	0.083	9.017 ***	0.137
Observations	5,279		5,227		4,918		5,107		6,484		5,559	
R-squared	0.253		0.262		0.286		0.209		0.219		0.237	
Adjusted R^2	0.253		0.262		0.286		0.209		0.219		0.237	

Note: *** p<0.01, ** p<0.05, * p<0.1. The model was run using other independent variables in concerns of omitted variable bias. This table only choose to not include the other variables for clarity. The omitted independent variables are available in previous cross-sectional OLS result list.

Table 6 compares three cases. In case 1, the model is regressed using levels of habits without health status or occupation.[45] Case 2 regresses the model together with health status and Case 3 regresses the model together with occupation. Case 1 removes only individual constant fixed effect(FE) while keeping endogeneity in error term remains. Case 2 eliminates individual constant FE and unobserved occupation related characteristics. Case 3, on the other hand, minimizes individual FE and unobserved health related characteristics.[46] Using this model, we can glean some insight into the bias due to proxies while noticing the effect of difference levels of health-related habits, verifying our assumptions about interaction between health or occupation with health-related activities.

Rummaging through the result lines in education and experience lines, health status but not occupation will have influence on their returns. As the same scenario in Mincer Model, one year of experience can have similar or even strong positive effects on earnings, with returns to education and experiences all highly statistically significant with any proxies.

But for health-related activities listed, health, surprisingly, does not have a huge correlation with them.[47] The only observable difference when including health status as a proxy is on exercising, with one hour of exercising varying up to 50% from that resulted without health status proxy.

On the other hand, still surprisingly, occupation proxy does have influence on activities like alcohol consumption, tea drinking, etc. This is quite noteworthy, but rational if we take social expectation and professional norms into consideration. Tea drinking needs time and money and only those with a light, intellectual and well-paid job can afford it. Similarly, alcohol drinking can be directly influenced by occupation. For instance, managers might drink more through socialization and farmer may drink for relaxation.

As for each health-related variable, the controlled dummy variable is level of 0, indicating no occurrence of the behavior, we can properly interpret the results as the level at which the behavior will cause statistically significant effects on the income.

Smoking and exercising are measured in continuous variable form for sake of original structure of database from CHNS. While continuous measure of smoking is insignificant, one additional hour of exercising per week is associated with 1.8% higher income in case 1 ($\beta = 0.018, P < 0.01$).

For alcohol and tea drinking, only the highest levels are associated with significant influences in income relative to 0 level. People who drink alcohol everyday, the highest level in the regression, are associated with 8.1% higher income than non-drinkers, which is possibly caused by the fact that higher incomers will socialize a lot and this can help increasing wages. Drinking

[45] Here, levels of habits indicate the different amount of dose a individual will take in that activities. Each level is annotated by L n, for example, L1 of heavily drinking alcohol means drinking alcohol almost every day

[46] Previous studies have shown that occupation dummies can be precise proxy for unobserved job-related characteristics, as shown in J. de BEYER and J. B. KNIGHT, "The Role of Occupation in the Determination of Wages," Oxford Economic Papers 41, no. 1 (January 17, 1989): pp. 595-618, accessed September 1, 2021. *https://doi.org/10.1093/oxfordjournals.oep.a*041916.

[47] as further established in the simple correlation results in the extension section of Yuxi Xiao, Haizheng Li, and Belton M. Fleisher, "The Earnings Effects of Health and Health-Related Activities: A Panel Data Approach," Applied Economics 47, no. 14 (July 2015): pp. 1407-1423, accessed September 3, 2021. *https://doi.org/*10.1080/00036846.2014.1000521.

Table 6: Fixed Effect Regression Results

Variables	Overall		With Health Status Proxy		With Occupation Proxy	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Education and Experience						
Schooling	0.110***	0.005	0.078***	0.006	0.111***	0.005
Experience	0.152***	0.006	0.114***	0.006	0.141***	0.005
Experience, square	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
Habits						
Smoking (Hundred Stick/Week)	0.001	0.000	0.001	0.000	0.000	0.000
Heavily Drinking Alcohol						
L1 Almost every day	0.081**	0.036	0.079**	0.035	0.061*	0.035
L2 3-4 times a week	0.073*	0.038	0.066*	0.037	0.058	0.037
L3 Once or twice a week	0.048	0.030	0.050*	0.030	0.037	0.029
L4 Once of twice a month	0.040	0.030	0.032	0.030	0.025	0.030
L5 No more than once a month	0.039	0.037	0.019	0.037	0.036	0.036
Missing Data	0.079	0.064	0.028	0.065	0.077	0.061
Tea Drinking						
L1 Almost every day	0.046**	0.022	0.048**	0.022	0.038*	0.022
L2 4-5 times a week	0.077*	0.043	0.068	0.043	0.066	0.042
L3 2-3 times a week	0.059*	0.035	0.044	0.035	0.053	0.035
L4 No more than once a week	0.032	0.055	0.021	0.054	0.048	0.054
L5 2-3 time a month	0.018	0.069	0.011	0.070	0.005	0.070
L6 Only once in a month	0.117	0.151	0.110	0.150	0.142	0.152
L7 Virually no	0.064	0.109	0.047	0.110	0.063	0.107
Missing Data	-0.082	0.094	0.189**	0.096	-0.075	0.095
Adequately Exercise (Hour/Week)	0.018***	0.005	0.009**	0.005	0.015***	0.004
Coffee Drinking						
L1 Almost every day	0.059	0.097	0.065	0.099	0.050	0.096
L2 4-5 times a week	-0.134	0.140	-0.146	0.133	-0.136	0.150
L3 2-3 times a week	0.042	0.066	0.039	0.067	0.017	0.066
L4 No more than once a week	-0.038	0.055	-0.017	0.056	-0.024	0.052
L5 2-3 time a month	0.101	0.084	0.124	0.084	0.110	0.081
L6 Only once in a month	0.066	0.106	0.088	0.107	0.094	0.104
L7 Virually no	0.078	0.080	0.068	0.078	0.038	0.085
Missing Data	-0.093	0.093	0.128	0.096	-0.064	0.094
Health Status						
Excellent or good			0.257***	0.056		
Fair			0.178***	0.056		
Missing(Health Status)			0.582***	0.058		
Occupation						
Technical Worker					0.083**	0.039
Manager					0.102**	0.041
Office Staff					0.061*	0.035
Farmer					-0.400***	0.033
Worker					0.117***	0.031
Missing(Occupation)					-0.631***	0.035
Constant	4.928***	0.102	5.799***	0.143	5.236***	0.102
Observations	32421		32421		32403	
R-squared	0.173		0.185		0.208	
Adjusted R^2	0.173		0.185		0.208	

Note: *** p<0.01, ** p<0.05, * p<0.1; Urban residence and sexual identity is omitted in the chart for they are not regressed in the fixed effect analysis due to the fact of low mobility and constant sexual identity.

tea, as always, is more than a simple health improvement activity in China. People may drink tea for mental wellness and recovery.[48] Thus, it is reasonably that professions like manager, which are associated with high income normally, can create more chance for the individual to drink tea, as drinking tea every day associates with 4.6% higher income.

Holistically, mild level of any form of health-related activities will not generate statistically significant influence on income.

Table 7 shows the fixed effect results when I make dummy variable for each year and include year as a proxy in the regression model.

The benefit is to include them in the panel data regression to control for the effect of any factors that vary over time but are the same for all individuals, potentially some economy-wide trends in wages aren't due to health behaviors. This way the coefficients on the health-related variables are arguably more likely due to a change in health behavior and not some omitted factor like a general trend in wages in the Chinese economy during this time period.

Compared with table 3, the new regression using years as dummies, in fact, yields little variation for coefficients of health-related activities, which provide evidence that general trends' influence, as a bias, is mitigated already.

Surprisingly, the results are noteworthy for coefficients of education. The education coefficient change from $\beta = 0.078$ in the original FE model to $\beta = 0.037$, indicating some factors that vary over time, same for all individuals and have association with education. As proposed before in the section of Literature Review, the national government has increasingly invest in education and reshape the workforce by providing university course at night and free skill-related online courses.

[48] Michael Hurwitz, "Why Do Chinese People Love Tea So Much?," Yoyo Chinese, 2018, accessed September 4, 2021. *https://yoyochinese.com/blog/learn-mandarin-chinese-tea-culture-why-is-tea-so-popular*

Table 7: Fixed Effect Regression Results With Year as Proxy

Variables	Year 2000 - 2015 Coef.	Std. Err.	Variables	Continue Coef.	Std. Err.
Education and Experience			Coffee Drinking		
Schooling	0.037	0.034	Yes	0.030 *	0.033
Experience	0.061 *	0.034	Missing(Coffee)	-0.081	0.130
Experience, square	-0.001 ***	0.000	Occupation		
Health Status			Technical Worker	0.081 **	0.039
Excellent or good	0.252 ***	0.056	Manager	0.093 **	0.041
Fair	0.177 ***	0.056	Office Staff	0.053	0.034
Missing(Health Status)	0.291 ***	0.103	Farmer	-0.424 ***	0.033
Habits			Worker	0.107 ***	0.031
Smoking			Missing(Occupation)	-0.595 ***	0.035
Yes	0.020	0.031	Additional Information		
Missing(Smoking)	-0.067	0.175	Being Female	---	---
Heavily Drinking Alcohol			Living in Urban	---	---
Yes	0.034	0.024	Year		
Missing(Alcohol)	0.013	0.058	2004	*0.058*	*0.135*
Tea Drinking			2006	*0.220*	*0.201*
Yes	0.043 **	0.019	2009	*0.582**	*0.309*
Missing(Tea)	0.266 *	0.157	2011	*0.717**	*0.373*
Adequately Exercise			2015	*0.749*	*0.523*
Yes	0.026 **	0.021	Constant	7.377 ***	0.959
Missing(Exercise)	-0.048	0.062	Observations	32574	
			R-squared	0.220	
			Adjusted R^2	0.220	

Note: *** p<0.01, ** p<0.05, * p<0.1

7 Conclusion

7.1 Advantages and Limitations

In this research, I improved previous study on health-related activities' effects on income by including the missing variables in the model, which will significantly decrease the standard error for sake of more observations included in the database.

Also, for the first time in researches, I try to delve deep into the gender differences on the influences of health-related activities on income, which can be of much use in understanding the underprivileged status of females in the job market.

This study used interaction terms for residence as well, in order to show the effect of residence in income. Also, I removed health status and occupation in the extension section in comparison in order to show the effect of unobserved health characteristics and unobserved job characteristics, trying to precisely contributing the influences.

Further, all health-related activities are assigned into new continuous or dummy variables in order to capture the effect of different levels of habits. From the results, we can find the effect of consuming one more unit in the continuous variable and find the lowest level at which the habits can show statistical significant effect on earnings.

While this study may offer some quite new findings, which I summarize below, we also acknowledge its limitations. Health status, an important independent variable in the model, is self-reported and each individual can have differing standard for this variable.

Finally, there are most likely some unobserved, individual specific and correlating with health habits independent variable that vary through time. For example, people may have differing levels of family influence, which is not reported and differing among the population. Family influence, or family control over the individual, definitely correlates with observations of health-related activities and have influence on income. For this, I cannot use statistical methods like panel data estimation to eliminate the bias by not including this time-varying factor in the regression.

7.2 Main Findings

In basic OLS regression by year and fixed effect estimation, drinking tea, adequately exercising and drinking coffee all will improve the individual's income, by 1% to 11% while alcohol drinking, counter-intuitively, is associated with income in cross-sectional study but this effect doesn't carry out to the fixed effect model. Smoking is detrimental to income, but the statistics are insignificant in fixed effect model.

Most notably, in the comparison model that determine whether health-related activities affect health through health status, current health situations are shown to not be pathways for the influence of exercising, tea or coffee, the statistically significant independent variables. Benefits can be caused by reduced mental stress, however.[49] The health-related activities can

[49] Mayo Clinic, "Exercise and Stress: Get Moving to Manage Stress," Mayo Foundation for Medical Education and Research, August 18, 2020, accessed September 10, 2021.
https : //www.mayoclinic.org/healthy − lifestyle/stress − management/in − depth/exercise − and − stress/art − 20044469

be determined jointly with health status too. Finally and most trustfully, direct effects of the health-related activities may not manifest immediately after the action. It may take some and thus the proxy of health status will appear to have no correlation with these activities.

Surprisingly, occupation yield statistically significant correlation with health-related activities. It was explained in the results section that these might be achieved due to socialization.

The most innovate and groundbreaking results are in the gender interaction parts. Women benefit more from drinking tea and less from drinking coffee, as supported by investigations in female biological differences.[50] Also, females benefit more from quitting smokings.

Lastly, only top levels(doing everday) of the habits will generate statistically significant effects on earnings, indicating further that mild levels of habits have no correlation with earnings, which is great in supporting that health-related activities are determinants of income, for if income level can determine different level of health-related habits, all possible levels should be statistically significant enough to yield correlation with income, which didn't occur here.

Reference List

1. Heien, D. "The Relationship between Alcohol Consumption and Earnings." Journal of Studies on Alcohol, 1996. https://www.jsad.com/doi/10.15288/jsa.1996.57.536.

2. Medgyesi, Márton, and István György Tóth. "Income, Wealth, Employment, and Beyond." Europe's Income, Wealth, Consumption, and Inequality, 2021, 177–217. https://doi.org/10.1093/oso/9780197545706.003.0005.

3. Inder, Brett. "Estimating Long-Run Relationships in Economics." Journal of Econometrics 57, no.1-3 (1993): 53–68. https://doi.org/10.1016/0304-4076(93)90058-D

4. Jaccard, James, and Robert Turrisi. Interaction Effects in Multiple Regression. Thousand Oaks, New York: Sage, 2003.

5. Arrow, Kenneth J. The Theory of Discrimination. Princeton, N.J.: Princeton University Press, 1971.

6. Project, Borgen. "Importance of Girls' Education around the World." The Borgen Project. Borgen Project, September 15, 2019. https://borgenproject.org/importance-of-girls-education/.

7. Notanubun, Zainuddin. "The Effect of Organizational Citizenship Behavior and Leadership Effectiveness on Public Sectors Organizational Performance: Study in the Department of Education, Youth and Sports in Maluku Province, Indonesia." Public Organization Review 21, no. 1 (2020): 1–18. https://doi.org/10.1007/s11115-020-00475-4.

8. Pitarakis, Jean-Yves. "On the Bias of the OLS Estimator in a Nonstationary Dynamic Panel Data Model." Statistics & Probability Letters 38, no. 2 (June 1998): 145–50. https://doi.org/10.1016/s0167-7152(97)00165-x

9. Abadir, Karim M. "Ols Bias in a Nonstationary Autoregression." Econometric Theory 9, no. 1 (March 1993): 81–93. https://doi.org/10.1017/s0266466600007349

10. Blackburn, McKinley L., and David Neumark. "Are Ols Estimates of the Return to Schooling Biased Downward? Another Look." The Review of Economics and Statistics 77, no. 2 (May 1995): 217. https://doi.org/10.2307/2109861

11. Sapra, S.K. "Asymptotic Properties of Restricted Maximum-Likelihood Estimator of Parameters in Censored Regression Model." Econometric Theory 4, no. 3 (August 1988): 535–36. https://doi.org/10.1017/s0266466600013463

[50] Allen, Alicia M., Cheryl Oncken, and Dorothy Hatsukami. "Women and Smoking: The Effect of Gender on The Epidemiology,Health Effects, and Cessation of Smoking." Current Addiction Reports 1, no. 1 (2014): 53–60, accessed September 11, 2021. *https : //doi.org/*10.1007/*s*40429`013`0003`6

12. Neumark, David, and Sanders Korenman. "Sources of Bias in Women's Wage Equations: Results Using Sibling Data." The Journal of Human Resources 29, no. 2 (1992): 379–450. https://doi.org/10.3386/w4019

13. Gupta, Anviti, and Alok Gupta. "Women Empowerment Initiatives and Socio-Economic Development." Women Empowerment and Well-Being for Inclusive Economic Growth, January 2021, 106–19. https://doi.org/10.4018/978-1-7998-3737-4.ch007

14. Kabeer, Naila. "Gender Equality and Women's Empowerment: A Critical Analysis of the Third MILLENNIUM Development Goal 1." Gender & Development 13, no. 1 (May 2005): 13–24. https://doi.org/10.1080/13552070512331332273

15. Malatesta, Victor J., Robert H. Pollack, Terri D. Crotty, and Lelon J. Peacock. "Acute Alcohol Intoxication and Female Orgasmic Response." The Journal of Sex Research 18, no. 1 (February 1982): 1–17. https://doi.org/10.1080/00224498209551130

16. Tekin, Erdal. "Employment, Wages, and Alcohol Consumption in Russia." Southern Economic Journal 71, no. 2 (October 2004): 397. https://doi.org/10.2307/4135298

17. Frijters, Paul, and Aydogan Ulker. "Robustness in Health Research: Do Differences in Health Measures, Techniques, and Time Frame Matter?" Journal of Health Economics 27, no. 6 (May 2008): 1626–44. https://doi.org/10.1016/j.jhealeco.2008.06.003

18. Cavalli, Lodovica, and Alessandra Tavani. "Coffee Consumption and Its Impact on Health." Beverage Impacts on Health and Nutrition, February 2016, 29–47. https://doi.org/10.1007/978-3-319-23672-8_3

19. Gravelle, Hugh, and Matt Sutton. "Income, Relative Income, and Self-Reported Health in Britain 1979-2000." Health Economics 18, no. 2 (February 2009): 125–45. https://doi.org/10.1002/hec.1354

20. Hershbein, Brad, Melissa Schettini Kearney, and Luke Pardue. "College Attainment, Income Inequality, and Economic Security: A Simulation Exercise." AEA Papers and Proceedings, May 2020. https://doi.org/10.3386/w26747

21. Cunningham, Cindy M., and Robert D. Mohr. "Using Tools to Distinguish General and Occupation-Specific Skills." Journal for Labour Market Research 53, no. 1 (December 16, 2019). https://doi.org/10.1186/s12651-019-0256-1

22. Huikari, Sanna, Hanna Junttila, Leena Ala-Mursula, Timo Jämsä, Raija Korpelainen, Jouko Miettunen, Rauli Svento, and Marko Korhonen. "Leisure-Time Physical Activity Is Associated with Socio-Economic Status beyond Income-Cross-Sectional Survey of the Northern FINLAND Birth Cohort 1966 Study." Economics & Human Biology 41 (May 2021): 100969. https://doi.org/10.1016/j.ehb.2020.100969

23. Hurley, Jeremiah, Emmanouil Mentzakis, and Marjan Walli-Attaei. "Inequality Aversion in Income, Health, and Income-Related Health." Journal of Health Economics 70 (July 2020): 102276. https://doi.org/10.1016/j.jhealeco.2019.102276

24. Boon, Niels. "Health Potential for Functional Green Teas?" International Journal for Vitamin and Nutrition Research 78, no. 6 (December 1, 2008): 275–81. https://doi.org/10.1024/0300-9831.78.6.275

25. Reardon, Sean F., and Kendra Bischoff. "Income Inequality and Income Segregation." Inequality in the 21st Century, January 15, 2018, 426–35. https://doi.org/10.4324/9780429499821-73

26. Wheeler, Christopher H. "Urban Decentralization and Income Inequality: Is Sprawl Associated with Rising Income Segregation across Neighborhoods?" FRB of St. Louis Working Paper, no. 2006-037A (May 25, 2006). https://doi.org/10.20955/wp.2006.037

27. Hogan, Richard. "Class, Race, Gender, and Life Chances." Inequality and Stratification, 2015, 175–96. https://doi.org/10.4324/9781315662794-17

28. Khan, Naghma, and Hasan Mukhtar. "Tea Polyphenols for Health Promotion." Life Sciences 81, no. 7 (July 26, 2007): 519–33. https://doi.org/10.1016/j.lfs.2007.06.011

29. Gao, Yanyan. "Urban Bias, Rural-Urban Income Gap and Agricultural Growth: The Resource-Diverting Effect of Rural-Urban Income Gap in China." SSRN Electronic Journal, 2010. https://doi.org/10.2139/ssrn.1857269

30. Zhang, Yingqiang, and Tor Eriksson. "Inequality of Opportunity and Income Inequality in Nine Chinese Provinces, 1989–2006." China Economic Review 21, no. 4 (December 2010): 607–16. https://doi.org/10.1016/j.chieco.2010.06.008

31. Xiao, Yuxi, Haizheng Li, and Belton M. Fleisher. "The Earnings Effects of Health and Health-Related Activities: A Panel Data Approach." Applied Economics 47, no. 14 (2015): 1407–23. https://doi.org/10.1080/00036846.2014.1000521

32. Basu, Deepankar. "Bias of Ols Estimators Due to Exclusion of Relevant Variables and Inclusion of Irrelevant Variables." Oxford Bulletin of Economics and Statistics 82, no. 1 (November 19, 2019): 209–34. https://doi.org/10.1111/obes.12322

33. Khan, Naghma, and Hasan Mukhtar. "Tea and Health: Studies in Humans." Current Pharmaceutical Design 19, no. 34 (2013): 6141–47. https://doi.org/10.2174/1381612811319340008

34. Ridgeway, Cecilia L. "Interaction and the Conservation of Gender Inequality: Considering Employment." American Sociological Review 62, no. 2 (April 1997): 218. https://doi.org/10.2307/2657301

35. Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data. Cambridge, Mass, Mass.: MIT Press, 2011.

36. Conti, Gabriella. "Early Endowments, Education, and Health." Academia.edu, 2011. https://www.academia.edu/5594807/Early_endowments_education_and_health

37. Kenkel, Don, and Ping Wang. "Are Alcoholics in Bad Jobs?" National Bureau of Econoc Research working paper, 1999. https://doi.org/10.3386/w6401

38. "A Literature Review of the Costs of Smoking and Drinking." The Costs of Poor Health Habits, 1991, 137–42. https://doi.org/10.4159/harvard.9780674422261.c9

39. Duflo, Esther. "Women Empowerment and Economic Development." Journal of Economic Literature 50, no. 4 (2012): 1051–79. https://doi.org/10.1257/jel.50.4.1051

40. Gabbuti, Giacomo. "Labor Shares and Inequality: Insights from Italian Economic History, 1895–1970." European Review of Economic History 25, no. 2 (2020): 355–78. https://doi.org/10.1093/ereh/heaa010

8 Acknowledgement

Special thanks here toward Professor Mark Foley. His guidance ignites me and leads me step by step to our final goals. He leads through a vigorous and valuable memory of research. I also would like to thank my parents for providing support me all the way through my academical life.