博士論文

Essays in Empirical Health Economics

(実証医療経済学に関する研究)

Masato Oikawa
及川 雅斗

I am indebted to Daiji Kawaguchi for his advice and encouragement. I am also grateful to Michihito Ando, Eric Bonsang, Taiyo Fukai, Yuji Genda, Yoko Ibuka, Nobuyuki Izumida, Ayako Kondo, Wataru Kureishi, Nozomu Muto, Haruko Noguchi, Fumio Ohtake, Tsunao Okumura, Hideo Owan, Ryuichi Tanaka, Takahiro Toriyabe, Masanori Tsuruoka, Mutlu Yuksel, and the participants at the conferences and the seminars where my studies were presented for their helpful comments. Editing assistance has been provided by Philip C. MacLellan. All remaining errors are my own. I also sincerely thank for Hidehiko Ichimura, Hiroyuki Motegi, and Yoshinori Nishimura for their support and encouragement. Finally, I wish to thank my family for their supports and encouragement. This work is supported by JSPS KAKENHI Grant Number JP17J07578. This work is also supported by JSPS KAKENHI Grant Number 15H05692 (Principal Investigator: Ichimura Hidehiko). The use of data in this paper is approved by Ministry of Health, Labour and Welfare under this grant.

Data Acknowledgement

The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

The ELSA was made available through the UK Data Archive. ELSA was developed by a team of researchers based at the NatCen Social Research, University College London and the Institute for Fiscal Studies. The data were collected by NatCen Social Research. The funding is provided by the National Institute of Aging in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

This paper uses data from SHARE Waves 1, 2, 3 (SHARELIFE), 4 and 5 (`DOIs:10.6103/SHARE.w1.260,10.6103/SHARE.w2.260,10.6103/SHARE.w3.100,10.6103/SHARE.w4.111,10.6103/SHARE.w5.100`), see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005 -028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG00 5842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_ 04-064$) and from various national funding sources is gratefully acknowledged (see `www.share-project.org`).

The Japanese Study of Aging and Retirement (JSTAR) was conducted by the Research Institute of Economy, Trade and Industry (RIETI), Hitotsubashi University, and the University of Tokyo.

The vital statistics is provided by the Ministry of Health, Labour and Welfare.

Table of Contents

Chapter 1: Introduction

Numerous developed countries have faced the same problems of a decreasing birthrate and an aging population. As population ages, the cost of medical expenditure is increasing and it erodes the country' s budget. Japan is one of the most aging societies in the world. In 2014, the total medical expenditure was about 40.8 trillion yen and was about 8.3% of the gross domestic product. In the medical expenditure, its 35.5% was paid for the elderly aged over 75. [1] Reducing the cost of medical expenditure especially for the elderly is an important policy issue. In order to be efficient in policy making, it is very important to analyze the people's health condition and behaviors, which are the mechanisms behind their health.

In this dissertation, we analyzed the health related behaviors and the consequent health outcomes. This dissertation consists of three studies. In Chapters 2 and 3, we focus on the elderly people and analyze the effects of health related policies on behaviors and consequent health outcomes. In Chapter 4, we analyze the effects of a natural disaster on the pregnancy outcome, which is an important determinant of their health in life cycle.

In Chapter 2, I analyzed effects of health checkup on health outcomes and behaviors focusing on the heterogeneous effects depending on education. In Japan, the checkup policy was reconstructed based on the scientific evidence in April 2008 and would become more effective. I use this variation to estimate the effects of health checkup. Since the checkup is mandatory for salaried workers but voluntary for self-employed workers, the participation rate of the checkup is significantly higher among salaried workers. In other words, salaried workers have more proportion of individuals affected by the policy reform. Using this institutional setting, I regard salaried workers as treatment group and self-employed workers as control group and employ a difference-in-differences approach. According to the estimated results, by the policy reform, university graduates with relatively high obesity risk significantly decrease the Body Mass Index and some diagnosed health problems although there are no significant

[1] See more details on `https://www.mhlw.go.jp/english/wp/wp-hw11/dl/02e.pdf`

changes among non-university graduates with the risk. Additionally, there are significant changes in the health behaviors such as physical activity and energy intake only among the university graduates.

In Chapter 3, we analyzed the effect of retirement on cognitive function; specifically, the hypothesis from human capital theory that because cognitive investment increases a worker's wage, workers may invest in their cognitive ability more than retirees, contributing to a post-retirement decline in cognitive function. While this topic is of great interest to health economics, we show that the method of analysis of some previous studies is not valid for examining this effect, and we propose an alternative method that addresses this concern. Further, our estimates indicate that retirement has only a weak effect on cognitive ability in a wide range of analyzed countries and heterogeneous groups. Therefore, according to our analysis, policies that have been widely adopted in developed countries to delay retirement, such as increasing the pensionable age, appear to have little detrimental affect on post-retirement cognitive ability. This chapter is based on Nishimura and Oikawa (2017a)

In Chapter 4, we analyzed the effects of the Fukushima-Daiichi Nuclear Power Plant accident on pregnancy outcomes of the babies in-utero on the date of the accident in Fukushima prefecture. The radioactive substances has been released into outside of the plant because of the meltdowns caused by the earthquake with a giant tsunami and has influenced throughout Japan. In this paper, we focus on the effects of the Fukushima-Daiichi accident on pregnancy outcomes. According to the estimated results, we found the negative effects of the Fukushima-Daiichi accident on the pregnancy outcomes among boys. We also found the heterogeneous effects depending on the timing of conception. The effects among the boys who affected by the accident during their 3rd trimester are robust with respect to the various measure of birth outcomes.

Chapter 2: Effects of The Health Checkup on The Health Outcomes and Behaviors: Heterogeneous Effects Depending on Education

1 Introduction

Estimating the effect of education on economic as well as non-economic outcomes has attached much of attention of economists. One important issue is the relationship between education and health conditions and, from Grossman (1972), there are numerous studies that investigate the relationship theoretically and empirically. Grossman (2006) reviews the theory and empirical evidence regarding the education-health gradient and Eide and Showalter (2011) and Grossman (2015) review the recent empirical studies. While most of the empirical studies focus on the causal effect of education on health, less attention has been paid to the mechanism behind the causal relationship. Grossman (2015) pointed out that the mechanisms of the education effects on health and health behaviors is one of the future research in this topic and understanding these mechanisms should have implications for the health related policies.

One hypothesis that explain the mechanism behind the causal relationship is that higher educated people more efficiently input investments into a health production function, this is known as the allocative efficiency hypothesis. In terms of allocative efficiency, higher educated people more quickly respond to a new health information. Some previous studies found that individuals with higher education more quickly response to health information using the variation from the new approved drug use (Lleras-Muney and Lichtenberg, 2002), medical research publication (Price and Simon, 2009), and a report that announced the smoking risk, the 1964 Surgeon General Report on Smoking and Health (de Walque, 2010;

Aizer and Stroud, 2010).

In some of the studies, the distance between information and behaviors is far and it is difficult to discuss more details of the mechanism. There are at least two possible path for the heterogeneous response to health information depending on education. One is that the higher educated have a preference to new health information and quickly access the information and the other is that the higher educated can process the new health information more efficiently and quickly implement. Analyzing these mechanisms is important for policy implementation. If the higher educated have a preference to new health information, expansion of information provision improve th health conditions among the lower educated individuals. On the other hand, if the higher educated can process the new health information more efficiently, the guidance from professionals helps to process the new information and improve health conditions among the lower educated individuals. In this paper, I use a variation of health information from a Japanese health checkup policy to be clear the mechanism behind the causal relationship between education and health. In Japan, the health checkup is mandatory for salaried workers and almost all of them have the checkup and receive health information. Therefore, since most of them have opportunity to access health information, by examining the educational heterogeneity of response to health information from the checkup, we can discuss the possibility that the higher educated can process the new health information more efficiently.

There are some studies analyzing the effects of health checkup on health outcomes and behaviors. In these studies, while the studies found the evidence for the effects on physician visit (Iizuka et al., 2017), medication (Kim et al., 2017), medical expenditure (Hackl et al., 2015; Iizuka et al., 2017), and fat intakes(Zhao et al., 2013), there is few evidence for the health outcomes (Kim et al., 2017; Inui et al., 2017). Additionally, there is few paper focusing on the heterogeneous effects depending on the educational attainment except for Zhao et al. (2013). Zhao et al. (2013) analyzed the effects of negative health information on consequent nutrition intakes and the heterogeneous effects depending on income and education. They regarded the diagnosis of hypertension around the biomaker threshold as random and employed the regression discontinuity design (RDD) to estimate the effect of the diagnosis of hypertension on nutrition intakes using a Chinese survey data and found that the exceeding the threshold decreases the individual fat intake. However, after dividing the whole sample by education, they found that the estimates are statistically insignificant in all educational groups.

In this paper, I analyze the effects of introduction of the Specific Health Checkups and Specific Health Guidance (SHC-SHG), which is a Japanese checkup system introduced on April 2008, on the health outcomes and behaviors focusing on the heterogeneity of the education. Since the checkup is mandatory for salaried workers but voluntary for self-employed workers, there is a difference in participation rate of the checkup between salaried and self-employed workers. Therefore, salaried workers have higher proportion of individuals affected by the policy reform and the treatment intensity is larger for salaried workers. Basic idea of the identification strategy is to compare the before-after change of the group with the higher participation rate before policy reform and that of the group with the lower, that is the difference-in-differences (DID) approach. Additionally, the panel structure of the dataset used in this paper allows me to control for the individual unobserved heterogeneity. Using the DID approach, I estimate effects of the policy reform on not only health behaviors but also health outcomes.

According to the estimated results, by the policy reform, university graduates with relatively high obesity risk significantly decrease the Body Mass Index (BMI), which is a measure of obesity, and some diagnosed health problems although there is no significant changes among the non-university graduates with high obesity risk. Additionally, there are significant changes in the health behaviors such as exercise and nutrition intakes among the university graduates. By considering the educational heterogeneity, effects of checkup are significantly estimated in both health behaviors and outcomes. These results suggest that there is the educational heterogeneity of the response to the health information even if there is less heterogeneity to access the health information.

The results are not consistent with the results of Zhao et al. (2013). Since the checkup analyzed in this paper includes a health guidance by professionals in addition to the information provision from checkup, it is possible that the treatment intensity of the checkup is stronger than that for Zhao et al. (2013). This point is consistent with Kim et al. (2017), which found the evidence for improvement of health conditions due to a health screening combined with further intervention but no evidence for that without further intervention. Additionally, as authors argued, the sample size for higher education group relatively small and there is a possibility that the estimation is not precise due to the small sample size.

The remainder of this paper is organized as follows: section 2 explains the institutional setting; section 3 discusses identification strategy and estimation model; section 2 explains the data used in this paper and some descriptive statistics; section 5 discusses the estimation results; section 6 discusses some remarks on results; and section 5 concludes this research.

2 Institutional Background

2.1 Health Checkup in Japan

In Japan, health checkups are provided by laws as part of health promotion policy. Employer are obligated to implement a health checkup for salaried workers by a law, Industrial Safety and Health Act, and by the law, salaried workers are obligated to undergo the checkup. Therefore, almost all of salaried workers receive the health checkup. Additionally, they receive unified contents which is established by the law. On the other hand, local governments provide a health checkup for residents aged over 40 and contents of checkup are almost same as that for salaried workers. People who is not salaried worker have opportunity to receive the health checkup. Thus, all middle aged or older aged Japanese have opportunity to receive health checkup and it is mandatory for salaried workers and voluntary for others such as self-employed workers.

However, sufficient improvement in health is not observed while there are publicly provided health checkups. In the mid-term evaluation of Health Japan 21, which is a health promotion policy, there are increases in patients of lifestyle-related diseases, which have a large proportion in medical expenditure and are caused by the individuals' life-styles. For example, there are increases in diabetics and pre-diabetes and in the proportion of obese males. The council of governments summarized problems of the checkups: not precise screening process of patients of the lifestyle related diseases; and the quality of the health guidance (Ministry of Health and Welfare, 2014). As a result of these discussion, for a prevention of these non-communicable diseases, a new health checkup system, the Specific Health Checkups and Specific Health Guidance, was introduced in April, 2008. The checkup system gives the information about objective health risk evaluated based on the scientific evidence and/or the guidance for health behaviors by professionals to the participants.

2.2 Specific Health Checkups and Specific Health Guidance

The purpose of the newly introduced system is to screen the patients of the metabolic syndrome, which is a condition increasing the risk of lifestyle-related diseases such as heart disease, stroke, and diabetes, and to improve their health conditions. The system focuses on people insured by public health insurance and their dependent aged between 40 and 74. Since, Japan introduces a universal health care insurance system, the target population cover almost all of the Japanese. The system is divided into two parts: the first part is the specific health checkups; and the second is the specific health guidance. In the first part, participants have the health checkups whose purpose is screening of the actual and potential patients of the metabolic syndrome. The content of the checkup is determined by the scientific evidence for identifying the patients of the metabolic syndrome. The checkup includes body measurements, blood tests, and questionnaires about such as smoking and medication histories. As a result of the checkup, examinees are assigned the eligibility status of the health guidance based on their health risk. Therefore, by receiving the guidance status, participants are able to know about their own health risk precisely.

In the part of the specific health guidance, participants with higher health risk receive the health guidance about their health behaviors by the experts, such as doctors and registered dietitian, for improving their health condition. There are two types of guidance status, the motivational support and the active support, and, participants with lower risk are only provided the information about their health conditions and the basic knowledge of the lifestyle-related disease. The participants are able to obtain more information by receiving the guidance. There is the heterogeneity in the treatment intensity depending on the guidance status even in the participants of the checkup.

The screening process is divided into two parts: the body measurement; and the additional risk factors. The procedure of the guidance status is as follows:

- First, examinees are divided by their girth of abdomen, examinees whose abdomen is over the criteria (male:85cm, female:90cm) are assigned to the group A.
- Second, even though their girth of abdomen is under the criteria (not in group A), when their value of body mass index (BMI) is above 25, they are assigned to the group B .
- Additionally, examinees who are in the group A or B are evaluated their risk level by additional four risk factors, the high blood sugar, the lipid abnormality, the high blood pressure, and smoking history. [2]
- In the group A, examinees with more than two risk factors receive the active support

[2] Smoking history is counted only when examinees have either other risk factors.

guidance, and examinees with one receive the motivational support guidance, and examinees without risk factor are provided the information about their health and do not receive any guidance.

- Similarly, in the group B, examinees with more than three risk factors receive the active support guidance, and examinees with one or two receive the motivational support guidance, and examinees without risk factor are provided the information.
- And examinees who are not in the group A and B are provided the information about their health and do not receive any guidance.

Figure 1 summarizes the procedure. In the body measurement part, the criteria of BMI is more strict than the that of girth of abdomen. Additionally, if participants have at least one additional risk factor, then they have the eligibility of receiving the guidance.

Due to the checkup policy reform, participants are able to know about the information about their precise health risk evaluated based on the scientific evidence and knowledge about the health behaviors from professionals. In this paper, I try to analyze effects of this information improvement on health outcomes and behaviors focusing on the heterogeneity in the effects depending on education levels. In the next section, I will discuss detail of the identification strategy.

3 Identification Strategy

In this paper, I use the heterogeneity in the participation rate of checkup to estimate effects of the checkup policy reform on health outcomes and behaviors. As mentioned before, the checkup is mandatory for salaried workers but voluntary for other groups. Most of salaried workers have the health checkup and the participation rate among salaried workers is higher than that for self-employed workers. Therefore, salaried workers have higher proportion of individuals affected by the policy reform because of the difference in the participation rates. Due to this heterogeneity, the treatment intensity should be larger for salaried workers than that for other individuals. Basic idea of the identification strategy is to compare the before-after change of the group with higher treatment intensity and that of group with lower treatment intensity, that is the difference-in-difference (DID) approach. I regard salaried workers before the policy reform as treatment group and self-employed workers as control group and employ DID approach. In this paper, I do not include individuals who do not work in the control group because not working individuals such as retired or disabled may have different trend compared to salaried workers after controlling for observable and unobservable characteristics.

For the validity of this DID approach, there are some remarks. One is a possibility that self-employed workers change their checkup behavior by the policy reform. As I mentioned above, the checkup is voluntary for self-employed workers while is mandatory for salaried workers. If self-employed workers begin to have checkup after policy reform and the participation rate becomes close to that for salaried workers, the difference in the participation rate between salaried and self-employed workers shrinks and above DID approach does not work well. However, the participation rate does not change significantly after policy reform. Figure 2 shows the participation rate of health checkup by employment status and there is no significant jump around 2008 in both salaried and self-employed workers.

Additionally, the common trend assumption is important for the causal interpretation of DID approach. Namely, change of health condition among salaried workers must be the same as that for self-employed workers in a case without policy reform. While many studies using DID approach checked the trend of the target outcomes before policy introduction for discussing the validity of common trend assumption, I cannot check the pre-trends because the dataset used in this paper has only one period before the policy reform. However, fortunately, the dataset is a longitudinal data and includes rich information about demographic, economic, and health related variables. Therefore, I can control for the individual observable characteristics and the time-invariant unobserved individual heterogeneity. Additionally, I estimate a placebo regression using the health conditions less related to the newly introduced system. I will explain the details of dataset and the placebo regression in later parts. Under the common trend assumption, the DID estimator represents the difference in policy reform effects, which is deducted by participation rate, between salaried and self-employed workers. If signs of effect are same between salaried and self-employed workers, the DID estimator is interpreted as the lower bound of treatment effects for salaried workers in terms of absolute value.

3.1 Estimation Model

I estimate the following equation to control for the bias from the heterogeneous trends between salaried and self-employed workers:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 SalariedWork07_i + \beta_2 After_t \\ &\quad + \beta_3 SalariedWork07_i \cdot After_t + x'_{it}\delta + \theta_i + \epsilon_{it} \end{aligned} \tag{1}$$

where i and t is indices of individual and time, respectively. The dependent variable y_{it} represents health outcomes such as BMI and diagnosed health problems and health investment behaviors such as physical activity and energy intake. $SalariedWork07_i$ takes one if respondent is a salaried worker before policy reform, i.e., 2007. $After_t$ takes one after the policy reform, i.e., 2009, 2011, and zero for 2007. The vector, x_{it}, is a set of control variables and that includes age, age squared, marital status, number of children, household income, house ownership, hours of work, stress condition at workplace. The parameter, θ_i, is the unobserved individual fixed effects and the parameter, ϵ_{it}, is an unobserved error term. In equation (1), the parameter β_3 is correspond to the DID estimate and a target parameter of this paper. I estimate the equation (1) in both university graduates and non-university graduates and compare the DID estimates to discuss the heterogeneity of effects of the policy reform.

As mentioned above, the individuals with the higher health risk receive not only the checkup but also the guidance. Moreover, individuals with the lower risk do not need to change the behaviors because they are already healthy. In these reasons, there is a possibility that the treatment effects are heterogeneous depending on individuals' potential heath conditions. In the estimation, I divide the sample by using the BMI value before the policy reform to take into consideration this heterogeneity.

For the causal interpretation of the DID estimation, the common trend assumption is necessary. To control for the heterogeneity of trends between salaried and self-employed workers, I add some observable variables into the estimation model and apply the fixed effects estimation using the panel structure of the dataset I use. I add the demographic variables into the model such as age, marital status and so on. Additionally, Ssnce some previous studies show the relationship between health and economic conditions (Case et al., 2002; Chetty et al., 2016; Semyonov et al., 2013), I also use the economic conditions as control variables. The dataset includes the financial crisis of 2008 and this may affect the individuals' workplace and the economic condition of their residential region heterogeneously. Therefore, I use the job status, workplace related variables (hours worked, physical stress at workplace, job stress at workplace, occupation-year fixed effects), and time-variant regional characteristics as the controls. Additionally, I use two prefecture level macro economic variables, GDP and income per capita as a regional characteristics. The accumulation of health stock until middle age and the preference in health may cause the heterogeneity in the trends. The fixed effects estimation controls for the time-invariant unobserved individual heterogeneity, which causes the potential bias from the time-invariant unobserved characteristics.

While control variables and fixed effects estimation control for the observable characteristics and the time-invariant unobserved heterogeneity, the model does not address time-variant health shocks. For addressing this issue, I run a placebo regression. The dataset includes the information about the diagnosis of some diseases. As a placebo regression, I regress the same estimation model using the diagnosis condition of the disease which is less related to the metabolic syndrome as a dependent variables.

4 Data

This paper uses the Japanese Study of Aging and Retirement (JSTAR). The JSTAR is a biennial panel survey of elderly Japanese aged over 50 and a sister datasets of the Health and Retirement Study in the U.S., the English Longitudinal Survey on Aging in the England, and the Survey on Health, Aging, and Retirement in Europe in European countries. The JSTAR conducted since 2007 and the 2007 JSTAR is conducted at five cities in Japan and the sample cities were added at the 2nd and 3rd waves of the JSTAR. I use the original five cities sample because the sample allows me to use the data before and after 2008, the year of the introduction of the checkups. The JSTAR includes the information about the demographics, the labor force status, the economic variables, the health investment behaviors, and the health outcomes. These information allows me to analyze the policy impact of the checkups system on health outcomes and behaviors. [3]

I use the samples surveyed from the 2007 JSTAR because the analysis needs the data from pre- and post- policy reform and samples who answer both the main survey and nutrition survey. I restrict the analysis sample to the males aged between 50 and 60 because the JSTAR includes the sample aged over 50 and most of the Japanese people retire at the age of 60. I regard the individuals who have university degree or more as the higher educated

[3] Please see `http://www.rieti.go.jp/en/projects/jstar/index.html` for more detailed information about the JSTAR.

group and those with less than university degree as the lower educated group.

First, I show the basic characteristics of salaried and self-employed workers in both the university graduates and the non-university graduates. Table 1 shows the average values of some observable characteristics by the educational groups and the employment status at pre-period. According to the Table 1, there seems to be no significant difference in characteristics between the salaried workers and the self-employed workers. One important issue is differences in working hours. Among the non-university graduates, self-employed workers work three more hours than salaried workers. The difference in the hours worked causes the heterogeneity in the leisure time between the employed and the non-employed and in the health investment behaviors which need the time such as exercising. Therefore, in the DID estimation, it is important to control for the hours worked. Additionally, among the university graduates, there is the differences in the proportion of individuals whose BMI is over 25 between the salaried workers and the self-employed workers. Therefore, trends might be heterogeneous because the baseline BMI conditions are different between salaried and self-employed workers. For addressing this issue, I divide the sample by health condition before the policy reform.

I use the BMI value before the policy reform to divide the sample by health condition by health before the policy reform. In the new health checkup system, people with higher health risk have the eligibility of having the health guidance and have additional treatment. The treatment effects may be heterogeneous due to the potential condition before policy reform. For addressing this issue, I control for the pre-treatment health condition. Namely, I divide the samples by the health condition before the policy reform and apply the DID estimation. To identify the guidance eligibility, we need the information about the girth of abdomen, BMI, the blood pressure readings, blood glucose level, and level of triglyceride. However, at the pre-treatment period, 2007, the JSTAR only have the information about BMI. Therefore, I use the BMI at pre-treatment period to divide the sample. Since the 2009 JSTAR includes the information about some health conditions for identifying the guidance eligibility such as the girth of abdomen, BMI, and the blood pressure readings, I construct the eligibility status of the health guidance and discuss the criteria of the BMI value for the sample division. [4] Figure 3 shows the proportion of having eligibility of the health guidance over BMI. According to Figure 3, the probability of having eligibility increases as BMI becomes higher and there is a jump of the probability at a BMI value 23.5. The probability of having guidance eligibility is about 40 % at BMI value 23 and 70 % at BMI value 23.5. The probability increases by about 30 percentage points by exceeding BMI value 23.5. Therefore, I use the value, 23.5 for the criteria of the sample division to control for potential health risk.

5 Estimation Results

5.1 The Effects on Health Conditions

In this section, I show the estimation results of Equation 1. First, I discuss the effects of the policy reform on health condition using BMI as a measurement of health condition. Table 2 shows the estimation results of effects on level of BMI and is divided by two panels. Panel A of Table 2 reports estimated effects among individuals whose BMI before policy reform is more than or equals to 23.5, that is individuals with higher obesity risk and higher probability of having health guidance. On the other hand, Panel B shows the results for individuals whose BMI before policy reform is less than 23.5. Columns (1) and (2) are the results for the university graduates and I use two types of control variables. Column (1) consists only basic control variables, and Column (2) is basic controls and prefecture-level macroeconomic variables. Similarly, Columns (3) and (4) are the results for the non-university graduates. I apply fixed effects estimation for all models, that is controlling for the individual

[4] The eligibility that I construct is not accurate because of the lack of the information about blood glucose level and level of triglyceride which is the determinant of risk factors.

time-invariant heterogeneity, and report only β_3 of equation 1, the DID estimates.

According to Table 2, in the individuals whose BMI before policy reform is above 23.5, the DID estimates for the university graduates are negatively and statistically significantly estimated at 1% level and robust for control patterns. The DID estimate ranges from -1.52 to -1.47 after controlling for the observable characteristics and unobserved time-invariant heterogeneity. (Panel A, Columns (1) and (2)) Since the average value of BMI for salaried workers before policy reform is 25.8, due to the policy reform, the value of BMI is reduced to about 24.3. I interpret this reduction using a medical study, Tsugane et al. (2002). Tsugane et al. (2002) analyzed the relationship between BMI and all-cause mortality for middle aged Japanese and found that the mortality for male has the U-shape and the bottom of the U-shape is at BMI value, 23.0-24.9. Therefore, according to Tsugane et al. (2002), the reduction of BMI among the university graduates is interpreted as the improvement of BMI condition. On the other hand, there are no statistically significant effects among the non-university graduates. These results suggest that the higher education has important role for improvement of the BMI conditions. Additionally, in the individuals with low obesity risk (Panel B), the DID estimates of both university graduates and non-university graduates are small and statistically insignificant. Individuals with low health risk do not need to change their behavior because they are already healthy. Moreover, they are likely to have only information provision. Therefore, treatment effects for individuals with lower obesity risk can be expected to be weaker than those with higher risk and the results are consistent with this discussion.

As another measure of health condition, I use the doctor diagnosed health conditions. The JSTAR include the information about the condition of diseases diagnosed by doctor. Using these information, I construct the number of the metabolic syndrome related diseases diagnosed by doctor. [5] Columns (1) and (2) of Table 3 show the estimation results of effects

[5] I regard heart disease, high blood pressure, hyperlipidemia, cerebral stroke, diabetes, and cancer as metabolic syndrome related diseases.

on metabolic syndrome related diseases. In Table 3, I use the same control variables as Columns (2) and (4) of Table 2 and apply fixed effects estimation. According to Table 3, the DID estimate for metabolic syndrome related diseases is negative and statistically significant among university graduates with high obesity risk (Column (1) in Panel A). On the other hand, there are no statistically significant coefficients among university graduates with low obesity risk (Column (1) in Panel B) and non-university graduates (Column (2) in Panels A and B). These results are consistent with the results for level of BMI and suggest that the decrease in BMI among the university graduates with high obesity risk improve consequent health problems.

The estimated results show the improvement of health conditions among the university graduates with obesity risk. However, as mentioned above, in the estimation, the model cannot control for the time-variant unobserved health shocks while the model control for time-variant observables and time-invariant unobserved heterogeneity. If there is a positive health shock for only salaried workers, we cannot attribute the DID estimate to the effect of policy reform or the health shock. In this paper, I check the existence of differential health trends depending on employment status by running same regression for diagnosed health condition less related to metabolic syndrome. The newly introduced system focus on the metabolic syndrome and the diagnosed condition of those diseases should be less affected by the policy reform. I construct a variable for disease which is less related to the metabolic syndrome. [6]

Columns (3) and (4) of Table 3 shows the estimation results of effects on other diseases. The DID estimates are statistically insignificant except for non-university graduates with lower obesity risk (Column (3) in Panels A and B and Column (4) in Panel A). At least, these results do not suggest the existence of heterogeneous trends of health shocks between

[6] The diseases consists all the other diseases asked in the JSTAR than the metabolic syndrome related diseases such as chronic lung disease, asthma, liver disease, ulcer or other gastrointestinal disorder, joint disorder, femoral neck fracture, osteoporosis, eye disease, ear disorder, bladder disorder, Parkinson' s disease, depression, dementia, and skin disorder.

salaried and self-employed workers among university graduates and non-university graduates with higher obesity risk. On the other hand, the DID estimate is negative and statistically significant among non-university graduates with low obesity risk (Column (4) in Panel B). Therefore, it needs to take care about the interpretation of results among non-university graduates with low obesity risk.

Tables 2 and 3 suggest the importance of the education for the improvement of BMI value and the metabolic syndrome related disease by the checkup policy reform. On the other hand, the mechanism behind the improvement is not clear. Therefore, to be clear about the mechanism, in the next subsection, I estimate effects on the health investment behaviors.

5.2 The Effects on Health Investment Behaviors

In this subsection, I analyze effects of the policy reform on health investment behaviors. In the newly introduced checkup, individuals with higher health risk receive the guidance, which is based on scientific evidence and is tailored to the physical condition of them, from professionals. By receiving the guidance, they would update the knowledge of productivity of health investments and change their behavior to appropriate levels. As health investment behaviors related to the BMI improvement, I focus on a physical activity and an energy intake. Casazza et al. (2013) discussed scientific evidence for some beliefs about obesity by reviewing medical studies. According to Casazza et al. (2013), there are scientific evidences for physical activity and reducing energy intakes to promote weight loss and increase health. Therefore, I estimate effects of policy reform on physical activities and energy intakes for discussing the mechanism behind health improvements.

Table 4 shows the estimation results for physical activity and energy intake. Since the JSTAR conducts the question about physical activities and the nutrition surveys, I use the information to construct the variables of health investments. As a physical activity variable, I construct a dummy which is equal to one if the respondent does at least one of following exercises, doing exercise more than 90 minutes in a weekday, doing exercise more than 90 minutes in a weekend, and walking more than 90 minutes. I use the value of the energy intake directly from the nutrition survey data and in the estimation. In Table 4, I use the same controls as Columns (2) and (4) of Table 2 and include individual fixed effects.

According to Table 4, the estimated coefficient for physical activity is positive and statistically significant only in university graduates with high obesity risk before policy reform. (Column (1) in Panel A) The magnitude of coefficient is about 0.4 and implies that the policy reform increases the probability of physical activity by about 40 percentage points. Since the proportion of doing physical activity before policy reform among salaried workers is about 47 %, after the policy reform, the result implies that almost salaried workers do physical activity. Additionally, in university graduates with high obesity risk before policy reform, the coefficient on energy intake is positively and significantly estimated. (Column (3) in Panel A) The estimate implies that salaried workers increase their energy intake by about 523 kcal after policy reform and it reaches about 2760 kcal, which is close to a level of required energy intake for sedentary work with commuting, shopping, and light sports. [7] Since, among the salaried workers, the level of energy intake before policy reform is close to a level for sedentary life-style, the DID estimate implies that salaried workers increase their

[7] According to National Institute of Health and Nutrition (2010), the EER is determined by multiplying the basal metabolic rate (BMR) and the physical activity level (PAL). By following Ganpule et al. (2007) which developed prediction equations of BMR using body measurements, I calculate the BMR and it is about 1527 kcal per day among salaried workers before policy reform in university graduates with higher potential obesity risk.

The basal metabolic rate (BMR) is calculated by $BMR = 0.1238 + 0.0481 \times Weight(kg) + 0.0234 \times Height(cm) - 0.0138 \times Age - 0.5473$ (this is equation for males), with 74.6kg, 170.0cm, and 54.3 years old which are average weight, height, and age for treated before policy reform.

By following the PAL used in National Institute of Health and Nutrition (2010), the EER are about 2290, 2670, and 3050 kcal for low, moderate, and high PAL. I use three values of the PAL used in National Institute of Health and Nutrition (2010), 1.5, 1.75, and 2.0 for low, moderate, and high levels.

The low physical activity corresponds to sedentary, the moderate activity for sedentary work with commuting, shopping, and light sports, and the high activity for works with vigorous physical activity or high-intensity leisure-time physical activity.

energy intake from the level with sedentary life-style to the level with moderate physical activity. Moreover, according to Figure 4, [8] the estimated coefficients on the dummy which takes one when the energy intakes is below 1900 kcal and the dummy for above 3100 kcal are negatively and statistically significant. (Panel (a)) Therefore, these results also suggest that the energy intake is close to the appropriate level.

The estimated results among the university graduates are consistent with medical evidence about obesity. Enough physical activity helps long-term weight maintenance while keeping on reducing energy intake does not work well in the long-term. (Casazza et al., 2013) Hill et al. (2012) argued that balancing energy intake and energy expenditure at a high level is more feasible ways to reduce obesity than doing only food restriction. Therefore, the increasing in physical activity and changing in energy intake as reaching the appropriate level seem to be one mechanism behind the health improvement among the university graduates observed in Tables 2 and 3.

On the other hand, there are no significant changes among the non-university graduates with high health risk. Additionally, the magnitude of coefficients are smaller than those for the university graduates discussed above and statistically insignificant. (Columns (2) and (4) in Panel A of Table 4) These results suggest the importance of education for health improvement by having the checkup and the guidance.

6 Discussion

In the previous section, I analyze the educational heterogeneity in the response to the policy reform by comparing the DID estimates of university graduates and non-university graduates. We need some additional remarks for the comparison even under the common

[8] Figure 4 shows the DID estimates on energy intakes by education and health condition before policy reform. I estimate the same model of Table 4 using the dummy variable which is equal to one if a respondent takes in the energy in a value range.

trend assumption. One possible concern is the educational heterogeneity in the difference in participation rate between salaried and self-employed workers. Suppose that the participation rate for self-employed workers is close to that for salaried workers among non-university graduates, that is the difference in participation rate is close to zero, and suppose that there is enough difference among the university graduates. Then, the DID estimate for the non-university graduates is smaller than that for the university graduates even if treatment effects have same magnitude for each education group because there is no difference in participation rate, which is a source of difference in treatment intensity, among non-university graduates. For comparing effects of policy reform using the DID estimates, the differences in participation rate of checkup should be same between the university graduates and the non-university graduates. In Section 2, I will confirm the differences in participation rate for both the university graduates and non-university graduates.

Table 1 also shows the checkup ratio before the policy reform by education and employment status. Among the university graduates, while almost all the salaried workers have the checkup (93.2%), the participation rate of self-employed workers is about 62.5% and the difference is the about 31 percentage points. Among the non-university graduates, the participation rate is about 84.4% for salaried workers and about 52.5% for self-employed workers and the difference in the checkup ratio is about 32 percentage points. These results suggest that the differences in the participation ratio are almost same level between the university graduates and the non-university graduates.

Additionally, there are the significant differences in participation rate between salaried and self-employed workers after controlling for some observable characteristics. Table 5 shows the estimation results with regressing the checkup participation on the employment status and other observable characteristics. According to Table 5, in both educational groups, there are the statistically significant differences in the checkup ratio between the salaried workers and the self-employed workers. (university graduates:0.306 and non-university graduates: 0.323)

After controlling for some observable characteristics, the magnitudes of the difference are almost same between the university graduates and non-university graduates. These results suggest that the difference in the checkup ratio among the salaried workers and the self-employed workers is due to the institutional settings rather than the other characteristics. Additionally, since the difference in participation rate is almost same level among both the university graduates and the non-graduates, if the identification assumptions discussed above are satisfied, we can interpret the difference in the DID estimates as the heterogeneity in the treatment effects rather than the heterogeneity in the checkup ratio due to the educational difference.

Additionally, it is difficult to interpret the comparison of DID estimates when there is a heterogeneous effect depending on employment status. As mentioned above, the newly introduced system has health checkup and health guidance and set of treatments received from the system is different depending on health risk of participant. Thus, it is possible that treatment effects are heterogeneous depending on health risk. Suppose that the self-employed with lower education is unhealthy and receive both checkup and guidance while the self-employed with higher education is healthy and receive only checkup. Then, self-employed workers with lower education receive more treatments than those with higher education and it is possible that the treatment effect for self-employed with higher education is stronger than that for the university graduates. Moreover, effects may be weak among individuals with low health risk because they are already healthy and do not have incentive to change their behaviors. In these, case, since, the DID estimate is the difference in effects between salaried and self-employed workers under common trend assumption, the DID estimate for non-university graduates is smaller than that for university graduates even when treatment effects are same across salaried workers with higher and lower education and the difference in participation rate are same between the university graduates and non-university graduates. Therefore, the difference in potential health risk by employment status and education make it difficult to interpret the comparison of DID estimates. In the estimation, I divided the sample by health condition before policy reform and address this potential issue.

7 Conclusion

This paper examines the effects of Japanese checkup policy reform on health outcomes and behaviors focusing on the heterogeneity of the education. Since, in Japan, the health checkup is mandatory for salaried workers but voluntary for self-employed workers, there are difference in the participation rate of health checkup between salaried and self-employed workers. I use the difference in the intensity of participation rate of the health checkup to control to estimate effects of policy reform. I regard salaried workers as treatment group and self-employed workers as control group and employ the DID approach.

According to the estimated results, the university graduates with relatively high risk of obesity significantly improve the BMI level and diagnosed metabolic syndrome related diseases although there is no significant change among the non-university graduates with high obesity risk. Moreover, the results shows the increase in physical activity and the increase in energy intake which is close to moderate energy requirement level among the university graduates. These results suggest that the university graduates improve their health behaviors and consequent health outcomes by the checkup policy reform.

Chapter 3: Mental Retirement: Evidence from Global Aging Data

1 Introduction

Due to rising life expectancies and declining birthrates associated with economic development, many industrialized countries are now facing the problem of an aging population. In 2015, there were 900 million people over 60 years of age worldwide, and this number is expected to continue to grow rapidly. As a country's population ages, the cost of social security and welfare increases, eroding the country's budget, and so numerous developed countries have introduced retirement-related policies such as pension system reform in order to reduce the cost of social security and social welfare to a sustainable level. Pension reforms in developed countries are mostly targeted at delaying retirement, and the United States, United Kingdom and Korea, for example, have decided to increase the age of pension eligibility, while Japan has already done so. The relationship between social security and retirement in developed countries has attracted a fair amount of attention in economics (Gruber and Wise (1998)), and one of the key factors for policymakers in evaluating the effects of these reforms is the health of retirees. An active and extended work life can be seen as beneficial to the health of the elderly because it might lead to a reduction in the often rapid growth in medical expenses throughout retirement.

However, longer life expectancies are associated with increased prevalence of chronic diseases such as dementia. According to the 2015 World Alzheimer Report [9] , the global cost of dementia has increased from USD 604 billion in 2010 to USD 818 billion in 2015, an increase of 35.4 percent. In the US, the total monetary cost of dementia in 2010 was estimated to be between 157 and 215 billion dollars, with about 11 billion dollars of this cost paid

[9] See https://www.alz.co.uk/research/world-report-2015 for further details.

by Medicare (Hurd et al. (2013)). Although the rising cost of dementia and other chronic diseases associated with the elderly make the relationship between retirement and health of great research and policy interest, to date, there is no consensus among the studies of the past two decades either on the mechanism by which retirement affects health or even on the direction of its impact. While some studies conclude that retirement has a positive impact on either mental or physical health, others conclude that retirement has either a negative effect or none at all. [10]

Just as with the effect of retirement on health, the discussion in the literature of the effects of retirement on cognitive function is also ambiguous. While Adam et al. (2006) find a positive effect of occupational activities on the cognitive function of the elderly in Europe and Coe et al. (2012) find a positive relationship between retirement duration and cognitive functioning but only for blue-collar workers, other studies of Europe find either no clear relation at all (Coe and Zamarro (2011)) or a negative effect (Mazzonna and Peracchi (2012)) between retirement and cognitive function. Exploring the theoretical foundations for their empirical study, Rohwedder and Willis (2010) discuss two hypotheses explaining why retirement might cause cognitive function to decline. The first, the "unengaged lifestyle hypothesis", is a mental retirement effect in which cognitive decline may result from a worker lacking cognitive stimulation after retirement. The second hypothesis, an "on-the-job" retirement effect, is based on the human capital production function (Ben-Porath (1967)), which relates inputs such as one's current stock of human capital and investments in schooling or

[10] Kerkhofs and Lindeboom (1997), among the first to suggest an endogeneity in decisionmaking regarding retirement and health, find in their study of the Netherlands using the fixed effect (FE) method that the Hopkins Symptom Checklist health index can be improved by taking early retirement. But when Lindeboom et al. (2002), also applying an FE method to Dutch data, extend the Kerkhofs and Lindeboom (1997) study to other indices such as the Mini Mental State Examination test on cognitive ability and the CES-D test of depressing feelings, they find different results. In another early study, Charles (2004) examines the causal effect of retirement on health by focusing on the subjective wellbeing of retirees using an instrumental variables (IV) approach. Since then, numerous studies have analyzed the effect of retirement on various health indices, including Bound and Waidmann (2007), Coe and Lindeboom (2008), Dave et al. (2008), Neuman (2008), Johnston and Lee (2009), Latif (2011), Coe and Zamarro (2011), Kajitani (2011), Behncke (2012), Bonsang et al. (2012), Mazzonna and Peracchi (2012), Hernaes et al. (2013), Bingley and Martinello (2013), Hashimoto (2013), Insler (2014), Kajitani et al. (2014), Hashimoto (2015), and Kajitani et al. (2016a).

on-the-job training to one's skill output. This hypothesis posits that a worker's incentive to continue investing in human capital before retirement will depend on one's expected age of retirement, with a worker expecting to retire later having a much greater incentive to invest human capital before retirement because of the potential economic returns (in increased wages) accruing from this additional capital. When a worker retires, this incentive is removed, and so the "on-the-job" retirement effect presumes that workers engage in more cognitive investment behavior than retirees, leading to cognitive decline in retirement.

In their empirical analysis, Rohwedder and Willis (2010) show a negative relationship between retirement and cognitive function, but their simple regression analysis does not control for such important factors as age and education. In their re-examination of the Rohwedder model, Bingley and Martinello (2013) include years of education and gender variables and find a weaker but still negative estimated effect, implying that the results of this model are sensitive to the specific controlled characteristics that are included. Similarly, Kajitani et al. (2014) argue that there exists a heterogeneity in cognitive deterioration related to the characteristics of the occupation and, further, Kajitani et al. (2016a) and Kajitani et al. (2016b) suggest that cognitive function is negatively affected by retirement duration and working hours.

In light of the above discussion, the goal of this study is to examine the "mental retirement effect" to determine whether or not there exists a causal effect of retirement on cognitive function. We do this in two steps. First, we examine the validity of the cross-sectional estimation procedure adopted in the literature and also the influence of the set of analyzed countries on the measured effects of retirement on cognitive function. Second, using a simple econometric model, we re-examine the mental retirement effect and the "on-the-job" retirement effect (Rohwedder and Willis (2010)) in the U.S. and other countries. To our knowledge, this is the first analysis that treats retirement endogenously in interpreting its effect on cognitive ability. Additionally, we investigate several potential sources of heterogeneity, including individual characteristics and time spent on leisure activities, that are either not covered in the literature or that are suggested as areas for future research. Finally, drawing upon the medical literature, we also examine the effect of body mass index (BMI) and fat intake on the heterogeneity of the effect of retirement on cognitive function.

Our estimates indicate that retirement has a weak effect on post-retirement cognitive scores in a wide range of analyzed countries and heterogeneous groups. Additionally, in our checks of the movement of cognitive scores around retirement age, we found no evidence that cognitive scores decline sharply around retirement age, or that retirement age influences how fast one's cognitive score declines. Therefore, our findings suggest that retirement policies in developed countries aimed at increasing the pensionable age do not substantially influence the cognitive abilities of the elderly.

The remainder of this paper is organized as follows: section 2 describes the dataset, section 3 discusses the wide-ranging effects of retirement on cognitive function, and section 4 examines the validity of a cross-sectional cross-country analysis. Section 5 discusses the main results from our dynamic analysis method, and section 6 concludes the paper and discusses the scope for future research.

2 Data

In order to conduct our analysis of the effect of retirement on cognitive ability, this study utilizes data on health and retirement from numerous countries, including the US Health and Retirement Study (HRS), China Health and Retirement Longitudinal Study (CHARLS), English Longitudinal Study of Ageing (ELSA), Korean Longitudinal Study of Aging (KLoSA), Survey of Health, Ageing, and Retirement in Europe (SHARE), and Japanese Study of Ageing and Retirement (JSTAR). These datasets are all panel surveys of individuals aged around 50 or over, and HRS family datasets are constructed so that the questions in the HRS family studies are as similar as possible to the original questions in the HRS. All of the datasets include a rich variety of variables to capture dimensions of life in terms of family background and economic, health, social and work status.

For data on cognitive ability, we used the cognitive function scores in the HRS and other related datasets, which included immediate and delayed word recall, a word recall summary score, serial 7s and backwards counting. The word recall test occurs in two rounds, with the respondent asked to recall as many words as possible from a 10 word list first immediately and then again after a given period of time. The score of the immediate and delayed word recall test is the number of words that were recalled correctly and a word recall summary score of between 0 and 20 is obtained from the sum of the two rounds. The serial 7s test asks the respondent to subtract 7 from the prior number beginning with 100 for 5 trials and from this, a score between 0 and 5 is obtained. The backwards counting test asks the respondent to count backwards for 10 continuous numbers from 20, and the original score obtained from this test is 2 if successful on the first try, 1 if successful on the second, and 0 if not successful on either try. However, because of the difficulty in interpreting the estimated coefficient of the original score, we adjusted this test score to indicate one when the respondent is successful on the first try and 0 otherwise. For our analysis in section 3, we used only the word recall summary score, while in sections 4 and 5, we used all types of scores.

Cognitive function scores are summarized in Tables 6 and 7, which show the descriptive statistics of the age group from 60 to 69 in all countries and the United States, respectively. From Table 6, we can see that cognition scores are not the same level in all countries, with scores in China and European countries comparatively lower than those of the U.S., U.K., Korea, [11] and Japan. In Table 7, we can see that females have a higher score than males in the word recall summary score, while males have higher score than females in serial 7. Additionally, those who are highly educated (i.e. university graduates) have a higher score

[11] For each test, the maximum test score in KLoSA is different from that of the other studies.

than those with lower levels of education in all cognitive scores.

As explained in detail in sections 4 and 5, in this study, we analyzed the effect of retirement on cognitive function in two ways. First, we performed a cross-sectional, cross-country analysis using two cohorts (2004 and 2010) of the cross sectional datasets of the HRS, ELSA, and SHARE datasets as well as CHARLS 2011 and JSTAR 2009. We used JSTAR 2009 because no survey was conducted in 2010 and the word recall questions in 2011 were asked only to people older than 65. Unfortunately, we were not able to use the Korean Longitudinal Study of Ageing (KLoSA) for this initial analysis because the test score questions were not comparable with other datasets.

Next, we performed a dynamic analysis of the long-term variation of retirement behavior for certain countries for which detailed information on the age of pension eligibility was available. When available, we used harmonized datasets, [12] but when the variables were not available in the harmonized datasets, we used the variables of the original datasets. Table 8 describes the datasets used for the analyses reported in each section of this paper.

Also note that in this paper, following Rohwedder and Willis (2010) and Bingley and Martinello (2013), we used pensionable age as an instrumental variable, and section 4 reports the results of our analysis when performing the cross-sectional cross-country analysis using the pensionable age for all countries. However, the pensionable age variable used by Rohwedder and Willis (2010) and Bingley and Martinello (2013) were based on data from the OECD *Pensions at a Glance* and the US *Social Security Programs throughout the World: Europe, 2004*, and these data for some countries are partly incorrect (see Appendix (A.1) for an explanation and description of how we corrected for this). In section 5, we report the results

[12] The Gateway to Global Aging Data (http://gateway.usc.edu) provides harmonized versions of data from international aging and retirement studies (e.g., HRS, ELSA, SHARE, KLoSA, and CHARLS), with all the variables of each dataset aiming to have the same items and follow the same naming conventions in order to enable researchers to conduct cross-national comparative studies. The program code to generate the harmonized datasets from the original datasets is provided by the Center for Global Ageing Research, USC Davis School of Gerontology and the Center for Economic and Social Research (CESR). Some variables, such as measures of assets and income, are imputed by this code.

of our analysis using only the pensionable ages confirmed to be correct.

3 Retirement and Decline in Cognitive Function

3.1 Discussion

One of the goals of this study is to analyze heterogeneity in the effect of retirement on cognitive function. In this section, we discuss which characteristics correlate with the difference in cognitive scores between retirees and non-retirees, and establish that there are factors other than basic individual characteristics such as gender and job characteristics that correlate with this difference.

Figures 5 and 6 report the differences in the relationship of the scores for the Serial 7s and Word Recall Summary tests of cognitive functioning between retired and non-retired individuals in Japan, the U.S., South Korea, China, Germany and France using two definitions of retirement: "not working for pay" and "self-reported retiree". [13] First, as the results are similar for both definitions of retirement, the influence of the retirement definition is weak. Next, we see that the difference in cognitive scores between retired and non-retired individuals is generally extremely small in all of the analyzed countries, and while the relationship is similar across countries, heterogeneity does exist. For example, while the serial 7s score in the U.S. is lower for those who are retired than those not retired, the results for China are

[13] "Not working for pay" indicates that a respondent is not working for wages or other type of payment, while "self-reported retiree" means that a respondent reported his status to be retired. We constructed these two variables from the "r@lbrf" variable reported in the RAND HRS dataset. In the HRS, "r@lbrf" takes seven values, and in this paper, we define a respondent as a *self-reported retiree* if r@lbrf' is "partly retired," "retired," "disabled" or "not in labor force". In other words, the difference between *not working for pay* and *self-reported retiree* is whether unemployed respondents are included or excluded. Page 1033 of the Rand HRS data codebook (http://hrsonline.isr.umich.edu/modules/meta/rand/randhrsm/randhrsM.pdf) explains in detail the variable "r@lbrf" used in all the harmonized datasets in this study (e.g., Harmonized SHARE, Harmonized ELSA), which follows numerous studies in the literature (e.g. Rohwedder and Willis (2010), Coe and Zamarro (2011), Bonsang et al. (2012), Bingley and Martinello (2013)) that use these two similar definitions of retirement. For example, Rohwedder and Willis (2010) consider the respondent retired if "not working for pay", and in Bonsang et al. (2012), a respondent is considered retired if s/he self-reports "not working".

the opposite. We can thus conclude that the cognitive function scores between retired and non-retired people have a heterogeneous relationship which depends on the specific countries analyzed. Further, this relationship does not seem to change according to the specific cognitive test used as, for example, the results for retirees and non-retirees in the U.S. and China for the word recall summary score are the same as those for the Serial 7s test described above: the score of retired individuals is lower than that of the non-retired in the U.S. but higher in China. Japan is similar to China in the serial 7s score. In sum, since the demographic profile of each country is different, it is possible that this is the source of the differences across countries in the relationship of overall average cognitive scores between retirees and non-retirees.

Next, we explore the apparent differences in the characteristics of the U.S. and China further by comparing the cognitive levels of retirees in both countries, using the two definitions of retired person described above and the Serial 7s and Word Recall Summary scores measuring cognitive functioning (Figures 7, 8, 9 and 10). We look at gender, education, job type and wealth. [14] While we can observe in Figures 7 through 10 that there seems to be some heterogeneity in individual characteristics between the U.S. and China, it is important to note that any difference in cognitive scores between retirees and non-retirees in these figures is not the effect of retirement on cognitive function since the former is endogenous. However, we can say that unobserved heterogeneity influences the difference in cognitive score between retirees and non-retirees among different countries. In summary:

- In each country, differences in characteristics such as gender, education and wealth explain the difference in cognitive function between retirees and non-retirees and, moreoever, the differences between the scores are heterogeneous with respect to the specific characteristic analyzed. We also note that the influence of retirement definition on the difference in the cognitive scores between retirees and non-retirees is weak.

- It is possible that there exist characteristics other than gender, education, and wealth that might contribute to observed differences in the scores between retirees and non-retirees as, indeed, does the endogeneity of retirement. However, it is possible that these factors strongly correlate with the country of residence for, as we have seen, in China, the cognitive function scores (either serial 7s or word recall) of retirees are larger than those of non-retirees for all characteristics while the relationship is the opposite in

[14] In the literature, heterogeneity such as gender or job type is important for explaining the effect of retirement on cognitive functioning. Coe et al. (2012), for example, estimates the effect of retirement on cognitive function for two job types (white-collar and blue-collar). In Figures 7, 8, 9 and 10, we also separated respondents into two job categories for the U.S. using the occupation code for the job with longest reported tenure. However, for China, we were not able to separate the job category into white-collar and blue-collar in the same way because the information on the job category of retirees was not available. As a result, we did not use the cognitive scores based on job types in China.

the U.S. Nonetheless, any unobserved factors other than gender and education are important, as they could potentially be causing an observed inverse relationship between the scores of retirees and non-retirees.

As we have discussed in this section, factors other than individual characteristics such as gender, education, and wealth are important for explaining the difference in cognitive function scores between retirees and non-retirees. This means that when we consider the effect of retirement on cognitive function, we have to also consider the potential influence of any unobserved heterogeneity on the difference in cognitive function scores. We consider this point further in the next section through a critical review of the literature.

4 Preliminary Analysis: Validation Test of Cross-Sectional Cross-Country Analysis

In the previous section, we discussed the heterogeneity of the difference in cognitive scores between retirees and non-retirees among different countries. In this section, we further consider this point by determining the validity of the cross-sectional analysis in previous studies through a critical review of the literature. We find that an estimation strategy based on cross-sectional analysis lacks robustness in that the estimated results are sensitive to the chosen set of countries analyzed. In other words, the specific countries chosen can unduly influence the final results.

4.1 Identification Strategy of Cross-Sectional Cross-Country Analysis

In this section, we investigate the robustness of the estimation strategy using cross-country variations in the age of pension eligibility. Since the goal of this research is to estimate the effect of retirement on cognitive function, the target of our identification strategy is to exclude any endogeneity bias that may be present in the retirement variable. Our analysis is carried out in two stages: first, we perform a cross-sectional cross-country analysis; and, second, a dynamic analysis of individual countries using panel data. In the first stage, the identification strategy is to use the variation in the age of pension eligibility among different countries in a specific year, which varies by country. We can use this exogenous variation to control for retirement endogeneity by simultaneously analyzing different countries with different pension eligibility ages.

Now, turning to the related literature, Rohwedder and Willis (2010), Coe and Zamarro (2011), and Bingley and Martinello (2013) are the studies analyzing the effect of retirement on cognitive function that are most relevant to this paper. Focusing on Rohwedder and Willis (2010), the age of pension eligibility used in the study is based on external data sources: the OECD *Pensions at a Glance* and US Social Security Administration *Social Security Programs throughout the World: Europe, 2004*. However, in our review of pension age data using primary data sources, we find that the data reported in the secondary sources used by Rohwedder and Willis (2010) are partly incorrect. Accordingly, in our analysis, we use a corrected version of the pension eligibility ages in Rohwedder and Willis (2010), as described in detail in Appendix (A.1).

According to our analysis, the estimated effect of retirement on cognitive function is heterogeneous among different sets of analyzed countries. Therefore, it is important to analyze this effect for each country rather than for groups of countries because it is possible that any unobserved heterogeneity might not be fully controlled for by using a cross-sectional cross-country analytic methodology. Consequently, in order to omit any potential individual unobserved heterogeneity, in the second stage of our analysis, we estimate the effect of retirement on cognitive function by using the dynamic variation in individual retirement behavior. Before moving to a description of our second-stage dynamic analysis, in the next section we describe the first-stage framework by which we came to our critical conclusion about the validity of the cross-sectional cross-country methodology used in most other studies.

4.2 Analysis Framework

For the first stage of our analysis, we update Rohwedder and Willis (2010) and Bingley and Martinello (2013) using a corrected dataset and adding a robust set of control variables. To begin, note that Rohwedder and Willis (2010) estimate the model below, using HRS, SHARE, and ELSA data for 2004 and restricting the analyzed sample to ages 60-64:

$$\begin{aligned} cognition\_score_i &= \beta_0 + \beta_1 notwork_i + \epsilon_{1i} \\ notwork_i &= \alpha_0 + \alpha_1 1\{age_i \geq A_i^{eb}\} + \alpha_2 1\{age_i \geq A_i^{fb}\} + \epsilon_{2i} \\ A_i^{eb}&: \text{age of eligibility for early retirement benefit} \\ A_i^{fb}&: \text{age of eligibility for full retirement benefit} \end{aligned} \tag{2}$$

where $notwork_i$ is an indicator equal to one when a respondent is not working for pay in the survey year, $cognition\_score_i$ is the word recall summary score (range: 0-20) and age_i is the respondent's age. Note that the model does not include any control variables.

Bingley and Martinello (2013), following Rohwedder and Willis (2010), also estimate Equation (2), but include an education (years of schooling) control variable. In our specification, we also include other control variables in addition to educational level, including gender, education and wealth. We also check the sensitivity of our results according to the control variables included. Our estimation specification, considering observed respondent heterogeneity, is shown in Equation 3:

$$cognition\_score_i = \beta_0 + \beta_1 notwork_i + \gamma' x_i + \epsilon_{1i} \tag{3}$$

$$notwork_i = \alpha_0 + \alpha_1 1\{age_i \geq A_i^{eb}\} + \alpha_2 1\{age_i \geq A_i^{fb}\} + \eta' x_i + \epsilon_{2i}$$

$$A_i^{eb}\text{: age of eligibility for early retirement benefit}$$

$$A_i^{fb}\text{: age of eligibility for full retirement benefit}$$

where x_i is a set of individual characteristics that we incorporate as control variables. These characteristics are unobserved in Equation (2) and so potentially could have produced the estimated differences in cognitive function among retirees. These characteristics could also have been correlated with the retirement variable, thus introducing bias into the estimated effect of retirement on health. Further, Rohwedder and Willis (2010) and Bingley and Martinello (2013) do not use any estimation weights for cross-country analysis in their estimations and also do not adjust the estimation according to population size. In our analysis, we incorporate an estimation weight based on UN data. [15]

With respect to control variables, it is rather difficult to assess which variables should be included in the estimation model, as the literature is wide-ranging and ambiguous. The public health literature discusses the relationship between behavioral factors (physical activity, lifestyle habits and leisure time activity) and cognitive function (Dik et al. (2003), Scarmeas and Stern (2003), Wilson et al. (2003), Nyberg et al. (2012), Raji et al. (2016), Satizabal et al. (2016)), and McEwen and Sapolsky (1995) and Sindi et al. (2016) indicate a relationship between stress and cognitive function. Additionally, Nyberg et al. (2000) suggest that gender differences influence cognitive function, and Satizabal et al. (2016) find that the incidence of dementia has declined over the last three decades, but cannot find a factor that explains this phenomenon. There are also numerous studies that discuss the relationship between

[15] See the website http://data.un.org/Default.aspx for more detail on the UN *A World of Information* data. Our methodology for calculating the estimation weights are described in Appendix A.4.

social factors and cognitive function. After considering all of the possible control variables discussed in the literature, in our study, we included demographic factors such as gender, family structure, economic variables, and country of residence in the estimation model to control for fundamental social determinants of human behavior. In section 5, we also use a simple economic model to discuss how social factors influence cognitive function.

To summarize the analysis of this section, we find the following conclusions:

- We find a significant effect of changing the sample set of countries analyzed, which suggests that the effects of retirement on cognitive function are also heterogeneous among different groups even if the analyzed groups have similar ages.

- We have shown that the effect of control variables for individual heterogeneity cannot be ignored. However, in our cross-country cross-sectional analysis specification, we found that the magnitude of the effect of retirement on cognitive function was similar to the magnitude estimated in some of related literature (Rohwedder and Willis (2010), Bingley and Martinello (2013)) even though we included control variables for individual heterogeneity while the other studies did not.

- Including a corrected instrumental variable influences the final results substantially when we compare our results to the instrumental variable estimates of Rohwedder and Willis (2010) and Bingley and Martinello (2013), showing that the effect of correcting the IV is not weak.

Finally, before we turn to our results, we note that the claims in Rohwedder and Willis (2010) of a negative relationship between average cognitive score and percent eligible for early public pension benefits may be overextended. In our Figure 11 below, we have used 2010 data to replicate and update Figure 6 in Rohwedder and Willis (2010) (pp. 134-135), which is based on a 2004 dataset. Like Rohwedder and Willis (2010), we also find an apparent negative relationship between average cognitive score and percent eligible for early public pension benefits, but this relationship may not be robust and, in any case, it may be only a correlation rather than a causal relationship.

The next section reports the main results of our stage 1 cross-sectional cross-country analysis. Secondary results are reported in Appendix A.2.

4.3 Results: Importance of the Set of Countries Analyzed

In this section, we focus on the importance of the choice of countries analyzed. Table 9 reports the results of our estimates of equations (2) and (3) for four groups of countries based on linguistic regions: *Latin* (France, Spain Portugal, and Italy), *Slavic* (Estonia, Slovenia, Poland, Hungary, and Czech Republic), *Germanic* (U.K., Netherlands, Germany, Denmark, Belgium, Sweden, Austria, and Switzerland), and *New SHARE and East Asia* (Japan, China, Czech Republic, Poland, Hungary, Portugal, Slovenia, and Estonia). We also include *Original without Greece*, which is the original set of countries analyzed by Rohwedder and Willis (2010), minus Greece, as it was not included in the 2010 survey. Since many countries have been included in the HRS sister surveys since 2002, for this analysis, we used the 2010 dataset.[16] Further, in order to accomodate the variation in the age of pension eligibility in as many countries as possible, for this analysis, we chose an age range of 60-69. Table 9 presents only the main results, omitting the coefficients of the country dummies and other control variables. For details of our specification of pensionable ages and estimation weights, see sections A.1) and A.4 of the Appendix.

From Table 9, we can see a large degree of heterogeneity in the estimated results. The coefficients for some country groups are negative and significant (*Original without Greece*: -0.608 (OLS); -3.940 (IV2) although the DWH test is not rejected; *Germanic*: -0.333 (OLS), and *Latin*: -0.362 (OLS)) while the *New SHARE and East Asia* coefficients are significant and positive (0.413 (OLS)) and those of *Slavic* are not significant (-0.138 (OLS)).

[16] JSTAR data is 2009, and CHARLS is 2011.

To summarize the results of our preliminary analysis:

- The choice of countries analyzed largely influences the estimated result. Therefore, we need to pay attention to country heterogeneity when we analyze the effect of retirement on cognitive function (Table 9);
- When important control variables are omitted, unobserved heterogeneity has a large influence on the estimated result (Tables 29 and 31);
- The definition of retirement does not have a large influence on the estimated results[17], even though the specific definition chosen does occasionally drive different conclusions (Table 34);
- Other factors such as differences in age or cohort do not seem to be important (Tables 33 and 34).

The implications of this finding that country heterogeneity largely influences the estimated results poses two problems for an identification strategy in a cross-sectional cross-country analysis:

- If the country heterogeneity of a given variable is large, even when endogeneity bias is small, the policy implications for an individual country may not be transparent;
- If the bias created by unobserved variables that are correlated with the retirement dependent variable differs among countries, then this cannot be eliminated through a common set of control variables in a cross-sectional cross-country analysis. If this is the case, it is difficult to estimate the effect of retirement on cognitive function because we cannot isolate how the estimated parameter is influenced by bias for each country.

[17]Kajitani et al. (2013) also report that the sensitivity of the definition of retirement definition is weak.

Based on this deficiency of cross-sectional cross-country analysis, we have chosen to analyze the effect of retirement on cognitive function in a single country through a dynamic model that also controls for unobserved individual heterogeneity. In the next section, we describe our dynamic strategy for estimating the effect of retirement on cognitive function in several countries individually, choosing only countries for which the age of pension eligibility is confirmed to be correct (see Appendix A.1 for a full discussion). We also analyze the influence of heterogeneity of transition behavior (leisure activity) before and after retirement and the influence of individual heterogeneity. We end with a discussion of the validity of cross-country cross-sectional analysis based on a comparison of our findings using that approach versus dynamic analysis.

5 Dynamic Analysis

Our analysis in the previous section identified two potential problems with the cross-country cross-sectional estimation strategy used in the previous literature (Rohwedder and Willis (2010), Coe and Zamarro (2011), and Bingley and Martinello (2013)) in estimating the effect of retirement on cognitive function. As such, in this section we describe another identification strategy to omit any potential unobserved heterogeneity of individual characteristics. Before we proceed to the estimation, we discuss the source of heterogeneity in the effect of retirement on cognitive function. In the related literature, Rohwedder and Willis (2010) suggests that it is possible that a difference in activity during leisure time influences cognitive function after retirement, raising this as a topic for future work. Bonsang et al. (2012) also suggest that increased social interaction may be an important factor enhancing cognitive reserve. In the next section, we introduce a simple framework to consider these points.

5.1 The Source of Heterogeneity in the Effect of Retirement on Cognitive Function: A Simple Theoretical Analysis

In this section, we investigate the hypothetical mechanism causing differences in cognitive function scores between retirees and non-retirees, using a simple economic model based on Grossman (1972). Rohwedder and Willis (2010) present a similar idea about the mechanism by which cognitive function decreases after retirement, and our interpretations of the *mental retirement effect* and the *on-the-job* retirement effect are drawn from this model. Mazzonna and Peracchi (2012) model the effect of retirement on cognitive function as well, but in their specification, retirement is exogenous and there is no asset accumulation. Further, the utility function is formulated from cognitive investment. However, in our specification, we formalize the utility function with cognitive ability because it is a health asset that is increased through cognitive investment. To our knowledge, ours is the first analysis of this model to treat retirement as an endogenous variable.

Equation (4) is a simple dynamic model with two cognitive abilities, and represents the maximization problem of an elderly person:

$$
\begin{aligned}
&\max_{\{c_t, l_t, i^f_{Wt}, i^j_{Wt}, i^f_{Lt}, i^j_{Lt}\}^T_{t=50}} \sum_{t=50}^{T} \beta^{t-50} u(c_t, \tilde{l}_t, a^f_t, a^j_t) \\
&(50 \leq t \leq T) \\
s.t.\ & A_{t+1} = (1+r)A_t + P(l_t, R, Pension_t) + y_t - c_t - G(i^f_{Wt}, i^j_{Wt}, i^f_{Lt}, i^j_{Lt}) \\
& a^f_t = A_f(t, i^f_{Wt}, i^f_{Lt}, X_{ft}) \\
& a^j_t = A_j(t, i^j_{Wt}, i^j_{Lt}, X_{jt}) \\
& \tilde{l}_t = l_t - L(i^f_{Lt}, i^j_{Lt}) \\
& y_t = y(a^f_t, a^j_t, t, l_t)
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
& l_t \in \{0, 0.5, 1\} \\
& (1 - l_t) \cdot i^{mMax}_{Wt} \geq i^m_{Wt} \geq 0 \ \ (m = f, j) \\
& c_t \geq 0, \tilde{l}_t \geq 0, i^m_{nt} \geq 0 (m = f, j)(n = W, L), A_{t+1} \geq 0
\end{aligned} \tag{5}
$$

where c_t is consumption, $\tilde{l}_t$ is final consumption of leisure time, a^f_t is fundamental cognitive ability, a^j_t is job specific cognitive ability, l_t is leisure time, A_t is assets, R is pensionable age and $Pension_t$ is pension payment. Cognitive investment at work and leisure are included through four variables: fundamental cognitive investment at workplace (i^f_{Wt}), fundamental cognitive investment during leisure time (i^f_{Lt}), job specific cognitive investment at workplace (i^j_{Wt}), and job specific cognitive investment during leisure time (i^j_{Lt}). X_{ft} and X_{jt} represent technological factors of fundamental and job specific cognitive ability, and $A_f(\cdot)$ and $A_j(\cdot)$ are the production functions of fundamental and job specific cognitive ability. Finally, $P(\cdot)$, $G(\cdot)$, $L(\cdot)$ and $y(\cdot)$ are functions for pension payment, cost of cognitive investment, reduced time by cognitive investment, and income, and $i_{Wt}^{mMax}(m = f, j)$ are the maximum values of cognitive investment during work time.

We assume that "fundamental cognitive ability", which is the target of our analysis, is a basic cognitive ability such as calculation, reading, or memorization while "job specific cognitive ability" is the cognitive ability required for a specific job, such as a computing skill. The following are the important structures of model (4):

- The elderly can undertake cognitive investment at the workplace only when they work, and the maximum amount of this investment depends on leisure($(1 - l_t) \cdot i_{Wt}^{mMax} \geq i_{Wt}^{m} \geq 0(m = f, j)$). When the elderly enjoy their leisure time, they can invest in their cognitive ability, but these investments reduce ($\tilde{l}_t = l_t - L(i_{Lt}^{f}, i_{Lt}^{j})$) the final amount of leisure consumed, $\tilde{l}_t$.

- We assume that, for the elderly, $y(a_{t+1}^{f} = \alpha_1, a_{t+1}^{j} = \alpha_2, t+1, l_{t+1} = \alpha_3) - y(a_t^{f} = \alpha_1, a_t^{j} = \alpha_2, t, l_t = \alpha_3) < 0$, or that aging lowers income. Although the elderly may continue to input the same level of leisure time and have the same level of cognitive ability, income continues to decrease during aging. This is an effect of health on income in that aging reduces the incentive to work.

- We assume that the elderly do not receive a pension if they are younger than pensionable age (i.e. $P(l_t, R, Pension_t) = 0$ if $t \leq R$). Also, because of the liquidity constraint ($A_t \geq 0$), the incentive to work increases when the age of pension eligibility, R, increases.

- We also assume that it is possible that the elderly have a preference for either fundamental or job-specific cognitive ability, so that it is possible that $\frac{\partial u(c_t, \tilde{l}_t, a_t^{f}, a_t^{j})}{\partial a_t^{m}} > 0(m = f, j)$.

This structure creates an incentive to invest in the cognitive ability which is an important property. Additionally, it creates two potential benefits for workers (income and the opportunity to invest in one's cognitive abilities at the workplace) because, in our model, the elderly can perform cognitive investment at the workplace only when they work. Finally, elderly who have a preference for cognitive ability have the incentive to invest in their cognitive abilities during leisure time, which also provides an incentive for these elderly to invest in their cognitive abilities even after retirement.

- The marginal utility of investing in one's cognitive ability is $\frac{\partial u(c_t,\tilde{l}_t,a_t^f,a_t^j)}{\partial a_t^m}\frac{\partial a_t^m}{\partial i_{nt}^m}(m = f,j)(n = W,L)$. In other words, the factors of the cognitive ability production functions $A_f(\cdot)$ and $A_j(\cdot)$ are important for deciding the amount of investment when they influence the marginal productivity of the production function($\frac{\partial a_t^m}{\partial i_{nt}^m}(m = f,j)$).

In order to discuss the theoretical mechanism more concretely, we parameterize model (4), with the specification presented here one example among many possibilities. The details of the parameterization are explained in Appendix (A.3). In what follows, we discuss only the hypothesis of why the effect of retirement on cognitive ability differs. Our parameterization of the utility, pension payment, and cognitive ability functions are as follows:

- $u(c_t,\tilde{l}_t,a_{ft},a_{jt}) = c_t^{\gamma_1}\tilde{l}_t^{\gamma_2}a_{ft}^{\gamma_3}a_{jt}^{1-\gamma_1-\gamma_2-\gamma_3}$
- $P(l_t,R,Pension_t) = 1\{l_t \geq 0.5\}1\{t \geq R\}Pension_t$
- $A_m(t,i_{Wt}^m,i_{Lt}^m,X_{mt}) = \alpha_1 i_{Wt}^m + \alpha_2 i_{Lt}^m + \alpha_3 Hetro_1 + \alpha_4 Hetro_2 + A_{m0}\exp(-\alpha_5 t)(m = f,j)$

In our benchmark model, we set the parameters to $\gamma_3 = 0.0, 1-\gamma_1-\gamma_2-\gamma_3 = 0, R = 70, \alpha_5 = 0.05$ and simulated the economic behavior of 5,000 agents after solving the dynamic programming. Subsequently, initial assets A_0 and initial cognitive abilities $A_{m0}(m = f,j)$ were drawn from a distribution, and the influence of the average value of A_{m0} of the initial distribution is presented in Figures 13-16 below. In all figures, the vertical axis indicates the average value of each variable for all agents, and the horizontal axis represents age, beginning at age 50. Our main findings of this analysis are summarized below:

- Influence of different preferences for fundamental cognitive ability: Figure 13 shows the effect of a change in the parameter for fundamental cognitive ability in the retiree's utility function. $\gamma_3 = 0.0$ in our benchmark model is changed to $\gamma_3 = 0.2$ in the "with preference" case. In the "with preference" case, when elderly start to retire around age 70, this is accompanied by a steep increase in fundamental cognitive investment during leisure time, but this does not occur in the "without preference" case. This difference in cognitive investment behavior occurs because in the "without preference" case, the incentive to increase leisure is large, and so this group of elderly retires earlier, decreases its investment in fundamental cognitive ability at the workplace, and also does not increase it during leisure after retirement. We can see, then, that the effect of retirement on cognitive function is due to different cognitive investment behaviors during leisure time and at the workplace both before and after retirement. The change in cognitive function after retirement that is caused by a change in cognitive investment behaviour, or lifestyle, is known as the "mental retirement effect" (Rohwedder and Willis (2010)). From the lower right panel of Figure 13, we see a great divergence in fundamental cognitive investment after retirement (age 72) due to heterogeneity in preferences for cognitive ability. This heterogeneity in preferences, in turn, causes heterogeneity in the mental retirement effect.

- Influence of different initial cognitive ability level: Figure 14 shows the effect of a change in the average value of A_{m0} in the initial distribution to a lower average value in the "high initial ability" case as compared to the benchmark. We see that this change produces no difference in cognitive investment behavior either before or after retirement. This indicates that it is not heterogeneity in initial cognitive ability but differences in cognitive investment behavior based on differences in preferences that creates heterogeneous effects of retirement on cognitive function.

- Influence of different technology in cognitive ability production function: Figure 15 shows the effect of changing the α_5 parameter from 0.05 in the benchmark case to 0.025 in the high tech case. The decrease in cognitive function by the increase in age becomes lower when $\alpha_5 = 0.025$ compared to $\alpha_5 = 0.05$. We find that higher technology raises fundamental cognitive ability and fundamental cognitive investment in the workplace, but lowers leisure time and has no effect on leisure time fundamental cognitive investment. Overall, the source of heterogeneity in the effect of retirement on cognitive function is clear because the difference in cognitive investment behavior is large.

- Influence of the age of pension eligibility on investment activity: In Figure 16, we see that a lowering of the age of pension eligibility from 70 to 65 causes the steep jump in average leisure time to occur 5 years earlier, at the age of retirement, as expected, but cognitive investment behavior at the workplace also sharply decreases at the pensionable age. Thus, retirement behavior is strongly influenced by whether an elderly person has arrived at their pensionable age or not. We can therefore use pensionable age as an instrumental variable to control for the endogeneity of cognitive investment behaviors, and we do incorporate this into our empirical estimation strategy described in the next section. Additionally, in the upper right panel of Figure 16, we see that a change in the age of pension eligibility causes heterogeneity in the age of retirement and, further, that the change in the age of pension eligibility also causes heterogeneity in fundamental cognitive investment at the workplace (upper left panel). This is Rohwedder and Willis (2010)'s "on-the-job" retirement effect.

Thus far in this section, we have discussed our simulation of various sources of heterogeneity on the effect of retirement on cognitive investment behavior before and after retirement both at the workplace and during leisure time. We close this section by relating this analysis to the public health literature. Numerous public health studies have focused on various determinants of cognitive function such as lifestyle habits or on the relationship between cognitive ability and human behaviors including physical activity or lifestyle habits such as leisure time activity (Dik et al. (2003), Scarmeas and Stern (2003), Wilson et al. (2003), Nyberg et al. (2012), Raji et al. (2016), Satizabal et al. (2016)) and find a heterogeneous effect on cognitive function due to varied cognitive investment behaviors during leisure time. Drawing on this public health literature, we consider the effect of this heterogeneity in leisure time activities in the empirical section below.

5.2 Estimation Strategy

In section 5.3, we use the dynamic variation of retirement behavior on cognitive function. Then, we analyze the effect of **whether or not a respondent retires** on cognitive functioning. Figure 17 shows the target of our analysis, and we perform this analysis for the U.S., England, France, Germany, Denmark, Korea and Japan – countries for which data on the age of pension eligibility is available and has been confirmed to be correct.

In Figure 18 we can see that there are two retirement stages in all of the countries studied. Between the ages of 50 to 70, many people begin retiring within a relatively short period of time, as seen by the steep slope of the retirement curves during this period in all countries. By age 70 to 80, however, almost all elderly have retired, and the slope of the retirement curve is quite flat. Most of the elderly in this latter group have been retired for some time. For our investigation, though, we analyze the effect on retirees who have recently retired, or those retirees characterized by the steep slope of the retirement curve as demarcated by the vertical lines in Figure (18).

Our empirical analysis is based on investigating the three sources of heterogeneity discussed in section 5.1 above that can produce differences in leisure time cognitive investment behaviors: cognitive ability preferences, initial cognitive ability, and technological factors in the cognitive ability production function. For our empirical analysis, we considered heterogeneity in the initial cognitive score (initial cognitive ability), activities during leisure time, and individual characteristics (e.g. gender, which is a technological factor of the cognitive ability production function). Our data on the time consumed in leisure activities both before and after retirement are from the Consumption and Activities Mail Survey (CAMS) administered by HRS, which describes activity patterns and "how specific types of activities are affected by health, family, and economic transitions in later life and, in turn, how activities affect health and well-being." [18] Our analysis did not uncover heterogeneity in preferences

[18] For further details, see https://ssl.isr.umich.edu/hrs/filedownload2.php?d=522.

for cognitive ability but we did observe differences in initial cognitive ability and some technological factors during the transition in leisure time activity before and after retirement. Next, we separated the sample depending on the heterogeneity of observable characteristics, which allowed us to control for the direct effect by controlling the heterogeneity of cognitive investment behaviors. However, some characteristics we used to separate the sample (e.g., BMI) are difficult to interpret, as we could not determine which factor (preference, initial cognitive ability, technology) BMI describes. This is a limitation of the analysis.

5.3 Retirement Analysis

5.3.1 Analysis Framework

As discussed in section 4, the effect of retirement on cognition differs substantially among countries, and so our strategy is to analyze the effect on each country individually instead of in a cross-sectional, cross-country analysis. The countries analyzed include the U.S., England, France, Germany, Denmark, Korea, and Japan because correct information on the age of pension eligibility and a sufficient number of dataset waves for dynamic analysis are available. The identification strategy in this section is to use the variation of whether a respondent arrives at the pension eligibility age to analyze the effect of **whether a respondent retires** on cognitive function. We derive the following equation from the fundamental cognitive ability equation in model (4):

$$a_t^f = A_f(t, i_{Wt}^f, i_{Lt}^f, X_{ft}) = \alpha_0 + \alpha_1 i_{Wt}^f + \alpha_2 i_{Lt}^f + \gamma' X_{ft} + \epsilon_{ft} \tag{6}$$

Now, let $a_t^f = cognition\_score_{it} + \tilde{\epsilon}_{1t}$, $\alpha_1 i_{Wt} + \alpha_2 i_{Lt} = \beta retire_{it} + \tilde{\epsilon}_{2t}$, $cognition\_score_{it}$ is cognitive test scores, and $retire_{it}$ is an indicator of whether a respondent i retires in period t. In other words, in this specification, cognitive score is a proxy of cognitive ability and retirement status is a proxy of investment activity in enhancing cognitive ability. Substituting the above into equation (6), we obtain the following:

$$cognition\_score_{it} = \alpha_0 + \beta retire_{it} + \gamma' X_{ft} + \epsilon_{ft} - \tilde{\epsilon}_{1t} + \tilde{\epsilon}_{2t}.$$

Our estimation equations are the following:

$$cognition\_score_{it} = \beta_0 + \beta_1 retire_{it} + \gamma' x_{it} + a_{1i} + \lambda_{1t} + \epsilon_{1it} \tag{7}$$

$$retire_{it} = \alpha_0 + \alpha_1 1\{age_{it} \geq A_i^{eb}\} + \alpha_2 1\{age_{it} \geq A_i^{fb}\}$$

$$+\alpha_1 1\{age_{it} \geq A_i^{eb}\} age_{it} + \alpha_2 1\{age_{it} \geq A_i^{fb}\} age_{it} + \eta' x_{it} + a_{2i} + \lambda_{2t} + \epsilon_{2i} \tag{8}$$

$$\tag{9}$$

where A_i^{eb} and A_i^{fb} are the ages of eligibility for early and full retirement benefits, $retire_{it}$ is an indicator equal to one when a respondent retires at period t, λ_{1t} and λ_{2t} are time fixed effects, a_{1i} and a_{2i} are individual fixed effects, and x_{it} are control variables at period t. As discussed in section 4.2., it is difficult to determine which specific control variables should be included in the estimation model. We included demographic factors such as gender, family structure, and economic variables in the estimation model to control for the fundamental social factors that influence human behavior.

There are two common ways of defining whether a respondent is retired. The first definition of retirement is based on the person's self-reported retirement status, with a respondent being *retired* when the "self-reported retiree" variable is equal to one. This is the definition of "self-reported retiree" based on the "r@lbrf" variable described in footnote 13 of section 3.1, and is used in several studies (e.g. Coe et al. (2012) and Mazzonna and Peracchi (2012)).

The second definition of retirement, which is most commonly used in the literature[19] is that a respondent is retired when he or she no longer works for pay.

Our definition of retirement, which we call "complete retirement", is the intersection of both of these common definitions; that is, a person is completely retired when he or she is both a self-reported retiree and is no longer working for pay. This takes care of some problems with both of the former definitions. Specifically, if a person self reports as retired, it is still possible that the person might be doing incidental work for pay and so might be continuing to invest in his cognitive abilities, while those who are not working for pay may be unemployed rather than retired and so may also be investing more than a retiree would.

In order to capture the heterogeneous transition pattern of cognitive investment during leisure both before and after retirement, we chose an age range for this analysis of 50-79 for respondents in all countries except Japan, where we chose 50+ because it is not unusual for people to still work even at age 80 (Figure 18). Finally, we chose for our analysis sample only those who were not "completely retired" at least once in this analysis, because we wanted to omit the respondents who both had not worked for pay and who had retired at an early age for this analysis of retirement transition.

All analyzed countries survey social activities, which we used to obtain information about each respondent's transition before and after retirement. However, in the U.S., the Consumption and Activities Mail Survey (CAMS) provides detailed information on the amount of time spent doing specific activities (Table 10). Although most activities show a clear difference before and after retirement, hours of watching TV shows the largest change. Computer use is another activity distinctly affected by retirement, as the elderly who still work use computers in their office. As such, we focus on the heterogeneity of of time spent watching TV and engaged in other social activities before and after retirement in the U.S.

Additionally, some studies (Eskelinen et al. (2008) and Devore et al. (2009)) report that

[19]For example, Rohwedder and Willis (2010), Coe and Zamarro (2011), Bonsang et al. (2012), Bingley and Martinello (2013), Hashimoto (2013) and Hashimoto (2015).

there is a relationship between fat intake and cognitive function, and so we also considered variations in the Body Mass Index (BMI) and the amount of fat intake. For this, we obtained data from the 2013 dataset of the Health Care and Nutrition Study. Since that survey year is different from our analysis year, the amount of fat intake in 2013 is a proxy for the amount of fat intake in other years. In addition to the relationship between fat intake and cognitive function, it is also possible that the amount of fat intake during the lifetime of the respondent forms a technological factor of cognitive decline, while BMI is also a proxy for a potential technological factor of cognitive decline. These two heterogeneities considered in this study are not analyzed in extant studies and so are a contribution of this paper.

5.3.2 Results: The US and Other Countries

In this section, we first discuss the results obtained for the U.S. and then compare these results with those from other countries. Table 11 shows the results for the U.S., and we can see that our instruments are valid, with all of the coefficients significantly different from zero. Tables 12, 13 and 14 show the results of the effect of retirement on cognitive function in various groups. In Table 12, we report individual characteristics representing heterogeneity in the technology of the cognitive ability production function: Gender (male/female), Education (low/high), and Occupation (blue collar/white collar). Table 13 relates cognition to the health measures of Body Mass Index (whether BMI > 25 in the 2013 survey or not) and fat intake (whether it is more than the median in the 2013 survey or not). As discussed in section 5.2, these characteristics were added to incorporate the relation between cognition and health found in the health economics literature, although it is not clear what these characteristics describe. Table 14 reports the influence of heterogeneity in initial cognitive ability (initial scores at 1st interview) and in differences in preference for cognitive ability as seen in changes in social activities (i.e. whether social activity decreases or increases after retirement), and having a spouse at 1st interview (which is also an indicator of leisure time activity).

In general, we found that the effect of retirement on cognitive scores was weak even though we did find evidence of variation in cognition scores among heterogeneous groups. These results from Tables 12, 13 and 14 can be summarized as follows:

- The effect of retirement on the Word Recall Summary score is negative for both males and females, but the magnitude is small (WR Summary scores in columns 1 and 2 of Table 12 show males = -0.137 and females = -0.164). The effect of retirement on Serial 7s score is negative only for females, and the magnitude is also small. Among occupations, the negative effect of retirement is stronger for white collar than blue collar workers (columns 5 and 6). We also see a negative effect of retirement on cognition for most groups in both the Immediate and Delayed Word Recall, though the magnitude is small and shows little variation.

- The negative effect of retirement on cognition is stronger for those with higher BMI and fat intake (columns 1-4, Table 13).

- While the related literature suggests that heterogeneity in post-retirement activities, particularly social activities, is important, our results (Table 14) show no evidence of that. For example, there is no evidence that the cognitive score is influenced by either differences in initial cognitive scores (columns 1 and 2) or a change in social activities after retirement (columns 3 and 4). We did find, however, a stronger negative effect of retirement on cognition for retirees with no spouse (columns 5 and 6).

Next, we compare the results from the U.S. to those of other countries, and discuss whether any systematic difference can be found due to the heterogeneity of activity after retirement and individual characteristics. We also analyze the effect of retirement on cognitive scores in countries other than the U.S. to see if similar effects can be found within a given group in numerous countries. Here we report only the main results. For a detailed description, see Appendix A.5.

- As in the U.S., the effect of retirement on cognitive scores according to gender is weak in many countries (Figure 19);

- Like the U.S., there is no evidence in other countries that the cognitive score decline is heterogeneous according to changes in social activities after retirement (Table 44);

- There is also no systematic heterogeneity of the effect of retirement on cognitive scores in countries other than the U.S. among those having or not having a spouse at 1st interview, BMI above or below 25, and high or low fat intake (Tables 45, 46, 47 and 48).

To sum up, the effect of retirement on cognitive scores is weak in all countries, including the U.S., and the heterogeneity of individual characteristics and activity after retirement analyzed in this paper was found to be not important. In the next section, we discuss another characteristic which we found explains the heterogeneity of the effect of retirement on cognitive scores.

5.3.3 Retirement Timing

Although we did not find that the heterogeneity of individual characteristics was important, we did find evidence that retirement timing may cause the systematic heterogeneity of the effect of retirement on cognitive scores. For this analysis, we chose only elderly in the U.S. and divided the respondents aged 58-69 into two groups. Table 15 shows a negative effect of retirement on the cognitive scores only of the older female group aged 64-69, though the magnitude is small. Figure 20 shows the relationship between the coefficient of the fixed effect model and the average retirement age in each country. We estimated the fixed effect model by using the original sample of those aged 50-79 in each country, and according to Figure 20, there is an negative correlation between the magnitude of the estimated coefficients and the average retirement age in the sample.

Finally, Figure 21 shows the relationship between average cognitive scores and age among three groups of retirees in each country (early, mid, and late retirement), which is summarized as the following:

- Retirement does not have a strong effect on the cognitive score, as there is no large change around the retirement age in any of the three retirement groups. In addition, the timing of retirement does not seem to influence how fast cognitive scores decline.

- While in the U.S., cognitive scores decrease sharply as the respondents become older, the effect is less pronounced in other countries.

- In the U.S., the initial cognitive score of people who retire late is higher than for other groups. This is consistent with Rohwedder and Willis (2010)'s ”on-the-job” retirement effect, as it is possible that those who retire late try to increase their cognitive function before retirement in order to delay their retirement.

In summary, although our analysis found that the effect of retirement on cognitive scores (Rohwedder and Willis (2010)'s “mental retirement effect) was weak in many countries, in the U.S. at least, it is possible that the ”on-the-job” retirement effect (Rohwedder and Willis (2010)) might exist. If cognitive function sharply declines as one ages, this could provide a strong incentive to increase one's cognitive ability before retirement for those who want to work at a relatively advanced age.

5.3.4 Discussion: Cross-Country Cross-Sectional Analysis and Dynamic Analysis

In this section, we discuss the validity of the cross-sectional analysis adopted in the literature versus the dynamic analysis of this paper. After controlling for individual characteristics, we found a negative effect of retirement on word recall score in all countries, although the magnitude of the effect was small, with the estimated effect of the US being -0.154 (Table 35). The estimated results without controls found, however, some problematic results, such as a coefficient of “**All countries**” of 3.728 (column 12 of Table 9). The results were not problematic for all country groups, however, for specifications for country groups with controls. For example, the coefficients of “**Latin**”, “**Slavic**” and “**Germanic**” were -0.362 (OLS), -0.138 (not significant)(OLS) and -0.333 (OLS); small in magnitude and negative. It thus seems that cross-country cross-sectional analysis may be unduly affected by strong relationships within individual countries. Such problems as those discussed in section 4.3 can be avoided through a dynamic analysis.

6 Conclusion

This study estimated the effect of retirement on cognition. The main findings of the paper are:

- In our analysis of the validity of cross-sectional cross-country analysis, we found that the robustness of the results is weak because the estimated results are sensitive to the heterogeneity of the set of analyzed countries. In particular, the effect of retirement on cognitive function within a subset of the analyzed countries can unduly influence the final conclusion.

- In our analysis of the relationship between retirement and cognition, we found that:

  – the "mental retirement effect" (Rohwedder and Willis (2010)) is weak in many countries;

  – individual characteristics such as job category, educational level and social activity after retirement are not important in producing the heterogeneity of the effect of retirement on cognitive scores;

  – there is evidence to suggest that this effect may be produced instead by the timing of retirement.

Comparing our results to those of related studies (Table 16), we found the effect of retirement on cognitive ability to be weak in all countries. When we observed the scores of cognitive ability tests before and after retirement around retirement age in the U.S., the U.K. and SHARE countries, there was no clear decline in scores on tests of cognition before and after retirement. This suggests that government policies within these countries to delay retirement through such measures as increasing the age of pension eligibility might not greatly influence the cognitive ability of the elderly after retirement. We did find, however, that BMI and fat intake were important determinants of the effect of retirement on cognitive function heterogeneity in the US. Additionally, our results suggest that in the U.S. in particular, it is possible that there is an "on-the-job" retirement effect (Rohwedder and Willis (2010)) whereby those who retire at an advanced age might try to increase their cognitive function before retirement in order to delay their retirement. Further analysis of this point remains important future work.

We also found that engaging in social activities may not be an adequate proxy for cognitive investment behaviors, and that elderly do not substantially change their leisure activities before and after retirement. This leaves us with the important question as to what kind of activity might constitute a cognitive investment behavior. Additionally, in this study, we were only able to analyze groups and countries where we could find correct and available data on the age of pension eleigibility in order to use it as an instrumental variable. Another instrumental variable would potentially allow us to expand our analysis.

Chapter 4: The Effect of Nuclear Accident on Pregnancy Outcomes: Evidence from Fukushima-Daiichi in Japan

1 Introduction

On March 11, 2011, the pacific northeast coast of Japan was hit by a large-scale earthquake, the Great East Japan Earthquake. The Fukushima-Daiichi nuclear power plant was damaged by the earthquake with a giant tsunami. It has caused the nuclear power plant to have three meltdowns and the radioactive substance has been released into the outside of the power plant. While the radioactive substance has influenced various economic activities in Japan, the extent of the impact has not been clarified so far. Estimating effects of an accident of nuclear power plant gives us an information about a potential cost of constructing nuclear power plant, which is useful when a government discusses the construction of nuclear power plants as a part of the energy policy.

Historically, there are two big accidents of nuclear power plant before the accident of Fukushima-Daiichi: Three Mile Island in 1979 and Chernobyl in 1986. The effects of the accidents on various outcomes has been paid attention by economists. By the accident of Chernobyl, there are significant decreases in subsequent cognitive ability of prenatal children in Sweden (Almond et al., 2009) and significant changes in behaviors of parents whose children were exposed to the shock in Austria (Halla and Zweimüller, 2014). There is a study analyzing the effects of the accident of Chernobyl on shaping environmental law. (Kahn, 2007) Furthermore, some studies have analyzed the impacts of the accident of Three Mile Island on the financial market, residential property values, and the change in the risk of unplanned outages. (Nelson, 1981; Gamble and Downing, 1982; Bowen et al., 1983; Chen, 1984; Barrett et al., 1986; David et al., 1996)

After the Fukushima-Daiichi accident, while economists have started to analyze the effects of the Fukushima-Daiichi accident on various economic activities (Ando et al., 2017; Bauer et al., 2013; Coulomb and Zylberberg, 2016; Fink and Stratmann, 2015; Hanaoka et al., 2018; Hayashi and Hughes, 2013; Hong et al., 2013; Huang et al., 2013; Kawaguchi and Yukutake, 2017; Nishimura and Oikawa, 2017b; Tajima et al., 2016; Tanaka and Managi, 2016; Yamane et al., 2013; Zhu et al., 2016), less attention has been paid to the effects on child health and development, especially, pregnancy outcomes. [20] The health condition of children, which affects their life cycle, is an important research topic in the field of economics. Economists have analyzed the effects of environmental factors on pregnancy outcomes, children' health ,and developments. (Almond et al., 2009; Black et al., 2013; Currie et al., 2009; Currie and Walker, 2011; Currie and Schwandt, 2016; Halla and Zweimüller, 2014) For assessing the cost of the nuclear power plant, the effects of the Fukushima-Daiichi accident on pregnancy outcome should be an important information.

There are some medical studies which analyzed the effects of the Fukushima-Daiichi nuclear power plant on the pregnancy outcomes. While the previous studies have found the negative effects of environmental pollution on pregnancy outcomes and child development, the negative effects of the Fukushima-Daiichi accident on children in-utero on the date of accident are less observed. (Hayashi et al., 2016; Ishii et al., 2017; Leppold et al., 2017; Yasuda et al., 2017) The studies compare the proportion of babies with adverse pregnancy outcomes by variables such as residential region and birth cohort, and do not use the variation of the level of radioactive contamination. By the accident, people who have more knowledge about health of babies may stop getting pregnant and may move to less contaminated area. The selection of getting pregnant and the selective migration of parents may make the estimation and interpretation the estimates difficult. Additionally, some studies used a survey data and the parents with the babies affected by the accident seriously could not response the survey.

[20] Yamamura (2016) analyzed the effects of the accident on the obesity condition of children aged over 5.

If this is the case, the effects may be underestimated.

In this paper, we analyzed the effects of the Fukushima-Daiichi nuclear power plant accident on pregnancy outcomes in Fukushima prefecture using a microdata of babies from a population level dataset. We focus on the babies conceived before the date of the Fukushima-Daiichi accident. By using the population level data and focusing on the babies, we tried to identify the effects of the accident with the biases discussed above. We combined the microdata of pregnancy outcomes with the level of radioactive contamination by using the municipality of the place where the notification of birth was submitted and estimated the effects of level of radioactive contamination on pregnancy outcomes in Fukushima prefecture. In the estimation, we control for the damage of the earthquake to identify the effects of the nuclear power plant accident. Additionally, we control for the unobserved regional heterogeneity, the unobserved birth cohort specific effects, and the maternal and child characteristics which may affect birth outcomes in the model. We divided samples by the trimester of pregnancy on the date of the accident to analyze the heterogeneous effects depending on the timing affected by the radioactive contamination.

According to the estimated results, there are negative effects of the Fukushima-Daiichi accident on pregnancy outcomes such as the premature birth and the low birth weight among boys, especially, the boys who affected by the accident during their 3rd trimester on the date of the accident. On the other hands, there is no significant impact among girls, systematically. There is no systematic relationship between radioactive contamination and the number of births except for the boys suffered from the accident during their 2nd trimester.

The remainder of this paper is organized as follows: section 2 explains the datasets we used in this paper; section 3 explains the estimation model; section 4 discusses the estimation results; and section 5 concludes this research.

2 Data

In this section, we described the two datasets used in this paper. For pregnancy outcomes, we used the vital statistics conducted by Japan's Ministry of Health, Labour and Welfare. Since the parents with newborn baby have to submit the notification of live birth to local governments, the dataset covers all of the newborn babies. The dataset includes the information about newborn babies and parents such as birth weight, duration of pregnancy, birth date, place of birth, age of parents, the municipalities of these babies' registered residence addresses after their birth, and the municipality of the place where parents submitted the notification of birth.

In addition to pregnancy outcomes, for the contamination level of radioactive substances, we used the dataset from the results of deposition of radioactive Cesium of the airborne monitoring survey by prefecture, which is conducted by the Ministry of Education, Culture, Sports, Science and Technology. The dataset includes the soil deposition density of radioactive cesium, cesium-134 and cesium-137, as of May 31, 2012 at the median points of the *quarter grid squares* (approximately 250 m $\times$ 250 m) and covers 22 prefectures. We used the deposition density of the sum of cesium-134 and cesium-137 as the measure of radioactive contamination. [21] We used the median of contamination level within municipalities as the measurement of radioactive contamination and matched the contamination data with the pregnancy outcome by using the municipality of the place submitted the notification of birth.

Figure 22 shows the distribution of soil contamination in Fukushima prefecture. In Figure 22, we divide the contamination level of the sum of cesium-134 and cesium-137 into four levels and plot for each level. The level (i), the brightest plot, indicates the area which contamination level is less than $10kBq/m^2$. Similarly, the levels (ii), (iii), and (iv) are for the areas which contamination levels are between $10kBq/m^2$ and $100kBq/m^2$, between

[21] Please see `https://emdb.jaea.go.jp/emdb/en/portals/b1020201/` for more details.

$100kBq/m^2$ and $200kBq/m^2$, and more than or equal to $200kBq/m^2$, respectively. A marker, "X", indicates the location of the Fukushima-Daiichi nuclear power plant.

According to Figure 22, the land of Fukushima prefecture is more contaminated than other area, especially, near the Fukushima-Daiichi nuclear power plant, and there is a variation of radioactive contamination in Fukushima prefecture. In Fukushima prefecture, about 90 % of the area were contaminated with the level more than $10kBq/m^2$ and about 23 % of the were contaminated with over $200kBq/m^2$ (panel (a)) while, in other prefectures, the proportion of the area with less than $10kBq/m^2$ is about 86 % (panel (b)).

In this paper, we used the newborn babies who are conceived between 1/1/2009 and 3/10/2011 and their parents submitted the birth notification to the local government in Fukushima prefecture. We restricted the sample to all live singleton birth. Additionally, we picked the municipalities which more than 50 % of the area were within a 20-km radius from the Fukushima-Daiichi nuclear power plant and dropped all of the babies who resided in these municipalities because these municipalities were under evacuation order on the next day of the accident.

Figures 23 and 24 show the distribution of pregnancy outcomes. In these figures, we plotted the differences in average pregnancy outcomes for each municipality between babies in-utero on the accident and babies conceived before 6/11/2010, almost all of the babies conceived before 6/11/2010 were born before 3/11/2011. Since the geographical features which may be correlated with the level of radioactive contamination may affect the pregnancy outcomes, for eliminating the bias from these, we plotted the differences in average outcomes. As the measure of pregnancy outcomes, we used the gestation length and birth weight. As the color of plot becomes darker, the change of pregnancy outcome becomes worse. We divided babies in-utero on the accident into three groups by their conceived date. One of the groups includes babies conceived between 12/11/2010 and 3/10/2011, within three months prior to 3/11/2011, and they are suffered from the accident during their first trimester ("1st trimester"). The second group consists babies conceived between 9/11/2010 and 12/10/2010 and they got caught in the earthquake during their second trimester ("2nd trimester"). The third group is babies conceived between 6/11/2010 and 9/11/2010 got the earthquake during their third trimester ("3rd trimester").

According to Figures 23, there is a negative relationship between the radioactive contamination and pregnancy outcomes among boys. In the municipalities near the Fukushima-daiichi nuclear power plant, which are the more contaminated by radioactive substance, the probability of low birth weight among boys is higher than the western area of Fukushima prefecture, especially, the boys who are suffered from the accident during their 3rd trimester. Similarly, according to Figures 24, in the municipalities near the Fukushima-daiichi nuclear power plant, the probability of premature birth among the boys who are suffered from the accident during their 2nd trimester is higher than the western area of Fukushima prefecture.

There is a remark while the figures show the negative relationship between the radioactive contamination and pregnancy outcomes. The radioactive contamination may be correlated with the damage from the earthquake wchich may negatively affect the pregnancy outcomes. If the correlation between the radioactive contamination and the earthquake damage is positive, the effects are negatively biased and overestimated. For identifying effects of the radioactive contamination and the earthquake damage, we estimated an econometric model. Additionally, we add maternal and child characteristics variables which may affect birth outcomes in the model. In the next section, we explain the details of the model.

3 Estimation Model

We estimate the following equation,

$$
\begin{aligned}
y_{ijym} &= \beta_0 + \beta_1 T1_{ym} + \beta_2 T2_{ym} + \beta_3 T3_{ym} \\
&\quad + T1_{ym} \times f(Cs_j) + T2_{ym} \times f(Cs_j) + T3_{ym} \times f(Cs_j) + x_i\delta + \eta_j + \phi_{ym} + u_{ijym} \qquad (10)
\end{aligned}
$$

where i, j, y, and m are indices of individual, municipality, year of conception, and month of conception, respectively. The dependent variable y_{ij} represents pregnancy outcomes such as the length of gestation and the birth weight. The variable $T1_i$ is a dummy variable which takes value one if a baby experienced the Great East Japan Earthquake during their first trimester. Similarly, variables $T2_i$ and $T3_i$ are dummy variables for second and third trimesters on 3/11/2011. The variable $f(Cs_j)$ is a median of the contamination level of sum of Cesium 134 and Cesium 137 in a municipality j and $f(Cs_j)$ is a function of the contamination level. The vector x_i is a set of control variables and that includes the measure of an earthquake damage, mother's age dummy variables, birth order dummy variables, an indicator variable which takes value one if fathers' age is missing, and dummy variables which indicate household's main job at birth [22] . As the measure of earthquake damage, we used the number of damaged residences per household in each municipality. The parameter η_j is residence region fixed effects and captures the unobserved municipality specific heterogeneity. The parameter ϕ_{ym} is conception year and month fixed effects. The parameter u_{ijym} is an error term.

[22] The categories of household's main job are farming, self-employed, employed, others, and unemployed. In the estimation, we use the unemployed dummy as a baseline.

4 Estimation Results

Table 17 shows the estimated results of effects of the radioactive contamination on pregnancy outcomes by gender of baby. In Table 17, we used the categorical contamination levels as the function of the contamination level. We used three contamination levels. The levels, "$0 \leq$ Cs $<$ 100k" corresponds to the levels (i) and (ii), "$100 \leq$ Cs $<$ 200k" for level (iii), and "200k $\leq$ Cs" for level (iv) in Figure 22. Since the babies whose notification of birth is submitted in the municipalities with level (i) is few in number, we combined the levels (i) and (ii). The coefficients of "$= 1$ if t trimester on $3/11 \times = 1$ if l" ($t =$ 1st, 2nd, and 3rd and $l =$ $100 \leq$ Cs $<$ 200k and 200k $\leq$ Cs) are the estimates that we are interested. In the estimation, we used the level "$0 \leq$ Cs $<$ 100k" as the baseline of the category dummy variables.

According to Table 17, there are negative effects on pregnancy outcomes among boys, especially, the boys who are affected by the accident during their 3rd trimester and are in the area which contamination level is over $200kBq/m^2$ while, among girls, there is no significant effects on pregnancy outcomes, systematically. The estimated coefficient of the cross term of the 3rd trimester dummy and the $200k \leq Cs$ dummy is positive and statistically significant for the probabilities of the low birth weight. (Column (4)) The estimate is about 0.051 after controlling for the earthquake damage and maternal and child characteristics. Since the the probability of low birth weight among babies conceived before 9 months prior to the accident is 0.071, the probability increases by about 72 % in the area which contamination level is over $200kBq/m^2$. Similarly, the coefficient for the gestation length is negatively estimated and that for the probability of the premature birth is positively estimated. (Columns (1) and (2)) According to the coefficients, the gestation length decreases by about 0.9 % and the probability of the premature birth increases by about 88 % in the area which contamination level is over $200kBq/m^2$. Similarly, the probability of the low birth weight increases by about 50% among boys who are affected by the accident during their 1st trimester and the probability of the premature birth increases by about 57 % among boys who are affected by the accident during their 2nd trimester. Among boys, there are negative and significant impact of radioactive contamination on pregnancy outcomes after controlling for the earthquake damage and other observable characteristics.

4.1 Discussion

There is some remarks for interpreting the results. One potential concern is the possibility of a selection of pregnancy depending on the radioactive contamination. Suppose that the mothers who have more knowledge about health of babies have healthier children and are likely to stop getting pregnant because of the radioactive contamination. Then, in the area contaminated by radioactive substance, the average pregnancy outcome becomes worth because of the selection of getting pregnant even though effects of radioactive substance on pregnancy outcomes are weak. In this case, the estimated coefficient of radioactive contamination has downward bias. Since, in the estimation, we used the babies who are conceived before the Fukushima-daiichi nuclear power plant accident, we can ignore the bias from these selection of pregnant.

Other concern is the selective migration of mothers. If the mothers with unborn babies whose health condition is not good moved from the area with serious contamination to the area with less contamination, there is a positive bias for estimated coefficients on pregnancy outcomes. In this case, the estimated coefficients are interpreted as the lower bound of the effects of radioactive contamination. Additionally, if wealthy families, who have more access to the optional treatment for their unborn child, are likely to move to the area with less contamination, the coefficients are overestimated and we cannot interpret the estimated coefficients as causal sense.

For discussing the selective migration of mothers, we analyzed the relationship between the radioactive contamination and the number of births in a municipality. We estimated the following equation,

$$
\begin{aligned}
N\_birth_j = \alpha_0 + \alpha_1 1\{100k \leq Cs_j < 200k\} + \alpha_2 1\{200k \leq Cs_j\} \\
+ \alpha_3 Lag\_N\_birth_j + x_j\delta + e_j
\end{aligned}
\tag{11}
$$

where j is the index of municipality. The dependent variable, N_birth_j is the number of births in municipality j. We used the same categorical level of radioactive contamination as Table 17, $1\{100k \leq Cs_j < 200k\}$ and $1\{200k \leq Cs_j\}$. We add the one year lag of the number of births in municipality j to control for the difference in population size among municipalities. The vector, x_j, includes the number of damaged residences per household in municipality j to control for the earthquake damage. In the estimation, we divided the sample by the gender of child and the pregnancy date.

Table 18 shows the estimation results of the relationship between radioactive contamination and the number of births. According to Table 18, there is no systematic relationship between radioactive contamination and the number of births except the boys suffered from the accident during their 2nd trimester. The number of births decreases in the area which contamination level is above $100kBq/m^2$ among the boys who are suffered from the accident during their 2nd trimester. Therefore, we need to be careful to interpret the effects among the boys who are suffered from the accident during their 2nd trimester.

5 Conclusion

In this paper, we analyzed the effects of the Fukushima-Daiichi nuclear power plant accident on pregnancy outcomes in Fukushima prefecture using the microdata of the vital statistics. We combined the vital statistics with the survey of radioactive contamination and estimated the effects of level of radioactive contamination on pregnancy outcomes in Fukushima prefecture. In the estimation, we control for the damage of the earthquake to identify the effects of the nuclear power plant accident. Additionally, we control for the unobserved regional heterogeneity, the unobserved birth cohort specific effects, and the maternal and child characteristics which may affect birth outcomes in the model.

According to the estimated results, there are negative effects of the Fukushima-Daiichi accident on pregnancy outcomes such as the premature birth and the low birth weight among boys, especially, the boys who affected by the accident during their 3rd trimester on the date of the accident. On the other hands, there is no significant impact among girls, systematically. There is no systematic relationship between radioactive contamination and the number of births except for the boys suffered from the accident during their 2nd trimester.

References

Adam, S., Bay, C., Bonsang, E., Germain, S., and Perelman, S. (2006). Occupational activities and cognitive reserve: a frontier approach applied to the Survey on Health, Ageing, and Retirement in Europe (SHARE). *CREPP Working Paper 2006/05*.

Aizer, A. and Stroud, L. (2010). Education, Knowledge and the Evolution of Disparities in Health. Technical report, National Bureau of Economic Research, Cambridge, MA.

Almond, D., Edlund, L., and Source:, M. P. (2009). Chernobyl' s Subclinical Legacy : Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *Quarterly Journal of Economics*, 124(4):1729–1772.

Ando, M., Dahlberg, M., and Engström, G. (2017). The risks of nuclear disaster and its impact on housing prices. *Economics Letters*, 154:13–16.

Barrett, W. B., Heuson, A. J., and Source:, R. W. K. (1986). The Effect of Three Mile Island on Utility Bond Risk Premia: A Note. *The Journal of Finance*, 41(1):255–261.

Bauer, T. K., Braun, S., and Kvasnicka, M. (2013). Distant Event, Local Effects? Fukushima and the German Housing Market. *Kiel Working Papers*, (1857).

Behncke, S. (2012). Does Retirement Trigger Ill Health? *Health economics*, 21:282–300.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *The Journal of Political Economy*, 75(4):352–365.

Bingley, P. and Martinello, A. (2013). Mental retirement and schooling. *European Economic Review*, 63:292–298.

Black, S. E., Bütikofer, A., Devereux, P. J., and Salvanes, K. G. (2013). This Is Only a Test? Long-Run Impacts of Prenatal Exposure to Radioactive Fallout. *NBER Working Paper No. 18987*.

Bonsang, E., Adam, S., and Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of Health Economics*, 31(3):490–501.

Bound, J. and Waidmann, T. (2007). Estimating the Health Effects of Retirement. *University of Michigan Retirement Research Center Working Papers*.

Bowen, R. M. ., Richard P . Castanias, and Lane A . Daley (1983). Intra-Industry Effects of the Accident at Three Mile Island. *The Journal of Financial and Quarterly Analysis*, 18(1):87–111.

Casazza, K., Fontaine, K. R., Astrup, A., Birch, L. L., Brown, A. W., Bohan Brown, M. M., Durant, N., Dutton, G., Foster, E. M., Heymsfield, S. B., McIver, K., Mehta, T., Menachemi, N., Newby, P., Pate, R., Rolls, B. J., Sen, B., Smith, D. L., Thomas, D. M., and Allison, D. B. (2013). Myths, Presumptions, and Facts about Obesity. *New England Journal of Medicine*, 368(5):446–454.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origin of the gradient. *American Economic Review*, 92(5):1308–1334.

Charles, K. K. (2004). Is Retirement Depressing?: Labor Force Inactivity and Psychological Well being in Later Life. *Research in Labor Economics*, 23:269–299.

Chen, C. R. (1984). The structural stability of the market model after the Three Mile Island accident. *Journal of Economics and Business*, 36(1):133–140.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the United States, 2001-2014. *JAMA - Journal of the American Medical Association*, 315(16):1750–1766.

Coe, N. B., Gaudecker, H.-M. V., Lindeboom, M., and Maurer, J. (2012). The Effect Of Retirement On Cognitive Functioning. *Health Economics*, 21:913–927.

Coe, N. B. and Lindeboom, M. (2008). Does Retirement Kill You? Evidence from Early Retirement Windows. *IZA Discussion Papers*, 93.

Coe, N. B. and Zamarro, G. (2011). Retirement effects on health in Europe. *Journal of Health Economics*, 30(1):77–86.

Coulomb, R. and Zylberberg, Y. (2016). Rare events and risk perception : evidence from the Fukushima accident Renaud Coulomb and Yanos Zylberberg Grantham Research Institute on Climate Change and the Environment. *Grantham Research Institute on Climate Change and the Environment Working Paper*, (229).

Currie, J., Neidell, M., and Schmieder, J. F. (2009). Air pollution and infant health: Lessons from New Jersey. *Journal of Health Economics*, 28(3):688–703.

Currie, J. and Schwandt, H. (2016). The 9/11 Dust Cloud and Pregnancy Outcomes: A Reconsideration. *Journal of Human Resources*, 51(4):805–831.

Currie, J. and Walker, R. (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1):65–90.

Dave, D., Rashad, I., and Spasojevic, J. (2008). The Effects of Retirement on Physical and Mental Health Outcomes. *Southern Economic Journal*, 75(2):497–523.

David, P. A., Maude-Griffin, R., and Rothwell, G. (1996). Learning by accident? Reductions in the risk of unplanned outages in U.S. nuclear power plants after Three Mile Island. *Journal of Risk and Uncertainty*, 13(2):175–198.

de Walque, D. (2010). Education, Information, and Smoking Decisions. *Journal of Human Resources*, 45(3):682–717.

Devore, E. E., Stampfer, M. J., Breteler, M. M. B., Rosner, B., Kang, J. H., Okereke, O., Hu, F. B., and Grodstein, F. (2009). Dietary fat intake and cognitive decline in women with type 2 diabetes. *Diabetes Care*, 32(4):635–640.

Dik, M. G., Deeg, D. J., Visser, M., and Jonker, C. (2003). Early Life Physical Activity and Cognition at Old Age. *Journal of Clinical and Experimental Neuropsychology*, 25(5):643–653.

Eide, E. R. and Showalter, M. H. (2011). Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review*, 30(5):778–791.

Eskelinen, M. H., Ngandu, T., Helkala, E. L., Tuomilehto, J., Nissinen, A., Soininen, H., and Kivipelto, M. (2008). Fat intake at midlife and cognitive impairment later in life: A population-based CAIDE study. *International Journal of Geriatric Psychiatry*, 23(7):741–747.

Fink, A. and Stratmann, T. (2015). U.S. housing prices and the Fukushima nuclear accident. *Journal of Economic Behavior and Organization*, 117:309–326.

Gamble, H. B. and Downing, R. H. (1982). Effects of Nuclear Power Plants on Residential Property Values. *Journal of Regional Science*, 22(4):457–478.

Ganpule, A. A., Tanaka, S., Ishikawa-Takata, K., and Tabata, I. (2007). Interindividual variability in sleeping metabolic rate in Japanese subjects. *European Journal of Clinical Nutrition*, 61(11):1256–1261.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255.

Grossman, M. (2006). Chapter 10 Education and Nonmarket Outcomes. *Handbook of the Economics of Education*, 1(06):577–633.

Grossman, M. (2015). The relationship between health and schooling : What ' s new ? *Nordic Journal of Health Economics*, 3(1):7–17.

Gruber, J. and Wise, D. A. (1998). Social Security and Retirement: An International Comparison. *American Economic Review, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association*, 88(2):158–163.

Hackl, F., Halla, M., Hummer, M., and Pruckner, G. J. (2015). The Effectiveness of Health Screening. *Health Economics*, 24(8):913–935.

Halla, M. and Zweimüller, M. (2014). Parental Response to Early Human Capital Shocks: Evidence from the Chernobyl Accident. *Institute for the Study of Labor (IZA) Working paper*, (7968).

Hanaoka, C., Shigeoka, H., and Watanabe, Y. (2018). Do risk preferences change? Evidence from the Great East Japan Earthquake. *American Economic Journal: Applied Economics*, 10(2):298–330.

Hashimoto, H. (2013). Health Consequences of Transitioning to Retirement and Social Participation: Results based on JSTAR panel data. *RIETI Discussion Paper Series 13-E-078*.

Hashimoto, H. (2015). Impacts of Leaving Paid Work on Health, Functions, and Lifestyle Behavior: Evidence from JSTAR panel data. *RIETI Discussion Paper Series 15-E-114*.

Hayashi, M., Fujimori, K., Yasumura, S., Goto, A., and Nakai, A. (2016). Obstetric Outcomes in Women in Fukushima Prefecture during and after the Great East Japan Earthquake and Fukushima Nuclear Power Plant Accident: The Fukushima Health Management Survey. *Open Journal of Obstetrics and Gynecology*, 06(12):705–713.

Hayashi, M. and Hughes, L. (2013). The policy responses to the Fukushima nuclear accident and their effect on Japanese energy security. *Energy Policy*, 59:86–101.

Hernaes, E., Markussen, S., Piggott, J., and Vestad, O. L. (2013). Does retirement age impact mortality? *Journal of Health Economics*, 32(3):586–598.

Hill, J. O., Wyatt, H. R., and Peters, J. C. (2012). Energy balance and obesity. *Circulation*, 126(1):126–132.

Hong, S., Bradshaw, C. J. A., and Brook, B. W. (2013). Evaluating options for the future energy mix of Japan after the Fukushima nuclear crisis. *Energy Policy*, 56:418–424.

Huang, L., Zhou, Y., Han, Y., Hammitt, J. K., Bi, J., and Liu, Y. (2013). Effect of the Fukushima nuclear accident on the risk perception of residents near a nuclear power plant in China. *Proceedings of the National Academy of Sciences*, 110(49):19742–19747.

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., and Langa, K. M. (2013). Monetary Costs of Dementia in the United States. *The New England Journal of Medicine*, 368(14):1326–1334.

Iizuka, T., Nishiyama, K., Chen, B., and Eggleston, K. (2017). Is Preventive Care Worth the Cost? Evidence from Mandatory Checkups in Japan. *NBER Working Paper Series*.

Insler, M. (2014). The Health Consequences of Retirement. *Journal of Human Resources*, 49(1):195–233.

Inui, T., Ito, Y., Kawakami, A., Ma, X. X., Nagashima, M., and Zhao, M. (2017). Empirical Study on the Utilization and Effects of Health Checkups in Japan. *RIETI Discussion Paper Series*, 17-E-082.

Ishii, K., Goto, A., Ota, M., Yasumura, S., and Fujimori, K. (2017). Pregnancy and Birth Survey of the Fukushima Health Management Survey: Review of 4 Surveys Conducted Annually after the Disaster. *Asia-Pacific Journal of Public Health*, 29:56S–62S.

Johnston, D. W. and Lee, W.-S. (2009). Retiring to the good life? The short-term effects of retirement on health. *Economics Letters*, 103(1):8–11.

Kahn, M. E. (2007). Environmental disasters as risk regulation catalysts? The role of Bhopal, Chernobyl, Exxon Valdez, Love Canal, and Three Mile Island in shaping U.S. environmental law. *Journal of Risk and Uncertainty*, 35(1):17–43.

Kajitani, S. (2011). Working in old age and health outcomes in Japan. *Japan and The World Economy*, 23:153–162.

Kajitani, S., Sakata, K., and McKenzie, C. (2013). Does the Definition of Retirement Matter in Estimating the Effects of Retirement on Cognitive Functioning? *20th International Congress on Modelling and Simulation, Adelaide, Australia, 1-6 December 2013*, pages 153–162.

Kajitani, S., Sakata, K., and McKenzie, C. (2014). Occupation, Retirement and Cognitive Functioning. *Discussion Paper Series Graduate School and School of Economics Meisei University*, Discussion Paper Series, No.27.

Kajitani, S., Sakata, K., and McKenzie, C. (2016a). Occupation, retirement and cognitive functioning. *Ageing and Society*, May 2016:1–29.

Kajitani, S., Sakata, K., and McKenzie, C. (2016b). Use It Too Much and Lose It? The Effect of Working Hours on Cognitive Ability. *Melbourne Institute Working Paper*, No. 7/16.

Kawaguchi, D. and Yukutake, N. (2017). Estimating the residential land damage of the Fukushima nuclear accident. *Journal of Urban Economics*, 99:148–160.

Kerkhofs, M. and Lindeboom, M. (1997). Age related health dynamics and changes in labour market status. *Health economics*, 6(4):407–423.

Kim, H. B., Lee, S. A., and Lim, W. (2017). Knowing is not half the battle: impacts of information from the national health screening program in Korea. *IZA Discussion Paper*, (10650).

Latif, E. (2011). The impact of retirement on psychological well-being in Canada. *Journal of Socio-Economics*, 40(4):373–380.

Leppold, C., Nomura, S., Sawano, T., Ozaki, A., Tsubokura, M., Hill, S., Kanazawa, Y., and Anbe, H. (2017). Birth outcomes after the Fukushima Daiichi nuclear power plant disaster: A long-term retrospective study. *International Journal of Environmental Research and Public Health*, 14(5).

Lindeboom, M., Portrait, F., and Van Den Berg, G. J. (2002). An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Economics*, 11(6):505–520.

Lleras-Muney, A. and Lichtenberg, F. (2002). The Effect of Education on Medical Technology Adoption: Are the More Educated More Likely to Use New Drugs. Technical Report 9185, National Bureau of Economic Research, Cambridge, MA.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710.

McEwen, B. S. and Sapolsky, R. M. (1995). Stress and Cognitive Function. *Current Opinion in Neurobiology*, 5:205–216.

Ministry of Health, L. and Welfare (2014). Annual Health, Labour and Welfare Report 2013-2014 (in Japanese).

National Institute of Health and Nutrition (2010). Dietary Reference Intakes for Japanese -2010-.

Nelson, J. P. (1981). Three Mile Island and Residential Property Values: Empirical Analysis and Policy Implications. *Land Economics*, 57(3):363–372.

Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, 29(2):177–201.

Nishimura, Y. and Oikawa, M. (2017a). Mental Retirement: Evidence from Global Aging Data. *MPRA Paper No. 84555*.

Nishimura, Y. and Oikawa, M. (2017b). The Effect of Nuclear Accidents on Land Prices: Evidence from Fukushima-Daiichi in Japan. *SSRN No.3057221*.

Nyberg, L., Habib, R., and Herlitz, A. (2000). Brain activation during episodic memory retrieval: Sex differences. *Acta Psychologica*, 105:181–194.

Nyberg, L., Lovden, M., Riklund, K., Lindenberger, U., and Backman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, 16(5):292–305.

Price, J. and Simon, K. (2009). Patient education and the impact of new medical research. *Journal of Health Economics*, 28(July 2001):1166–1174.

Raji, C., Merrill, D., Eyre, H., Mallam, S., Torosyan, N., Erickson, K., Lopez, O., Becker, J., Carmichael, O., Gach, H., Thompson, P., Longstreth, W., and Kuller, L. (2016). Longitudinal Relationships between Caloric Expenditure and Gray Matter in the Cardiovascular Health Study. *Journal of Aalzheimers Disease*, 52:719–729.

Rohwedder, S. and Willis, R. J. (2010). Mental Retirement. *Journal of Economic Perspectives*, 24(1):119–138.

Satizabal, C. L., Beiser, A. S., Chouraki, V., Chene, G., Dufouil, C., and Seshadri, S. (2016). Incidence of Dementia over Three Decades in the Framingham Heart Study. *The New England Journal of Medicine*, 374:523–532.

Scarmeas, N. and Stern, Y. (2003). Cognitive Reserve and Lifestyle. *Journal of Clinical and Experimental Neuropsychology*, 25(5):483–506.

Semyonov, M., Lewin-Epstein, N., and Maskileyson, D. (2013). Where wealth matters more for health: The wealth-health gradient in 16 countries. *Social Science and Medicine*, 81:10–17.

Sindi, S., Hagman, G., Håkansson, K., Kulmala, J., Nilsen, C., Kåreholt, I., Soininen, H., Solomon, A., and Kivipelto, M. (2016). Midlife Work-Related Stress Increases Dementia Risk in Later Life: The CAIDE 30-Year Study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 00(00):gbw043.

Tajima, K., Yamamoto, M., and Ichinose, D. (2016). How do agricultural markets respond to radiation risk? Evidence from the 2011 disaster in Japan. *Regional Science and Urban Economics*, 60:20–30.

Tanaka, K. and Managi, S. (2016). Impact of a Disaster on Land Price: Evidence From Fukushima Nuclear Power Plant Accident. *The Singapore Economic Review*, 61(01):164003 (15 pages).

Tsugane, S., Sasaki, S., and Tsubono, Y. (2002). Under- and overweight impact on mortality among middle-aged Japanese men and women: a 10-y follow-up of JPHC study cohort I. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, 26:529–537.

Wilson, R. S., Barnes, L. L., and Bennett, D. A. (2003). Assessment of Lifetime Participation in Cognitively Stimulating Activities. *Journal of Clinical and Experimental Neuropsychology*, 25(5):634–642.

Yamamura, E. (2016). Impact of the Fukushima nuclear accident on obesity of children in Japan (2008-2014). *Economics and Human Biology*, 21:110–121.

Yamane, F., Ohgaki, H., and Asano, K. (2013). The Immediate Impact of the Fukushima Daiichi Accident on Local Property Values. *Risk Analysis*, 33(11):2023–2040.

Yasuda, S., Kyozuka, H., Nomura, Y., Fujimori, K., Goto, A., Yasumura, S., Hata, K., Ohira, T., and Abe, M. (2017). Influence of the Great East Japan Earthquake and the Fukushima Daiichi nuclear disaster on the birth weight of newborns in Fukushima Prefecture: Fukushima Health Management Survey. *Journal of Maternal-Fetal and Neonatal Medicine*, 30(24):2900–2904.

Zhao, M., Konishi, Y., and Glewwe, P. (2013). Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption. *Journal of Health Economics*, 32(2):367–385.

Zhu, H., Deng, Y., Zhu, R., and He, X. (2016). Fear of nuclear power? Evidence from Fukushima nuclear accident and land markets in China. *Regional Science and Urban Economics*, 60:139–154.

Tables

Table 1: The Differences in The Characteristics between Salaried and Self-Employed workers by Education

	University Graduates			Non University Graduates		
	(1) **Salaried Workers**	(2) **Self-Employed Workers**	**(1)-(2)**	(3) **Salaried Workers**	(4) **Self-Employed Workers**	**(3)-(4)**
Having health checkup (%)	93.21	62.50	30.71***	84.41	52.52	31.89***
Hours worked (hours/week)	48.37	45.98	2.39	46.38	49.67	-3.30***
Demographics						
Age	54.47	55.35	-0.87*	55.88	56.14	-0.25
Marriage	0.91	0.90	0.01	0.84	0.91	-0.06*
Num of children	1.88	2.00	-0.12	1.96	2.12	-0.16
Economic variables						
Household income (10k USD)	6.82	7.52	-0.71	4.83	5.15	-0.32
Own house	0.77	0.78	-0.01	0.74	0.84	-0.10**
Health condition						
Over weight ($BMI \geq 25$)	0.23	0.37	-0.13*	0.25	0.27	-0.02
Under weight ($BMI < 18.5$)	0.02	0.06	-0.04	0.01	0.01	0.00
Preference for health						
Interested in own health	0.83	0.86	-0.03	0.87	0.83	0.04
Have confidence in own health	0.33	0.35	-0.02	0.34	0.32	0.02

[1] $*p < .1$, $**p < .05$, $***p < .01$.

Table 2: The effects on BMI by education and health condition before policy reform

	Dependent variable: BMI			
	University Graduates		**Non University Graduates**	
	(1)	(2)	(3)	(4)
Panel A: BMI(2007)≥23.5				
Salaried Worker(2007)=1 ×After=1	-1.515***	-1.468***	-0.016	0.012
	(0.380)	(0.457)	(0.386)	(0.386)
Observations	157	157	356	356
Average value of outcome for treated at pre-period	25.8	25.8	25.65	25.65
Panel B: BMI(2007)<23.5				
Salaried Worker(2007)=1 ×After=1	-0.137	-0.136	0.227	0.197
	(0.408)	(0.416)	(0.243)	(0.243)
Observations	169	169	414	414
Average value of outcome for treated at pre-period	21.29	21.29	21.39	21.39
Control variables	✓	✓	✓	✓
Macroeconomic variables		✓		✓

[1] All specification are estimated using FE and clustered robust standard errors in parentheses.
[2] Control variable includes age, age squared, mariage dummy, number of children, household income, house ownership, hours worked, physical stress at workplace, job stress at workplace, and occupation-year fixed effects.
[3] * $p < .1$, ** $p < .05$, *** $p < .01$

Table 3: The effects on diagnosed diseases by education

	Number of Metabolic syndrome related diseases		**Number of Other diseases**	
	(1) **University Graduate**	(2) **Non University Graduate**	(3) **University Graduate**	(4) **Non University Graduate**
Panel A: BMI(2007)≥23.5				
Employee(2007)=1 ×After=1	-0.645***	0.232	-0.276	0.083
	(0.235)	(0.155)	(0.225)	(0.095)
Number of Observations	157	355	157	353
Average value of outcome for treated at pre-period	0.78	0.59	0.43	0.35
Panel B: BMI(2007)<23.5				
Employee(2007)=1 ×After=1	-0.265	-0.128	-0.019	-0.400**
	(0.258)	(0.091)	(0.427)	(0.161)
Number of Observations	168	414	168	414
Average value of outcome for treated at pre-period	0.44	0.46	0.29	0.42
Control variables	✓	✓	✓	✓
Macroeconomic variables	✓	✓	✓	✓

[1] All specification are estimated using FE and clustered robust standard errors in parentheses.
[2] Control variable includes age, age squared, mariage dummy, number of children, household income, house ownership, hours worked, physical stress at workplace, job stress at workplace, and occupation-year fixed effects.
[3] * $p < .1$, ** $p < .05$, *** $p < .01$

Table 4: The effects on physical activity and energy intake by education

	Exercising /Walking		Energy intake (kcal)	
	(1)	(2)	(3)	(4)
	University Graduate	**Non University Graduate**	**University Graduate**	**Non University Graduate**
Panel A: BMI(2007)≥23.5				
Employee(2007)=1 ×After=1	0.399*	0.153	523.941*	16.193
	(0.204)	(0.148)	(275.013)	(144.758)
Number of Observations	157	352	157	356
Average value of outcome for treated at pre-period	0.47	0.4	2239.07	2098.16
Panel B: BMI(2007)<23.5				
Employee(2007)=1 ×After=1	0.205	0.061	112.219	-102.438
	(0.195)	(0.166)	(246.419)	(212.417)
Number of Observations	168	410	169	414
Average value of outcome for treated at pre-period	0.43	0.54	2160.62	2210.2
Control variables	✓	✓	✓	✓
Macroeconomic variables	✓	✓	✓	✓

[1] All specification are estimated using FE and clustered robust standard errors in parentheses.
[2] Control variable includes age, age squared, mariage dummy, number of children, household income, house ownership, hours worked, physical stress at workplace, job stress at workplace, and occupation-year fixed effects.
[3] * $p < .1$, ** $p < .05$, *** $p < .01$

Table 5: The Relationship between The Health Checkup and Some Observable Characteristics before Policy Reform (2007)

Dependent variable: = 1 if having a checkup		
	University Graduates (1)	Non University Graduates (2)
Salaried worker	0.306***	0.323***
	(0.081)	(0.050)
Hours worked	-0.002	0.001
	(0.002)	(0.002)
Age	-0.449	-0.310
	(0.280)	(0.251)
Age squared	0.004	0.003
	(0.003)	(0.002)
Marriage	0.050	0.001
	(0.096)	(0.069)
Number of children	-0.016	-0.015
	(0.028)	(0.025)
HH income (10k USD)	0.008	0.001
	(0.006)	(0.007)
Own house	0.101	0.094*
	(0.065)	(0.052)
Interested in own health	-0.007	0.047
	(0.058)	(0.061)
Have confidence in own health	-0.015	-0.040
	(0.052)	(0.039)
Observations	184	465
City-FEs	✓	✓

[1] Standard errors in parentheses.
[2] $*p < .1$, $**p < .05$, $***p < .01$.

Table 6: Summary Statistics of Cognition Scores (Age 60 -69) around 2010

	Obs.	Mean	S.D.	Min	Max
HRS					
Word Recall Summary Score	5057	10.33	3.23	0	20
Immediate Word Recall	5057	5.64	1.56	0	10
Delayed Word Recall	5057	4.68	1.89	0	10
Serial 7s	5057	3.53	1.65	0	5
ELSA[1]					
Word Recall Summary Score	3593	11.17	3.32	0	20
Immediate Word Recall	3592	6.17	1.67	0	10
Delayed Word Recall	3593	5.01	1.93	0	10
SHARE[2]					
Word Recall Summary Score	18998	9.31	3.40	0	20
Immediate Word Recall	19025	5.33	1.68	0	10
Delayed Word Recall	19019	3.97	2.03	0	10
Serial 7s	18576	3.76	1.74	0	5
JSTAR					
Word Recall Summary Score	1463	10.10	3.00	0	20
Immediate Word Recall	1501	5.27	1.49	0	10
Delayed Word Recall	1471	4.80	1.85	0	10
Serial 7s	1508	4.10	1.20	0	5
CHARLS					
Word Recall Summary Score	3838	6.89	3.16	0	18
Immediate Word Recall	3890	3.91	1.60	0	10
Delayed Word Recall	3856	2.95	1.87	0	10
Serial 7s	3880	3.11	1.88	0	5
KLoSA					
Word Recall Summary Score[3]	2253	4.74	1.31	0	6
Immediate Word Recall[3]	2253	2.68	0.67	0	3
Delayed Word Recall[3]	2253	2.06	0.96	0	3
Serial 7s	2253	3.83	1.57	0	5

[1] ELSA does not include Serial 7s scores.

[2] Calculated using weight.

[3] KLoSA's Word Recall Scores are not comparable with other datasets.

Table 7: Summary Statistics: The US (Age:60-69) at 2010

	Obs.	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max
	Male					**Female**				
Word Recall Summary Score	2038	9.70	3.14	0	20	3019	10.76	3.21	0	20
Immediate Word Recall	2038	5.37	1.56	0	10	3019	5.82	1.54	0	10
Delayed Word Recall	2038	4.32	1.82	0	10	3019	4.92	1.90	0	10
Serial 7s	2038	3.76	1.55	0	5	3019	3.38	1.69	0	5
	Not University Graduate					**University Graduate**				
Word Recall Summary Score	3819	9.86	3.14	0	20	1236	11.81	3.05	0	20
Immediate Word Recall	3819	5.42	1.53	0	10	1236	6.33	1.45	0	10
Delayed Word Recall	3819	4.43	1.84	0	10	1236	5.45	1.82	0	10
Serial 7s	3819	3.27	1.70	0	5	1236	4.33	1.12	0	5
	White Collar					**Blue Collar**				
Word Recall Summary Score	2889	11.04	3.13	0	20	1027	9.36	3.05	1	19
Immediate Word Recall	2889	5.95	1.50	0	10	1027	5.17	1.50	1	10
Delayed Word Recall	2889	5.08	1.85	0	10	1027	4.17	1.76	0	10
Serial 7s	2889	3.80	1.51	0	5	1027	3.25	1.69	0	5

Table 8: Datasets Used in this Study

	Wave	Year
Cross Sectional Analysis (Section 4)		
HRS	7,10	2004,2010
SHARE	1,4	2004,2010
ELSA	2,5	2004,2010
JSTAR	2	2009
CHARLS	1	2011
Dynamic Analysis (Section 5)		
HRS	3-10	1996-2010
SHARE[1]	1-5	2004-2012
ELSA	1-6	2002-2014
JSTAR	1-4	2007-2013
KLoSA	1-4	2006-2012

[1]: Only Denmark, France and Germany are analyzed.

Table 9: Effect of Choice of Countries Analyzed (Sample Aged 60-69)

	Latin		**Slavic**		**Germanic**		**New SHARE /East Asia**		**Original /witout Greece**		**All countries**	
	(1) OLS	(2) IV2	(3) OLS	(4) IV2	(5) OLS	(6) IV2	(7) OLS	(8) IV2	(9) OLS	(10) IV2	(11) OLS	(12) IV2
Panel A: without controls												
1st stage												
$1\{age \geq PAE\}$		0.235***		0.051***		0.060***		0.474***		0.136***		0.328***
		(0.062)		(0.014)		(0.018)		(0.012)		(0.010)		(0.008)
$1\{age \geq PAN\}$		0.167***		0.243***		0.328***		0.123***		0.221***		0.120***
		(0.014)		(0.028)		(0.020)		(0.021)		(0.010)		(0.008)
2nd stage												
Not working for pay	-0.823***	-2.396***	-0.449**	0.119	-0.661***	-1.116***	0.833***	1.247***	-1.153***	-2.463***	0.531***	3.728***
	(0.159)	(0.621)	(0.220)	(0.839)	(0.137)	(0.412)	(0.104)	(0.225)	(0.069)	(0.245)	(0.071)	(0.210)
Observations	3620	3620	6086	6086	8802	8802	10315	10315	16746	16746	27061	27061
DWH p-value		0.006		0.251		0.409		0.000		0.000		0.000
Panel B: with controls												
1st stage												
$1\{age \geq PAE\}$		0.079		-0.020		0.013		0.091***		0.002		0.052***
		(0.063)		(0.026)		(0.048)		(0.024)		(0.020)		(0.017)
$1\{age \geq PAN\}$		0.037*		0.108***		0.091***		0.118***		0.052***		0.091***
		(0.020)		(0.031)		(0.028)		(0.027)		(0.015)		(0.014)
2nd stage												
Not working for pay	-0.362**	-1.613	-0.138	2.917	-0.333**	-0.867	0.411***	2.793*	-0.608***	-3.940*	-0.030	-0.081
	(0.159)	(3.594)	(0.231)	(2.513)	(0.151)	(2.147)	(0.110)	(1.585)	(0.072)	(2.336)	(0.072)	(0.958)
Observations	3620	3620	6086	6086	8802	8802	10315	10315	16746	16746	27061	27061
DWH p-value		0.752		0.204		0.801		0.558		0.146		0.960

[1] Standard errors in parentheses, * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] "Latin" shows the estimated results including countries such as France, Spain Portugal, and Italy. "Slavic" includes only European countries: Estonia, Slovenia, Poland, Hungary, and Czech Republic. "Germanic" includes European countries as well: the U.K., the Netherlands, Germany, Denmark, Belgium, Sweden, Austria, and Switzerland. "New SHARE/East Asia" includes Japan, China, Czech Republic, Poland, Hungary, Portugal, Slovenia, and Estonia. "Original without Greece" includes the countries in the "original" set(the set of analyzed countries used by Rohwedder and Willis (2010)) without Greece.
[3] All specifications are estimated with the sampling weight to adjust the population size of each country.
[4] In Panel B, we also include the demographic variables (age, age squared, female dummy, the dummy which takes one if university graduate or more, marriage dummy, number of children), economic variables (household income, house ownership, total wealth), the country dummy variables, and the interaction term of the economic variables and the country dummy variables (e.g., $wealth \times country(j)$).

Table 10: Time Spent on Various Activities Before and After Retirement (Hours)

Activities	**(1) Not retired**	**(2) Retired**	**(3) (2)-(1)**
Working for pay	3.9	0.1	-3.8
Using the computer	1.7	0.6	-1.1
Watching TV	2.5	3.4	0.9
Walking	1.1	0.8	-0.3
Attending social activities	2	2.2	0.2
Reading newspapers	0.6	0.8	0.2
Listening to music	1.1	0.9	-0.2
House cleaning	0.6	0.8	0.2
Preparing meals and cleaning-up afterwards	0.8	1	0.2
Sleeping and napping	6.7	6.6	-0.1
Visiting in-person with friends	1	1.1	0.1
Washing, ironing, or mending clothes	0.3	0.4	0.1
Yard work or gardening	0.3	0.4	0.1
Playing cards or games, or solving puzzles	0.1	0.2	0.1
Reading books	0.5	0.6	0.1
Praying or meditating	0.5	0.6	0.1
Shopping or running errands	0.5	0.6	0.1
Physically showing affection	0.5	0.4	-0.1
Treating or managing an existing medical condition	0.2	0.3	0.1
Participating in sports	0.3	0.3	0
Communicating by phone, letters, e-mail	0.8	0.8	0
Personal grooming	1	1	0
Caring for pets	0.4	0.4	0
Helping friends	0.2	0.2	0
Doing volunteer work	0.1	0.1	0
Attending religious services	0.1	0.1	0
Attending meetings of clubs or religious groups	0.1	0.1	0
Taking care of finances or investments	0.1	0.1	0
Attending concerts, movies	0	0	0
Singing or playing a musical instrument	0	0	0
Doing arts and crafts projects	0.1	0.1	0
Doing home improvements	0.1	0.1	0

Table 11: Effect of Pension Eligibility Age on Full Retirement

	US		
	(1) Full	(2) Male	(3) Female
$1\{age \geq A^{eb}\}$	0.078***	0.094***	0.065***
	(0.007)	(0.010)	(0.010)
$1\{age \geq A^{fb}\}$	1.103***	1.025***	1.188***
	(0.216)	(0.319)	(0.293)
$1\{age \geq A^{fb}\} \times age$	-0.016***	-0.015***	-0.017***
	(0.003)	(0.005)	(0.005)
Observations	86773	38848	47925

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 12: Heterogeneity of Observable Characteristics 1

	Gender		Education		Occupation	
	(1)	(2)	(3)	(4)	(5)	(6)
	Male	Female	Low	High	Blue collar	White collar
WR summary score						
Completely retired	-0.137***	-0.164***	-0.149***	-0.158***	-0.083	-0.169***
	(0.040)	(0.036)	(0.030)	(0.058)	(0.055)	(0.034)
Observations	38848	47925	65323	21433	19404	53512
DWH p-val	0.179	0.134	0.397	0.425	0.340	0.301
Model	FE	FE	FE	FE	FE	FE
Immediate WR						
Completely retired	-0.058***	-0.064***	-0.053***	-0.088***	-0.054*	-0.405**
	(0.020)	(0.018)	(0.015)	(0.030)	(0.028)	(0.203)
Observations	38848	47925	65323	21433	19404	53512
DWH p-val	0.306	0.948	0.852	0.131	0.530	0.097
Model	FE	FE	FE	FE	FE	FE-IV
Delayed WR						
Completely retired	-0.081***	0.451*	-0.098***	-0.070**	-0.029	-0.104***
	(0.024)	(0.252)	(0.019)	(0.035)	(0.033)	(0.021)
Observations	38848	47925	65323	21433	19404	53512
DWH p-val	0.175	0.023	0.236	0.931	0.314	0.715
Model	FE	FE-IV	FE	FE	FE	FE
Serial 7s						
Completely retired	-0.008	-0.415**	-0.321**	-0.039*	-0.048**	-0.338**
	(0.016)	(0.183)	(0.141)	(0.022)	(0.024)	(0.168)
Observations	38848	47925	65323	21433	19404	53512
DWH p-val	0.937	0.040	0.035	0.761	0.674	0.058
Model	FE	FE-IV	FE-IV	FE	FE	FE-IV
Backward counting						
Completely retired	-0.002	-0.001	-0.001	0.121**	-0.003	-0.000
	(0.003)	(0.003)	(0.002)	(0.055)	(0.005)	(0.002)
Observations	38848	47925	65323	21433	19404	53512
DWH p-val	0.167	0.552	0.505	0.019	0.675	0.262
Model	FE	FE	FE	FE-IV	FE	FE

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 13: Heterogeneity of Observable Characteristics 2

	Body Mass Index		Fat intake	
	(1)	(2)	(3)	(4)
	< 25	≥ 25	$< Median$	$\geq Median$
WR summary score				
Completely retired	-0.066	-0.191***	-0.077	-0.119*
	(0.049)	(0.033)	(0.064)	(0.064)
Observations	26866	59286	15217	15187
DWH p-val	0.769	0.724	0.466	0.717
Model	FE	FE	FE	FE
Immediate WR				
Completely retired	-0.025	-0.076***	-0.026	-0.063**
	(0.024)	(0.016)	(0.031)	(0.032)
Observations	26866	59286	15217	15187
DWH p-val	0.283	0.975	0.886	0.946
Model	FE	FE	FE	FE
Delayed WR				
Completely retired	-0.044	-0.115***	-0.052	-0.062
	(0.030)	(0.020)	(0.040)	(0.040)
Observations	26866	59286	15217	15187
DWH p-val	0.173	0.619	0.189	0.511
Model	FE	FE	FE	FE
Serial 7s				
Completely retired	-0.034	-0.026*	-0.588*	-0.823***
	(0.021)	(0.014)	(0.303)	(0.274)
Observations	26866	59286	15217	15187
DWH p-val	0.156	0.304	0.065	0.002
Model	FE	FE	FE-IV	FE-IV
Backward counting				
Completely retired	-0.004	-0.000	0.007	0.010*
	(0.004)	(0.003)	(0.005)	(0.005)
Observations	26866	59286	15217	15187
DWH p-val	0.158	0.322	0.153	0.875
Model	FE	FE	FE	FE

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 14: Heterogeneity in Initial Score and Change in Leisure Activities

	Initial scores		Change in social attendance		Having spouse at 1st interview	
	(1) $< Median$	(2) $\geq Median$	(3) Not increase	(4) Increase	(5) No	(6) Yes
WR summary score						
Completely retired	-0.238***	-0.179***	0.030	-0.165	-0.217***	-0.134***
	(0.053)	(0.043)	(0.120)	(0.110)	(0.059)	(0.030)
Observations	21704	36127	4384	4523	20053	66720
DWH p-val	0.790	0.229	0.193	0.978	0.912	0.492
Model	FE	FE	FE	FE	FE	FE
Immediate WR						
Completely retired	-0.671*	-0.090***	0.058	-0.046	-0.103***	-0.049***
	(0.349)	(0.020)	(0.061)	(0.058)	(0.029)	(0.015)
Observations	17879	39952	4384	4523	20053	66720
DWH p-val	0.087	0.266	0.134	0.257	0.407	0.884
Model	FE-IV	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.137***	-0.097***	-0.013	-0.119*	-0.113***	-0.087***
	(0.035)	(0.025)	(0.068)	(0.066)	(0.036)	(0.018)
Observations	19011	38820	4384	4523	20053	66720
DWH p-val	0.989	0.388	0.453	0.362	0.630	0.240
Model	FE	FE	FE	FE	FE	FE
Serial 7s						
Completely retired	-0.080***	-0.410*	-0.028	0.012	-0.050**	-0.023*
	(0.028)	(0.216)	(0.053)	(0.047)	(0.025)	(0.013)
Observations	19034	38797	4384	4523	20053	66720
DWH p-val	0.603	0.052	0.520	0.901	0.115	0.155
Model	FE	FE-IV	FE	FE	FE	FE
Backward counting						
Completely retired			-0.003	-0.016*	-0.002	0.050**
			(0.010)	(0.008)	(0.005)	(0.024)
Observations			4384	4523	20053	66720
DWH p-val			0.414	0.200	0.992	0.031
Model			FE	FE	FE	FE-IV

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 15: Effect of Retirement on Cognitive Function by Age Group

	Age 58-63		Age 64-69	
	(1) Male	(2) Female	(3) Male	(4) Female
WR summary score				
Completely retired	-0.106	0.055	-0.096	-0.268***
	(0.095)	(0.081)	(0.092)	(0.093)
Observations	11263	16596	10892	14951
DWH p-val	0.479	0.382	0.444	0.537
Model	FE	FE	FE	FE
Immediate WR				
Completely retired	-0.039	0.032	-0.039	-0.117**
	(0.049)	(0.041)	(0.049)	(0.046)
Observations	11263	16596	10892	14951
DWH p-val	0.438	0.405	0.743	0.720
Model	FE	FE	FE	FE
Delayed WR				
Completely retired	-0.067	0.024	-0.057	-0.147***
	(0.058)	(0.050)	(0.055)	(0.057)
Observations	11263	16596	10892	14951
DWH p-val	0.612	0.460	0.347	0.192
Model	FE	FE	FE	FE
Serial 7s				
Completely retired	0.070*	-3.546	-3.213	-0.018
	(0.040)	(2.749)	(2.028)	(0.041)
Observations	11263	16596	10892	14951
DWH p-val	0.194	0.078	0.024	0.460
Model	FE	FE-IV	FE-IV	FE
Backward counting				
Completely retired	-0.003	-0.007	-0.003	-0.743
	(0.008)	(0.007)	(0.007)	(0.607)
Observations	11263	16596	10892	14951
DWH p-val	0.859	0.773	0.778	0.078
Model	FE	FE	FE	FE-IV

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include age and age squared.

Table 16: Summary of Estimation Results in the Related Literature

	Lindeboom et al.	Rohwedder and Willis	Coe and Zamarro	Behncke	Bonsang et al.	Mazzonna and Peracchi	Coe, Gaudecker, Lindeboom and Maurer	Bingley and Martinello	Motegi, Nishimura and Oikawa
	2002, Health Economics	2010, J Econ Perspectives	2011, J Health Economics	2012, Health Economics	2012, J Health Economics	2012, European Economic Review	2012, Health Economics	2013, European Economic Review	2016
cognitive functioning	negative(MMSE (tests cognitive abilities))	negative	no	negative	negative	negative	positive (blue color) no (white color)	negative	negative (Word Recall, US), no (Word Recall, England, Germany, France, Denmark), positive (Word Recall, Korea), negative (Serial 7, US, Korea)
Method	FE method	IV method	IV method	Nonparametric matching	FE-IV method	IV method	Generalization of 2SLS	IV method	FE-IV method
Method (details)		IVs: pension eligibility age for early and full	IVs: eligibility age for early and full retirement	Using state pension eligibility age as IV	IVs: pension eligibility age	IVs: pension eligibility age for early and full	IVs: pension eligibility age (nonparametric regression of first stage regression)	IVs: pension eligibility age for early and full	IVs: pension eligibility age for early and full
Def. of Retirement		not having worked for pay in the last 4 weeks	someone who is not in the paid labor force	retired describes her current situation best and not in paid work was her activity in the last month	not having worked for pay in the last 1 year	max {0, current age-age as retirement} including unemployment elderly as retirement	interview year-retirement year (calculating by units of month and convert to the unit of year)	not having worked for pay in the last 4 weeks	not working for pay and self-reported retire
Controls(Demog.)	age, residential area, marital status, children' health		education, marital status, children	children, birth place, residential area	age	age and education	education, race, religion and age	age, sex, and education	age, sex, family strcture and education
Controls(Economic)			income	income					income, asset
Controls(Working.)	employment status		self employment	working hours, employment status					
Controls(Health)	health								
Data	Longitudinal Aging Study Amsterdam panel 92, 95, 98	HRS ELSA SHARE at 2004	SHARE 1st-2nd wave	ELSA 1st-3rd wave	HRS 1998~2008 6 waves	SHARE 2004, 06	HRS, only male elderly born after 1931	HRS ELSA SHARE 2004	HRS 1996-2010, SHARE 2004-2012, ELSA 2002-2014, JSTAR 2007-2013, KLoSA 2006-2012
Country	Netherlands	The U.S. The U.K. EU	EU	The U.K.	The U.S.	EU	The U.S.	The U.S. The U.K. EU	The US, The UK, France, Germany, Denmark, Korea, Japan

Table 17: Effects of Radioactive Contamination Level on Pregnancy Outcomes by Gender using Categorical Contamination Levels

	Boy				Girl			
	Gestation Length		Birth Weight		Gestation Length		Birth Weight	
	(1) Day	(2) < 37weeks	(3) Gram	(4) < 2500g	(5) Day	(6) < 37weeks	(7) Gram	(8) < 2500g
= 1 if 1st trimester on 3/11	-1.604	0.050*	-48.808	0.029	0.588	0.000	80.942	-0.022
	(1.248)	(0.028)	(57.525)	(0.038)	(1.757)	(0.027)	(64.854)	(0.040)
× = 1 if 100k ≤ Cs < 200k	0.073	-0.003	-20.572	0.004	-0.203	0.003	-7.773	0.007
	(0.420)	(0.009)	(13.074)	(0.010)	(0.392)	(0.008)	(17.196)	(0.010)
× = 1 if 200k ≤ Cs	-1.142	0.005	-42.817	0.035**	-1.087	0.013	-38.392	0.037
	(0.842)	(0.017)	(32.194)	(0.016)	(1.126)	(0.017)	(37.742)	(0.028)
= 1 if 2nd trimester on 3/11	0.124	0.014	-12.636	0.006	0.819	0.006	35.602	0.013
	(1.393)	(0.030)	(58.788)	(0.034)	(1.274)	(0.022)	(44.249)	(0.030)
× = 1 if 100k ≤ Cs < 200k	-0.279	0.015*	3.314	-0.009	-0.033	-0.001	-28.154*	-0.011
	(0.333)	(0.008)	(17.652)	(0.010)	(0.461)	(0.006)	(16.138)	(0.012)
× = 1 if 200k ≤ Cs	-0.554	0.028***	-23.189	0.004	-1.006	0.012	-66.394***	0.027
	(0.709)	(0.008)	(28.007)	(0.013)	(1.165)	(0.021)	(24.433)	(0.022)
= 1 if 3rd trimester on 3/11	-0.812	0.002	-10.865	-0.000	1.434	0.000	46.552	-0.003
	(1.043)	(0.025)	(49.074)	(0.019)	(1.059)	(0.016)	(40.484)	(0.025)
× = 1 if 100k ≤ Cs < 200k	-0.382	0.012	7.449	0.005	0.122	0.004	14.462	-0.020
	(0.500)	(0.012)	(24.316)	(0.013)	(0.565)	(0.009)	(25.813)	(0.013)
× = 1 if 200k ≤ Cs	-2.464***	0.043*	-51.912	0.051**	-0.138	-0.026*	31.409	-0.015
	(0.841)	(0.023)	(46.646)	(0.022)	(0.973)	(0.014)	(21.119)	(0.024)
N	17100	17100	17100	17100	15971	15971	15971	15971
Average outcomes among babies conceived before 6/11/2010	274.836	0.049	3057.160	0.071	275.802	0.039	2966.507	0.096
Conceived Year × Month FE	✓	✓	✓	✓	✓	✓	✓	✓
Residence region FE	✓	✓	✓	✓	✓	✓	✓	✓

[1] Standard errors are in parentheses and clustered robust by municipalities.
[2] Control variable includes mother's age dummy variables, birth order dummy variables, a indicator which takes value one if the father's information is missing, household's man job dummy variables (farming, self-employed, employed, others)(baseline: unemployed), and the number of damaged residences per household.
[3] * $p < .1$, ** $p < .05$, *** $p < .01$

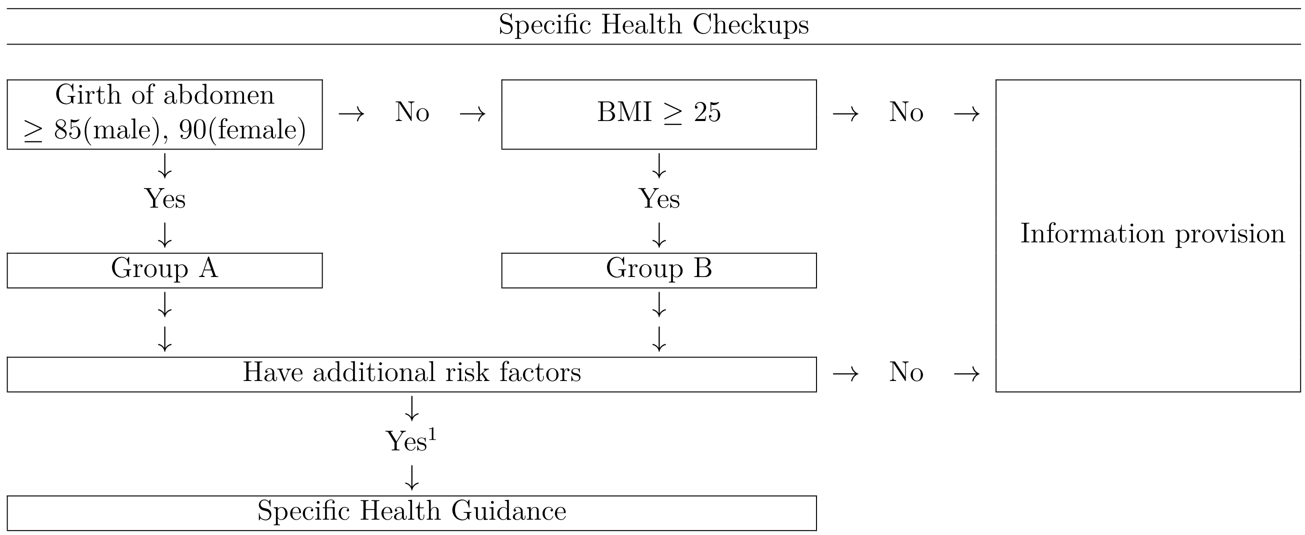
Table 18: The Relationship between Radioactive Contamination and Number of Births Excluding Stillbirth

	Boy			Girl		
	3rd trimester	2nd trimester	1st trimester	3rd trimester	2nd trimester	1st trimester
= 1 if 100k $\leq$ Cs $<$ 200k	3.843	-7.334*	-3.561	4.946	-1.397	-3.766*
	(2.798)	(3.880)	(3.840)	(4.027)	(3.093)	(2.063)
= 1 if 200k $\leq$ Cs	-2.300	-7.735*	-2.712	-3.820	-6.722	-5.395
	(5.011)	(3.864)	(3.928)	(5.326)	(4.341)	(3.638)
Constant	1.162	1.160	1.442	0.971	1.534*	1.121
	(0.998)	(0.948)	(1.356)	(1.155)	(0.858)	(0.970)
N	54	54	54	54	54	54

[1] Standard errors are in parentheses.
[2] Control variable includes the one year lag of the number of births in municipality and the number of damaged residences per household.
[3] * $p < .1$, ** $p < .05$, *** $p < .01$

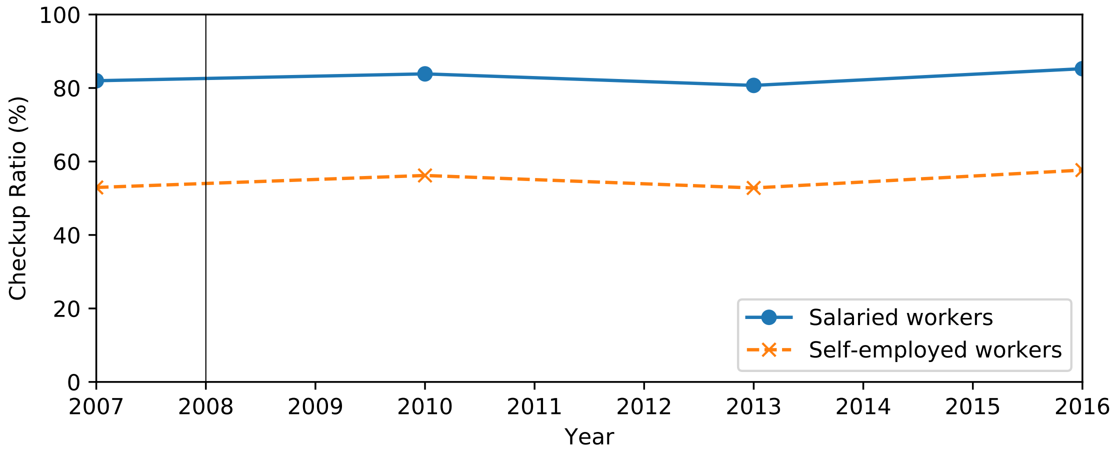
Figures

Figure 1: The screening procedure of the health guidance status



[1] The participants are assigned the level of the guidance depending on the number of additional risk factors. In Group A, examinees with more than two risk factors receive the active support guidance, and examinees with one receive the motivational support guidance. In Group B, examinees with more than three risk factors receive the active support guidance, and examinees with one or two receive the motivational support guidance

Figure 2: The Participation Rate of Health Checkups in Japan by Employment Status



[1] I calculate the lines using Comprehensive Survey of Living Conditions and use males aged over 20 for the figure.

Figure 3: The Relationship between BMI and The Eligibility of The Health Guidance (2009)

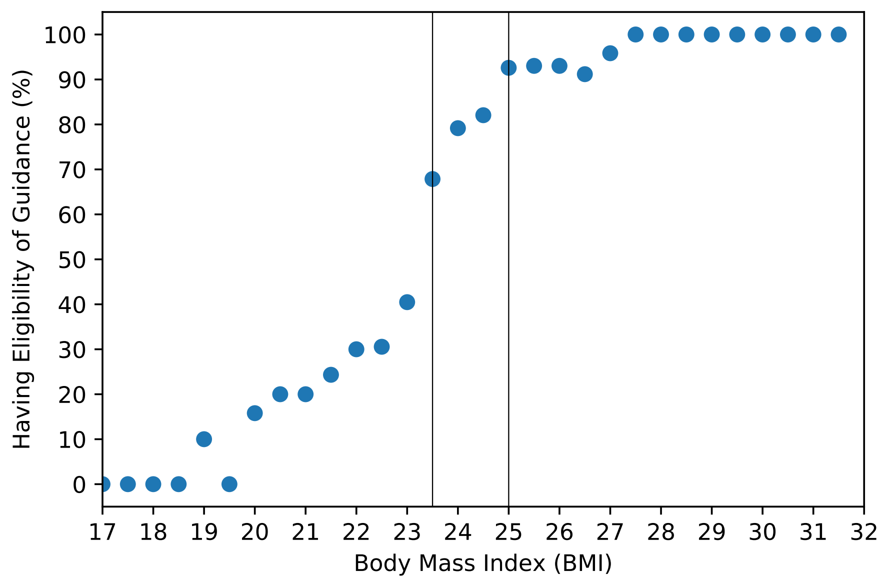
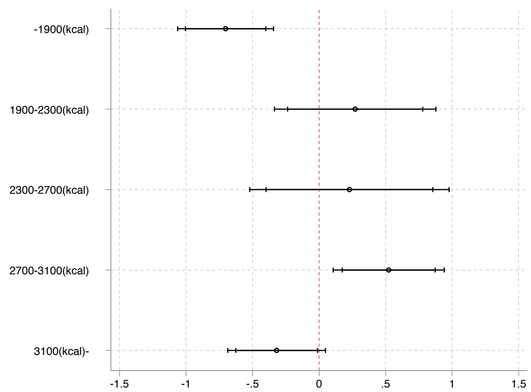
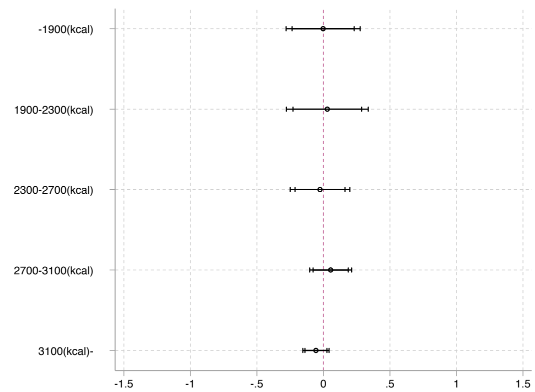


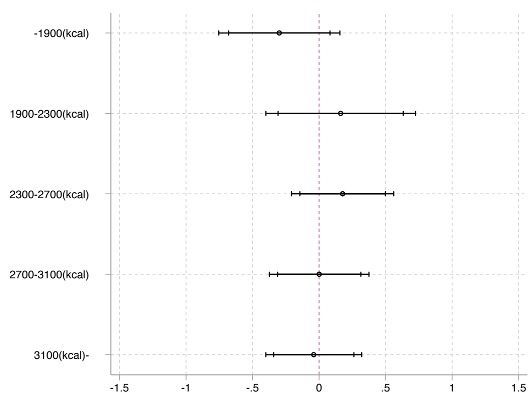
Figure 4: The effects on energy intakes



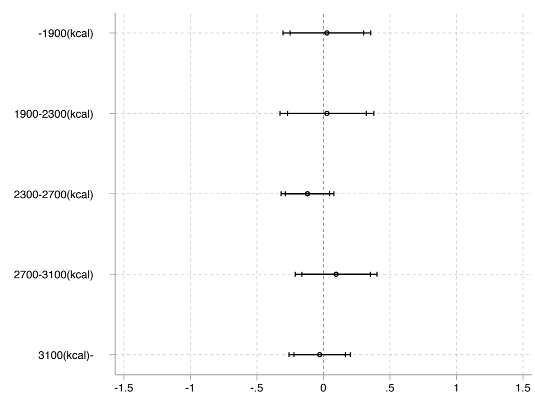
(a) University graduates, BMI(2007)≥23.5



(b) Non university graduates, BMI(2007)≥23.5



(c) University graduates, BMI(2007)<23.5



(d) Non university graduates, BMI(2007)<23.5

[1] This figure reports the DID estimates on energy intakes by education and health condition before policy reform. The plot corresponds to the estimated coefficient on an energy intake dummy variable. The bar is indicates confidence interval and the outer bar is the 95 % level confidence interval and inner bar is the 90 % level confidence interval. Label on the left indicates a range of energy intake which dummy variable takes value one. For example, ‘‘-1900(kcal)” means that the dependent variable of DID estimation is the dummy variable which takes value one when the level of energy intake is in the range between 0 and 1900 kcal.

Figure 17: The Target of our Analysis

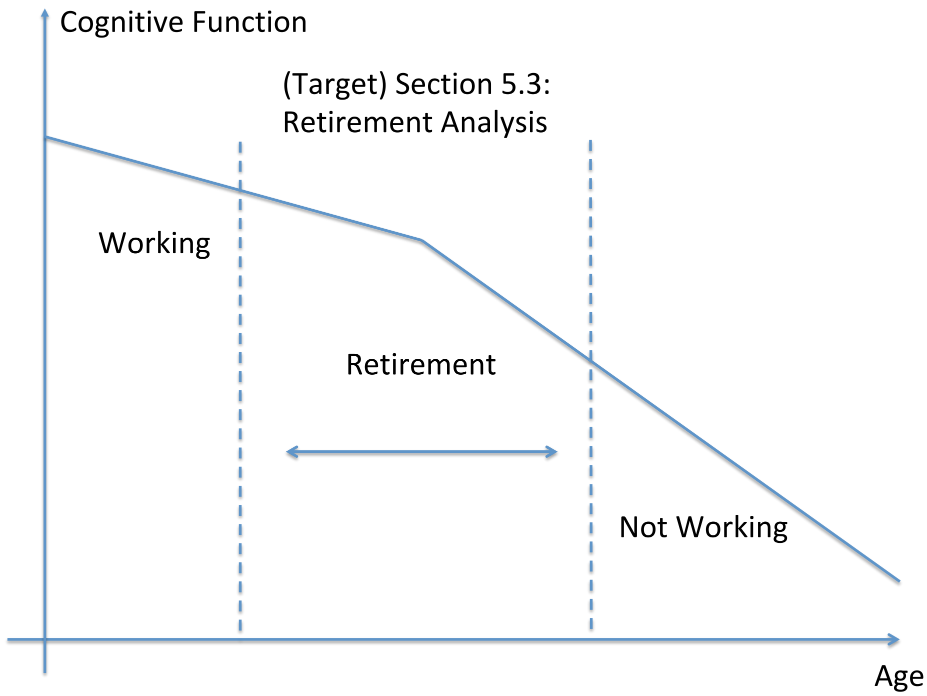
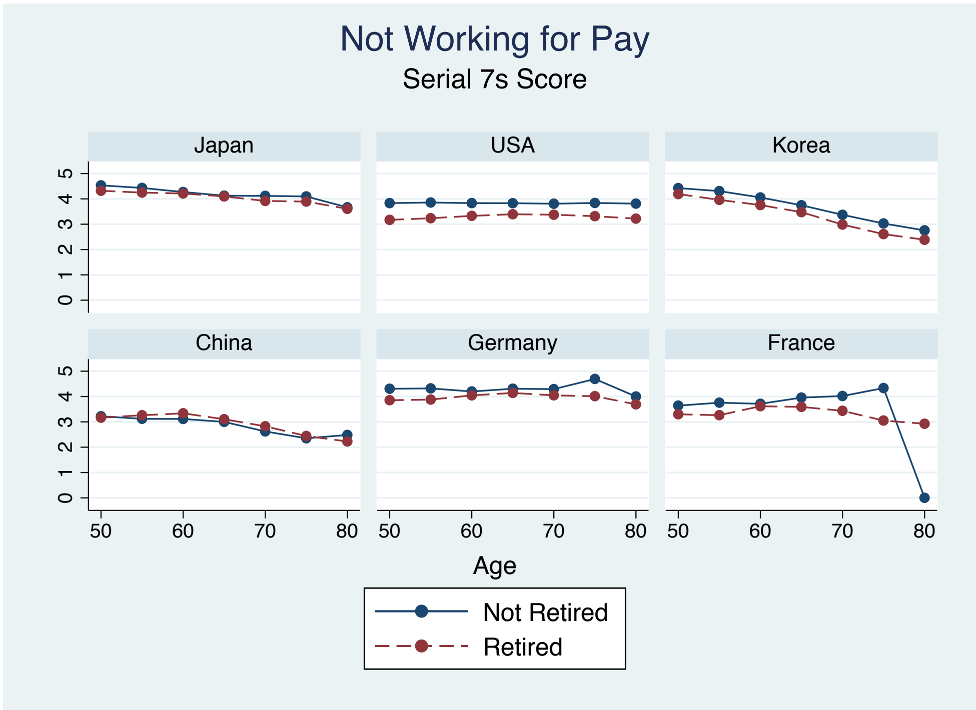


Figure 5: Serial 7s Score by Country (All Waves)



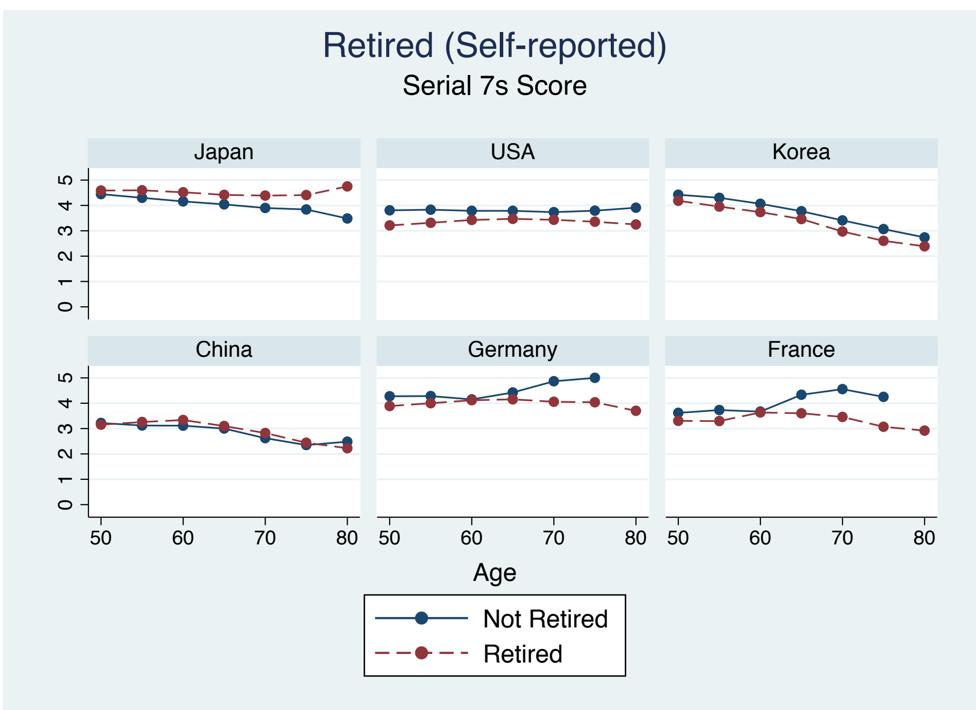
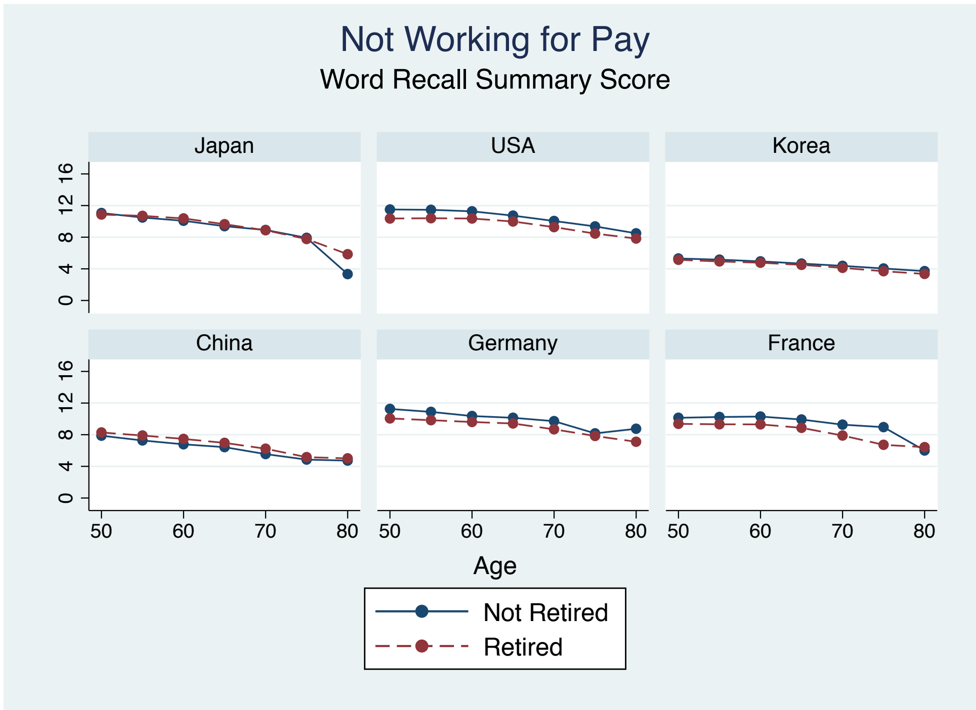


Figure 6: Word Recall Summary Score by Country (All Waves)



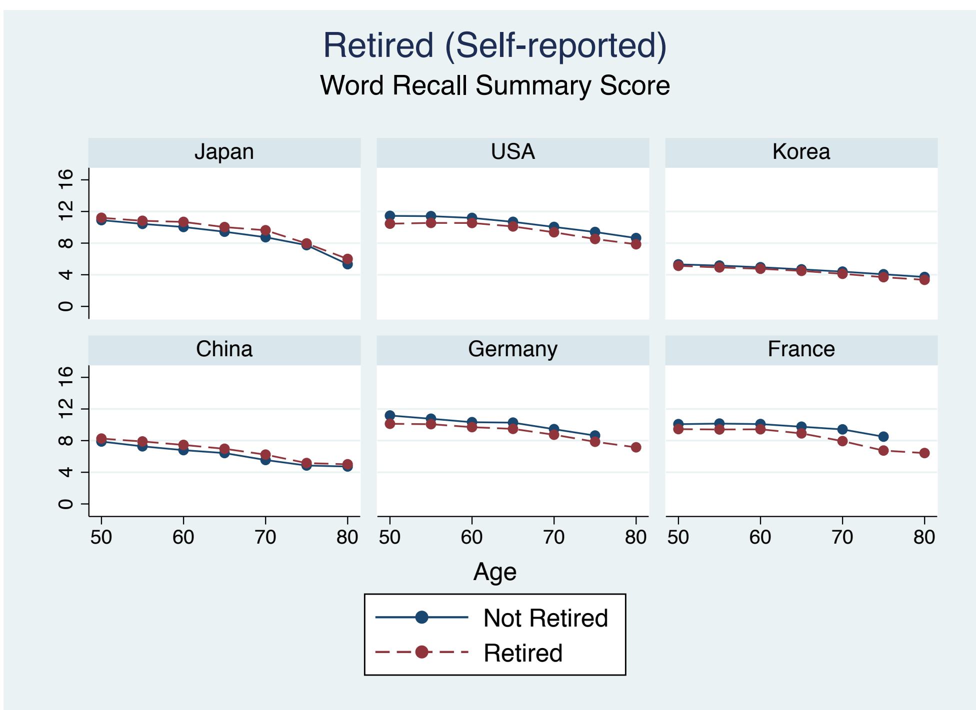


Figure 7: Serial 7 Scores in the U.S. and China (All Waves) by Gender, Education, Occupational Type and Wealth

Figure 8: Serial 7 Scores in the U.S. and China (All Waves) by Gender, Education, Occupational Type and Wealth

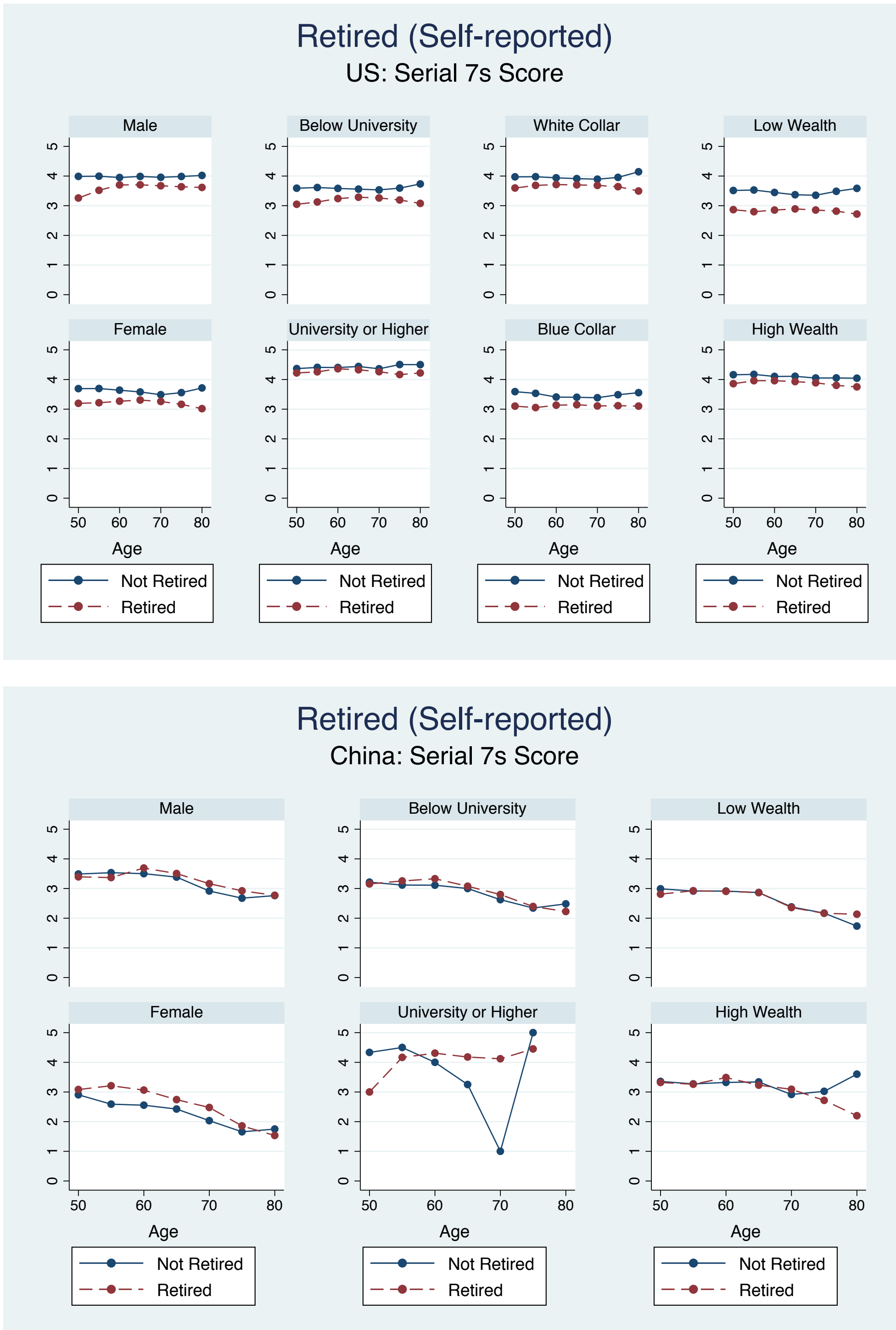


Figure 9: Word Recall Summary Scores in the U.S. and China (All Waves) by Gender, Education, Occupational Type and Wealth

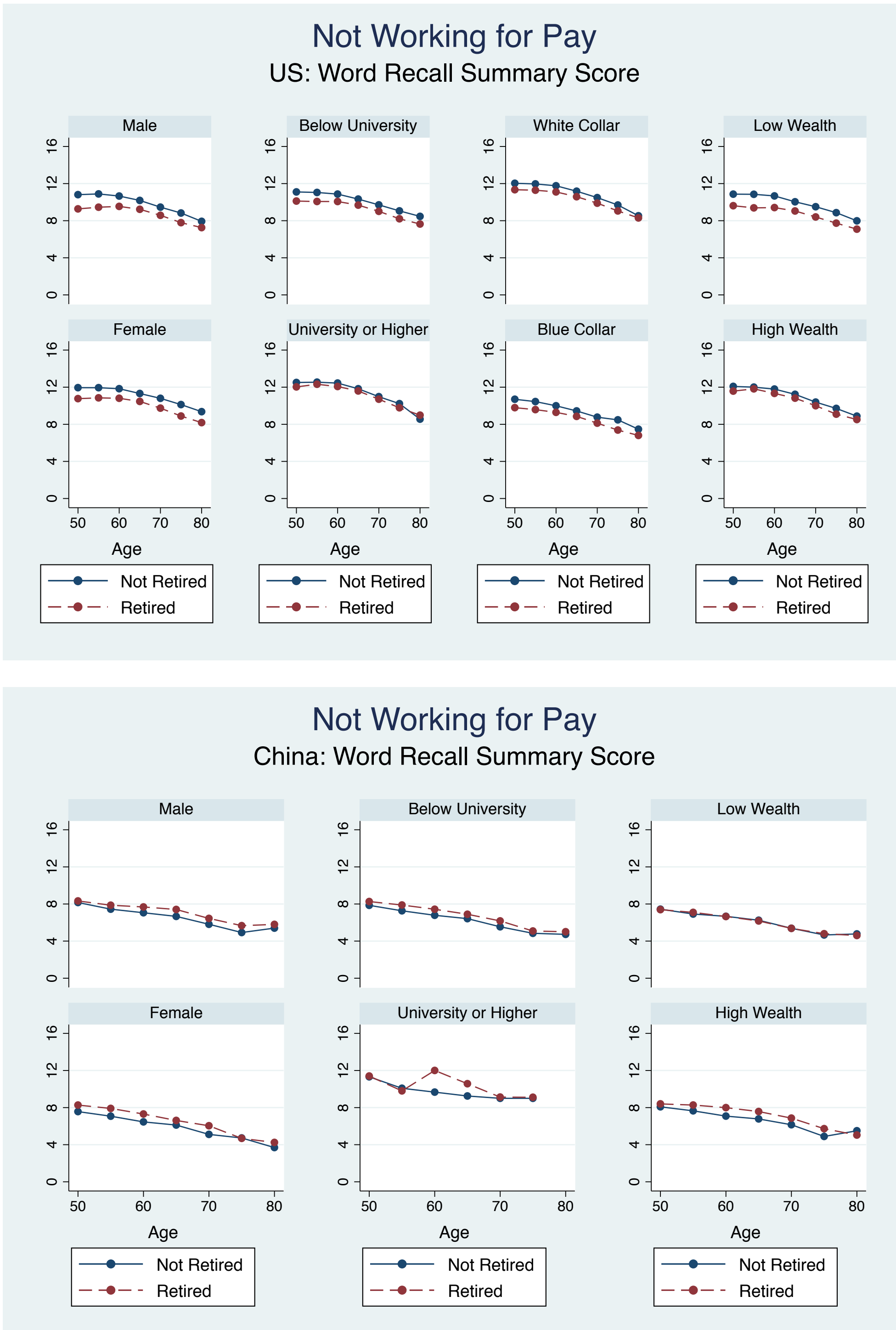


Figure 10: Word Recall Summary Scores in the U.S. and China (All Waves) by Gender, Education, Occupational Type and Wealth

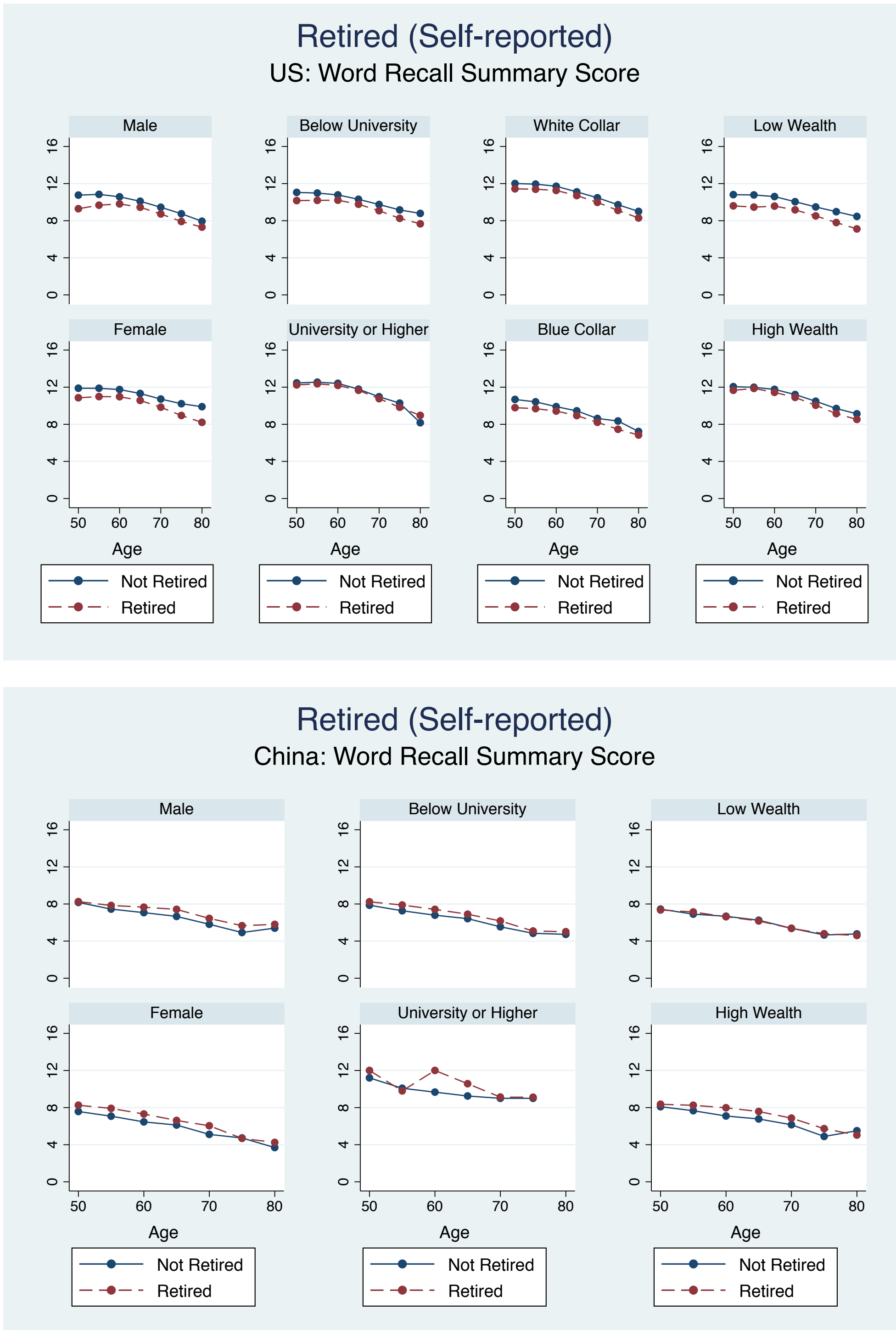
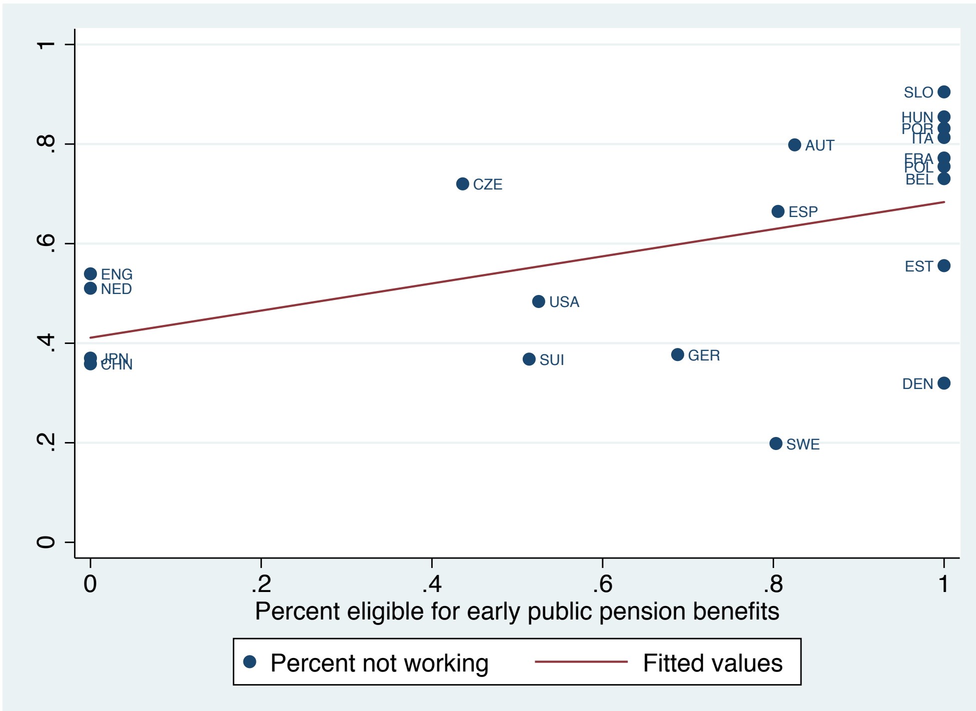


Figure 11: Replication of Rohwedder & Willis 2010 (Early Retirement)



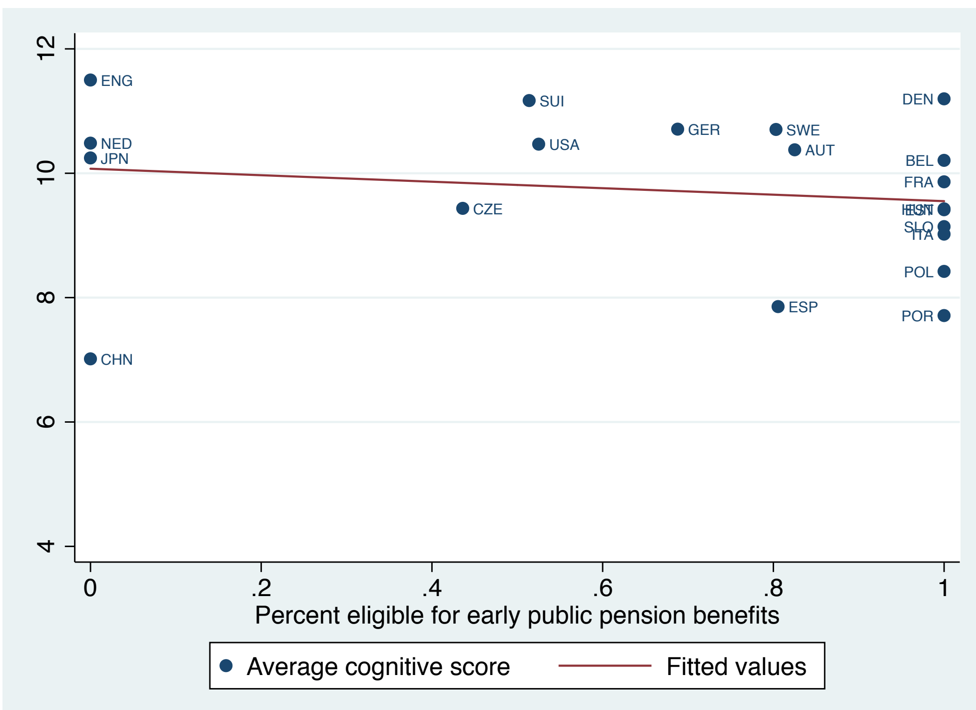


Figure 12: Replication of Rohwedder & Willis 2010 (Full Retirement)

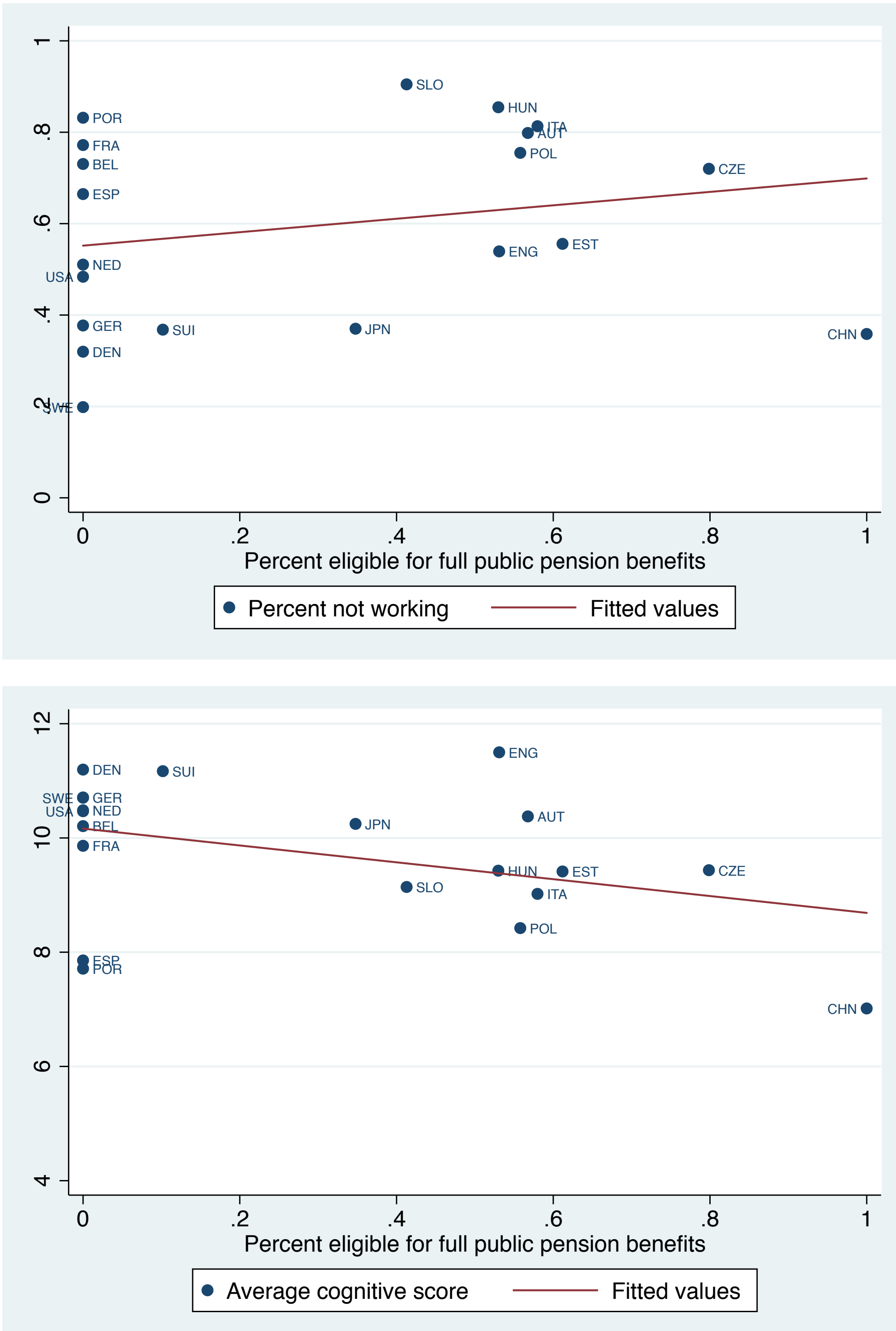


Figure 13: Influence of Different Preferences for Fundamental Cognitive Ability (With or Without)

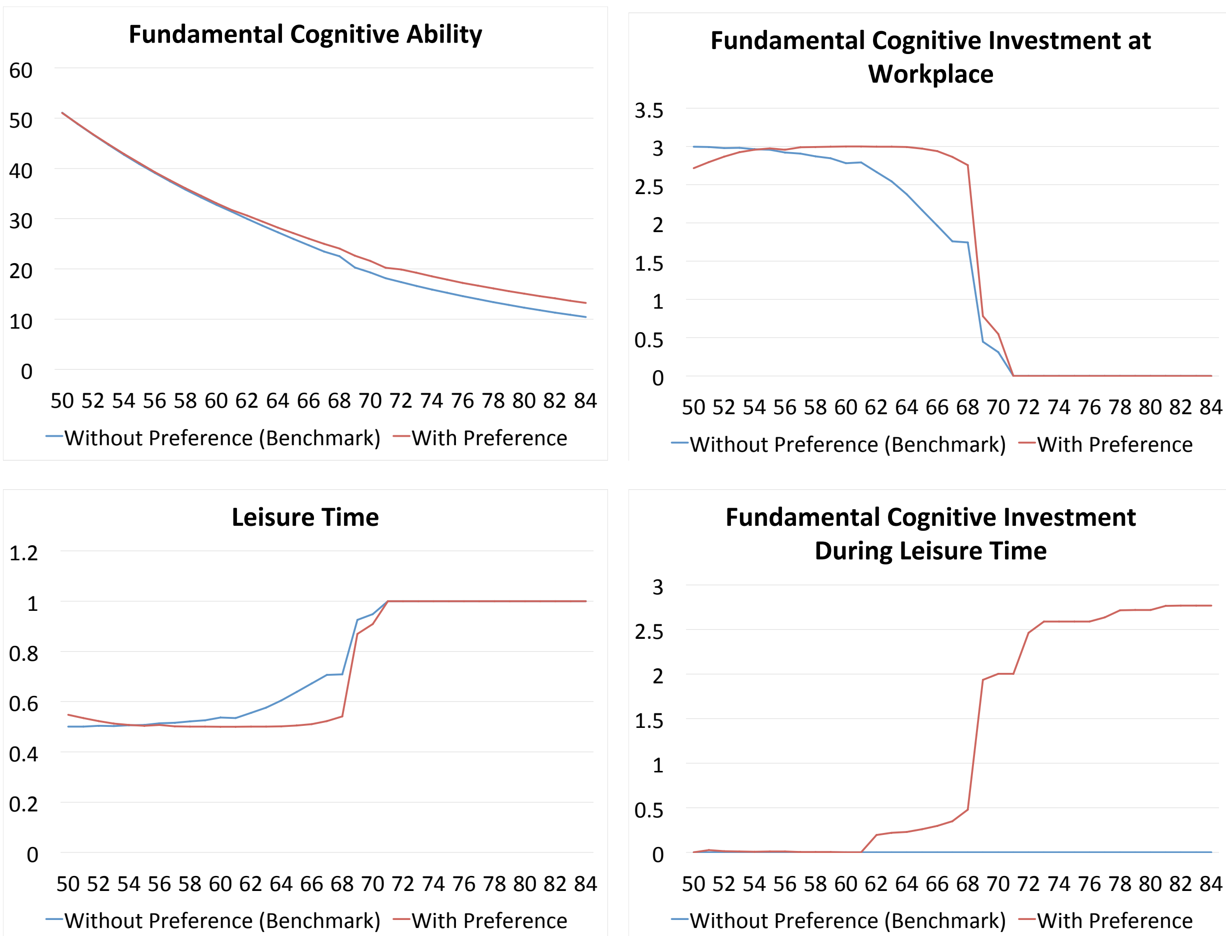


Figure 14: Influence of Different Initial Cognitive Ability Level (High or Low)

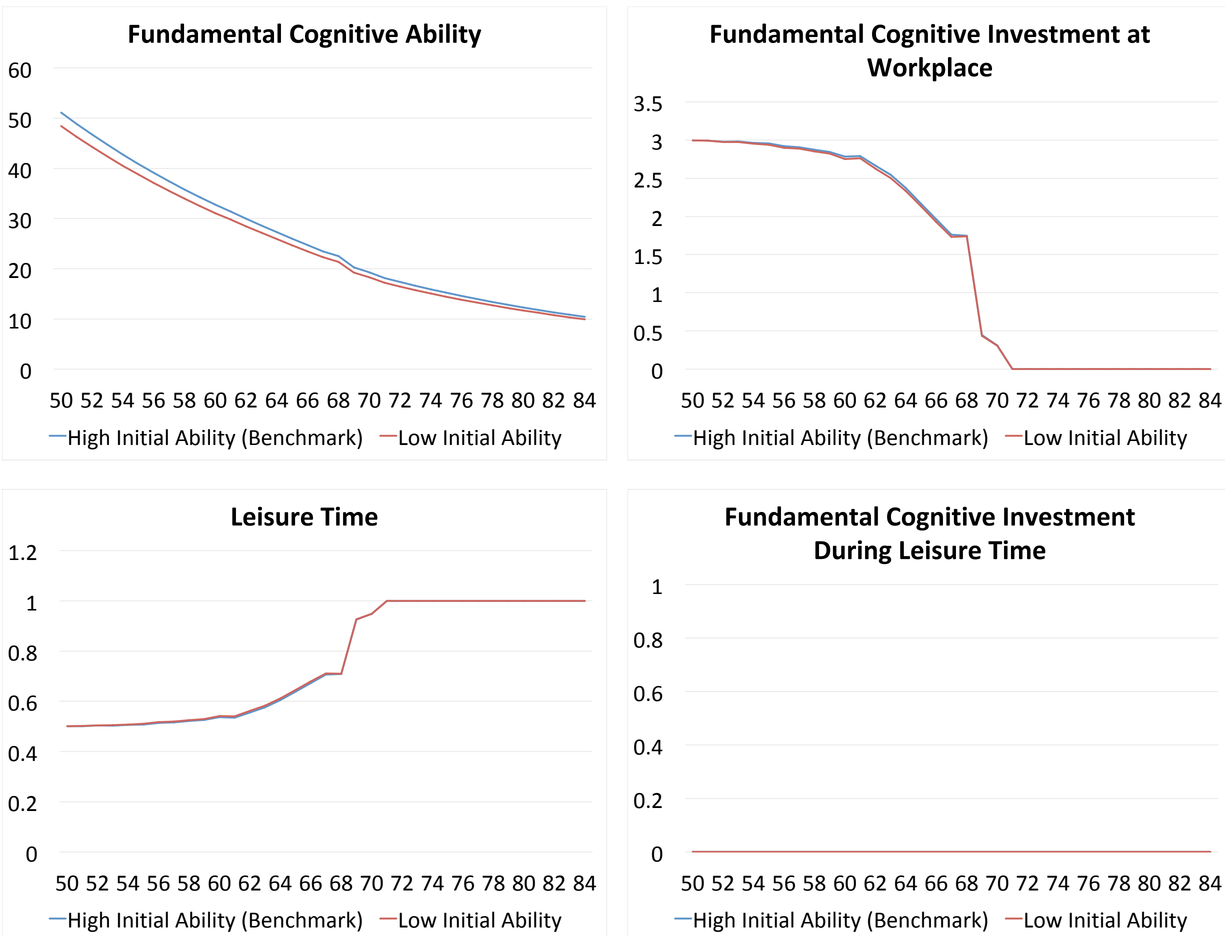


Figure 15: Influence of Different Technology in Cognitive Ability Production Function (High or Low)

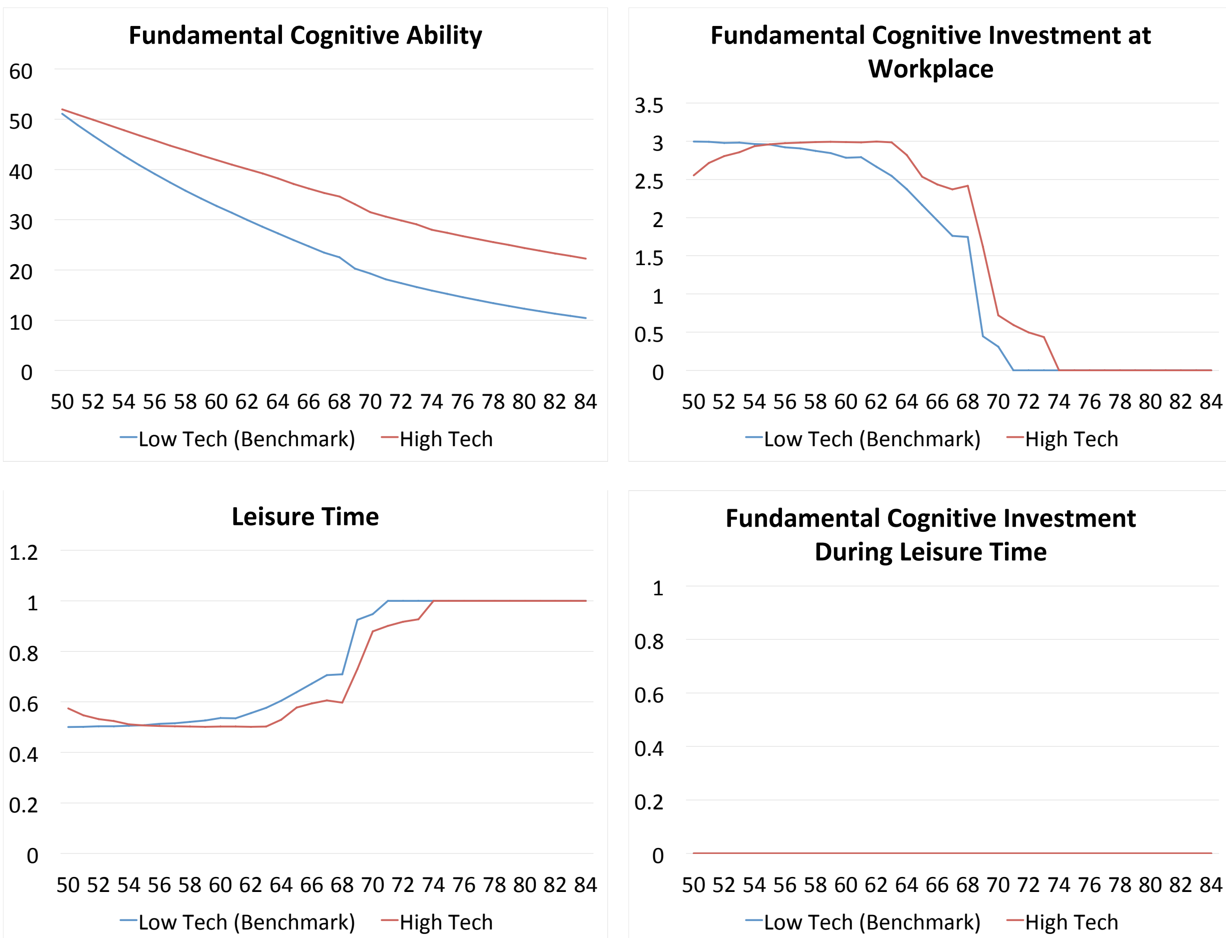


Figure 16: Influence of the Pension Eligibility Age (70 or 65)

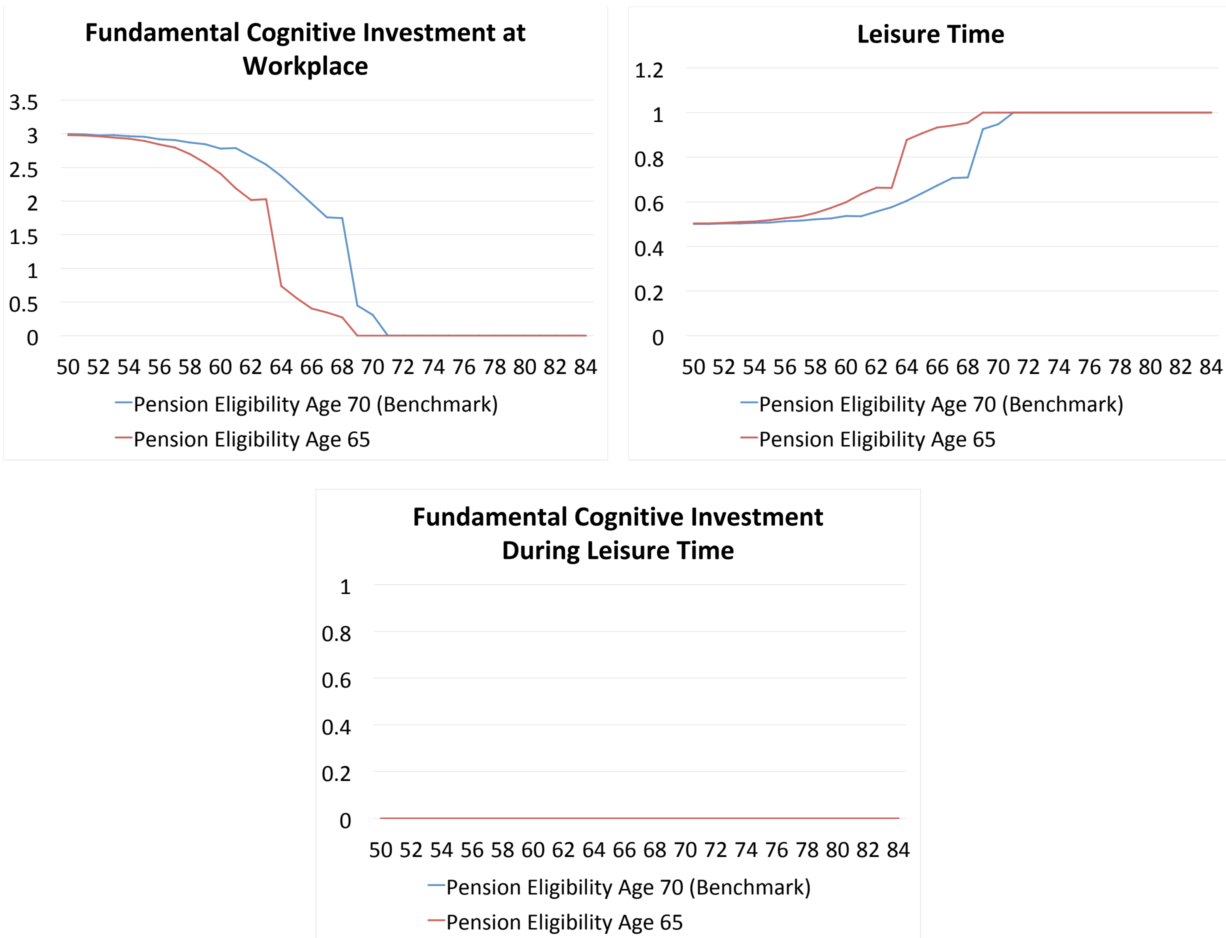


Figure 18: Proportion of Retired Elderly By Age and Country

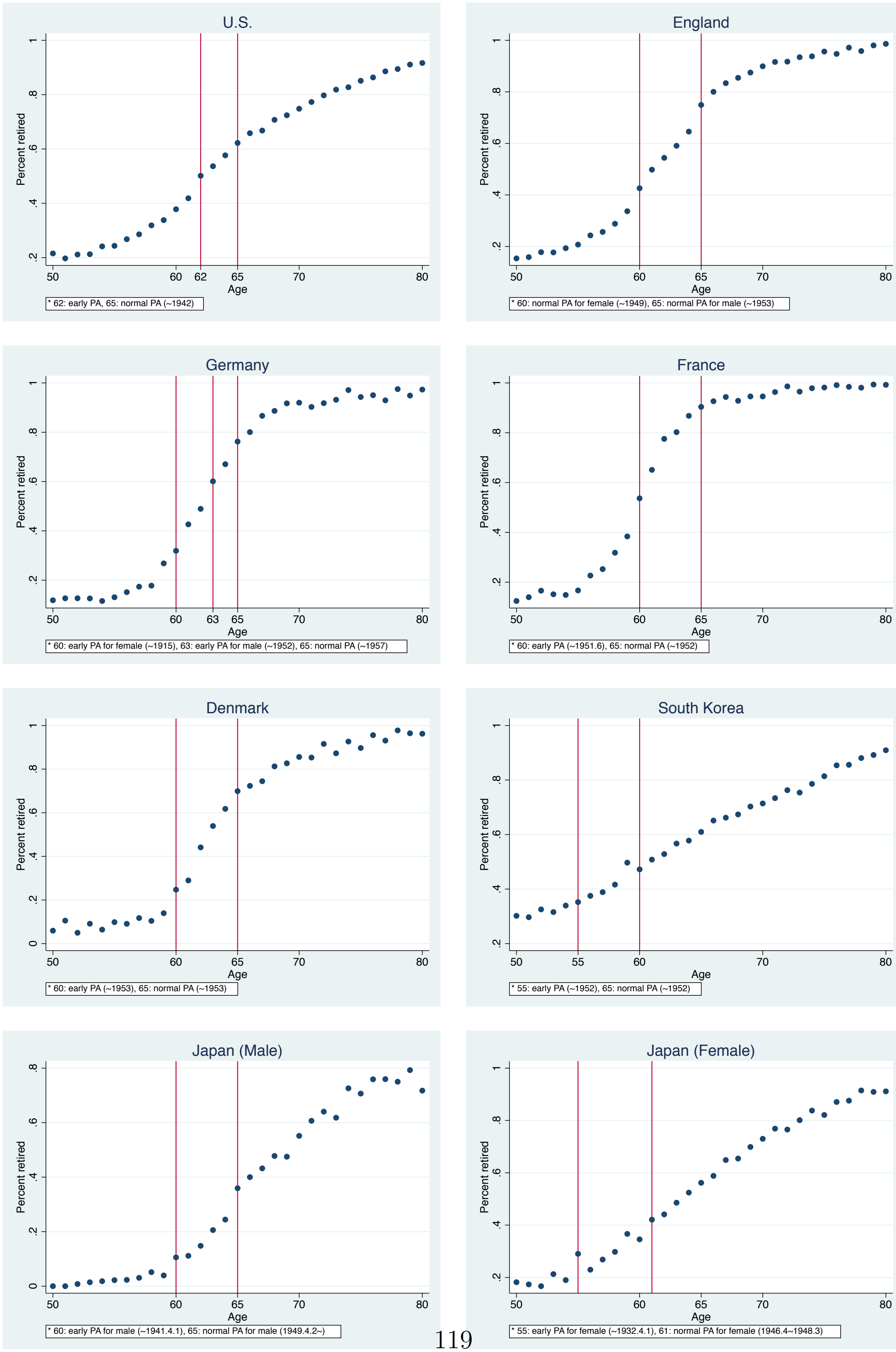
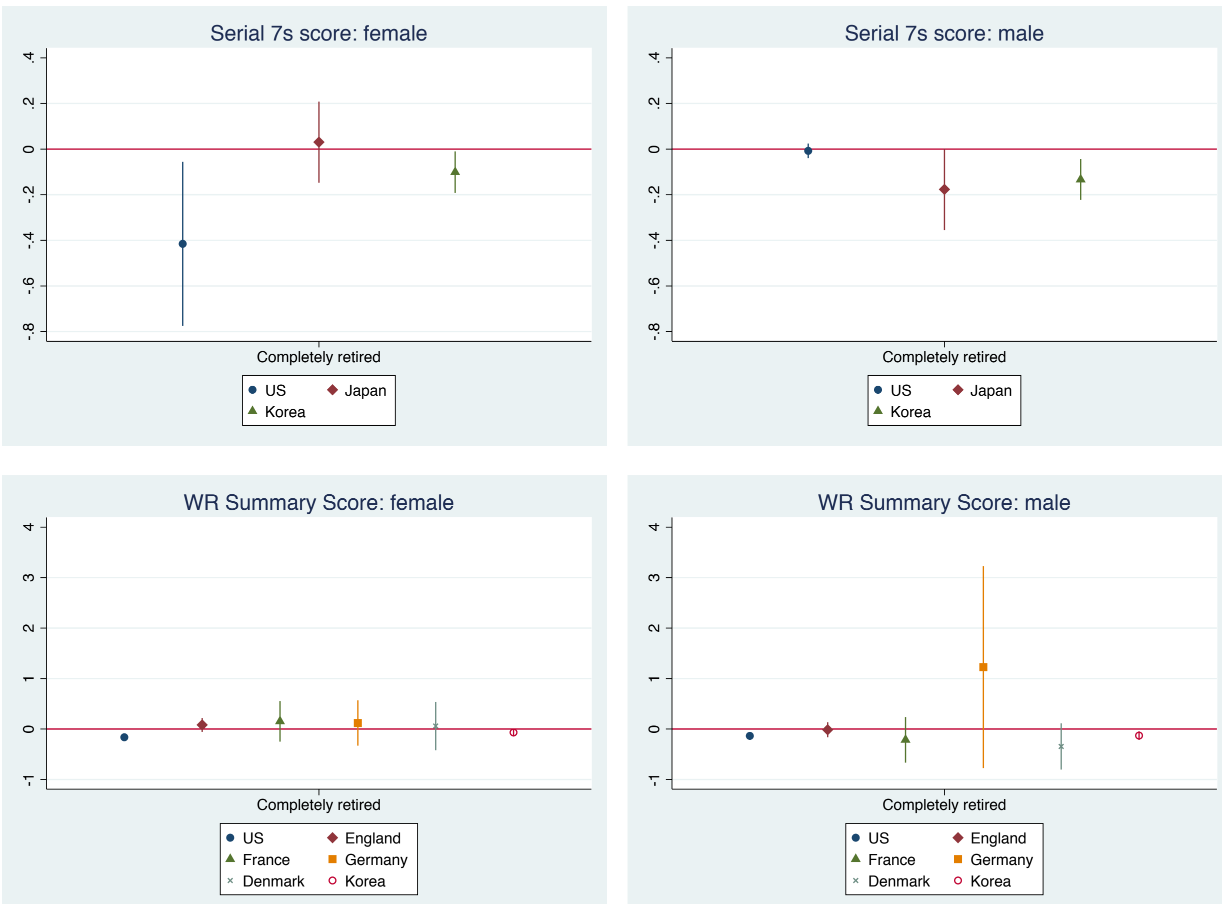


Figure 19: Effect of Retirement on Cognitive Score in Other Countries



[1] Plots indicate the estimated coefficient and bars for the 95% level confidence intervals.

Figure 20: Coefficients of the Fixed Effects (FE) Model by Country

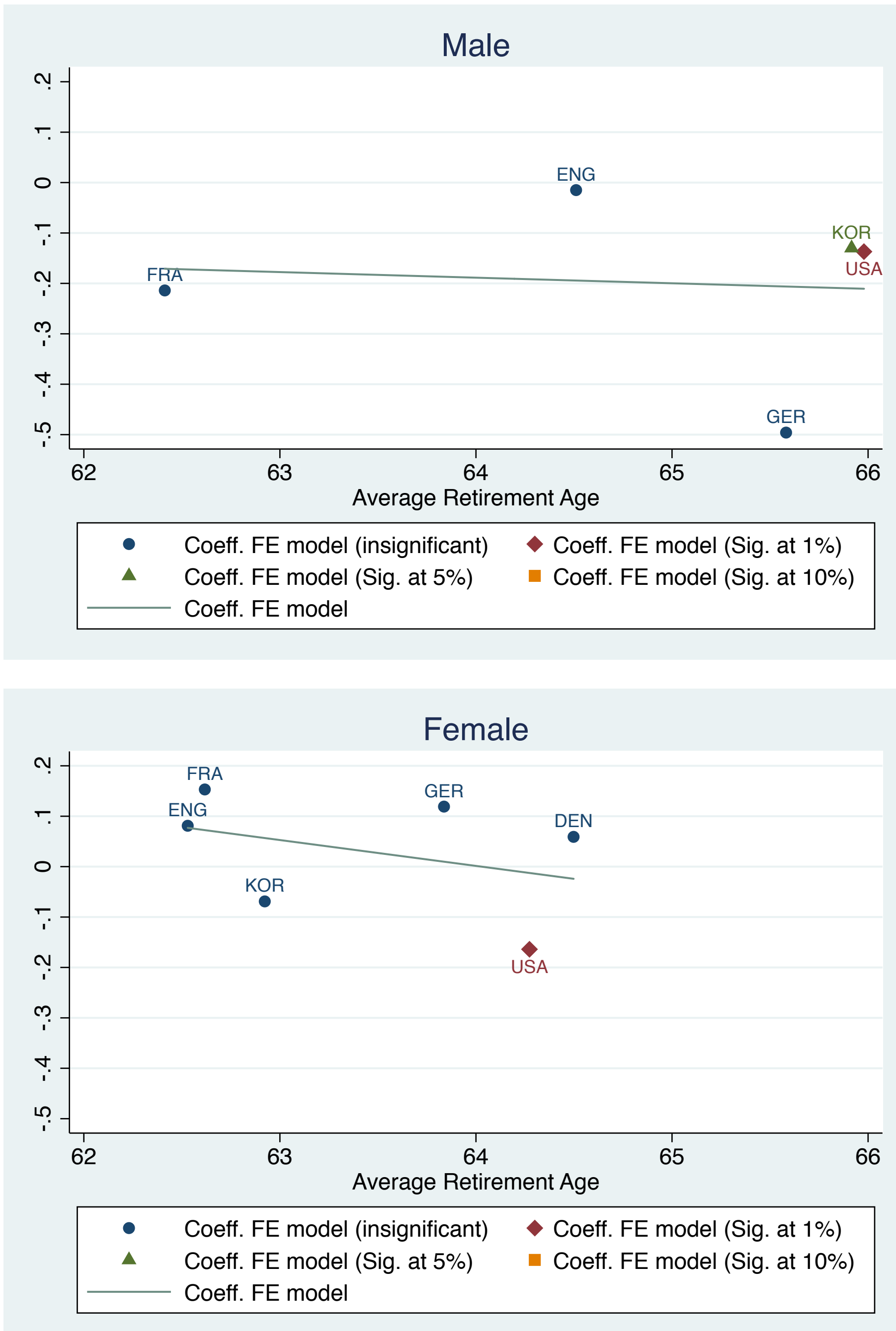
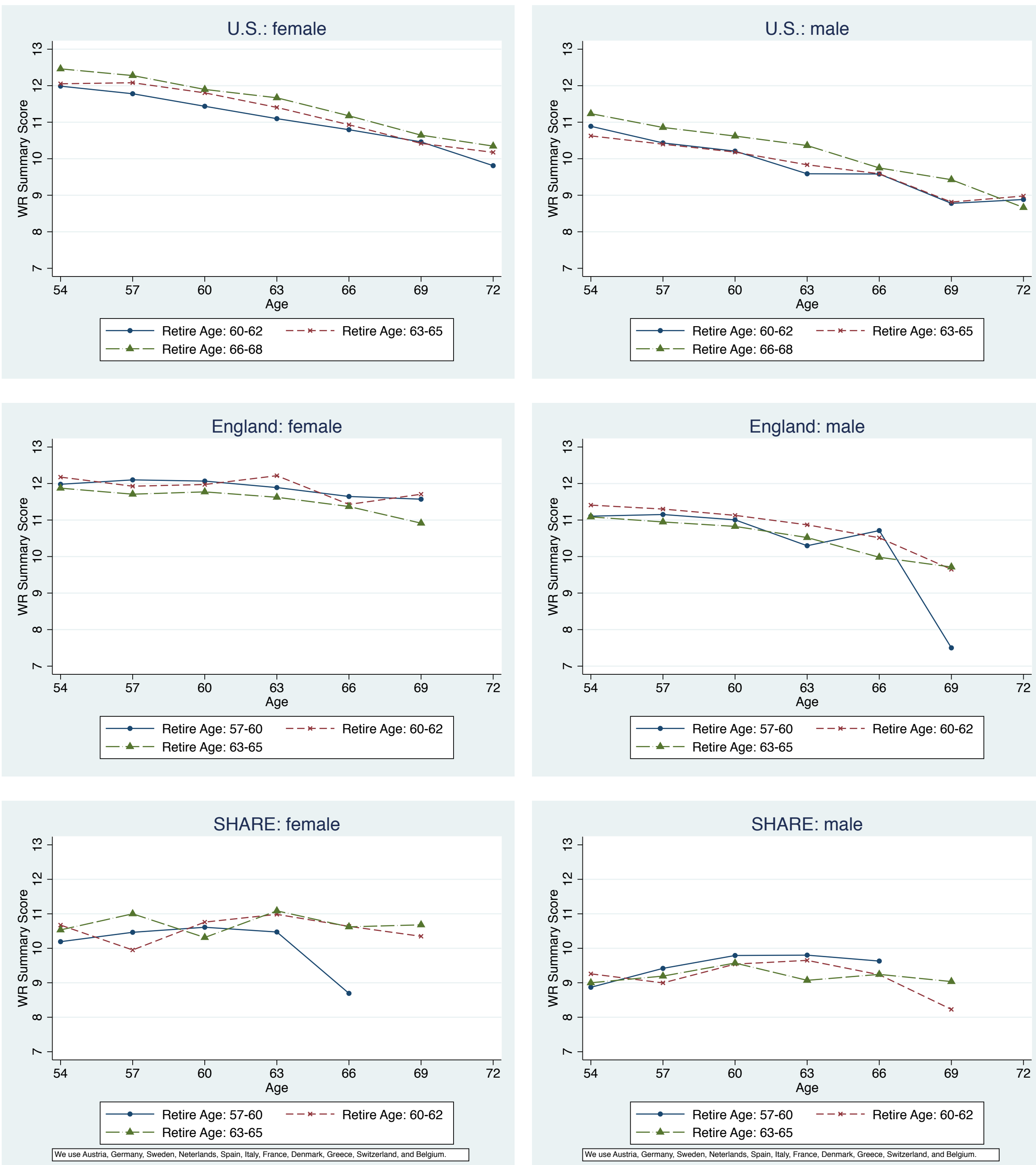
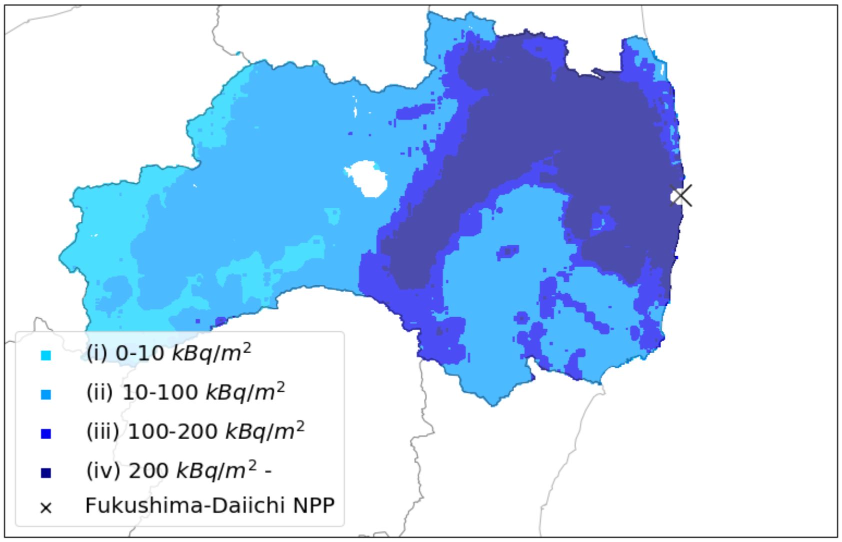
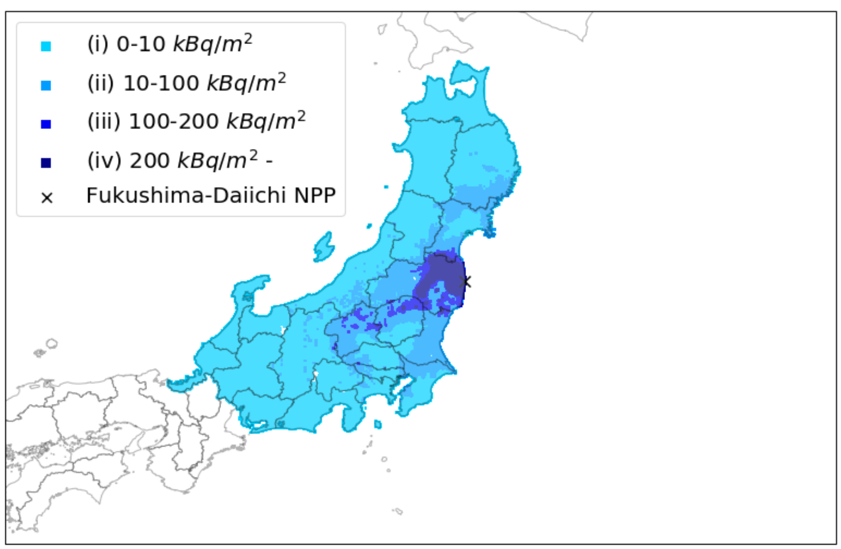


Figure 21: Cognitive Score and Age by Retirement Age





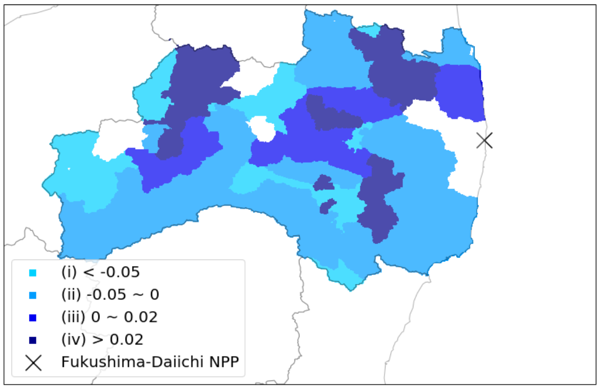
(g) Fukushima



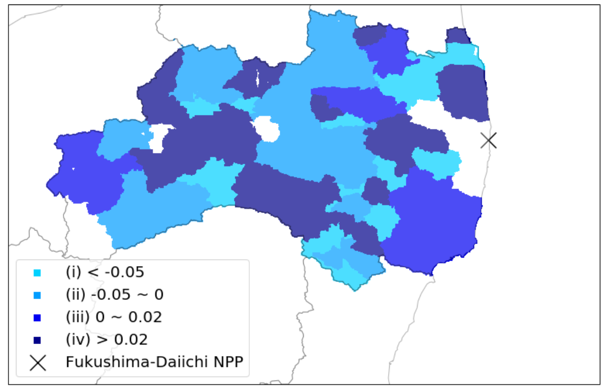
(h) All areas coverd in the dataset

Figure 22: The Distribution of Level of Soil Contamination by Radioactive Cesium

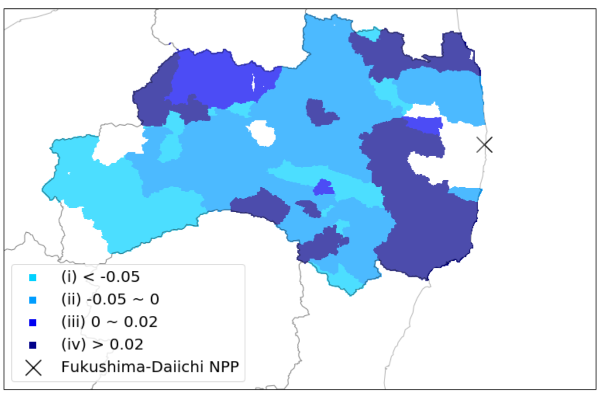
Figure 23: The Relationship between Radioactive Contamination and Low Birth Weight



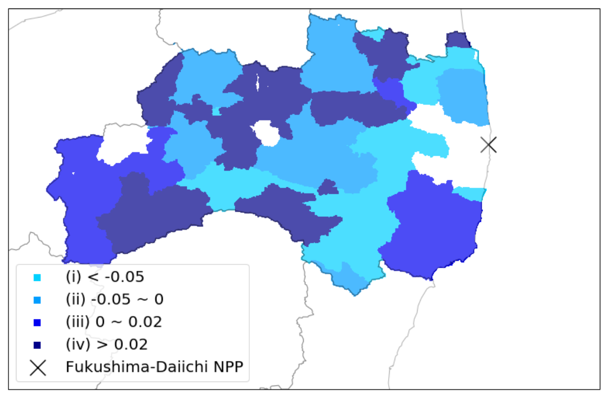
(a) 3rd trimester, Boy



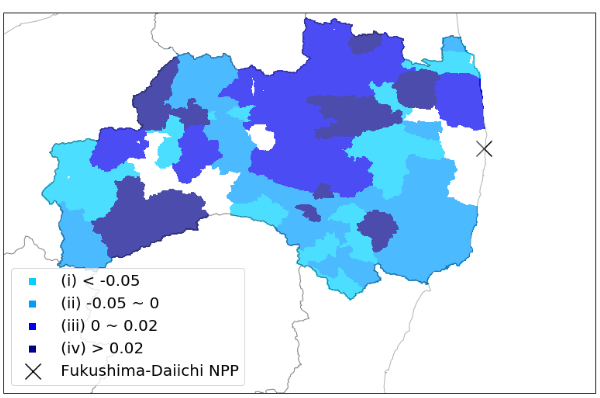
(b) 3rd trimester, Girl



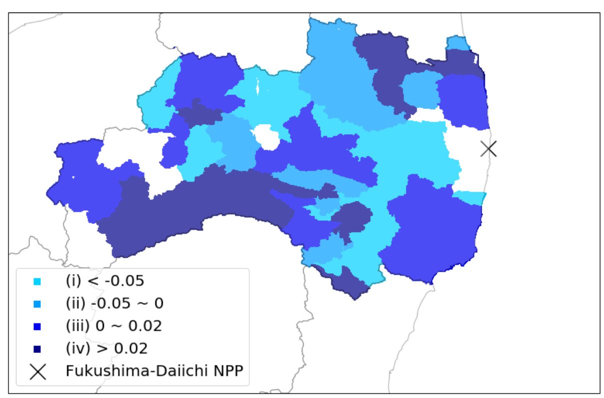
(c) 2nd trimester, Boy



(d) 2nd trimester, Girl

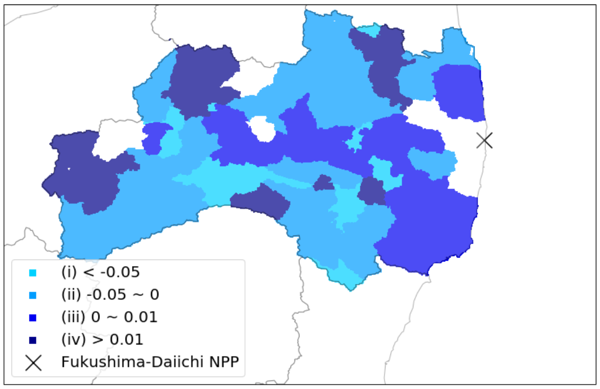


(e) 1st trimester, Boy

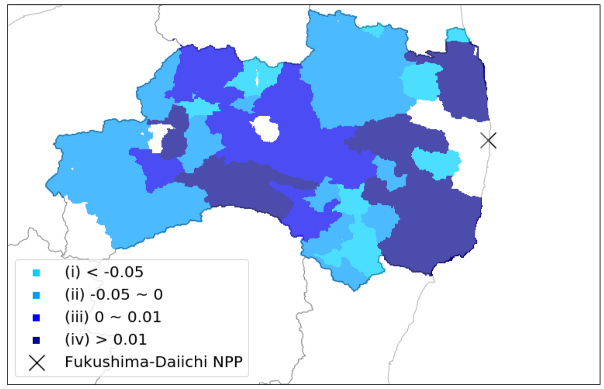


(f) 1st trimester, Girl

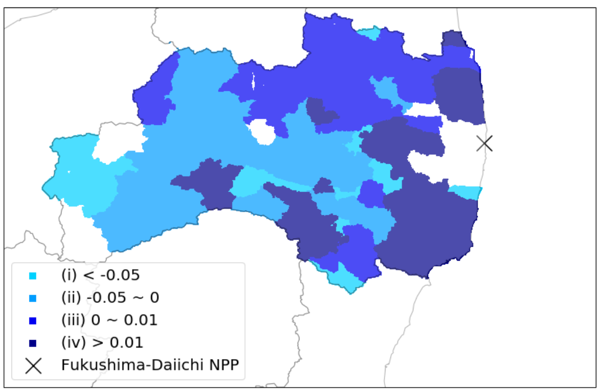
Figure 24: The Relationship between Radioactive Contamination and Premature Birth



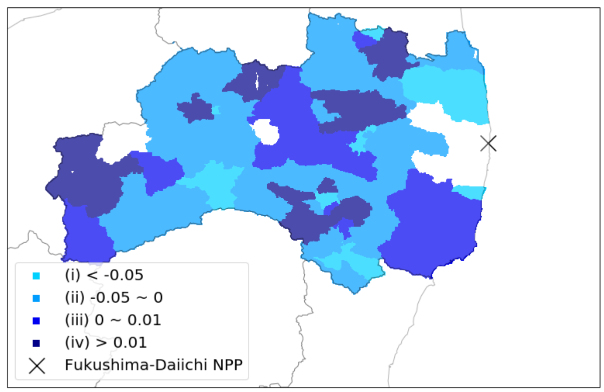
(a) 3rd trimester, Boy



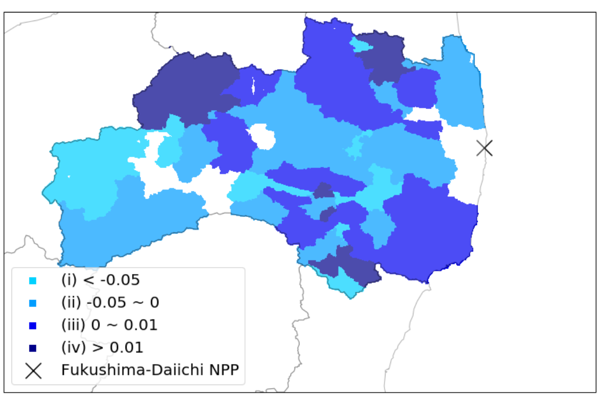
(b) 3rd trimester, Girl



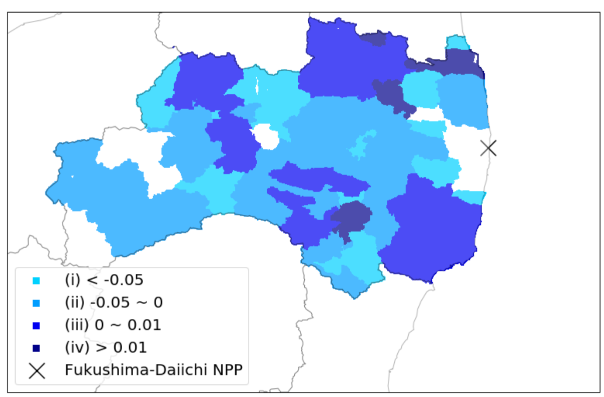
(c) 2nd trimester, Boy



(d) 2nd trimester, Girl



(e) 1st trimester, Boy



(f) 1st trimester, Girl

A Appendix of Chapter 3

A.1 Age of Pension Eligibility

This section explains our construction of the data for the age of pension eligibility. In our replication of Rohwedder and Willis (2010), we expanded the number of countries, updated the data from 2004 in the original to 2010 (the most recent available), and attempted to obtain data directly from each source. Rohwedder and Willis (2010) obtain their data on the age of pension eligibility from two sources[23]: the OECD *Pensions at a Glance* and the US Social Security Administration's *Social Security Programs throughout the World: Europe, 2004*. For our analysis, we obtained information from the Bureau of Labor Statistics in each country, either through their official website or through direct contact with the country's Bureau of Labor Statistics or Bureau of Statistics. If data were not available through these methods, we obtained the most recent information for each country from the OECD *Pensions at a Glance*, International Social Security Association's *Social Security Programs Throughout The World (Europe, Asia and the Pacific, and the Americas)* and the EU *Mutual Information System in Social Protection.* Despite our efforts to obtain data for as many countries as possible, detailed information about the age of pension eligibility was not available for many countries, and so these are excluded from our analysis. Correct and detailed pension eligibility age data was obtained for the U.S., England, Germany, France, Denmark, Switzerland, Czech Republic, Estonia, Japan, China, and Korea, and this data is summarized in the tables below, indicating the section of the paper in which it is used.

[23]See the online Appendix by Rohwedder and Willis (2010) for details.

A.2 Results of the Preliminary Analysis: Cross-Sectional Cross-Country Analysis

In this section, we describe the results of sensitivity analyses of previous studies that are not discussed in section 4. Here, we check the sensitivity of the estimated results of previous studies on the following points:

- Correcting the instrumental variables used in previous studies;
- Including control variables into the analysis of the previous studies;
- Using estimation weights which the previous studies did not use.

First, we restricted the sample to those aged 60-64, following Rohwedder and Willis (2010) and Bingley and Martinello (2013) and then we examined the effect of including other control variables and changing the instrumental variable (Table 29). We estimated the results using ordinary least squares (OLS) when the Durbin-Wu-Hausman (DWH) test was not rejected in the specification using IV, and when the DWH test was rejected, we supported the result of the specification using IV. The IV1 columns represent the results of our analysis when we used the same IV as Rohwedder and Willis (2010) and Bingley and Martinello (2013). The IV2 columns show the results when we substituted our IV, which we have confirmed to be correct.

Columns 1 and 3 in Table 29 present the results of the Rohwedder and Willis (2010) and Bingley and Martinello (2013) specifications [24] , and column 2 shows the results of the Bingley and Martinello (2013) specification when only the variable indicating university enrolment was changed. In the results reported in columns 2 and 3, we also verified the effect of the different definition of education level on the estimated coefficients.

[24] Bingley and Martinello (2013) impute the value of the years of schooling in the ELSA, but as we do not impute this value, the ELSA sample is omitted from our estimates reported in column 3.

Columns 4-7 report our estimates when we added basic individual characteristics variables not included in Rohwedder and Willis (2010) and Bingley and Martinello (2013). Column 4 controls for age effect, column 5 adds country dummies into column 2, and columns 6 and 7 add the other individual characteristics control variables into column 5. The results of our three sensitivity analyses listed above and shown in Table 29 can be summarized as follows:

- Correcting the IV: Changing the instrumental variable has an increasingly large effect when control variables for individual characteristics are also added. While there is only a small difference in the estimated effect of retirement on cognitive function between IV1 (Rohwedder and Willis (2010)) and IV2, which corrects the IV but does not include control variables, columns 6 and 7, which also include control variable for individual characteristics, show a large difference in the effect of retirement on cognitive function between specifications IV1 and IV2. This indicates that the estimated effect of retirement on cognitive function is influenced by the control variables included.

- Including Other Control Variables: The inclusion of country dummy variables causes a large change in the magnitude of coefficients, with the coefficients of column 2 significantly larger in magnitude than column 5. Additionally, the direction of the coefficients is negative in column 6, and the absolute value of the coefficient for the OLS result in column 7 is very small (-0.455) compared to the coefficient for IV2 in column 2 (IV2= -6.538). This shows that the omitted variable bias is significant in column 2. In sum, the results reported in Table 29 thus suggest that one's country is a significant contributor to the observed heterogeneity of the effect of retirement on cognitive function. For example, in column 7, the coefficients for Spain (OLS: -2.230) and Italy (OLS: -1.243) are negative, while the coefficient of the U.S. (OLS: 2.082) is positive.

- Including Estimation Weight: From Table 31, we can see that the effect of using estimation weights (see Appendix (A.4 for an explanation of our calculation methodology) is not insignificant. By including estimation weights, the influence of the U.S. and the U.K. increased because of their relatively large population size among the analyzed countries, and this increased the magnitude of the estimated negative effect of retirement on cognitive function.

Next, we discuss the weighted estimation results reported in Table 31 and comment on the difference in the definition of education level. Column 3 of Table 31 shows our estimated coefficiants using the same specification as Bingley and Martinello (2013)[25], and we can see that the effect of omitting the ELSA and weighting estimation is significant, for we obtained coefficients of -5.011 (IV1) and -5.138 (IV2), compared to -3.014 reported in Bingley and Martinello (2013). Recall also that while we omitted ELSA data from our analysis because data on education was not available.[26], Bingley and Martinello (2013) included ELSA by imputing years of schooling, and thus it seems that the difference in included countries may also be important.

Next, we report the results of our estimates using the most recent data available, and find that the estimated coefficients for 2010 and 2004 are almost the same, indicating that the effect of retirement on cognitive function remained strong in 2010, as it was in 2004. Table 33 shows the estimated results for the two different cohorts (people aged 60-69 in 2004 and again in 2010) using the same column 7 as reported in Table 31. Greece is omitted from the 2010 analysis because the country was not included in the 2010 SHARE data, but otherwise, the set of countries analyzed remained the same for the two cohorts. From Table 33, we can see that the effect of changing the year cohort is weak. The DWH tests in both columns “2004” and “2010” are rejected (IV2), but the OLS results are almost the same (2004: -0.468; 2010: -0.694) after controlling for the heterogeneity of the analyzed countries, and the coefficients

[25]See the specification “All” in Table 3 of Bingley and Martinello (2013).

[26]In our harmonized data set of analyzed countries which in Tables 29 and 31 we call the “original” set, we included a dummy variable indicating people with education above college degree, using a code provided by the Gateway to Global Aging Data (http://gateway.usc.edu), a project of the USC Center for Economic and Social Research (CESR) and funded by the US National Institute on Aging.

of the other control variables for the 2004 and 2010 OLS estimates are also similar. We can thus conclude that the effect of retirement on cognitive function was strong both in 2004 and in 2010.

Finally, Table 34 shows the effect of changing the surveyed age-group and the definition of retirement. As the results are similar to those of column 7 reported in Table 31, we find that the effects of these changes are not significant. The estimates in the “not working” columns are for retirement defined as “not working for pay”, “SR retirement” columns for retirement defined as “respondent reports a retired status” (the same definition as “self-reported retiree” as described in footnote 13 of section 3), and “complete retirement” for retirement defined as both “not working for pay” and “respondent reports a retired status.”

A.3 Parameterization: Model of Retirement and Cognitive Function Decline

In this section, we provide a detailed explanation of our parameterization of model (4). We have explained the utility function, pension payments, and cognitive production functions in section 5. The cost function of cognitive investment, the function of reduced time by cognitive investment, and the function of income are parameterized as follows:

- $G(i^f_{Wt}, i^j_{Wt}, i^j_{Lt}, i^j_{Lt}) = \beta_{Wf} i^f_{Wt} + \beta_{Wj} i^j_{Wt} + \beta_{Lf} i^f_{Lt} + \beta_{Lj} i^j_{Lt}$

- $L(i^f_{Lt}, i^j_{Lt}) = \alpha_f i^f_{Lt} + \alpha_j i^j_{Lt}$

- $y(a_{ft}, a_{jt}, t, l_t) = Y \cdot (a^{\eta_1}_{ft} a^{\eta_2}_{jt})(T - t)^{\eta_3}(1 - l_t)$

A.4 Weight

This section explains the procedure by which we calculated the estimation weights in section 4:

- First, we created the cells considering individual characteristics: age × gender × country of residence. The total number of cells was (The Number of Ages from 60 to 64, or 5) × (The Number of Genders, Male or Female, or 2) × (The Number of Countries of Residence).

- Next, in each cell, we calculated the population based on data from UN *World Information* data. [27] Using this procedure, all respondents were able to be assigned to a cell number.

- Finally, we constructed the estimation weight for each respondent i with characteristic k as follows, where B is the set of characteristics and T_k is the number of respondents in the (merged) dataset (for cross-country analysis) assigned to characteristic k.

[27] http://data.un.org/.

$$W_{ik} = \frac{1}{T_k} \frac{\Pr(\text{Cell Number} = k)}{\sum_{l \in B} \Pr(\text{Cell Number} = l)} \tag{12}$$

A.5 Results from Other Countries

In this section, we show the results from our analysis of other countries, estimated in various groups: Social attendance+ (the respondent increases social activity after retirement) and Social attendance- (not Social attendance+), PNR:Yes (the respondent has a spouse at the first response) and PNR:No (not PNR:Yes), BMI ≥ 25 (the BMI of the respondents is more than 25 at the first response) and BMI < 25 (not BMI ≥ 25), Fat+ (the amount of fat intake is more than the median at the first response) and Fat- (not Fat+).

Table 19: Pensionable Age in Section 4

	2004								2010			
	R & W (2010)				MNO(2015)				MNO(2015)			
	Early		Full		Early		Full		Early		Full	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
R & W original countries												
US	62	62	65	65	62	62	65	65	62	62	66	66
UK	65	60	65	60	*1	*1	65	60	*1	*1	65	60
Austria	61	56	65	60	61+6m	56+6m	65	60	62	57+6m	65	60
Germany	60	60	65	65	60	60	65	65	63	60	65	65
Sweden	61	61	65	65	61	61	65	65	61	61	65	65
Netherlands	60	60	65	65	*1	*1	65	65	*1	*1	65	65
Spain	60	60	65	65	61	61	65	65	61	61	65	65
Italy	57	57	65	60	57	57	65	60	59	57	65	60
France	60	60	60	60	60	60	65	65	60	60	65	65
Denmark	60	60	65	65	60	60	65	65	60	60	65+6m	65+6m
Greece	60	55	62	57	55	55	65	60	60	60	65	65
Switzerland	63	62	65	63	63	61	65	63	63	62	65	64
Belgium	60	60	65	63	60	60	65	63	60	60	65	65
Other Western countries												
Czechia									60	*1	62+2m	*2
Poland									60	55	65	60
Ireland									*1	*1	65	65
Hungary									60	60	62	62
Portugal									55	55	65	65
Slovenia									58	58	61	63
Estonia									60	58	63	61
Luxemberg									60	60	65	65
East Asian countries												
Japan									60	60	64	62
China									*3	*3	60 *3	55 *3

*1: No early retirement.

*2: Different among the number of children. 61(No child), 59y8m(1 child) 58+4m(2 children) , 57(3 or 4 children) , 55+8m(more than 5 children)

Table 20: Pension eligibility age in Section 5

Table 21: PEA: US

Birth cohort	PEA
Early PEA	
	62y0m
Normal PEA	
~ 1937.12	65y0m
1938.1 ~ 1938.12	65y2m
1939.1 ~ 1939.12	65y4m
1940.1 ~ 1940.12	65y6m
1941.1 ~ 1941.12	65y8m
1942.1 ~ 1942.12	65y10m
1943.1 ~ 1943.12	66y0m
1944.1 ~ 1944.12	66y0m
1945.1 ~ 1945.12	66y0m
1946.1 ~ 1946.12	66y0m
1947.1 ~ 1947.12	66y0m
1948.1 ~ 1948.12	66y0m
1949.1 ~ 1949.12	66y0m
1950.1 ~ 1950.12	66y0m
1951.1 ~ 1951.12	66y0m
1952.1 ~ 1952.12	66y0m
1953.1 ~ 1953.12	66y0m
1954.1 ~ 1954.12	66y0m
1955.1 ~ 1955.12	66y2m
1956.1 ~ 1956.12	66y4m
1957.1 ~ 1957.12	66y6m
1958.1 ~ 1958.12	66y8m
1959.1 ~ 1959.12	66y10m
1960.1 ~ 1960.12	67y0m

Table 22: PEA: UK

Birth cohort	PEA
Normal PEA: Male	
~ 1953.12	65y0m
1954.1 ~ 1954.12	66y0m
1955.1 ~ 1959.12	66y0m
1960.1 ~ 1960.12	67y0m
1961.1 ~	67y0m
Normal PEA: Female	
~ 1949.12	60y0m
1950.1 ~ 1950.12	61y0m
1951.1 ~ 1951.12	62y0m
1952.1 ~ 1952.12	63y0m
1953.1 ~	65y0m

Table 23: PEA: Germany

Birth cohort	PEA
Early PEA: Male	
~ 1952.12	63y0m
1953.1 ~ 1953.12	63y2m
1954.1 ~ 1954.12	63y4m
1955.1 ~ 1955.12	63y6m
1956.1 ~ 1956.12	63y8m
1957.1 ~ 1957.12	63y10m
1958.1 ~ 1958.12	64y0m
1959.1 ~ 1959.12	64y2m
1960.1 ~ 1960.12	64y4m
1961.1 ~ 1961.12	64y6m
1962.1 ~ 1962.12	64y8m
1963.1 ~ 1963.12	64y10m
1964.1 ~ 1964.12	65y0m
Early PEA: Female	
~ 1951.12	60y0m
Normal PEA	
~ 1946.12	65y0m
1947.1 ~ 1947.12	65y1m
1948.1 ~ 1948.12	65y2m
1949.1 ~ 1949.12	65y3m
1950.1 ~ 1950.12	65y4m
1951.1 ~ 1951.12	65y5m
1952.1 ~ 1952.12	65y6m
1953.1 ~ 1953.12	65y7m
1954.1 ~ 1954.12	65y8m
1955.1 ~ 1955.12	65y9m
1956.1 ~ 1956.12	65y10m
1957.1 ~ 1957.12	65y11m
1958.1 ~ 1958.12	66y0m
1959.1 ~ 1959.12	66y2m
1960.1 ~ 1960.12	66y4m
1961.1 ~ 1961.12	66y6m
1962.1 ~ 1962.12	66y8m
1963.1 ~ 1963.12	66y10m
1964.1 ~ 1964.12	67y0m

Table 24: PEA: France

Birth cohort	PEA
Early PEA	
~ 1951.6	60y0m
1951.7 ~ 1951.12	60y4m
1952.1 ~ 1952.12	60y9m
1953.1 ~ 1953.12	61y2m
1954.1 ~ 1954.12	61y7m
1955.1 ~ 1955.12	62y0m
1956.1 ~ .	62y0m
Normal PEA	
~ 1951.6	65y0m
1951.7 ~ 1951.12	65y4m
1952.1 ~ 1952.12	65y9m
1953.1 ~ 1953.12	66y2m
1954.1 ~ 1954.12	66y7m
1955.1 ~ 1955.12	67y0m
1956.1 ~ .	67y0m

Table 25: Pension eligibility age in Section 5

Table 26: PEA: Denmark

Birth cohort	PEA
Early PEA	
~ 1953.12	60y0m
1954.1 ~ 1954.6	60y6m
1954.7 ~ 1954.12	61y0m
1955.1 ~ 1955.6	61y6m
1955.7 ~ 1955.12	62y0m
1956.1 ~ 1956.6	62y6m
1956.7 ~ 1958.12	63y0m
1959.1 ~ 1959.6	63y6m
1959.7 ~ 1964.6	64y0m
1964.7 ~	64y0m
Normal PEA	
~ 1953.12	65y0m
1954.1 ~ 1954.6	65y6m
1954.7 ~ 1954.12	66y0m
1955.1 ~ 1955.6	66y6m
1955.7 ~ 1955.12	67y0m
1956.1 ~ 1956.6	67y0m
1956.7 ~ 1958.12	67y0m
1959.1 ~ 1959.6	67y0m
1959.7 ~ 1964.6	67y0m
1964.7 ~	67y0m

Table 27: PEA: Japan

Birth cohort	PEA
Normal PEA: Male	
~1941.4.1	60y0m
1941.4.2~1943.4.1	61y0m
1943.4.2~1945.4.1	62y0m
1945.4.2~1947.4.1	63y0m
1947.4.2~1949.4.1	64y0m
1949.4.2~1953.4.1	65y0m
1953.4.2~1955.4.1	65y0m
1955.4.2~1957.4.1	65y0m
1957.4.2~1959.4.1	65y0m
1959.4.2~1961.4.1	65y0m
1961.4.2~	65y0m
Normal PEA: Female	
~1932.4.1	55y0m
1932.4.2~1934.4.1	56y0m
1934.4.2~1936.4.1	57y0m
1936.4.2~1937.4.1	58y0m
1937.4.2~1938.4.1	58y0m
1938.4.2~1940.4.1	59y0m
1940.4.2~1946.4.1	60y0m
1946.4.2~1948.4.1	61y0m
1948.4.2~1950.4.1	62y0m
1950.4.2~1952.4.1	63y0m
1952.4.2~1954.4.1	64y0m
1954.4.2~1958.4.1	65y0m
1958.4.2~1960.4.1	65y0m
1960.4.2~1962.4.1	65y0m
1962.4.2~1964.4.1	65y0m
1964.4.2~1965.4.1	65y0m
1965.4.2~	65y0m

Table 28: PEA: Korea

Birth cohort	PEA
Early PEA	
~ 1952.12	55y0m
1953.1 ~ 1956.12	56y0m
1957.1 ~ 1960.12	57y0m
1961.1 ~ 1964.12	58y0m
1965.1 ~ 1968.12	59y0m
1969.1 ~ .	60y0m
Normal PEA	
~ 1952.12	60y0m
1953.1 ~ 1956.12	61y0m
1957.1 ~ 1960.12	62y0m
1961.1 ~ 1964.12	63y0m
1965.1 ~ 1968.12	64y0m
1969.1 ~ .	65y0m

Table 29: The effect of instrumental variables and other control variables (without the coefficients of country)

	(1)		(2)		(3)		(4)		(5)			(6)			(7)		
	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
1st Stage Result																	
IV-early	0.183***	0.122***	0.171***	0.109***	0.190***	0.070***	0.161***	0.098***		0.084***	0.088***		0.014	-0.003		0.019	0.001
	(0.013)	(0.010)	(0.014)	(0.011)	(0.016)	(0.013)	(0.015)	(0.011)		(0.015)	(0.016)		(0.017)	(0.019)		(0.017)	(0.018)
IV-normal	0.160***	0.212***	0.136***	0.171***	0.161***	0.170***	0.166***	0.208***		0.041**	0.082***		0.055***	0.068***		0.056***	0.070***
2nd Stage Result																	
Not working for pay	-3.346***	-3.047***	-5.216***	-6.538***	-4.433***	-4.708***	-3.302***	-2.588***	-0.570***	-0.771	-2.483***	-0.527***	-0.717	-3.502*	-0.455***	-0.581	-3.217
	(0.319)	(0.373)	(0.415)	(0.574)	(0.404)	(0.630)	(0.347)	(0.409)	(0.071)	(0.940)	(0.931)	(0.072)	(2.116)	(2.114)	(0.073)	(1.960)	(2.060)
Univ			1.140***	0.948***					1.669***	1.640***	1.397***	1.618***	1.592***	1.217***	1.484***	1.472***	1.222***
			(0.119)	(0.141)					(0.081)	(0.156)	(0.157)	(0.082)	(0.297)	(0.299)	(0.086)	(0.205)	(0.216)
Years of schooling					0.239***	0.232***											
					(0.014)	(0.018)											
Female			1.694***	1.857***	1.569***	1.598***			1.040***	1.067***	1.298***	1.120***	1.146***	1.534***	1.106***	1.123***	1.479***
			(0.096)	(0.117)	(0.093)	(0.109)			(0.066)	(0.141)	(0.142)	(0.067)	(0.300)	(0.303)	(0.067)	(0.272)	(0.287)
Age							4.112	4.107				1.762	1.801	2.379	2.172	2.191	2.587
							(2.729)	(2.660)				(2.401)	(2.442)	(2.661)	(2.406)	(2.421)	(2.611)
Age squared							-3.310	-3.325				-1.497	-1.523	-1.904	-1.827	-1.838	-2.079
							(2.201)	(2.145)				(1.936)	(1.958)	(2.136)	(1.941)	(1.946)	(2.099)
Mariage												0.591***	0.593***	0.620***	0.356***	0.363**	0.510***
												(0.083)	(0.086)	(0.091)	(0.089)	(0.144)	(0.151)
N of children												-0.099***	-0.099***	-0.095***	-0.085***	-0.085***	-0.086***
												(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)
Income															0.129**	-0.010	-0.021
															(0.056)	(0.016)	(0.015)
Own house															0.581***	0.576***	0.475***
															(0.093)	(0.120)	(0.126)
Total wealth															0.006	0.003*	0.002
															(0.006)	(0.002)	(0.002)
Observations	8838	8838	8509	8509	7352	7352	8838	8838	8509	8509	8509	8447	8447	8447	8355	8355	8355
R^2	-0.071	-0.046	-0.260	-0.521	-0.046	-0.085	-0.067	-0.014	0.189	0.189	0.122	0.199	0.198	0.038	0.210	0.209	0.076
DWHchi2	57.548	31.994	144.107	142.550	102.651	46.607	48.327	19.207		0.044	4.607		0.023	2.365		0.010	2.089
DWHpval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.834	0.032		0.879	0.124		0.919	0.148

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$
All economic variables (e.g. Total wealth, Income) are measured in dollars.
In the specification (7), (country dummy) × (economic variable)
(e.g. (Total wealth) × (the U.S. dummy)) variables are also included.
The estimated coefficients of these cross terms are not presented.
The Belgium dummy is omitted.

Table 30: The effect of instrumental variables and other control variables (only the coefficients of country)

	(1)		(2)		(3)		(4)		(5)			(6)			(7)		
	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
Country dummy																	
2.US									1.958***	1.886***	1.274***	2.138***	2.066**	1.011	2.082***	2.035***	1.053
									(0.145)	(0.365)	(0.366)	(0.148)	(0.817)	(0.822)	(0.183)	(0.754)	(0.794)
3.UK									2.015***	1.960***	1.486***	1.992***	1.939***	1.157*	1.658***	1.630***	1.055**
									(0.158)	(0.308)	(0.309)	(0.157)	(0.623)	(0.627)	(0.207)	(0.482)	(0.513)
11.Austria									0.776***	0.787***	0.881***	0.799***	0.808***	0.936***	1.004***	1.008***	1.107***
									(0.232)	(0.237)	(0.242)	(0.232)	(0.250)	(0.260)	(0.324)	(0.332)	(0.348)
12.Germany									0.672***	0.647***	0.433*	0.617***	0.591*	0.212	0.562**	0.548*	0.257
									(0.184)	(0.219)	(0.226)	(0.183)	(0.344)	(0.354)	(0.235)	(0.319)	(0.337)
13.Sweden									1.185***	1.092**	0.303	1.158***	1.070	-0.224	1.177***	1.134	0.228
									(0.184)	(0.469)	(0.470)	(0.184)	(1.002)	(1.007)	(0.306)	(0.740)	(0.772)
14.Netherlands									0.731***	0.718***	0.603***	0.685***	0.671***	0.470*	0.917***	0.908***	0.714**
									(0.195)	(0.205)	(0.210)	(0.194)	(0.249)	(0.261)	(0.290)	(0.323)	(0.334)
15.Spain									-1.694***	-1.728***	-2.018***	-1.692***	-1.726***	-2.225***	-2.230***	-2.250***	-2.669***
									(0.213)	(0.266)	(0.275)	(0.213)	(0.436)	(0.449)	(0.284)	(0.422)	(0.453)
16.Italy									-1.109***	-1.120***	-1.211***	-1.174***	-1.186***	-1.361***	-1.243***	-1.245***	-1.297***
									(0.186)	(0.193)	(0.199)	(0.186)	(0.229)	(0.241)	(0.228)	(0.230)	(0.242)
17.France									-0.122	-0.125	-0.146	-0.112	-0.115	-0.162	0.239	0.235	0.146
									(0.203)	(0.203)	(0.213)	(0.201)	(0.205)	(0.224)	(0.260)	(0.268)	(0.287)
18.Denmark									1.332***	1.284***	0.870***	1.352***	1.306**	0.629	1.296***	1.280***	0.936**
									(0.234)	(0.326)	(0.336)	(0.232)	(0.566)	(0.579)	(0.306)	(0.399)	(0.438)
19.Greece									0.081	0.048	-0.235	0.083	0.052	-0.404	-0.527*	-0.542	-0.851**
									(0.192)	(0.248)	(0.253)	(0.193)	(0.398)	(0.404)	(0.275)	(0.356)	(0.383)
20.Switzerland									1.169***	1.091**	0.427	1.150***	1.076	-0.009	1.621***	1.581**	0.761
									(0.284)	(0.458)	(0.467)	(0.285)	(0.871)	(0.885)	(0.427)	(0.736)	(0.773)
Observations	8838	8838	8509	8509	7352	7352	8838	8838	8509	8509	8509	8447	8447	8447	8355	8355	8355
R^2	-0.071	-0.046	-0.260	-0.521	-0.046	-0.085	-0.067	-0.014	0.189	0.189	0.122	0.199	0.198	0.038	0.210	0.209	0.076
DWHchi2	57.548	31.994	144.107	142.550	102.651	46.607	48.327	19.207		0.044	4.607		0.023	2.365		0.010	2.089
DWHpval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.834	0.032		0.879	0.124		0.919	0.148

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$
All economic variables (e.g. Total wealth, Income) are measured in dollars.
In the specification (7), (country dummy) × (economic variable)
(e.g. (Total wealth) × (the U.S. dummy)) variables are also included.
The estimated coefficients of these cross terms are not presented.
The Belgium dummy is omitted.

Table 31: The effect of instrumental variables and other control variables using weight (without the coefficients of country)

	(1)		(2)		(3)		(4)		(5)			(6)			(7)		
	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
1st Stage Result																	
IV-early	0.210***	0.167***	0.197***	0.147***	0.203***	0.126***	0.185***	0.144***		0.085***	0.081***		0.010	-0.001		0.020	0.004
	(0.016)	(0.013)	(0.016)	(0.013)	(0.018)	(0.015)	(0.017)	(0.014)		(0.017)	(0.017)		(0.020)	(0.022)		(0.020)	(0.021)
IV-normal	0.186***	0.218***	0.160***	0.173***	0.160***	0.135***	0.190***	0.215***		0.052**	0.090***		0.068***	0.075***		0.066***	0.076***
2nd Stage Result																	
Not working for pay	-5.397***	-5.338***	-6.192***	-7.612***	-5.011***	-5.138***	-5.623***	-5.602***	-0.571***	-0.979	-2.903***	-0.532***	-3.826	-5.318**	-0.461***	-2.485	-5.056**
	(0.357)	(0.433)	(0.432)	(0.634)	(0.451)	(0.674)	(0.391)	(0.482)	(0.083)	(0.998)	(1.118)	(0.084)	(2.422)	(2.618)	(0.085)	(2.136)	(2.578)
Univ			1.149***	0.928***					1.724***	1.665***	1.386***	1.656***	1.205***	1.001***	1.538***	1.332***	1.069***
			(0.156)	(0.184)					(0.101)	(0.178)	(0.194)	(0.101)	(0.354)	(0.383)	(0.106)	(0.247)	(0.295)
Years of schooling					0.256***	0.253***											
					(0.016)	(0.020)											
Female			1.675***	1.851***	1.573***	1.587***			0.909***	0.963***	1.219***	0.989***	1.438***	1.641***	0.985***	1.249***	1.583***
			(0.120)	(0.143)	(0.111)	(0.124)			(0.080)	(0.159)	(0.168)	(0.081)	(0.354)	(0.386)	(0.081)	(0.301)	(0.363)
Age							5.954	5.956				6.526**	6.936**	7.122**	7.086**	7.260**	7.481**
							(3.749)	(3.744)				(2.871)	(3.234)	(3.574)	(2.882)	(3.016)	(3.509)
Age squared							-4.746	-4.748				-5.330**	-5.566**	-5.673**	-5.785**	-5.868**	-5.973**
							(3.023)	(3.019)				(2.316)	(2.602)	(2.876)	(2.324)	(2.430)	(2.826)
Mariage												0.551***	0.625***	0.659***	0.334***	0.461***	0.622***
												(0.097)	(0.120)	(0.131)	(0.106)	(0.172)	(0.204)
N of children												-0.133***	-0.128***	-0.126***	-0.121***	-0.122***	-0.124***
												(0.024)	(0.026)	(0.028)	(0.024)	(0.024)	(0.028)
Income															0.162*	-0.018	-0.030*
															(0.084)	(0.016)	(0.018)
Own house															0.549***	0.509***	0.457***
															(0.113)	(0.125)	(0.146)
Total wealth															0.006	0.002	0.001
															(0.006)	(0.002)	(0.002)
Observations	8838	8838	8509	8509	7352	7352	8838	8838	8509	8509	8509	8447	8447	8447	8355	8355	8355
R^2	-0.281	-0.272	-0.394	-0.715	-0.074	-0.093	-0.316	-0.313	0.222	0.219	0.124	0.232	0.041	-0.171	0.242	0.172	-0.116
DWHchi2	171.812	96.053	207.155	177.266	115.794	56.994	168.328	88.990		0.366	5.966		1.294	4.719		0.504	4.356
DWHpval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.545	0.015		0.255	0.030		0.478	0.037

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$
All economic variables (e.g. Total wealth, Income) are measured in dollars.
In the specification (7), (country dummy) × (economic variable)
(e.g. (Total wealth) × (the U.S. dummy)) variables are also included.
The estimated coefficients of these cross terms are not presented.
The Belgium dummy is omitted.

Table 32: The effect of instrumental variables and other control variables using weight (only the coefficients of country)

	(1)		(2)		(3)		(4)		(5)			(6)			(7)		
	IV1	IV2	IV1	IV2	IV1	IV2	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2	OLS	IV1	IV2
Country dummy																	
2.US									1.952***	1.798***	1.068**	2.104***	0.858	0.294	2.064***	1.305	0.342
									(0.150)	(0.395)	(0.442)	(0.152)	(0.934)	(1.012)	(0.188)	(0.821)	(0.991)
3.UK									1.977***	1.860***	1.307***	1.932***	0.994	0.569	1.614***	1.161**	0.586
									(0.160)	(0.329)	(0.362)	(0.160)	(0.713)	(0.770)	(0.209)	(0.527)	(0.634)
11.Austria									0.746***	0.767***	0.864***	0.751***	0.904***	0.974***	1.000***	1.093***	1.212***
									(0.235)	(0.241)	(0.251)	(0.236)	(0.276)	(0.292)	(0.325)	(0.348)	(0.378)
12.Germany									0.642***	0.589***	0.340	0.566***	0.115	-0.090	0.527**	0.310	0.034
									(0.186)	(0.226)	(0.245)	(0.186)	(0.389)	(0.423)	(0.235)	(0.326)	(0.381)
13.Sweden									1.142***	0.951*	0.051	1.114***	-0.424	-1.120	1.128***	0.432	-0.452
									(0.187)	(0.502)	(0.558)	(0.187)	(1.150)	(1.244)	(0.305)	(0.793)	(0.946)
14.Netherlands									0.701***	0.671***	0.531**	0.654***	0.412	0.302	0.915***	0.773**	0.592
									(0.195)	(0.209)	(0.219)	(0.195)	(0.277)	(0.299)	(0.291)	(0.329)	(0.364)
15.Spain									-1.671***	-1.746***	-2.102***	-1.686***	-2.288***	-2.561***	-2.253***	-2.605***	-3.052***
									(0.215)	(0.284)	(0.309)	(0.215)	(0.506)	(0.551)	(0.286)	(0.479)	(0.570)
16.Italy									-1.096***	-1.122***	-1.243***	-1.182***	-1.404***	-1.505***	-1.257***	-1.313***	-1.386***
									(0.188)	(0.198)	(0.207)	(0.188)	(0.253)	(0.271)	(0.225)	(0.233)	(0.256)
17.France									-0.139	-0.144	-0.168	-0.134	-0.171	-0.188	0.196	0.152	0.097
									(0.207)	(0.208)	(0.219)	(0.206)	(0.228)	(0.246)	(0.267)	(0.276)	(0.308)
18.Denmark									1.303***	1.205***	0.743**	1.320***	0.540	0.186	1.251***	1.003**	0.688
									(0.240)	(0.340)	(0.373)	(0.238)	(0.635)	(0.689)	(0.316)	(0.428)	(0.507)
19.Greece									0.026	-0.042	-0.364	0.012	-0.540	-0.789	-0.493*	-0.751*	-1.078**
									(0.198)	(0.260)	(0.282)	(0.199)	(0.466)	(0.501)	(0.286)	(0.401)	(0.471)
20.Switzerland									1.226***	1.066**	0.310	1.172***	-0.107	-0.685	1.665***	1.033	0.229
									(0.293)	(0.487)	(0.540)	(0.294)	(0.998)	(1.084)	(0.432)	(0.792)	(0.943)
Observations	8838	8838	8509	8509	7352	7352	8838	8838	8509	8509	8509	8447	8447	8447	8355	8355	8355
R^2	-0.281	-0.272	-0.394	-0.715	-0.074	-0.093	-0.316	-0.313	0.222	0.219	0.124	0.232	0.041	-0.171	0.242	0.172	-0.116
DWHchi2	171.812	96.053	207.155	177.266	115.794	56.994	168.328	88.990		0.366	5.966		1.294	4.719		0.504	4.356
DWHpval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.545	0.015		0.255	0.030		0.478	0.037

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

All economic variables (e.g. Total wealth, Income) are measured in dollars.
In the specification (7), (country dummy) × (economic variable)
(e.g. (Total wealth) × (the U.S. dummy)) variables are also included.
The estimated coefficients of these cross terms are not presented.
The Belgium dummy is omitted.

Table 33: The effect of the difference in the cohort groups using weight (Sample aged from 60 to 64)(Original without Greece)

	2004			2010	
	OLS	IV1	IV2	OLS	IV2
1st Stage Result					
IV-early		0.021	0.005		0.002
		(0.020)	(0.022)		(0.023)
IV-normal		0.061**	0.070***		0.088***
		(0.026)	(0.024)		(0.025)
2nd Stage Result					
Not working for pay	-0.468***	-2.064	-5.131*	-0.694***	-6.379**
	(0.086)	(2.222)	(2.916)	(0.091)	(2.574)
Univ	1.534***	1.369***	1.051***	1.360***	0.670*
	(0.107)	(0.258)	(0.331)	(0.107)	(0.342)
Female	0.996***	1.200***	1.590***	1.173***	1.800***
	(0.082)	(0.306)	(0.397)	(0.082)	(0.325)
Age	7.714***	7.897***	8.247**	-0.345	2.480
	(2.929)	(3.019)	(3.593)	(2.954)	(4.189)
Age squared	-6.292***	-6.394***	-6.592**	0.298	-1.759
	(2.362)	(2.431)	(2.890)	(2.384)	(3.350)
Mariage	0.345***	0.443**	0.631***	0.383***	0.647***
	(0.108)	(0.174)	(0.219)	(0.110)	(0.187)
N of children	-0.123***	-0.125***	-0.127***	-0.071***	-0.080**
	(0.024)	(0.024)	(0.028)	(0.026)	(0.033)
Income	-0.006	-0.016	-0.030	0.055	-0.002
	(0.033)	(0.017)	(0.019)	(0.071)	(0.018)
Own house	0.550***	0.517***	0.454***	0.489***	0.327**
	(0.115)	(0.126)	(0.152)	(0.112)	(0.160)
Total wealth	0.004*	0.002	0.001	-0.001	-0.004
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Observations	7987	7987	7987	9195	9195
R^2	0.243	0.200	-0.124	0.173	-0.438
DWHchi2		0.197	3.507		7.515
DWHpval		0.657	0.061		0.006

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

	2004			2010	
	OLS	IV1	IV2	OLS	IV2
Country dummy					
2.US	2.062***	1.464*	0.315	-0.080	-1.262**
	(0.188)	(0.851)	(1.115)	(0.210)	(0.600)
3.UK	1.612***	1.254**	0.567	1.248***	0.690*
	(0.209)	(0.544)	(0.706)	(0.236)	(0.392)
11.Austria	1.002***	1.075***	1.214***	0.282	0.885*
	(0.325)	(0.346)	(0.383)	(0.293)	(0.453)
12.Germany	0.525**	0.353	0.024	-0.323	-1.493*
	(0.235)	(0.331)	(0.409)	(0.455)	(0.775)
13.Sweden	1.124***	0.575	-0.481	0.322	-1.720
	(0.305)	(0.821)	(1.057)	(0.442)	(1.087)
14.Netherlands	0.914***	0.802**	0.586	0.083	-1.044
	(0.291)	(0.330)	(0.377)	(0.289)	(0.637)
15.Spain	-2.256***	-2.534***	-3.067***	-2.654***	-2.752***
	(0.286)	(0.487)	(0.618)	(0.292)	(0.389)
16.Italy	-1.260***	-1.305***	-1.391***	-1.515***	-1.301***
	(0.225)	(0.232)	(0.259)	(0.310)	(0.400)
17.France	0.197	0.161	0.094	-0.583**	0.028
	(0.267)	(0.274)	(0.311)	(0.271)	(0.438)
18.Denmark	1.249***	1.054**	0.677	0.826**	-0.476
	(0.316)	(0.429)	(0.536)	(0.365)	(0.801)
20.Switzerland	1.661***	1.161	0.202	0.884***	-1.399
	(0.432)	(0.813)	(1.037)	(0.287)	(1.094)
Observations	7987	7987	7987	9195	9195
R^2	0.243	0.200	-0.124	0.173	-0.438
DWHchi2		0.197	3.507		7.515
DWHpval		0.657	0.061		0.006

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 34: The effect of the difference in the definition of retirement and the surveyed age-group using weight

	Retired (Self-reported)				Not Working for Pay				Completely retired			
	2004		2010		2004		2010		2004		2010	
	OLS	IV2	OLS	IV2	OLS	IV2	OLS	IV2	OLS	IV2	OLS	IV2
Age group: 60-64												
1st Stage Result												
IV-Early-bi		0.037		0.062**		0.003		0.022		0.010		0.032
		(0.023)		(0.025)		(0.023)		(0.025)		(0.023)		(0.025)
IV-Normal-bi		0.107***		0.098***		0.044**		-0.003		0.073***		0.033
		(0.021)		(0.022)		(0.021)		(0.023)		(0.022)		(0.024)
2nd Stage Result												
retirement	-0.274***	-0.216	-0.393***	1.239	-0.439***	-1.415	-0.583***	14.847	-0.398***	-0.702	-0.530***	3.577
	(0.092)	(1.401)	(0.107)	(1.547)	(0.088)	(3.615)	(0.110)	(19.854)	(0.089)	(2.147)	(0.108)	(4.303)
Observations	8078	8078	9239	9239	8095	8095	9299	9299	8076	8076	9213	9213
R^2	0.171	0.171	0.104	0.064	0.172	0.155	0.105	-3.669	0.172	0.171	0.106	-0.170
DWHchi2		0.008		1.051		0.077		2.926		0.018		1.106
DWHpval		0.930		0.305		0.781		0.087		0.892		0.293
Age group: 60-69												
1st Stage Result												
IV-Early-bi		0.070***		0.082***		0.029		0.022		0.027		0.028
		(0.021)		(0.021)		(0.021)		(0.021)		(0.021)		(0.021)
IV-Normal-bi		0.123***		0.139***		0.069***		0.058***		0.089***		0.084***
		(0.012)		(0.011)		(0.012)		(0.013)		(0.013)		(0.013)
2nd Stage Result												
retirement	-0.275***	0.777	-0.319***	0.411	-0.451***	1.527	-0.480***	1.241	-0.422***	1.167	-0.443***	0.490
	(0.074)	(0.704)	(0.087)	(0.720)	(0.067)	(1.328)	(0.081)	(1.849)	(0.067)	(1.033)	(0.081)	(1.263)
Observations	15830	15830	16858	16858	15852	15852	16945	16945	15827	15827	16823	16823
R^2	0.168	0.155	0.111	0.105	0.170	0.113	0.112	0.070	0.170	0.132	0.113	0.100
DWHchi2		2.264		0.890		2.383		0.631		2.477		0.390
DWHpval		0.132		0.345		0.123		0.427		0.116		0.532

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 35: Heterogeneity in Full Retirement (Word Recall)

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
WR summary score						
Completely retired	-0.154***	0.067	0.018	-0.144	-0.200	2.326
	(0.027)	(0.056)	(0.167)	(0.201)	(0.195)	(1.421)
Observations	86773	23923	3998	2365	3497	13437
DWH p-val	0.642	0.305	0.549	0.693	0.224	0.019
Model	FE	FE	FE	FE	FE	FE-IV
Immediate WR						
Completely retired	-0.061***	0.033	-0.063	-0.090	-0.134	-0.033**
	(0.014)	(0.030)	(0.090)	(0.110)	(0.105)	(0.015)
Observations	86773	23930	4000	2365	3497	14127
DWH p-val	0.602	0.194	0.903	0.482	0.674	0.371
Model	FE	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.093***	0.034	0.074	-0.054	1.762	1.898*
	(0.016)	(0.034)	(0.101)	(0.116)	(1.266)	(1.132)
Observations	86773	23936	4004	2365	3498	13437
DWH p-val	0.256	0.634	0.436	0.996	0.097	0.014
Model	FE	FE	FE	FE	FE-IV	FE-IV

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 36: Heterogeneity in Full Retirement (Word Recall)

	(1) US	(2) Korea	(3) Japan
Serial 7s			
Completely retired	-0.279**	-0.122***	-0.072
	(0.127)	(0.032)	(0.064)
Observations	86773	14129	3791
DWH p-val	0.049	0.518	0.561
Model	FE-IV	FE	FE
Backward counting			
Completely retired	-0.001		
	(0.002)		
Observations	86773		
DWH p-val	0.121		
Model	FE		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

Table 37: Heterogeneity in Retirement by Gender (Word Recall)

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
Panel A: Male						
WR summary score						
Completely retired	-0.137***	-0.015	-0.214	1.227	-0.346	-0.130***
	(0.040)	(0.076)	(0.230)	(1.021)	(0.233)	(0.043)
Observations	38848	13032	1949	1205	1920	8149
DWH p-val	0.179	0.347	0.961	0.069	0.393	0.436
Model	FE	FE	FE	FE-IV	FE	FE
Immediate WR						
Completely retired	-0.058***	0.002	-0.231*	0.550	-0.169	-0.052**
	(0.020)	(0.041)	(0.123)	(0.535)	(0.131)	(0.022)
Observations	38848	13035	1951	1205	1920	8149
DWH p-val	0.306	0.982	0.498	0.078	0.495	0.656
Model	FE	FE	FE	FE-IV	FE	FE
Delayed WR						
Completely retired	-0.081***	-0.017	0.023	-0.150	1.737	-0.079**
	(0.024)	(0.047)	(0.141)	(0.159)	(1.111)	(0.032)
Observations	38848	13041	1953	1205	1921	8149
DWH p-val	0.175	0.124	0.645	0.128	0.049	0.477
Model	FE	FE	FE	FE	FE-IV	FE
Panel B: Female						
WR summary score						
Completely retired	-0.164***	0.081	0.153	0.119	0.059	-0.069*
	(0.036)	(0.070)	(0.205)	(0.228)	(0.244)	(0.041)
Observations	47925	13283	2206	1365	1816	5978
DWH p-val	0.134	0.121	0.883	0.907	0.938	0.956
Model	FE	FE	FE	FE	FE	FE
Immediate WR						
Completely retired	-0.064***	0.018	-0.004	0.040	-0.024	-0.015
	(0.018)	(0.039)	(0.108)	(0.124)	(0.131)	(0.021)
Observations	47925	13288	2206	1366	1816	5978
DWH p-val	0.948	0.113	0.640	0.706	0.759	0.125
Model	FE	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	0.451*	0.063	0.140	0.074	0.083	-0.054*
	(0.252)	(0.042)	(0.125)	(0.135)	(0.138)	(0.031)
Observations	47925	13288	2208	1365	1816	5978
DWH p-val	0.023	0.276	0.536	0.540	0.914	0.302
Model	FE-IV	FE	FE	FE	FE	FE

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 38: Heterogeneity in Retirement by Gender (Numeracy)

	(1) US	(2) Korea	(3) Japan
Panel A: Male			
Serial 7s			
Completely retired	-0.008	-0.133***	-0.177*
	(0.016)	(0.046)	(0.091)
Observations	38848	8151	2149
DWH p-val	0.937	0.154	0.548
Model	FE	FE	FE
Backward counting			
Completely retired	-0.002		
	(0.003)		
Observations	38848		
DWH p-val	0.167		
Model	FE		
Panel B: Female			
Serial 7s			
Completely retired	-0.415**	-0.101**	*0.030*
	(0.183)	(0.046)	*(0.091)*
Observations	47925	5978	*1642*
DWH p-val	0.040	0.577	*0.778*
Model	FE-IV	FE	*FE*
Backward counting			
Completely retired	-0.001		
	(0.003)		
Observations	47925		
DWH p-val	0.552		
Model	FE		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 39: Heterogeneity in Retirement by Education (Word Recall)

	(1)	(2)	(3)	(4)	(5)	(6)
	US	England	France	Germany	Denmark	Korea
Panel A: Low education						
WR summary score						
Completely retired	-0.149***	0.064	0.041	-0.068	0.061	2.792
	(0.030)	(0.062)	(0.195)	(0.195)	(0.219)	(1.925)
Observations	65323	17761	2603	1682	1870	12120
DWH p-val	0.397	0.125	0.919	0.897	0.306	0.024
Model	FE	FE	FE	FE	FE	FE-IV
Immediate WR						
Completely retired	-0.053***	0.399**	-0.071	-0.111	0.052	-0.035**
	(0.015)	(0.169)	(0.101)	(0.106)	(0.120)	(0.016)
Observations	65323	17765	2603	1683	1870	12683
DWH p-val	0.852	0.022	0.791	0.809	0.709	0.574
Model	FE	FE-IV	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.098***	0.042	0.108	0.038	1.253*	2.466
	(0.019)	(0.037)	(0.120)	(0.117)	(0.694)	(1.609)
Observations	65323	17771	2606	1682	1871	12120
DWH p-val	0.236	0.684	0.919	0.943	0.058	0.010
Model	FE	FE	FE	FE	FE-IV	FE-IV
Panel B: High education						
WR summary score						
Completely retired	-0.158***	-0.003	-0.032	-0.216	-0.401	-0.061
	(0.058)	(0.116)	(0.284)	(0.379)	(0.253)	(0.091)
Observations	21433	5443	1262	882	1858	1439
DWH p-val	0.425	0.167	0.861	0.166	0.693	0.964
Model	FE	FE	FE	FE	FE	FE
Immediate WR						
Completely retired	-0.088***	0.025	-0.174	-0.064	-0.300**	-0.036
	(0.030)	(0.064)	(0.150)	(0.197)	(0.136)	(0.042)
Observations	21433	5446	1264	882	1858	1439
DWH p-val	0.131	0.118	0.753	0.489	0.483	0.486
Model	FE	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.070**	-0.025	0.138	1.064	-0.101	-0.025
	(0.035)	(0.069)	(0.175)	(0.748)	(0.155)	(0.073)
Observations	21433	5446	1265	882	1858	1439
DWH p-val	0.931	0.400	0.998	0.091	0.892	0.702
Model	FE	FE	FE	FE-IV	FE	FE

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 40: Heterogeneity in Retirement by Education (Numeracy)

	(1) US	(2) Korea	(3) Japan
Panel A: Low education			
Serial 7s			
Completely retired	-0.321**	-0.134***	*-0.082*
	(0.141)	(0.035)	*(0.069)*
Observations	65323	12685	*3221*
DWH p-val	0.035	0.549	*0.597*
Model	FE-IV	FE	*FE*
Backward counting			
Completely retired	-0.001		
	(0.002)		
Observations	65323		
DWH p-val	0.505		
Model	FE		
Panel B: High education			
Serial 7s			
Completely retired	-0.039*	0.002	0.023
	(0.022)	(0.076)	(0.188)
Observations	21433	1439	558
DWH p-val	0.761	0.507	0.487
Model	FE	FE	FE
Backward counting			
Completely retired	0.121**		
	(0.055)		
Observations	21433		
DWH p-val	0.019		
Model	FE-IV		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 41: Heterogeneity in Retirement by Occupation (Word Recall)

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
Panel A: Blue collar						
WR summary score						
Completely retired	-0.083	0.814**	-0.721*	-1.046**	-0.565	*-0.134****
	(0.055)	(0.355)	(0.404)	(0.484)	(0.373)	*(0.038)*
Observations	19404	12907	915	580	696	*8542*
DWH p-val	0.340	0.037	0.861	0.523	0.572	*0.176*
Model	FE	FE-IV	FE	FE	FE	*FE*
Immediate WR						
Completely retired	-0.054*	0.584***	-0.595***	-0.622**	-0.288	*-0.046***
	(0.028)	(0.197)	(0.229)	(0.248)	(0.209)	*(0.020)*
Observations	19404	12910	915	581	696	*8542*
DWH p-val	0.530	0.003	0.652	0.782	0.618	*0.510*
Model	FE	FE-IV	FE	FE	FE	*FE*
Delayed WR						
Completely retired	-0.029	0.064	-0.088	-0.438	-0.275	*-0.088****
	(0.033)	(0.044)	(0.230)	(0.290)	(0.211)	*(0.029)*
Observations	19404	12914	916	580	697	*8542*
DWH p-val	0.314	0.455	0.447	0.468	0.185	*0.194*
Model	FE	FE	FE	FE	FE	*FE*
Panel B: White collar						
WR summary score						
Completely retired	-0.169***	0.007	0.181	-0.177	0.059	-0.049
	(0.034)	(0.074)	(0.240)	(0.262)	(0.246)	(0.050)
Observations	53512	13316	2570	1520	2785	4794
DWH p-val	0.301	0.614	0.404	0.745	0.566	0.917
Model	FE	FE	FE	FE	FE	FE
Immediate WR						
Completely retired	-0.405**	0.012	0.055	-0.094	-0.024	-0.013
	(0.203)	(0.041)	(0.124)	(0.135)	(0.130)	(0.023)
Observations	53512	13321	2572	1520	2785	4794
DWH p-val	0.097	0.550	0.379	0.772	0.225	0.300
Model	FE-IV	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.104***	-0.003	0.127	-0.083	0.083	-0.036
	(0.021)	(0.044)	(0.150)	(0.156)	(0.140)	(0.041)
Observations	53512	13323	2572	1520	2785	4794
DWH p-val	0.715	0.763	0.565	0.793	0.965	0.443
Model	FE	FE	FE	FE	FE	FE

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 42: Heterogeneity in Retirement by Occupation (Numaracy)

	(1) US	(2) Korea	(3) Japan
Panel A: Blue collar			
Serial 7s			
Completely retired	-0.048**	*-0.115****	*-0.270***
	(0.024)	*(0.044)*	*(0.106)*
Observations	19404	*8544*	*1383*
DWH p-val	0.674	*0.543*	*0.292*
Model	FE	*FE*	*FE*
Backward counting			
Completely retired	-0.003		
	(0.005)		
Observations	19404		
DWH p-val	0.675		
Model	FE		
Panel B: White collar			
Serial 7s			
Completely retired	-0.338**	-0.086*	0.032
	(0.168)	(0.049)	(0.086)
Observations	53512	4794	2177
DWH p-val	0.058	0.333	0.452
Model	FE-IV	FE	FE
Backward counting			
Completely retired	-0.000		
	(0.002)		
Observations	53512		
DWH p-val	0.262		
Model	FE		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 43: Heterogeneity in Retirement by Leisure Activities 1

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
Panel A: Social activity increases						
WR summary score						
Completely retired	-0.165	-2.528**	*0.805*	*0.474*	-0.073	*0.372***
	(0.110)	(1.185)	*(0.721)*	*(0.523)*	(0.787)	*(0.155)*
Observations	4523	1833	*261*	*192*	166	*384*
DWH p-val	0.978	0.029	*0.072*	*0.419*	0.496	*0.205*
Model	FE	FE-IV	*FE*	*FE*	FE	*FE*
Immediate WR						
Completely retired	-0.046	-1.120*	*0.258*	*-0.544*	-0.342	*0.207****
	(0.058)	(0.645)	*(0.389)*	*(0.335)*	(0.362)	*(0.071)*
Observations	4523	1833	*261*	*192*	166	*384*
DWH p-val	0.257	0.084	*0.176*	*0.569*	0.559	*0.020*
Model	FE	FE-IV	*FE*	*FE*	FE	*FE*
Delayed WR						
Completely retired	-0.119*	-1.428**	*0.548*	*1.018****	0.269	*0.165*
	(0.066)	(0.713)	*(0.395)*	*(0.321)*	(0.487)	*(0.121)*
Observations	4523	1834	*261*	*192*	166	*384*
DWH p-val	0.362	0.042	*0.038*	*0.425*	0.573	*0.590*
Model	FE	FE-IV	*FE*	*FE*	FE	*FE*
Panel B: Social activity decreases						
WR summary score						
Completely retired	0.030	0.011	-0.675**	-0.336	-0.476*	*-0.015*
	(0.120)	(0.096)	(0.261)	(0.328)	(0.282)	*(0.055)*
Observations	4384	6488	918	670	879	*3089*
DWH p-val	0.193	0.443	0.538	0.666	0.598	*0.187*
Model	FE	FE	FE	FE	FE	*FE*
Immediate WR						
Completely retired	0.058	0.041	-0.424***	0.002	-0.243	*-0.040*
	(0.061)	(0.055)	(0.149)	(0.177)	(0.150)	*(0.030)*
Observations	4384	6488	918	671	879	*3089*
DWH p-val	0.134	0.413	0.886	0.974	0.340	*0.182*
Model	FE	FE	FE	FE	FE	*FE*
Delayed WR						
Completely retired	-0.013	-0.032	-0.260	-0.339*	-0.236	*0.025*
	(0.068)	(0.058)	(0.163)	(0.190)	(0.170)	*(0.040)*
Observations	4384	6489	920	670	880	*3089*
DWH p-val	0.453	0.594	0.378	0.525	0.981	*0.392*
Model	FE	FE	FE	FE	FE	*FE*

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 44: Heterogeneity in Retirement by Leisure Activities 2

	(1) US	(2) Korea	(3) Japan
Panel A: Social activity increases			
Serial 7s			
Completely retired	0.012	*0.327*	*0.430*
	(0.047)	*(0.214)*	*(0.283)*
Observations	4523	*384*	*172*
DWH p-val	0.901	*0.302*	*0.782*
Model	FE	*FE*	*FE*
Backward counting			
Completely retired	-0.016*		
	(0.008)		
Observations	4523		
DWH p-val	0.200		
Model	FE		
Panel B: Social activity decreases			
Serial 7s			
Completely retired	-0.028	*-0.134***	-0.150
	(0.053)	*(0.057)*	(0.134)
Observations	4384	*3090*	774
DWH p-val	0.520	*0.833*	0.440
Model	FE	*FE*	FE
Backward counting			
Completely retired	-0.003		
	(0.010)		
Observations	4384		
DWH p-val	0.414		
Model	FE		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 45: Heterogeneity in Retirement by Having a Partner 1

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
Panel A: Not having spouse						
WR summary score						
Completely retired	-0.217***	-0.033	-0.022	-0.180	*-0.207*	*-0.100*
	(0.059)	(0.118)	(0.288)	(0.520)	*(0.342)*	*(0.081)*
Observations	20053	5020	1064	390	*730*	*1727*
DWH p-val	0.912	0.577	0.354	0.562	*0.070*	*0.688*
Model	FE	FE	FE	FE	*FE*	*FE*
Immediate WR						
Completely retired	-0.103***	-0.028	-0.079	-0.088	*-0.158*	*-0.062*
	(0.029)	(0.063)	(0.154)	(0.263)	*(0.187)*	*(0.044)*
Observations	20053	5023	1064	390	*730*	*1727*
DWH p-val	0.407	0.309	0.137	0.656	*0.895*	*0.325*
Model	FE	FE	FE	FE	*FE*	*FE*
Delayed WR						
Completely retired	-0.113***	-0.018	0.057	-0.092	*-0.049*	*-0.038*
	(0.036)	(0.074)	(0.189)	(0.304)	*(0.202)*	*(0.061)*
Observations	20053	5024	1064	390	*730*	*1727*
DWH p-val	0.630	0.920	0.894	0.760	*0.004*	*0.261*
Model	FE	FE	FE	FE	*FE*	*FE*
Panel B: Having spouse						
WR summary score						
Completely retired	-0.134***	0.064	0.031	-0.083	-0.112	2.659*
	(0.030)	(0.057)	(0.180)	(0.185)	(0.197)	(1.582)
Observations	66720	21295	3091	2180	3006	11803
DWH p-val	0.492	0.344	0.423	0.291	0.389	0.011
Model	FE	FE	FE	FE	FE	FE-IV
Immediate WR						
Completely retired	-0.049***	0.024	-0.093	-0.099	-0.075	-0.030*
	(0.015)	(0.032)	(0.095)	(0.100)	(0.108)	(0.016)
Observations	66720	21300	3093	2181	3006	12400
DWH p-val	0.884	0.216	0.167	0.465	0.219	0.164
Model	FE	FE	FE	FE	FE	FE
Delayed WR						
Completely retired	-0.087***	0.040	0.116	0.012	-0.040	1.936
	(0.018)	(0.034)	(0.107)	(0.109)	(0.113)	(1.191)
Observations	66720	21305	3097	2180	3007	11803
DWH p-val	0.240	0.666	0.928	0.242	0.755	0.016
Model	FE	FE	FE	FE	FE	FE-IV

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 46: Heterogeneity in Retirement by Having a Partner 2

	(1) US	(2) Korea	(3) Japan
Panel A: Not having spouse			
Serial 7s			
Completely retired	-0.050**	*-0.166**	*-0.036*
	(0.025)	*(0.085)*	*(0.140)*
Observations	20053	*1727*	*777*
DWH p-val	0.115	*0.126*	*0.935*
Model	FE	*FE*	*FE*
Backward counting			
Completely retired	-0.002		
	(0.005)		
Observations	20053		
DWH p-val	0.992		
Model	FE		
Panel B: Having spouse			
Serial 7s			
Completely retired	-0.023*	-0.113***	*-0.074*
	(0.013)	(0.035)	*(0.072)*
Observations	66720	12402	*3014*
DWH p-val	0.155	0.260	*0.715*
Model	FE	FE	*FE*
Backward counting			
Completely retired	0.050**		
	(0.024)		
Observations	66720		
DWH p-val	0.031		
Model	FE-IV		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).
[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).
[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 47: Heterogeneity in Retirement by BMI 1

	(1) US	(2) England	(3) France	(4) Germany	(5) Denmark	(6) Korea
Panel A: $BMI < 25$						
WR summary score						
Completely retired	-0.066	0.039	0.105	-0.063	*-0.243*	2.620
	(0.049)	(0.119)	(0.250)	(0.280)	*(0.253)*	(1.619)
Observations	26866	4901	2015	1024	*1672*	10121
DWH p-val	0.769	0.115	0.427	0.354	*0.348*	0.015
Model	FE	FE	FE	FE	*FE*	FE-IV
Immediate WR						
Completely retired	-0.025	0.058	-0.080	-0.023	*-0.127*	1.171
	(0.024)	(0.065)	(0.129)	(0.152)	*(0.129)*	(0.762)
Observations	26866	4902	2017	1024	*1672*	10121
DWH p-val	0.283	0.351	0.584	0.519	*0.952*	0.032
Model	FE	FE	FE	FE	*FE*	FE-IV
Delayed WR						
Completely retired	-0.044	-0.724*	0.161	-0.041	*-0.117*	1.449
	(0.030)	(0.430)	(0.149)	(0.162)	*(0.154)*	(1.088)
Observations	26866	4903	2020	1024	*1672*	10121
DWH p-val	0.173	0.088	0.400	0.373	*0.134*	0.077
Model	FE	FE-IV	FE	FE	*FE*	FE-IV
Panel B: $BMI \geq 25$						
WR summary score						
Completely retired	-0.191***	-0.031	-0.095	-0.155	-0.057	*-0.157****
	(0.033)	(0.075)	(0.191)	(0.229)	(0.230)	*(0.054)*
Observations	59286	12740	2118	1537	2042	*3392*
DWH p-val	0.724	0.134	0.675	0.530	0.563	*0.220*
Model	FE	FE	FE	FE	FE	*FE*
Immediate WR						
Completely retired	-0.076***	-0.031	-0.126	-0.145	-0.073	*-0.074***
	(0.016)	(0.041)	(0.104)	(0.122)	(0.133)	*(0.030)*
Observations	59286	12741	2118	1538	2042	*3392*
DWH p-val	0.975	0.143	0.869	0.571	0.117	*0.310*
Model	FE	FE	FE	FE	FE	*FE*
Delayed WR						
Completely retired	-0.115***	0.000	0.031	-0.014	0.013	*-0.083***
	(0.020)	(0.046)	(0.118)	(0.136)	(0.127)	*(0.041)*
Observations	59286	12746	2118	1537	2043	*3392*
DWH p-val	0.619	0.272	0.666	0.564	0.693	*0.028*
Model	FE	FE	FE	FE	FE	*FE*

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.

Table 48: Heterogeneity in Retirement by BMI 2

	(1) US	(2) Korea	(3) Japan
Panel A: $BMI < 25$			
Serial 7s			
Completely retired	-0.034	-0.119***	-0.033
	(0.021)	(0.037)	(0.078)
Observations	26866	10661	2796
DWH p-val	0.156	0.468	0.785
Model	FE	FE	FE
Backward counting			
Completely retired	-0.004		
	(0.004)		
Observations	26866		
DWH p-val	0.158		
Model	FE		
Panel B: $BMI \geq 25$			
Serial 7s			
Completely retired	-0.026*	*-0.116**	*-0.181*
	(0.014)	*(0.068)*	*(0.110)*
Observations	59286	*3393*	*975*
DWH p-val	0.304	*0.508*	*0.982*
Model	FE	*FE*	*FE*
Backward counting			
Completely retired	-0.000		
	(0.003)		
Observations	59286		
DWH p-val	0.322		
Model	FE		

[1] Standard errors in parentheses and * ($p < .1$), ** ($p < .05$), *** ($p < .01$).

[2] All specifications include demographic variables (age, age , marriage dummy, number of children), economic variables (household income, house ownership, total wealth), region dummies, year dummies, 1st survey dummy, and 2nd-4th survey dummy (only HRS).

[3] The green character indicates that the IVs do not work well in the 1st stage regression.