HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 16/13

What Explains the Difference in the Effect of Retirement on Health?: Evidence from Global Aging Data

Hiroyuki Motegi, Yoshinori Nishimura & Masato Oikawa

July 2016

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

What Explains the Difference in the Effect of Retirement on Health?: Evidence from Global Aging Data

Hiroyuki Motegi* Yoshinori Nishimura† Masato Oikawa‡

July 29, 2016

Abstract

This paper analyzes the reasons for differences in the effect of retirement on health estimated results in previous studies. We investigate these differences by focusing on the analysis methods used by these studies. Using various health indexes, numerous researchers have examined the effects of retirement on health. However, there are no unified views on the impact of retirement on various health indexes. Consequently, we show that the choice of analysis method is one of the key factors in explaining why the estimated results of the effect of retirement on health differ. Moreover, we re-estimate the effect of retirement on health by using a fixed analysis method controlling for individual heterogeneity and endogeneity of the retirement behavior. We analyze the effect of retirement on health parameters, such as cognitive function, self-report of health, activities of daily living (ADL), depression, and body mass index in eight countries. We find that the effects of retirement on self-report of health, depression, and ADL are positive in many of these countries.

JEL Classification Numbers: I00, I100, I120, I190, J260.
Keywords: aging, health, retirement, global aging data

*Graduate School of Economics, the University of Tokyo, Japan. Email: motegihiro@gmail.com
†Corresponding Author: Graduate School of Economics, the University of Tokyo, Japan. Email: nishimura.yy@gmail.com
‡Corresponding Author: Graduate School of Economics, the University of Tokyo, Japan. Email: masato.oikawa1991@gmail.com

1 Introduction

Retirement related policies, such as pension system reform, have become important for developed countries to sustain their social security systems. Numerous developed countries have faced the same problems of a decreasing birthrate and an ageing population. As population ages, the cost of social security and social welfare increases, eroding the country's budget. As such, developed countries have reformed their pension systems to reduce the cost of social security and social welfare. Moreover, many developed countries, such as the United States, the United Kingdom, and Korea have already decided to increase pension eligibility age for the next decades. Japan has already increased the pension eligibility age. These pension reforms in developed countries are expected to delay retirement. As Gruber and Wise (1998) discuss, the relationship between the social security system and retirement in developed countries generated a lot of attention in economics. When policy makers evaluate the effect of these reforms, health is a key factor. If working is beneficial for the health of the elderly, it would lead to reduced medical expenses and vice-versa.

Along with a growing interest in the effect of these retirement delaying policies, a number of studies have investigated the relation between retirement and health over the last two decades. [1] Using various health indexes, numerous researchers have examined the relationship between health and retirement. To the best of our knowledge, Kerkhofs and Lindeboom (1997) is one of the first papers suggesting endogenous decisions between retirement and health, and identifying the effect of retirement on health. They find that the Hopkins Symptom Checklist (HSCL) health index can be improved after early retirement in the Netherlands by applying FE methods. Lindeboom et al. (2002) extend Kerkhofs and Lindeboom (1997) study to other indices such as the mini-mental state examination (MMSE) test on cognitive ability, the Center for Epidemiological Studies-Depression (CES-D) test of depressing feelings, and others, and apply FE methods to Dutch data different from that of Kerkhofs and Lindeboom (1997). Charles (2004) is also one of the first investigations that analyze the causal effect of retirement on health focusing on subjective well-being (SWB) in economic literature by using instrumental variables (IVs).

Additionally, there are numerous other papers that study the effect of retirement on various health indexes (e.g., Bound and Waidmann, 2007; Coe and Lindeboom, 2008, Dave, Rashad, and Spasojevic, 2008; Neuman, 2008; Johnston and Lee, 2009; Latif, 2011; Coe and Zamarro, 2011; Kajitani, 2011; Behncke, 2012; Bonsang, Adam, and Perelman, 2012; Mazzonna and Peracchi, 2012; Hernaes et al., 2013; Bingley and Martinello, 2013; Hashimoto, 2013; Insler, 2014; Kajitani, Sakata, and McKenzie, 2014; Hashimoto, 2015; Kajitani, Sakata, and McKenzie, 2016). There are, however, no unified views on the impact of retirement on various health indexes. While some studies conclude that retirement has a positive impact on health defined as mental or physical health, other studies conclude that retirement has no or negative effect. Additionally, these results depend on characteristics such as gender and education.

The goal of this paper is to explain why the effect of retirement on health estimated results in the previous studies differ. One of the keys to understanding these differences is a better understanding of the path through which retirement influences health. If there is an important link between retirement and health (i.e., a mechanism through which retirement influences health outcomes), the effect of retirement on health could be heterogeneous. In fact, some researchers focus on the change in the health investment behaviors after retirement to explain why the effect of retirement on health estimated results in the previous studies differ (e.g., Zhao, Konishi, and Noguchi, 2013; Ayyagari,

[1] We omit the literature on the effect of health on retirement. However, a representative paper is McGarry (2004).

2014; Insler, 2014; Eibichi, 2015; Motegi, Nishimura, and Terada, 2016). On the other hand, we investigate the differences by focusing on the research framework. There is no research to focus on the research framework to explain why the effect of retirement on health estimated results in the previous studies differ.

According to our analysis, the analysis method is one of the determinants of these differences. By choosing an analysis methodology, we also comprehensively reexamine the effect of retirement on health in eight countries. We analyze five health indexes, such as self-reported health, depression, cognitive function, body mass index (BMI), and activities of daily living (ADL). Related literature does not seem to control for retirement endogeneity, while we control for individual heterogeneity and endogeneity of retirement behavior. By doing so, we show the comprehensive results of the effect of retirement on health. The rest of this paper is organized as follows: Section 2 reviews preceding studies; Section 3 discusses the data; Section 4 examines why the estimated results of the effect of retirement on health in previous studies differ from each other; Section 5 performs harmonized analysis on the effect of retirement on health; and Section 6 concludes this paper and discusses future research scope.

2 Literature Review

This section summarizes related studies, focusing on economic literature. As such, we introduce studies that examine the effect of retirement on health. The study by Kerkhofs and Lindeboom (1997) is one of the first to suggest an endogenous decision linking retirement and health regarding the effects of retirement on health. Using a fixed effects (FE) method, they find that, in the Netherlands, the HSCL health index can be improved after early retirement. Lindeboom et al. (2002) extended the study by Kerkhofs and Lindeboom (1997) with other measurement scales, such as the MMSE and the CES-D, with FE methods, using Dutch data, obtaining different results. Charles (2004) also conducted an early investigation analyzing the causal effects of retirement on health by focusing on SWB and through IV. Psychological and psychiatric literature boasts a large body of research on the correlation of retirement and SWB, but has paid scant attention to causal effects. [2]

Furthermore, Rohwedder and Willis (2010), who investigated the effects of retirement on cognitive abilities and compared micro data across the USA, the UK, and 11 European countries, found a negative influence of retirement on cognitive abilities. They suggest that institutional differences across countries, such as pensions, taxes, and disability policies, are also important in explaining the differences in health outcomes across countries. As such, Rohwedder and Willis (2010) gave an impetus to research on the effect of retirement on cognitive abilities, making possible studies such as those by Bonsang et al. (2012), Mazzonna and Peracchi (2012), Coe et al. (2012), and Bingley and Martinello (2013). Additionally, numerous other studies assessed the effects of retirement on other aspects of health. [3] Finally, Tables 1, 2, 3, and 4 show a summary of relevant studies, chosen based on the following criteria:

[2] Charles (2004) surveyed psychological research both theoretically and empirically.

[3] Bound and Waidmann (2007), Coe and Lindeboom (2008), Dave et al. (2008), Neuman (2008), Johnston and Lee (2009), Lee and Smith (2009), Latif (2011), Coe and Zamarro (2011), Behncke (2012), Hernaes et al. (2013), Fonseca et al. (2014) and Insler (2014) are representative papers. Furthermore, recently review papers have been published on the impact of retirement on health in other fields. For example, van der Heide et al. (2013) put retirement in the public health context, whereas Wang and Shi (2014) took up retirement in a psychological context.

- We choose all papers that analyze the effect of retirement on health which have been published by November 2015.
- We choose all working papers that have more than 50 citations on Google Scholar by November 2015.

In Tables 1, 2, 3, and 4, we show the category of health outcome, method, the definition of retirement, control variables information, dataset, the method of sample selection, and the surveyed country. Here, "positive" means the positive impact on a health status (better after retirement), "negative" means a negative impact worse after retirement, and "no" means no impact. According to Tables 1, 2, 3, and 4, there is no unifying result in all health indexes except the health index, which only a few studies analyze. Numerous studies analyze CES-D, self-report of health, ADL, and cognitive functioning. We consider why they obtain different results. We also add BMI to the analyzed indexes, although only two duties in our list use it. This is because we comprehensively analyze the effect of retirement on health indexes. In the Appendix (A.2. The Review of Additional Preceding Studies), we show the other indexes on illness. However, this paper does not focus on the health indexes of illness.

Table 1: Original index and Mental health 1

	Kerkhofs and Lindeboom	Lindeboom et al.	Charles	Bound and Waidmann	Coe and Lindeboom	Dave et al.	Neuman
	1997, Health Economics	2002, Health Economics	2004, Research in Labor Economics	2007, Univ. Michigan WP	2008, IZA DP	2008, Sourthern Economic Journal	2008, J of Labor Research
original index							
CESD		positive			no	negative	no (M) no(F)
SWB			positive			no(psychological problem	
SR health					positive (restricting within 2 years)	negative	positive(M) positive(F)
health fair poor				positive(M) positive(F)			
HSCL	positive						
Method	Fixed effect method	Fixed effect method	Instrumental variable method	pseudo RDD	Instrumental variable method	Fixed effect method	Instrumental variable method
Method (details)			IVs: Social security normal retirement age		IVs: Pension eligibility age	Restricting sample who has good health before retirement, and retire as of 62	IVs: public and private PEA for respondent and for spouse working more than 10 years
Def. of Retirement	Early retirement (elderly who retire as of 55 y/o)		not working for pay not seeking work not worked for a year		Report to be out of the labor force or not having any paid employment	complete retirement (retired and not working)	Working less than 1200 hours in a year
Controls(Demog.)	age, education	age, residential area, marital status, children' health	race, education, age, marital status		age, education, marital status, children	age, sex, race, marital status, education	age, education, race, whether parents living or not, children, marital status, region
Controls(Economic)						income, asset	financial status
Controls(Working.)	working status, occupation	employment status			job types (blue and white collar)		occupation
Controls(Health)	lifestyle habits	health	health conditions			lifestyle habits	early factors health behaviors
Data	CERRA 93, 95	Longitudinal Aging Study Amsterdam panel 92, 95, 98	HRS	ELSA 1st wave	HRS 1st-7th wave	HRS 1st-7th wave	HRS 1st-7th wave
Sample					male workers aged 55-70 years		
Country	Netherlands	Netherlands	The U.S.	The U.K.	The U.S.	The U.S.	The U.S.

Table 2: Original index and Mental health 2

	Johnston and Lee	Lee and Smith	Kajitani	Latif	Coe and Zamarro	Behncke	Fonseca et al.	Insler
	2009, Economics Letters	2009, J Population Aging	2011, Japan and the World Economy	2011, J Socio-Economics	2011, J Health Economics	2012, Health Economics	2014, J Population Aging	2014, J Human Resources
original index					positive	negative		positive (for the case of long term retirement)
CESD	positive(M)	no				no		
SWB				positive			no (EUROD)	
SR health			positive(=1 if "excellent" or "fairly good")		positive	negative		
Method	RDD	Two-limit Robit and Probit	Probit	Fixed effect method and FE-IV	Instrumental variable method	Nonparametric matching	Instrumental variable method	FE-IV
Method (details)	Using 65 years as kink points robustness check by changing bandwidth		1st stage: Tobit estimation with the employment status(self-employed or not) and marital status as IVs; 2nd stage: Probit estimation.	IVs: pension eligibility age	IVs: eligibility age for early and full retirement	Using state pension eligibility age as IV	IVs: pension eligibility age	IV; working expectations and preference derived from "workers' self-reported probabilities of working past ages 62 and 65."
Def. of Retirement	Retired from paid work	Answering retired from working, never worked, retired and unemployed	working hours per week	currentry not working due to retirement	Not in the paid labor force	retired describes her current situation best and not in paid work was her activity in the last month	Answered retired	short retirement; retire at period t, long term retirement; retire before period t-1. self reported retirement (robustness check; Are you currently working for pay)
Controls(Demog.)		sex, education, marital status, children	age, age-squared, education, householder, large city, year dummy	age, education, residential area, marital status	education, marital status, children	children, birth place, residential area	age, sex, education	sex, age, marital status, education, race
Controls(Economic)		income, asset			income	income		
Controls(Working.)		employment status	longest-held occupation dummy		self employment	working hours, employment status		asset
Controls(Health)		health condition and lifestyle habits	BMI category, illness of any member of the respondent's family	health conditions			disability and health conditions	
Data	Health Survey for England	KLoSA 1st wave	1990, 1993, and 1996 National Surveys of the Japanese Elderly	Canadian National Population Health Survey 1st-6th wave	SHARE 1st-2nd wave	ELSA 1st-3rd wave	SHARE 2004, 06, 10	HRS 1992˜2010
Sample			male over 60 years old					restricting elderly working more than 10 years
Country	The U.K.	Korea	Japan	Canada	European countries	The U.K.	European countries	The U.S.

Table 3: Cognitive functioning and Physical function 1

	Lindeboom et al.	Bound and Waidmann	Coe and Lindeboom	Dave et al.	Neuman	Johnston and Lee	Rohwedder and Willis
	2002, Health Economics	2007, Univ. Michigan WP	2008, IZA DP	2008, Sourthern Economic Journal	2008, J of Labor Research	2009, Economics Letters	2010, J Econ Perspectives
cognitive functioning	negative(MMSE (tests cognitive abilities))						negative
physical performance		no(M) negative(F)					
body nagi limitations		positive (M) positive(F)					
ADL			no	negative	no(M) positive(F)		
Body Mass Index						No	
Method	FE method	pseudo RDD	IV method	Fixed effect method	Iinstrumenta variable method	RDD	IV
Method (details)			IVs: pension eligibility age	Restricting sample who has good health before retirement, and retire as of 62	IVs: public and private PEA for respondent and for spouse working more than 10 years	Using 65 years as kink points robustness check by changing bandwidth	IVs: pension eligibility age for early and full
Def. of Retirement			people report to be out of the labor force or not having any paid employment	complete retirement (retired and not working)	elderly working less than 1200 hours in a year	Retired from paid work	not having worked for pay in the last 4 weeks
Controls(Demog.)	age, residential area, marital status, children' health		age, education, marital status, children	age, sex, race, marital status, education	age, education, race, whether parents living or not, children, marital status, region		
Controls(Economic)				income, asset	financial status		
Controls(Working.)	employment status		job types (blue and white collar)		occupation		
Controls(Health)	health			lifestyle habits	early factors health beaviors		
Data	Longitudinal Aging Study Amsterdam panel 92, 95, 98	ELSA 1st wave	HRS 1st-7th wave male workers aged 55-70 years	HRS 1st-7th wave	HRS 1st-7th wave	Health Survey for England	HRS ELSA SHARE at 2004
Sample						Male who do not have degree	
Country	Netherlands	The U.K.	The U.S.	The U.S.	The U.S.	The U.K.	The U.S.The U.K.European countries

Table 4: Cognitive functioning and Physical function 2

	Coe and Zamarro	Behncke	Bonsang et al.	Mazzonna and Peracchi	Coe, Gaudecker, Lindeboom and Maurer	Bingley and Martinello	Godard
	2011, J Health Economics	2012, Health Economics	2012, J Health Economics	2012, European Economic Review	2012, Health Economics	2013, European Economic Review	2016, J Health Economics
cognitive functioning	no	negative	negative	negative	positive (blue collor) no (white collor)	negative	
physical performance							
body nagi limitations							
ADL		negative					
Body Mass Index							positive(BMI,M), no(BMI,F)
Method	Iinstrumenta variable method	Nonparametric matching	FE-IV method	IV method	Generalization of 2SLS	IV method	FE-IV method
Method (details)	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 retirement age
Def. of Retirement	someone who is not in the paid labor force	retirede 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	self-declared current job situation (whether an individual is retired)
Controls(Demog.)	education, maritl status, children	children, birth place, residential area	age	age and education	education, race, religion and age	age, sex, and education	age, age squared, year dummy, living with partners or not
Controls(Economic)	income	income					
Controls(Working.)	self employment	working hours, employment status					
Controls(Health)							
Data	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	SHARE 2004, 2006, 2010.
Sample						Dropping elderly whose educational variables are missing and restricting 60˜64.	restricting 50-69
Country	EU	The U.K.	The U.S.	European countries	The U.S.	The U.S.The U.K.European countries	European countries

3 Data

This paper uses the Health and Retirement Study (HRS) [4] and other related datasets, such as the English Longitudinal Study of Ageing (ELSA), the Health Survey for England (HSE), the Survey of Health, Ageing, and Retirement in Europe (SHARE), and the Japanese Study of Ageing and Retirement (JSTAR). These are panel surveys of individuals 50 or older. These family datasets are constructed so that the questions in the HRS family studies are as similar to the original questions in the HRS as possible. They include a rich variety of variables to capture living aspects in terms of economic status, health status, family background, as well as social and work status. We subsequently explain all health indexes used.

Cognitive score: We use the cognitive function score in the HRS and other related datasets. In the HRS, we use the immediate word recall scores (first half of the word recall test), delayed word recall (second half of the word recall test), [5] and word recall summary score (immediate word recall plus delayed word recall). The word recall summary score is between 0 and 20. The immediate word recall and delayed word recall tests ask the respondent to recall as many words as possible from a list of 10 words. The score of immediate word recall and delayed word recall is the number of words from the 10-word list that were recalled correctly.
Self-report of health: In the HRS, there is a variable that indicates self-reported health conditions. The variable measures the categories of health self-reports as excellent, very good, good, fair, poor. The health categories are numbered from 1 (excellent) to 5 (poor). In all related datasets, the same variable is present. We convert the five values into two health statuses, poor health or not poor health. Additionally, in the ELSA and the SHARE, we can use another scale of self-assessed health: very good, good, fair, bad, and very bad. We also define the health self-report index of "bad health." [6]
ADL: This variable measures the change in the index for ADL. In the HRS and other related datasets, all respondents are asked to answer questions such as "Because of a health or memory problem do you have any difficulty with bathing or showering?" We use this information when calculating the ADL score.
Depression: In the HRS, there is a question targeting whether a respondent has symptoms of depression. For example, one of the statements is "Much of the time during the past week, you felt depressed." We use these questions when we calculate the CES-D score. In the HRS and other related datasets, there are similar questions. Additionally, we use another depression scale, EURO-D, which is available in all version of the SHARE. We mainly use the EURO-D scale in the SHARE because the CES-D scale is only available in waves 1 and 2 of the SHARE.
BMI: In the HRS and other related datasets, all respondents are asked to provide their weight and height, and BMI is calculated using this information. We use the value of BMI and create a dummy variable that takes the value 1 if the respondents BMI value is greater than or equal to 30.

We summarize all scores and values of these health indexes in Tables 5 and 6. In Table 5, we

[4] See the website at (http://hrsonline.isr.umich.edu) for more details on HRS.

[5] There are two rounds in the Word Recall tests. In the first round (Immediate Word Recall), there is a test to recall the number of words as much as possible. After a while, the second round starts. In the second round (Delayed Word Recall), a respondent is asked to recall the same words as much as possible.

[6] "Bad health" is a dummy variable that takes the value 1 if respondents assess their health as fair, bad, and very bad, and 0 otherwise.

show the descriptive statistics of the age group above 50 in all countries and the descriptive statistics for the USA in Table 6. According to Table 5, the scores and values are not at the same level in all countries, BMI in the US being higher than in other countries. In Table 6, we can observe characteristics of the cognitive function. Females have a higher score than males in the word recall summary score. Highly educated individuals have higher overall cognitive scores.

In Section 5, we perform a dynamic analysis for selected countries. We utilize both the pension eligibility age and the long-term variation of retirement behavior. Moreover, we choose the analyzed countries based on the availability of information regarding pension eligibility age. We mainly use the harmonized datasets. [7] However, when our preferred variables are not available in the harmonized datasets, we use the variables of the original datasets. In Table 7, we show a summary explaining which dataset we use in Section 5 of this paper.

More importantly, we use the pensionable age when we calculate our IVs. We explain this point in Appendix (A.1), while in section 5, we use only the pensionable age confirmed to be correct.

[7]The Gateway to Global Aging Data (http://gateway.usc.edu) provides harmonized versions of data from the international ageing and retirement studies (e.g., HRS, ELSA, SHARE, KLoSA). All variables of each dataset have the same items and follow the same naming conventions. The harmonized datasets 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). This code is used to input some variables, such as measures of assets and income.

Table 5: Summary Statistics of Cognition Scores (Age 50 or older) around 2010

	Obs.	Mean	S.D.	Min	Max
HRS					
Word Recall Summary Score	19681	9.61	3.41	0	20
Serial 7's Score	19681	3.41	1.68	0	5
Poor health	21029	0.09	0.28	0	1
ADL summary score (0-3)	20892	0.25	0.66	0	3
CESD summary score (0-8)	19480	1.51	2.03	0	8
BMI	20645	28.46	6.16	7	79
ELSA[*1]					
Word Recall Summary Score	9536	10.40	3.73	0	20
Poor health	9570	0.08	0.27	0	1
ADL summary score (0-3)	10087	0.26	0.63	0	3
CESD summary score (0-8)	9435	1.51	1.96	0	8
BMI[*2]	8230	28.26	5.30	15	71
SHARE[*3]					
Word Recall Summary Score	55472	8.91	3.76	0	20
Serial 7's Score	53332	3.78	1.75	0	5
Poor health	56790	0.13	0.33	0	1
ADL summary score (0-3)	56770	0.17	0.53	0	3
EURO-D summary score (0-12)	55229	2.58	2.31	0	12
BMI	54110	26.92	4.93	6	222
JSTAR					
Word Recall Summary Score	1690	9.56	3.04	0	20
Serial 7's Score	1740	4.16	1.18	0	5
Poor health	2263	0.03	0.17	0	1
ADL summary score (0-3)	2265	0.05	0.33	0	3
CESD summary score (0-8)	1865	2.11	1.75	0	8
BMI	2222	23.52	2.96	13	41
KLoSA					
Word Recall Summary Score[*4]	7231	4.48	1.57	0	6
Serial 7's Score	7231	3.57	1.76	0	5
Poor health	7649	0.24	0.43	0	1
ADL summary score (0-3)	7649	0.10	0.49	0	3
CESD summary score (0-7)	7596	2.64	1.95	0	7
BMI	7458	23.20	2.81	12	37

[*1]: No Serial 7's Score in ELSA.

[*2]: We use BMI in Wave 4 ELSA because Wave 5 ELSA does not include BMI.

[*3]: Calculated using weight.

[*4]: KLoSA's Word Recall Scores are not comparable with other dataset.

Table 6: Summary Statistics: The US (Age 50 or older) at 2010

	Obs.	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max
	Male					**Female**				
Word Recall Summary Score	8291	9.07	3.31	0	20	11390	10.01	3.42	0	20
Serial 7's Score	8291	3.66	1.57	0	5	11390	3.22	1.74	0	5
Poor health	8993	0.08	0.28	0	1	12036	0.09	0.29	0	1
ADL summary score (0-3)	8929	0.22	0.61	0	3	11963	0.27	0.70	0	3
CESD summary score (0-8)	8202	1.30	1.88	0	8	11278	1.67	2.12	0	8
BMI	8904	28.42	5.27	7	61	11741	28.49	6.75	9	79
	Not Univ. Graduate					**Univ. Graduate**				
Word Recall Summary Score	15286	9.18	3.32	0	20	4391	11.12	3.29	0	20
Serial 7's Score	15286	3.17	1.73	0	5	4391	4.21	1.18	0	5
Poor health	16441	0.10	0.30	0	1	4584	0.03	0.18	0	1
ADL summary score (0-3)	16332	0.29	0.70	0	3	4556	0.13	0.49	0	3
CESD summary score (0-8)	15116	1.67	2.10	0	8	4360	0.96	1.63	0	8
BMI	16103	28.69	6.30	7	79	4538	27.65	5.53	12	61
	White					**Blue**				
Word Recall Summary Score	8634	10.16	3.43	0	20	3187	8.52	3.27	0	20
Serial 7's Score	8634	3.65	1.59	0	5	3187	3.14	1.74	0	5
Poor health	9095	0.06	0.24	0	1	3528	0.10	0.30	0	1
ADL summary score (0-3)	9082	0.20	0.61	0	3	3528	0.27	0.68	0	3
CESD summary score (0-8)	8560	1.26	1.87	0	8	3147	1.49	1.98	0	8
BMI	8993	28.12	5.92	7	72	3491	28.57	5.68	11	59

Table 7: The datasets which we use in each section

	Wave	Year
Section 5 (The Harmonized Analysis)		
The HRS	3-11	1996-2011
The SHARE[*1]	1,2,4,5	2004-2006, 2010-2012
The ELSA	1-6	2002-2014
The JSTAR	1-4	2007-2013
The KLoSA	1-4	2006-2012

[*1]: We analyze only Denmark, France, Germany, and Switzerland.

4 Critical Literature Assessment

4.1 Targeted Literature

Our goal is to explain why the estimated results of the effect of retirement on health in previous studies differ. We investigate the difference by focusing on the research framework. First, we create pairs of related studies for each health index, based on the following criteria:

- We choose papers from Tables 1, 2, 3, and 4.
- We can replicate them by using the HRS, related studies (the Global Aging Data), and the HSE.
- We choose only published papers in Health Economics or Labor Economics.
- We choose published papers in journals with higher impact factor as much as possible.
- We choose only published papers that estimate a linear model to analyze the effect of retirement on health.

Based on these criteria, we choose the studies in Table 8, which we use in the next sections. In the subsequent section, we explain how we analyze why the effect of retirement on health differs.

Table 8: The Targeted Literature

	(1)Cognition	
	Bonsang et al. (2012)	**Coe and Zamarro (2011)**
Impact	Negative	No
Survey countries	United States	European countries
Dataset	HRS	SHARE
Index	Word recall	Word recall, Verbal fluency
Method	FE-IV	IV
Definition of Retirement	Retired for at least one year	Not working for pay
Control variables*1	Only Age variables	B, E, L, H

	(2)Self-report of health	
	Dave et al. (2008)	**Coe and Zamarro (2011)**
Impact	Negative	Positive
Survey countries	United States	European countries
Dataset	HRS	SHARE
Method	FE	IV
Definition of Retirement	Reporting retired and not working	Not working for pay
Control variables*1	B, E	B, E, L

	(3)Depression	
	Dave et al. (2008)	**Coe and Zamarro (2011)**
Impact	Negative	No
Survey countries	United States	European countries
Dataset	HRS	SHARE
Index	CESD	EUROD
Method	FE	IV
Definition of Retirement	Reporting retired and not working	Not working for pay
Control variables*1	B, E	B, E, L

	(4)ADL	
	Dave et al. (2008)	**Neuman (2008)**
Impact	Negative	No (Male)/Positive(Female)
Survey countries	United States	United States
Dataset	HRS	HRS
Method	FE	IV
Definition of Retirement	Reporting retired and not working	Work less than 1200 h per year
Control variables*1	B, E	B, E, H

	(5)Obesity	
	Johnston and Lee (2009)	**Godard (2016)**
Impact	No	Negative
Survey countries	England	European countries
Dataset	HSE	SHARE
Index	BMI	BMI$\geq$30
Method	RDD	FEIV
Definition of Retirement	Reporting retired	Reporting retired
Control variables*1	No	B

*1 B:Basic variables(Ex:Age, education), E:Economic variables(Ex:Income), L:Labor force status(Ex:Self-employed), H:Health variables(Ex:Number of ADLs).

4.2 Review 1

Having chosen the targeted studies, we first analyze the effect of the difference in each factor on the final results. Each study consists of certain factors, such as surveyed country, analysis method, retirement definition, etc. (see Table 8). These studies use various identification strategies, analysis methods, and definitions of retirement. As such, we analyze why the estimated results of the effect of retirement on heath in previous studies differ by focusing on the differences in these factors. In each pair of studies, we first replace only one factor (e.g., the estimation method). Finally, we replace all the factors, one by one, in the paired studies, as shown in Figure 1. By replacing each factor, we analyze the effect of each factor on the difference in the final results. There are five characteristics in each study: "index," "def. of retire," "controls," "method," "sample," and "survey country." The differences in these characteristics explain the different results on the effect of retirement on health. The details of these characteristics are as follows.

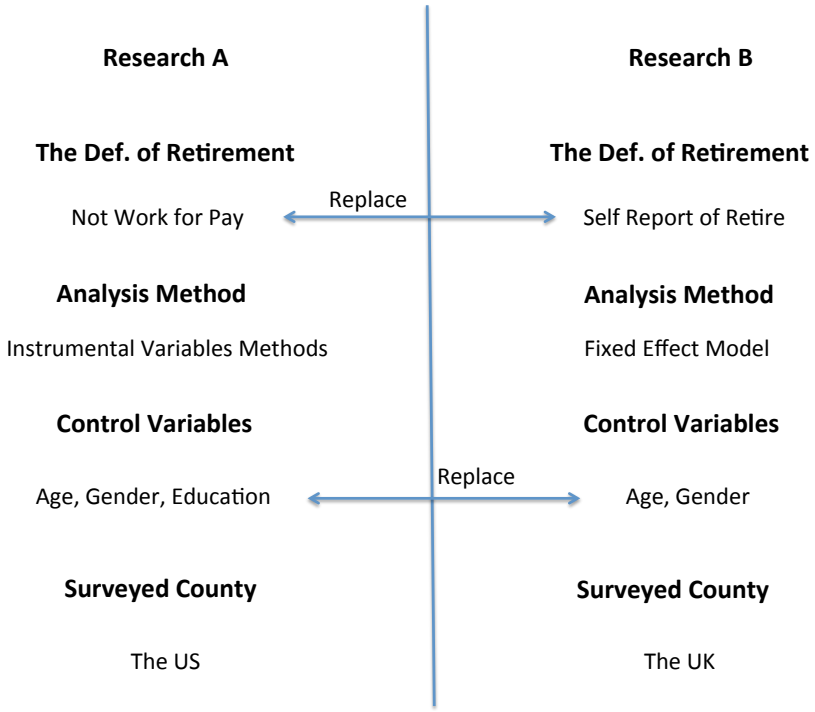
- Index: characteristics of the index used (e.g., CES-D versus EURO-D);
- Def. of retire: definition of retirement (e.g., retired for at least one year versus not working for pay);
- Controls: What the researchers include as control variables (e.g., only family structure variables versus family structure variables + economic variables);
- Method: analysis method (e.g., FE methods versus IV methods);
- Sample: sample selection method (e.g. only male versus full sample);
- Survey country: surveyed country (e.g., the USA versus France).

Here, we summarize our results.

- The sensitivity of replacing the definition of retirement is not strong.
- The sensitivity of replacing the analysis method is not weak. In almost all indexes, the estimated results change when replacing the analysis method.
- The sensitivity of replacing the surveyed country is also significant.
- The difference in the estimated results cannot be explained by only one-factor replacement.

In this section, by replacing only one factor, we have checked the sensitivity of each factor on the estimated results. According to our results, it is difficult to explain why the estimated results are different by replacing only one factor. In the next section, we provide another framework to explain why the estimated results in the previous studies differ.

Figure 1: Review 1



In the Appendix (A.3), we summarize the replication and replacement notes in this section. When we replicate and replace the analysis of related literature, we make some adjustments if needed (see section A.3 for details).

Cognitive score (Bonsang et al. (2012) versus Coe and Zamarro (2011)):

- According to Figure 9, when transplanting one factor from Bonsang et al. (2012) to Coe and Zamarro (2011), the replacement of the surveyed country yields the opposite results (negative-positive) and vice-versa. However, the sensitivity of replacing the control variables, the surveyed country, and the method are important.

Self-report of health (Dave et al. (2008) versus Coe and Zamarro (2011)):

- According to Figure 10, when transplanting one factor from Dave et al. (2008) to Coe and Zamarro (2011), the replacement of the analysis method and the surveyed country change from a negative effect to no effect and vice-versa. The sensitivity of replacing the index, the analysis method, the sample selection method, and the surveyed country are important.

ADL (Dave et al. (2008) versus Neuman (2008)):

- We discuss Figure 11. Transplanting one factor from Dave et al. (2008) to Neuman (2008), the replacement of the estimation method and the sample selection method change from a

Table 9: Cognitive score

	Bonsang et al. (2012)		**Coe and Zamarro (2011)**
Estimated result in the paper	-0.942***		-0.0390
Def. of retire		→	-1.244***
Controls	Our replication result	→	-1.189***
Method	-1.036***	→	-1.444***
Sample		→	-1.266*
Survey country		→	23.672**

	Coe and Zamarro (2011)		**Bonsang et al. (2012)**
Estimated result in the paper	-0.0390		-0.942***
Def. of retire		→	1.064
Controls	Our replication result	→	-3.248***
Method	0.995	→	6.468***
Sample		→	-0.035
Survey country		→	-2.649**

*[1] The red (blue) character indicates the positive (negative) impact.

negative effect to no effect, while replacing other factors does not produce such a difference, and vice-versa. This time, the replacement of each factor, except the definition of retirement, produces a change in the results, while the change in the estimation method produces the opposite result for female samples.

Depression (Dave et al. (2008) versus Coe and Zamarro (2011)):

- We discuss Figure 12. Transplanting one factor from Dave et al. (2008) to Coe and Zamarro (2011), the replacement of the estimation method and the surveyed country, [8] change from a negative effect to no effect, while replacing other factors does not produce such a difference, and vice-versa. This time, the replacement of each factor, except the control variables, produces a change in the results.

BMI (Godard (2016) versus Johnston and Lee (2009)):

- We discuss Figure 13. Transplanting one factor from Godard (2016) to Johnston and Lee (2009), the replacement of all factors except the definition of retirement and the control variables change from a negative effect to no effect, while replacing other factors does not produce such a difference, and vice-versa. This time, the replacement of each factor does not produce a change in the results.

4.3 Review 2

In the previous section, we have discussed the sensitivity of each factor on the estimated results. We have also found that there are multiple factors that explain why the estimated results are different. In this section, we propose another framework to explain why the estimated results are different. As such, we start from one study and arrive at another study, replacing factors one by one

[8] We also change the index of depression (from CES-D to EURO-D) when we change the surveyed country.

Table 10: Self-report of health

	Dave et al. (2008)		**Coe and Zamarro (2011)**
Estimated result in the paper	0.0268***		-0.3545**
Def. of retire		→	0.023***
Controls	Our replication result	→	0.025***
Method	0.025***	→	0.02
Sample		→	0.027***
Survey country		→	0.009

	Coe and Zamarro (2011)		**Dave et al. (2008)**
Estimated result in the paper	-0.3545**		0.0268***
Index		→	-0.011
Def. of retire		→	-0.187***
Controls	Our replication result	→	-0.234***
Method	-0.241*	→	-0.001
Sample		→	-0.226
Survey country		→	-0.123(Poor health)

*1 The red (blue) character indicates the positive (negative) impact.

Table 11: ADL

	Dave et al. (2008)		**Neuman (2008)**
Estimated result in the paper	0.0267***		-0.025(M)/0.101**(F)
Def. of retire		→	0.021***
Controls	Our replication result	→	0.029***
Method	0.043***	→	0.142
Sample		→	0.003(M)/0.004(F)

	Neuman (2008)		**Dave et al. (2008)**
Estimated result in the paper	-0.025(M)/0.101**(F)		0.0267***
Def. of retire		→	-0.03(M)/0.219***(F)
Controls	Our replication result	→	0.014(M)/0.082(F)
Method	-0.013(M)/0.211**(F)	→	0.029***(M)/0.042***(F)
Sample		→	0.01

*1 The red (blue) character indicates the positive (negative) impact.

Table 12: Depression

	Dave et al. (2008)		**Coe and Zamarro (2011)**
Estimated result in the paper	0.1157***		-0.0691
Def. of retire		→	0.165***
Controls	Our replication result	→	0.109***
Method	0.116***	→	-0.132
Sample		→	0.143***
Survey country		→	0.042(EURO-D)

	Coe and Zamarro (2011)		**Dave et al. (2008)**
Estimated result in the paper	-0.0691		0.1157***
Index		→	-0.141
Def. of retire		→	0.404
Controls	Our replication result	→	2.605***
Method	0.534	→	-0.049
Sample		→	1.009
Survey country		→	-0.195

*1 The red (blue) character indicates the positive (negative) impact.

Table 13: BMI

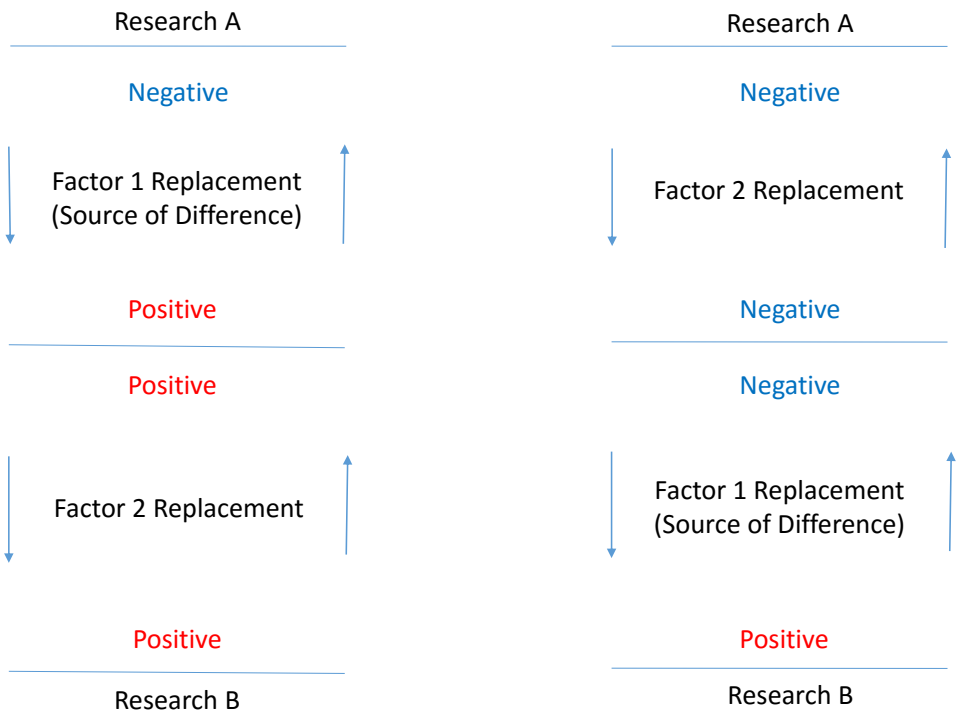
	Godard (2016)		**Johnston and Lee (2009)**
Estimated result in the paper	0.115**		0.092
Index		→	0.371
Def. of retire		→	0.122**
Controls	Our replication result	→	0.077***
Method	0.122**	→	0.077
Sample		→	0.072
Survey country		→	-0.386

	Johnston and Lee (2009)		**Godard (2016)**
Estimated result in the paper	0.092		0.115**
Index		→	-0.018
Def. of retire		→	0.118
Controls	Our replication result	→	-0.798
Method	0.118	→	0.728
Sample		→	0.235
Survey country		→	0.291

*1 The red (blue) character indicates the positive (negative) impact.

(see Figure 2). If the source of the difference in the effect of retirement on health exists, the result will change after we change this source as per Figure 2. We discuss the results in the following.

Figure 2: Review 2



Cognitive score (Bonsang et al. (2012) versus Coe and Zamarro (2011)):

- In Figure 14, we combine method, controls, and country, as these are the factors producing the change in the results in Review 1. We consider that these factors are important for explaining the difference in the effect of retirement on health between two different studies. The figure on the left shows the change in the estimation results when we change the order of replacing the block (method + controls + country). On the other hand, the right-hand figure shows the change in the estimation results when we change the order of replacing the retirement definition. We compare these cases as follows.
- In all patterns (A, B, C), we observed that the estimated results change after replacing the block (method + controls + country) (Negative $\rightarrow$ No)(left-hand figure). On the other hand, we do not observe any change just after replacing the definition of retirement (right-hand figure).

Self-report of health (Dave et al. (2008) versus Coe and Zamarro (2011)):

- In Figure 15, we show the same procedure as in Figure 14. The left-hand figure shows the change in the estimation results when we change the order of replacing the block (method

+ controls + country + index), as these factors (method + controls + country + index) produce the change in the results in Review 1. On the other hand, the right-hand figure shows the change in the estimation results when we change the order of replacing the retirement definition. We compare these cases as follows.

- In all patterns (A, B, C), we observed that the estimated results change after replacing the block (method + controls + country + index) (Negative → Positive)(left-hand figure). On the other hand, we do not observe any change just after replacing the definition of retirement except in pattern B (right-hand figure).

ADL (Dave et al. (2008) versus Neuman (2008)):

- In Figure 16, we show the same procedure as in Figure 14. The left-hand figure shows the change in the estimation results when we change the order of replacing the block (method + controls), as these factors (method + controls) produce the change in the results in Review 1. On the other hand, the right-hand figure shows the change in the estimation results when we change the order of replacing the retirement definition. We compare these cases as follows.

- In all patterns, changing both the estimation method and the difference in what the researcher uses as control variables produce a change in the results. In particular, in pattern C (left-hand figure), the change in method + controls produces the opposite impact for female samples. In patterns A and B, "sample" is also significant. The estimated results changes just after replacing "sample" (No → No (male) and Positive (female))(left-hand figure). As such, the definition of retirement seems to have no impact on the results (right-hand figure).

Depression (Dave et al. (2008) versus Coe and Zamarro (2011)):

- In Figure 17, we show the same procedure as in Figure 14. The left-hand figure in Figure 17 shows the change in the estimation results when we change the order of replacing the block (method + controls), as these factors (method + controls) produce the change in the results in Review 1. On the other hand, the right-hand figure shows the change in the estimation results when we change the order of replacing the retirement definition. We compare these cases as follows.

- In all patterns (A, B, C), we observe that the estimated results change after replacing the block (method + controls) (Negative → No). In pattern D, "country + index" is also significant. The estimated results changes just after replacing "country + index" (Negative → No)(left-hand figure). On the other hand, we do not observe any change just after replacing the retirement definition (right-hand figure).

BMI (Johnston and Lee (2009) versus Godard (2016)):

- In Figure 18, we show the same procedure as in Figure 14. The left-hand figure in Figure 18 shows the change in the estimation results when we change the order of replacing the block (method + controls + sample), as these factors (method + sample) produce the change in the results in Review 1. On the other hand, the right-hand figure shows the change in the estimation results when we change the order of replacing the index. There is no difference in the definition of retirement between Johnston and Lee (2009) and Godard (2016). Here, we replace the index, and compare these cases as follows.

- In all patterns (A, B), we observe that the estimated results change after replacing the block (method + controls + sample) (Negative → No). In patterns C and D, "country" is also significant. The estimated results changes just after replacing "country" (Negative → No)(left-hand figure). On the other hand, we do not observe any change just after replacing the index except for pattern A (right-hand figure).

Finally, we summarize our results.

- The choice of the estimation method seems to be the key factor for explaining the difference in the estimation results in all indexes. Additionally, the use of control variables is also important. What the researcher uses as control variables is also included in all health indexes. In all health indexes, the estimation method plus other factors (e.g., method + controls) changes in the estimation result.
- The influence of the difference in the surveyed country is also important for explaining the difference in the effect of retirement on health.
- Changes in the definition of retirement have a lower impact.

According to our results, the difference in the estimation method is a key factor in explaining why the estimated effects of retirement on health in preceding studies differ. It is intuitive that the sensitivity of the surveyed country chosen is strong. However, we do not consider this as problematic. On the other hand, a strong sensitivity of the analysis method choice is problematic because it is possible that we do not appropriately estimate the effect of retirement on health, depending on the choice of the analysis method. In some studies, it is possible that there remains room for further improvement. For example, Coe and Zamarro (2011) estimate the effect of retirement on cognitive function by using cross-sectional data. They use the exogenous variation of the pensionable age as an IV, the SHARE being their data source. As such, we can use a dynamic variation of the retirement behavior in the SHARE. Dave et al. (2008) only use FE and do not use an IV. Consequently, we can use the FE-IV method, often used in recent studies to estimate the effect of retirement on health indexes. [9]

[9] Bonsang et al. (2012), Insler (2014), and Godard (2016).

Table 14: Cognitive score

Pattern A	Pattern B	Pattern A	Pattern B
Bonsang et al. (2012)	**Bonsang et al. (2012)**	**Bonsang et al. (2012)**	**Bonsang et al. (2012)**
-0.942***	-0.942***	-0.942***	-0.942***
-1.036***(Replication)	-1.036***(Replication)	-1.036***(Replication)	-1.036***(Replication)
↓ Method + Controls + Country ↑	↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Method + Controls + Country ↑
-0.216	-1.244***	-1.244***	-0.216
↓ Def. of Retirement ↑	↓ Method + Controls + Country ↑	↓ Method + Controls + Country ↑	↓ Def. of Retirement ↑
-0.214	-0.214	-0.214	-0.214
↓ Sample ↑	↓ Sample ↑	↓ Sample ↑	↓ Sample ↑
0.995(Replication)	0.995(Replication)	0.995(Replication)	0.995(Replication)
-0.0390	-0.0390	-0.0390	-0.0390
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**

Pattern C	Pattern C
Bonsang et al. (2012)	**Bonsang et al. (2012)**
-0.942***	-0.942***
-1.036***(Replication)	-1.036***(Replication)
↓ Def. of Retirement ↑	↓ Method + Controls + Country ↑
-1.244***	-0.216
↓ Sample ↑	↓ Sample ↑
-1.825*	0.981
↓ Method + Controls + Country ↑	↓ Def. of Retirement ↑
0.995(Replication)	0.995(Replication)
-0.0390	-0.0390
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**

Table 15: Self-report of health

Pattern A	Pattern B	Pattern A	Pattern B
Dave et al. (2008)	**Dave et al. (2008)**	**Dave et al. (2008)**	**Dave et al. (2008)**
0.0268***	0.0268***	0.0268***	0.0268***
0.025***(Replication)	0.025***(Replication)	0.025*** (Replication)	0.025*** (Replication)
↓ Method + Controls + Country + Index ↑	↓ Sample ↑	↓ Def. of Retirement ↑	↓ Method + Controls + Country + Index ↑
-0.276*	0.027***	0.023***	-0.276*
↓ Sample ↑	↓ Method + Controls + Country + Index ↑	↓ Method + Controls + Country + Index ↑	↓ Def. of Retirement ↑
-0.183*	-0.183*	-0.226	-0.226
↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Sample ↑	↓ Sample ↑
-0.241*(Replication)	-0.241*(Replication)	-0.241*(Replication)	-0.241*(Replication)
-0.3545**	-0.3545**	-0.3545**	-0.3545**
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**

Pattern C	Pattern C
Dave et al. (2008)	**Dave et al. (2008)**
0.0268***	0.0268***
0.025***(Replication)	0.025***(Replication)
↓ Sample ↑	↓ Method + Controls + Country + Index ↑
0.027***	-0.276*
↓ Def. of Retirement ↑	↓ Sample ↑
0.051***	-0.187**
↓ Method + Controls + Country + Index ↑	↓ Def. of Retirement ↑
-0.241*(Replication)	-0.241*(Replication)
-0.3545**	-0.3545**
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**

Table 16: ADL

Pattern A	**Pattern B**	**Pattern A**	**Pattern B**
Dave et al. (2008)	**Dave et al. (2008)**	**Dave et al. (2008)**	**Dave et al. (2008)**
0.0267***	0.0267***	0.0267***	0.0267***
0.043***(Replication)	0.043***(Replication)	0.043***(Replication)	0.043***(Replication)
↓ Method + Controls ↑	↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Method + Controls ↑
-0.01	0.021***	0.021***	-0.01
↓ Def. of Retirement ↑	↓ Method + Controls ↑	↓ Method + Controls ↑	↓ Def. of Retirement ↑
0.01	0.01	0.01	0.01
↓ Sample ↑	↓ Sample ↑	↓ Sample ↑	↓ Sample ↑
-0.013(M)/0.211***(F)	-0.013(M)/0.211***(F)	-0.013(M)/0.211***(F)(Replication)	-0.013(M)/0.211***(F)(Replication)
-0.025(M)/0.101**(F)(Replication)	-0.025(M)/0.101**(F)(Replication)	-0.025(M)/0.101**(F)	-0.025(M)/0.101**(F)
Neuman (2008)	**Neuman (2008)**	**Neuman (2008)**	**Neuman (2008)**

Pattern C	**Pattern C**
Dave et al. (2008)	**Dave et al. (2008)**
0.0267***	0.0267***
0.043***	0.043*** (Replication)
↓ Def. of Retirement ↑	↓ Method + Controls ↑
0.021***	-0.01
↓ Sample ↑	↓ Sample ↑
0.062***(M)/0.084***(F)	-0.03(M)/0.219***(F)
↓ Method + Controls ↑	↓ Def. of Retirement ↑
-0.013(M)/0.211***(F)	-0.013(M)/0.211***(F)(Replication)
-0.025(M)/0.101**(F)(Replication)	-0.025(M)/0.101**(F)
Neuman (2008)	**Neuman (2008)**

Table 17: Depression

Pattern A	**Pattern B**	**Pattern A**	**Pattern B**
Dave et al. (2008)	**Dave et al. (2008)**	**Dave et al. (2008)**	**Dave et al. (2008)**
0.1157***	0.1157***	0.1157***	0.1157***
0.116***(Replication)	0.116***(Replication)	0.116***(Replication)	0.116***(Replication)
↓ Method + Controls ↑	↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Method + Controls ↑
0.274	0.165***	0.165***	0.274
↓ Def. of Retirement ↑	↓ Method + Controls ↑	↓ Method + Controls ↑	↓ Def. of Retirement ↑
0.282	0.282	0.282	0.282
↓ Sample ↑	↓ Sample ↑	↓ Sample ↑	↓ Sample ↑
-0.227	-0.227	-0.227	-0.227
↓ Country + Index ↑	↓ Country + Index ↑	↓ Country + Index ↑	↓ Country + Index ↑
0.534(Replication)	0.534(Replication)	0.534(Replication)	0.534(Replication)
-0.0691	-0.0691	-0.0691	-0.0691
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**

Pattern C	**Pattern D**	**Pattern C**	**Pattern D**
Dave et al. (2008)	**Dave et al. (2008)**	**Dave et al. (2008)**	**Dave et al. (2008)**
0.1157***	0.1157***	0.1157***	0.1157***
0.116***(Replication)	0.116***(Replication)	0.116***(Replication)	0.116***(Replication)
↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Method + Controls ↑	↓ Method + Controls ↑
0.165***	0.165***	0.274	0.274
↓ Sample ↑	↓ Sample ↑	↓ Sample ↑	↓ Sample ↑
0.259***	0.259***	-0.285	-0.285
↓ Method + Controls ↑	↓ Country + Index ↑	↓ Def. of Retirement ↑	↓ Country + Index ↑
-0.227	0.046	-0.227	0.374
↓ Country + Index ↑	↓ Method + Controls ↑	↓ Country + Index ↑	↓ Def. of Retirement ↑
0.534(Replication)	0.534(Replication)	0.534(Replication)	0.534(Replication)
-0.0691	-0.0691	-0.0691	-0.0691
Coe and Zamarro (2011)	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**	**Coe and Zamarro (2011)**

Table 18: BMI

Pattern A	Pattern B	Pattern A	Pattern B
Godard (2016)	**Godard (2016)**	**Godard (2016)**	**Godard (2016)**
0.115**	0.115**	0.115**	0.115**
0.122**(Replication)	0.122**(Replication)	0.122**(Replication)	0.122**(Replication)
↓ Method + Controls + Sample ↑	↓ Def. of Retirement ↑	↓ Index ↑	↓ Method + Controls + Sample ↑
0.002	0.122**	0.371	0.002
↓ Def. of Retirement ↑	↓ Method + Controls + Sample ↑	↓ Method + Controls + Sample ↑	↓ Index ↑
0.002	0.002	0.291	0.291
↓ Country ↑	↓ Country ↑	↓ Def. of Retirement ↑	↓ Def. of Retirement ↑
-0.018	-0.018	0.291	0.291
↓ Index ↑	↓ Index ↑	↓ Country ↑	↓ Country ↑
0.118(Replication)	0.118(Replication)	0.118(Replication)	0.118(Replication)
0.092	0.092	0.092	0.092
Johnston and Lee (2009)	**Johnston and Lee (2009)**	**Johnston and Lee (2009)**	**Johnston and Lee (2009)**

Pattern C	Pattern D	Pattern C	Pattern D
Godard (2016)	**Godard (2016)**	**Godard (2016)**	**Godard (2016)**
0.115**	0.115**	0.115**	0.115**
0.122**(Replication)	0.122**(Replication)	0.122**(Replication)	0.122**(Replication)
↓ Def. of Retirement ↑	↓ Def. of Retirement ↑	↓ Method + Controls + Sample ↑	↓ Method + Controls + Sample ↑
0.122**	0.122**	0.002	0.002
↓ Country ↑	↓ Country ↑	↓ Def. of Retirement ↑	↓ Def. of Retirement ↑
-0.386	-0.386	0.002	0.002
↓ Method + Controls + Sample ↑	↓ Index ↑	↓ Index ↑	↓ Country ↑
-0.018	-2.057	0.291	-0.018
↓ Index ↑	↓ Method + Controls + Sample ↑	↓ Country ↑	↓ Index ↑
0.118(Replication)	0.118(Replication)	0.118(Replication)	0.118(Replication)
0.092	0.092	0.092	0.092
Johnston and Lee (2009)	**Johnston and Lee (2009)**	**Johnston and Lee (2009)**	**Johnston and Lee (2009)**

5 The Harmonized Analysis on the Effect of Retirement on Health

5.1 Analysis Framework

Here, we use the FE-IV estimation method and estimate the impact of retirement on some health indexes for eight countries. Coe and Zamarro (2011) estimate the effect of retirement on cognitive function by using cross-sectional data, and use the cross-country variation of the pensionable age to control for retirement endogeneity, using SHARE. However, we use a dynamic variation of the retirement behavior, and control for retirement endogeneity by using the pensionable age in the surveyed countries. We also estimate the effect of retirement on health indexes in each country. While Dave et al. (2008) only use the FE, while we use the FE-IV method to estimate the effect of retirement on health indexes as follows:

$$health\_index_{it} = \beta_0 + \beta_1 retire_{it} + \gamma' x_{it} + a_{1i} + \lambda_{1t} + \epsilon_{1it} \quad (1)$$

$$\begin{aligned} retire_{it} &= \alpha_0 + \alpha_1 1\{age_{it} \geq A_i^{eb}\} + \alpha_2 1\{age_{it} \geq A_i^{fb}\} \\ &+ \alpha_1 1\{age_{it} \geq A_i^{eb}\} \cdot age_{it} + \alpha_2 1\{age_{it} \geq A_i^{fb}\} \cdot age_{it} + \eta' x_{it} + a_{2i} + \lambda_{2t} + \epsilon_{2i} \end{aligned} \quad (2)$$

A_i^{eb}: the early retirement benefit eligibility age

A_i^{fb}: the full retirement benefit eligibility age

Where $retire_{it}$ is an indicator which is equal to 1 when a respondent retires at period t. We use two retirement definitions. The first is "not work for pay," which means that a respondent retires if not

working for payment. The second definition is "complete retire," which is the same the retirement definition of Dave et al. (2008). λ_{1t} and λ_{2t} are time FE. a_{1i} and a_{2i} are individual FE. x_{it} are control variables at period t.

5.2 The Results

We discuss the estimated results only when the coefficients of IV in the first stage are significant. We also test the endogeneity of retirement with the Durbin-Wu-Hausman test. When we do not reject the null hypothesis, we support the results of FE model. We analyze only countries whose pensionable ages are confirmed to be correct, and discuss how to confirm each pensionable age in Appendix (A.1). We use the retirement definition of "not work for pay" in all countries except Korea and Japan. On the other hand, in Korea and Japan, we use the retirement definition of "complete retire." This is because, in Korea and Japan, we do not get the significant result in the first stage regression when we use the retirement definition of "not work for pay." We perform a robustness check with respect to the retirement definition in the next section.

- As per Table 19, in each health index, only Korea has the opposite effect compared to the USA. With respect to self-reported health and CES-D, in many countries, we observe a positive effect of retirement on health. However, only in Korea and the USA there is a significant effect on cognitive function. Nonetheless, there is an opposite effect (positive effect and negative effect) between these countries.
- As Table 20, there is a negative effect or no effect of retirement on BMI. However, in many countries, there is a positive effect of retirement on ADL.

Table 19: FEIV estimation 1

Poor health	US		England		Denmark		France		Germany		Switzerland		Japan[*1]		Korea[*1]	
Full sample																
FE	0.034***	DWH p-val	0.020***	DWH p-val	0.003	DWH p-val	-0.007	DWH p-val	-0.013	DWH p-val	0.004	DWH p-val			0.063***	DWH p-val
FE-IV	-0.138***	0.000	-0.097***	0.000	-0.054	0.294	-0.158***	0.016	-0.140*	0.080	0.003	0.997			0.071	0.996
Male																
FE	0.041***	DWH p-val	0.026***	DWH p-val	-0.009	DWH p-val	-0.001	DWH p-val	0.003	DWH p-val	-0.005	DWH p-val	-0.037***	DWH p-val	0.072***	DWH p-val
FE-IV	-0.119***	0.000	-0.061*	0.010	-0.013	0.962	-0.070	0.674	-0.056	0.585	-0.073	0.352	0.149	0.035	-0.060	0.362
Female																
FE	0.029***	DWH p-val	0.014***	DWH p-val	0.017	DWH p-val	-0.011	DWH p-val	-0.025	DWH p-val	0.010	DWH p-val			0.056***	DWH p-val
FE-IV	-0.159***	0.000	-0.154***	0.000	-0.080	0.144	-0.206***	0.015	-0.196*	0.070	0.011	0.939			0.409	0.623

CESD summary (0-8)	US		England		Denmark		France		Germany		Switzerland		Japan[*1]		Korea[*1*2]	
Full sample																
FE	0.193***	DWH p-val	0.084***	DWH p-val	-0.121	DWH p-val	0.020	DWH p-val	-0.035	DWH p-val	-0.064	DWH p-val			0.045	DWH p-val
FE-IV	-1.153***	0.000	-0.501***	0.000	-1.334***	0.005	-0.040	0.887	-0.194	0.731	-0.099	0.954			1.155	0.252
Male																
FE	0.194***	DWH p-val	0.043	DWH p-val	-0.158	DWH p-val	0.137	DWH p-val	0.047	DWH p-val	-0.004	DWH p-val	0.114	DWH p-val	0.105*	DWH p-val
FE-IV	-0.865***	0.000	-0.586***	0.001	-1.360**	0.075	0.881	0.224	-0.188	0.747	0.021	0.910	-2.795**	0.010	1.185	0.190
Female																
FE	0.189***	DWH p-val	0.116***	DWH p-val	-0.067	DWH p-val	-0.083	DWH p-val	-0.083	DWH p-val	-0.103	DWH p-val			-0.007	DWH p-val
FE-IV	-1.308***	0.000	-0.406*	0.027	-1.265**	0.027	-0.793	0.245	-0.746	0.376	-0.008	0.961			0.688	0.803

Word Recall	US		England		Denmark		France		Germany		Switzerland				Korea[*1*3]	
Full sample																
FE	-0.102***	DWH p-val	0.039	DWH p-val	-0.014	DWH p-val	0.140	DWH p-val	-0.176	DWH p-val	0.091	DWH p-val			0.037	DWH p-val
FE-IV	-0.091	0.936	0.358	0.236	0.635	0.373	-0.356	0.445	0.983	0.194	1.359	0.169			1.895**	0.007
Male																
FE	-0.092**	DWH p-val	-0.010	DWH p-val	-0.115	DWH p-val	-0.056	DWH p-val	-0.354	DWH p-val	-0.176	DWH p-val			0.046	DWH p-val
FE-IV	-0.781**	0.072	0.102	0.817	0.583	0.583	-0.478	0.647	1.002	0.221	1.166	0.403			1.316**	0.030
Female																
FE	-0.122***	DWH p-val	0.081	DWH p-val	0.060	DWH p-val	0.310	DWH p-val	-0.043	DWH p-val	0.269	DWH p-val			0.028	DWH p-val
FE-IV	0.354	0.164	0.255	0.644	0.688	0.464	-0.284	0.494	1.724	0.151	1.248	0.388			2.573	0.222

Standard errors in parentheses

All specifications include age, age squared, marital status(married or not), number of children, household income, housing(have own house or not), household total wealth, region dummy, and wave dummy.

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

The red (blue) character indicates the positive (negative) impact.

*1 The definition of retirement is "Complete retirement".

*2 In Korea, we use different scale of CESD summary score(0-7).

*3 In Korea, we use different scale of Word Recall summary score(0-6).

Table 20: FEIV estimation 2

BMI	US		England		Denmark		France		Germany		Switzerland		Japan*1		Korea*1	
Full sample																
FE	0.115***	DWH p-val	0.124**	DWH p-val	-0.035	DWH p-val	0.136	DWH p-val	-0.048	DWH p-val	0.072	DWH p-val			0.016	DWH p-val
FE-IV	1.406***	0.000	0.179	0.840	0.121	0.730	-0.056	0.645	-0.319	0.707	0.776	0.245			0.532	0.612
Male																
FE	0.079**	DWH p-val	0.176**	DWH p-val	0.037	DWH p-val	0.151	DWH p-val	-0.035	DWH p-val	0.080	DWH p-val	0.157**	DWH p-val	-0.092	DWH p-val
FE-IV	1.419***	0.000	0.880**	0.065	1.007	0.187	-0.027	0.941	-0.060	0.842	0.785	0.375	2.107**	0.011	-0.504	0.535
Female																
FE	0.153***	DWH p-val	0.069	DWH p-val	-0.108	DWH p-val	0.124	DWH p-val	-0.090	DWH p-val	0.076	DWH p-val			0.088*	DWH p-val
FE-IV	1.524***	0.000	-0.833	0.081	-0.686	0.406	-0.180	0.622	-0.917	0.296	0.796	0.438			4.777	0.179

BMI≥30	US		England		Denmark		France		Germany		Switzerland		Japan*1		Korea*1	
Full sample																
FE	0.006**	DWH p-val	-0.011	DWH p-val	0.007	DWH p-val	-0.007	DWH p-val	-0.013	DWH p-val	0.013	DWH p-val			0.006**	DWH p-val
FE-IV	0.069***	0.013	-0.004	0.881	0.064	0.382	0.005	0.652	0.026	0.628	0.121*	0.106			-0.074	0.257
Male																
FE	0.003	DWH p-val	-0.008	DWH p-val	0.022	DWH p-val	-0.015	DWH p-val	-0.015	DWH p-val	0.001	DWH p-val	0.002	DWH p-val	-0.001	DWH p-val
FE-IV	0.039	0.339	0.062	0.291	0.098	0.480	0.062	0.365	-0.068	0.653	0.143	0.343	0.040	0.652	-0.036	0.302
Female																
FE	0.009**	DWH p-val	-0.014	DWH p-val	-0.006	DWH p-val	0.000	DWH p-val	-0.012	DWH p-val	0.021*	DWH p-val			0.010**	DWH p-val
FE-IV	0.094***	0.016	-0.067	0.431	-0.004	0.969	-0.040	0.934	0.036	0.648	0.088	0.365			-0.037	0.865

ADL summary (0-5)	US		England		Denmark		France		Germany		Switzerland		Japan*1			
Full sample																
FE	0.066***	DWH p-val	0.054***	DWH p-val	0.036	DWH p-val	-0.022	DWH p-val	0.013	DWH p-val	-0.025	DWH p-val				
FE-IV	-0.804***	0.000	-0.284***	0.000	-0.093	0.144	0.064	0.436	-0.491***	0.003	-0.177	0.589				
Male																
FE	0.093***	DWH p-val	0.052***	DWH p-val	0.062	DWH p-val	-0.026	DWH p-val	-0.012	DWH p-val	-0.054*	DWH p-val	-0.064*	DWH p-val		
FE-IV	-0.535***	0.000	-0.349***	0.000	-0.085	0.369	0.143	0.679	-0.543**	0.013	-0.409	0.294	-0.551**	0.022		
Female																
FE	0.047***	DWH p-val	0.055***	DWH p-val	0.014	DWH p-val	-0.019	DWH p-val	0.036	DWH p-val	-0.005	DWH p-val				
FE-IV	-0.943***	0.000	-0.232***	0.000	0.007	0.539	0.000	0.129	-0.458**	0.019	-0.115	0.710				

Standard errors in parentheses

All specifications include age, age squared, marital status(married or not), number of children, household income, housing(have own house or not), household total wealth, region dummy, and wave dummy.

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

The red (blue) character indicates the positive (negative) impact.

*1 The definition of retirement is "Complete retirement".

Table 21: Robustness check: ADL (0-3)(Left) and Poor health (Self-report of health)(Right)

ADL(0-3)	Def. of retire	Controls			
		Pattern 1	Pattern 2	Pattern 3	Pattern 4
United States	Not work	-0.493***	-0.484***	-0.473***	-0.484***
	Complete retire	-0.323***	-0.318***	-0.310***	-0.284***
England	Not work	-0.173***	-0.166***	-0.149***	-0.146***
	Complete retire	-0.102***	-0.098***	-0.090***	-0.088***
Denmark	Not work	-0.114	-0.112	-0.119	-0.104
	Complete retire	-0.008	-0.007	-0.127**	-0.120**
France	Not work	-0.009	-0.009	-0.017	-0.018
	Complete retire	-0.009	-0.01	-0.018	-0.019
Germany	Not work	-0.294**	-0.252**	-0.309**	-0.315**
	Complete retire	-0.144**	-0.026	-0.180**	-0.180**
Switzerland	Not work	-0.004	-0.005	-0.017	-0.018
	Complete retire	0.019	-0.097*	0.018	0.017

Pattern 1 includes age and age squared.
Pattern 2 includes age, age squared, married and number of children(basic variables).
Pattern 3 includes basic variables and, HH income, housing and HH total wealth(economic variables).
Pattern 4 includes basic variables, economic variables and, region dummy and wave dummy.
* $p < .1$, ** $p < .05$, *** $p < .01$
The red (blue) character indicates the positive (negative) impact.

Poor health	Def. of retire	Controls			
		Pattern 1	Pattern 2	Pattern 3	Pattern 4
United States	Not work	-0.107***	-0.110***	-0.107***	-0.138***
	Complete retire	-0.082***	-0.084***	-0.082***	-0.105***
England	Not work	-0.094***	-0.095***	-0.098***	-0.097***
	Complete retire	-0.059***	-0.060***	-0.062***	-0.061***
Denmark	Not work	0.01	0.008	0.002	0.003
	Complete retire	0.009	0.007	0.006	0.008
France	Not work	-0.135**	-0.130**	-0.149***	-0.158***
	Complete retire	-0.013	-0.011	-0.076**	-0.082**
Germany	Not work	-0.006	-0.129*	-0.136*	-0.140*
	Complete retire	-0.025*	-0.023*	-0.023*	-0.023
Switzerland	Not work	0.002	0.002	0.002	0.004
	Complete retire	0.006	0.006	0.005	0.007

Pattern 1 includes age and age squared.
Pattern 2 includes age, age squared, married and number of children(basic variables).
Pattern 3 includes basic variables and, HH income, housing and HH total wealth(economic variables).
Pattern 4 includes basic variables, economic variables and, region dummy and wave dummy.
* $p < .1$, ** $p < .05$, *** $p < .01$
The red (blue) character indicates the positive (negative) impact.

Table 22: International comparison of the effect of retirement on health

	US	England	Denmark	France	Germany	Switzerland	Japan	South Korea
Self-report of health	+	+		+	+			-
Depression	+	+	+				+	
Cognition	-							+
BMI	-	-					-	
ADL	+	+			+	+	+	

Subsequently, we will check the sensitivity of the retirement definition and the pattern of control variables on the effect of retirement on health. We prepare two retirement definitions ("not work for pay" and "complete retire") and four control patterns ("Pattern 1," etc.). According to Table 21, the estimates are robust although we change the retirement definition and control variable patterns for each country. The results are not significant for some countries, but there is no opposite effect.

5.3 Discussion

We summarize our main results in Table 22. Our analysis method (FE-IV) is established in this section. According to Table 22, when we fix our analysis method, we do not obtain opposite results (positive effect and negative effect). For each health index, we obtain positive (negative) or no effects of retirement on health in all surveyed countries. The important point is that there is a heterogeneity of the effect of retirement on health, even if we fix our methods and control for retirement endogeneity.

6 Conclusion

We summarize our results in this paper as follows.

- Review 1:
  - The sensitivity of replacing the definition of retirement is not strong.
  - The sensitivity of replacing the analysis method is not weak. In almost all indexes, the estimated results change when replacing the analysis method.

  - The sensitivity of replacing the surveyed country is also significant.

- Review 2:
  - The choice of the estimation method seems to be the key factor for explaining the difference in the estimation results in all indexes. Additionally, what the researcher uses as control variables is also important. In all health indexes, the estimation method plus other factors (e.g., method + controls) changes in the estimation result. What the researcher uses as control variables is also included in all health indexes.
  - The influence of the difference in the surveyed country is also important for explaining the difference in the effect of retirement on health.
  - Changes in the definition of retirement have a lower impact.

We summarize our main results in Table 22, and fix our analysis method (FE-IV) in Section 5. According to Table 22, when we fix our analysis method, we obtain comparatively stable results. However, there is heterogeneity of the effect of retirement on health even if we fix our methods and control the endogeneity of retirement. As such, future work could answer on why is there heterogeneity of the effect of retirement on health among different countries.

A Appendix

A.1 Pension Eligibility Age

To obtain pensionable age, we use the information from the Bureau of Labor Statistics in each country. However, this information is not available for some countries. Subsequently, we contact the Bureau of Labor Statistics or Bureau of Statistics directly, and obtain the information if possible. If we cannot find any information in the previous step, we use the OECD Pensions at a Glance, social security programs throughout the world (Europe, Asia and the Pacific, and the Americas) and The EUs Mutual Information System in Social Protection (MISSOC) as data sources. However, we cannot obtain the detailed pension eligibility age for many countries. Finally, the correct pension eligibility ages are obtained for the USA, the UK, Germany, France, Denmark, Switzerland, Czech, Estonia, Japan, China, and Korea. We do not consider countries where this information is missing. In this paper, we analyze the USA, the UK, Denmark, France, Germany, Switzerland, Japan and Korea. We show the pension eligibility ages used in this paper, as per the following tables.

Table 23: Pension eligibility age in Section 5

Table 24: 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 25: 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 26: 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 27: 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 28: Pension eligibility age in Section 5

Table 29: 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 30: PEA: Switzerland

Birth cohort	PEA
Early PEA: Male	
~ 1924.12	63y0m
1925.1 ~ 1950.12	63y0m
Early PEA: Female	
~ 1937.12	60y0m
1938.1 ~ 1940.12	61y0m
1941.1 ~	62y0m
Normal PEA: Male	
~ 1924.12	65y0m
1925.1 ~ 1950.12	65y0m
Normal PEA: Female	
~ 1937.12	62y0m
1938.1 ~ 1940.12	63y0m
1941.1 ~	64y0m

Table 31: 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 32: 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

A.2 Additional Literature Review

Here, we show the rest of the results on the health indexes which we do not introduce in Section 2. We summarize the rest of the results on health indexes in Table 33.

Table 33: Illness

	Bound and Waidmann	Coe and Lindeboom	Dave et al.	Neuman	Johnston and Lee	Coe and Zamarro	Behncke	Hernaes et al.
	2007, Univ. Michigan WP	2008, IZA DP	2008, Southern Economic Journal	2008, J of Labor Research	2009, Economics Letters	2011, J Health Economics	2012, Health Economics	2013, J Health Economics
Metabolic Syndrome	positive(M) no(F)	positive (restricting within 4 years)	negative(diabetes)				negative(metabolic syndrome) no(diabetes)	
heart risk	no(M) no(F)	no(heart attack)	no(heart desease)		no		no(heart attack) negative(again heart attack and stroke)	
mortality		no				no		no(M&F)
SPBB score	positive(M) no(F)							
heart diabats diagnosis M	no(M) negative(F)							
chronic illness M	positive(M) positive(F)			no(M) no(F) chronic condition			negative	
plain M	positive(M) positive(F)							
high blood pressure		no	no		no(hypertension)			
cancer		no	no				negative	
mobility			negative	no(M) no(F)			negative(difficulty walking)	
illness			negative					
stroke			no					
arthritis			negative					
difference in self ratings				no(M) no(F)				
large muscle functions				no(M) no(F)				
# days ill					positive			
asthma					no			
arthritis					no		no	
depression					positive(mental health)	no(EUROD)		
angina							no	
stroke							negative	
psychiatric							no	
health stock							negative	
limitting long standing illness							negative	
seeing difficulties							no	
hearing difficulties							negative	
high C-reactive protein (>3mg/L)							no	
high fibrinogen (7>mmol/L)							negative	
low hemoglobin (<12g/dl)							negative	
Method	pseudo RDD	IV method	Fixed effect method	IV method	RDD	IV method	Nonparametric matching	IV method and hazard model
Method (details)		IVs: pension eligibility age	Restricting sample who has good health before retirement, and retire as of 62	IVs: public and private PEA for respondent and for spouse working more than 10 years	Using 65 years as kink points robustness check by changing bandwidth	IVs: eligibility age for early and full retirement	Using state pension eligibility age as IV	IVs: entitle retirement age
Def. of Retirement		people report to be out of the labor force or not having any paid employment	complete retirement (retired and not working)	elderly working less than 1200 hours in a year	Retired from paid work	someone who is not in the paid labor force	retire describes her current situation best and not in paid work was her activity in the last month	receving pension, other benefits or sharp drop of income
Controls(Demog.)		age, education, marital status, children	age, sex, race, marital status, education	age, education, race, whether parents living or not, children, marital status, region		education, marital status, children	children, birth place, residential area	education, faculty, marital status,
Controls(Economic)			income, asset	financial status		income	income	income, pension information
Controls(Working.)		job types (blue and white collar)		occupation		self employment	working hours, employment status	job industry
Controls(Health)			lifestyle habits	early factors health behaviors				
Data	ELSA 1st wave	HRS 1st-7th wave	HRS 92~05 7 wave	HRS 1992~2004 7 wave. Only elderly consecutive for 3 years	Health Survey for England	SHARE 1st-2nd wave	ELSA 1st-3rd wave	administrative data 1992~2010
Sample		male workers aged 55-70 years						
Country	The U.K.	The U.S.	The U.S.	The U.S.	The U.K.	European countries	The U.K.	Norway

A.3 Notes on Replication and Replacement

- Replication 1: In this subsection, we explain the details of replication and replacement procedures. Table 34 shows the table number in the original papers we replicate, the number of samples when we replicate the results, and our comments on the replication. In most cases, we can replicate the results in preceding literature with a number of samples similar to the original number of samples, except for Coe and Zamarro (2011). When we replicate Coe and Zamarro (2011), the number of samples is 7,066 for self-report of health and depression, and 6,637 for cognition, while the number of samples in the original paper are 5,282 and 4,926, respectively. It is possible that the difference in the number of samples is due to the difference in the version of the SHARE dataset. We use the SHARE 5.0.0 for all waves, while Coe and Zamarro (2011) use 2.0.1 SHARE 2004. However, the summary statistics (e.g., average value) calculated by our replicated samples are very similar to the original test statistics, and, therefore, we use the replication results with our replicated samples of Coe and Zamarro (2011).

- Replication 2: We exclude some control variables when we replicate Neuman (2008) because of data limitation. Neuman (2008) uses detailed regional information and the health status when a respondent is a child. We have generated these variables by using the Cross-Wave: Census Region/Division and Mobility File and Aging Trends and Effects (RELATE) Files. However, when we include these generated variables in the estimated model, the sample size significantly decreases. Therefore, we exclude these variables from the control variables in the Neuman (2008) replication.

Table 34: Notes on Replication

	Table we replicate	**Sample replication** (Original) → (Our replication)	**Comment**
Cognition			
Bonsang et al. (2012)	Table 1	54377 → 55564	
Coe and Zamarro (2011)	Table 6 (Memory)	4928 → 6637	
Self-report of health			
Dave et al. (2008)	Table 2 (Poor health)	NA (not mentioned) → 35594	
Coe and Zamarro (2011)	Table 5 (Bad health)	5282 →7066	
Depression			
Dave et al. (2008)	Table2 (Column 3)	NA (not mentioned) → 28420	
Coe and Zamarro (2011)	Table 5 (Euro-D)	5282 → 7066	
ADL			
Dave et al. (2008)	Table 2 (Column 3)	NA (not mentioned) → 30731	
Neuman (2008)	Table 3	7632 → 7655	We omit some control variables.
Obesity			
Johnston and Lee (2009)	Table 1 (Bandwidth 3)	2877 → 2876	
Godard (2016)	Table 9 (Obese)	3951 → 4059	

Tables 35 and 36 summarize the notes on the replacement procedures by each replacement factor. For example, (Bonsang et al. (2012) → Coe and Zamarro (2011)) describe the comments when we carry out the replacement procedure from Bonsang et al. (2012) to Coe and Zamarro (2011). (Controls) describes the comments when we replace control variables.

Table 35: Notes on Replacement 1

Cognition

Bonsang et al. (2012) → Coe and Zamarro (2011)

(Controls)

•We exclude some control variables Coe and Zamarro (2011) include because the variables are not available in all waves used in Bonsang et al. (2012). The problem is that the sample size significantly decreases when we include these variables.[*1]

(Sample)

•Coe and Zamarro (2011) use health condition variables to restrict the analyzed samples in the SHARE. Since some of these variables are not available in the HRS, we do not apply the same sample restriction procedure in Coe and Zamarro (2011).

Coe and Zamarro (2011) → Bonsang et al. (2012)

(Method and data)

•Since Coe and Zamarro (2011) use only wave 1 of the SHARE, we cannot directly apply the FE-IV estimation for the analysis framework of Coe and Zamarro (2011). Therefore, we use wave 1 and 2 of the SHARE for FE-IV estimation when replacing the method and the dataset.

Self-report of health

Dave et al. (2008) → Coe and Zamarro (2011)

(Method)

•Since Dave et al. (2008) use the FE estimation, they do not use the IVs. Therefore, when applying IV estimation to Dave et al. (2008), we use the same pensionable ages as Bonsang et al. (2012) for the IVs, because Dave et al. (2008) and Bonsang et al. (2012) analyze the USA and the data collection periods roughly overlap.

•We use age and age squared instead of the age dummy when we use the IV estimation. There is a multicollinearity between the IVs (takes the value 1 after a respondent reaches pensionable age) and the age dummy when applying the IV estimation.

Coe and Zamarro (2011) → Dave et al. (2008)

(Index)

•We use "Poor health" (included in wave 1 and 2) as the index for FE estimation because the European scale of self-report of health is asked only in the SHARE wave 1.

(Method and data)

•We use wave 1 and 2 in the SHARE for FE estimation when replacing the method and data because of the same reason in (method and data) of the cognition section.

(Controls)

•We exclude some control variables that are not asked in the SHARE[*2] and the health insurance variable that is asked in only several countries, when replacing the control pattern from Coe and Zamarro (2011) to Dave et al. (2008).

(Data)

•We use "Poor health" in the HRS because the European scale of self-report of health is not asked in the HRS when replacing the dataset from the SHARE to the HRS.

[*1] e.g., non-professional activities and physical activities.

[*2] e.g., race, religious preference.

Table 36: Notes on Replacement 2

Depression
Dave et al. (2008) → Coe and Zamarro (2011)
(Method)
•The same comments as in Self-report of health apply.
Coe and Zamarro (2011) → Dave et al. (2008)
(Method and data)
•The same comments as in Self-report of health apply.
(Controls)
•The same comments as in Self-report of health apply.
(Data)
•We use the CES-D in the HRS because the EURO-D is not asked in the HRS when replacing the dataset from the SHARE to the HRS.
ADL
Dave et al. (2008) → Neuman (2008)
(Method)
•When applying the estimation method by Neuman (2008), we use the same estimation equation and the IVs as Neuman (2008).

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 (English Longitudinal Study of Ageing) is funded by the US National Institute on Ageing and a consortium of UK government departments coordinated by the Office for National Statistics.

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: N211909, SHARE-LEAP: N227822, SHARE M4: N261982). 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.

References

Ayyagari, P. (2014). The Impact of Retirement on Smoking Behavior. *Eastern Economic Journal*, pages 1–18.

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

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

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*.

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

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.

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.

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.

Eibich, P. (2015). Understanding the Effect of Retirement on Health: Mechanisms and Heterogeneity. *Journal of Health Economics*, 43:1–12.

Fonseca, R., Kapteyn, A., Lee, J., Zamarro, G., and Feeney, K. (2014). A Longitudinal Study of Well-being of Older Europeans: Does Retirement Matter? *Journal of population ageing*, 7(1):21–41.

Godard, M. (2016). Gaining weight through retirement ? Results from the SHARE survey. *Journal of Health Economics*, 45:27–46.

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.

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*.

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

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

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.

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. (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. (2016). Occupation, retirement and cognitive functioning. *Ageing and Society*, May 2016:1–29.

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

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

Lee, J. and Smith, J. P. (2009). Work, Retirement, and Depression. *Journal of Population Ageing*, 2(1-2):57–71.

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.

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.

McGarry, K. (2004). Health and Retirement: Do Changes in Health Affect Retirement Expectations? *Journal of Human Resources*, 39(3):624–648.

Motegi, H., Nishimura, Y., and Terada, K. (2016). Does Retirement Change Lifestyle Habits? *Japanese Economic Review*, 67(2):169–191.

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

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

van der Heide, I., van Rijn, R. M., Robroek, S. J., Burdorf, A., and Proper, K. I. (2013). Is retirement good for your health? A systematic review of longitudinal studies. *BMC public health*, 13(1):1180.

Wang, M. and Shi, J. (2014). Psychological research on retirement. *Annual review of psychology*, 65:209–33.

Zhao, M., Konishi, Y., and Noguchi, H. (2013). Retiring for Better Health Evidence from Health Investment Behaviors in Japan. Mimeo.