WINPEC Working Paper Series No.E2301
August 2023

Sex Differences in the Impact of Retirement on Health: Evidence from 35 Countries

Koryu Sato and Haruko Noguchi

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Sex Differences in the Impact of Retirement on Health: Evidence from 35 Countries

Koryu Sato and Haruko Noguchi*

Abstract

Although many studies have explored the impact of retirement on health, their results were inconsistent. To provide a comprehensive view, this study examined the impact of retirement on health using harmonized longitudinal data from 35 countries. Fixed effects instrumental variable model revealed that women demonstrated improved cognitive function and physical independence after retirement. In both sexes, retirement improved self-rated health, but women indicated a larger effect than men. Consistently, retirement reduced physical inactivity and smoking among women, which was not observed among men. The observed sex differences in post-retirement health behaviors may induce heterogeneous effects on health. Given the global trend of increasing state pension age, the promotion of healthy behaviors could mitigate the potential adverse effects of delayed retirement on health.

Keywords: retirement, cognitive function, physical independence, self-rated health, state pension age, fixed effects instrumental variable

JEL classification: I10, J26, C26

---

* Koryu Sato is an assistant professor at the Department of Social Epidemiology, Graduate School of Medicine and School of Public Health, Kyoto University, and a doctoral student at the Graduate School of Economics, Waseda University, corresponding to Room 204, Science Frontier Laboratory, Yoshida-Konoecho, Sakyo-ku, Kyoto-shi, Kyoto 606-8315, Japan (sato.koryu.8i@kyoto-u.ac.jp). Haruko Noguchi is a professor at the Graduate School of Economics, Waseda University and Director of Waseda Institute of Social & Human

Capital Studies (WISH). This study was supported by the Japan Society for the Promotion of Sciences (grant number: 20K18931, 23H03164) and the Health Care Science Institute Research Grant. The funders had no role in the study design; in the collection, analysis, and interpretation of the date; in the writing of the report; or in the decision to submit the article for publication. There are no conflicts of interest to declare. This analysis used data or information from: the Harmonized SHARE dataset and Codebook, Version F as of June 2022; the Harmonized ELSA dataset and Codebook, Version G.2 as of July 2021; the Harmonized CRELES dataset and Codebook, Version A as of August 2016; the Harmonized MHAS dataset and Codebook, Version C as of September 2022; RAND HRS Longitudinal File 2018 (V2) and the Harmonized HRS dataset and Codebook, Version C as of January 2022; the Harmonized CHARLS dataset and Codebook, Version D as of June 2021; the Harmonized JSTAR dataset and Codebook, Version B as of August 2014; and the Harmonized KLoSA dataset and Codebook, Version D.2 as of July 2021. The survey harmonization was funded by the National Institute on Aging (grant number: R01 AG030153, RC2 AG036619, R03 AG043052) and conducted by the Gateway to Global Aging Data in collaboration with the research team of the surveys. The HRS was sponsored by the National Institute on Aging (grant number: NIA U01AG009740) and was conducted by the University of Michigan. The MHAS received support from the National Institute on Aging (grant number: R01 AG018016) in the United States and the Instituto Nacional de Estadística y Geografía (INEGI) in Mexico. The harmonized datasets and more information are available through the Gateway to Global Aging Data website (https://g2aging.org/). This study used publicly available data that have obtained informed consent from all participants and ethical approval from relevant local ethics committees. Thus, the Ethics Committee of Kyoto University exempted this study from review.

1. Introduction

To accommodate an aging population, many developed countries are increasing their state pension age (SPA).[1] These policy changes may influence population health because they delay the timing of retirement and change budget constraints and time allocations for health investments in later life (Grossman 1972). However, the impact of delayed retirement on health remains unclear, and there is a lack of consensus on this issue (Nishimura, Oikawa, and Motegi 2018; Garrouste and Perdrix 2022). We hypothesized that the inconsistent results in the previous literature stemmed from effect heterogeneity. When a subgroup adversely influenced by retirement dissimulates its beneficial effect in other subgroups, the average treatment effect of retirement in the population will be obscure. Therefore, this study examined the impact of retirement on health by employing an array of analyses targeting heterogeneity. This was accomplished through the utilization of harmonized longitudinal data derived from 35 countries.

Retirement is commonly an endogenous decision, as people in poorer health are more likely to retire. To address the potential downward bias, researchers often use SPA as an instrumental variable (IV) for retirement.[2] SPA seems to be valid because of its dual attributes; first, reaching the SPA will increase the probability of retirement (meeting the relevance condition); second, the SPA itself does not directly affect health outcomes (satisfying the exclusion restriction condition). Using the harmonized data, this study employed within- and between-country variations in SPA to identify the effects of retirement on health. Moreover, we benefitted from a panel structure of the longitudinal surveys and included fixed effects (FEs) of individuals,

---

[1] See Organisation for Economic Co-operation and Development (2021).

[2] The regression discontinuity design (RDD) using SPA as a threshold for the timing of retirement is another possible identification strategy. Ebeid and Oguzoglu (2023) used the nonparametric fuzzy RDD and examined the effect of retirement on the cognitive function. Although the method has the strength of no parametric assumptions, we did not use it because it cannot account for multiple thresholds representing an early retirement age and the panel structure of longitudinal surveys (i.e., incompatible with fixed effects).

countries, years, and interactions between countries and years. Thus, our model, grounded in fixed effects instrumental variable (FEIV), effectively accounted for unobserved time-invariant characteristics of individuals and countries and heterogeneous time trends across countries. Additionally, this study delved into the disparities in the effects of retirement, aiming to unveil the underlying cause of the inconsistent results encountered in the earlier academic discourse.

We found a discernible pattern wherein women exhibited improved cognitive function and physical independence after retirement. In both sexes, retirement improved self-rated health, but this effect was more pronounced among women compared to men. Remarkably, this trend extended to a reduction in physical inactivity and smoking among women, which was not observed within the male cohort. Encouragingly, these effects appeared a consistent homogeneity across various dimensions, including countries, educational backgrounds, and pre-retirement occupational characteristics.

This study contributes significantly to the existing literature in several ways. First, it suggests that the inconsistent results in previous studies can be attributable to sex heterogeneity in the effects of retirement. In particular, we found that retirement improved women's cognitive function, whereas men indicated an insignificant but detrimental effect of retirement. Intriguingly, in scenarios where a substantial proportion of male subjects constitutes the study population, their presence could potentially overshadow the beneficial consequences of retirement for women. This is consistent with many studies that showed no evidence of the association between retirement and cognitive function (Coe and Zamarro 2011; Coe et al. 2012; Romero Starke et al. 2019; Rose 2020). Our findings imply that the average treatment effect of retirement can vary depending on the sex composition prevalent within the study population.

Second, we present the potential mechanism of health disparities in the older population.

The consistent sex differences in health outcomes and behaviors suggest that post-retirement health behaviors can induce the heterogeneous effects of retirement on health in line with Eibich (2015). We showed that retirement is beneficial, especially for women; thus, delayed retirement owing to increasing SPA can deteriorate population health. However, increasing SPA seems inevitable in many developed countries given their imminent pension finance facing the challenges posed by the rapidly aging population. Our findings provide policymakers with valuable insights that the promotion of healthy behaviors can mitigate potential adverse effects of delayed retirement on health owing to the mounting SPA.

Third, we provide an encompassing perspective on the effects of retirement vis-à-vis health. A recent review showed that many studies indicated the detrimental effects of retirement on cognitive function, while evidentiary support for its influence on physical function remains inconclusive. Notably, retirement appears to yield beneficial consequences for self-rated health (Garrouste and Perdrix 2022). Paradoxically, this array of findings across diverse outcomes remains enigmatic, as self-rated health constitutes a strong predictor of both cognitive and physical impairments (Bond et al. 2006; Brenowitz et al. 2014; Idler and Benyamini 1997). Inconsistencies in prior literature may be owing to variations in statistical methodologies, study designs, retirement measurement modalities, outcome assessments, and contexts of various countries under examination. The discrepancy lacks a clear explanation, and thus a comprehensive cross-country investigation using consistent research methodologies is required. One exception is Nishimura, Oikawa, and Motegi (2018), who explored the effects of retirement on various health outcomes using data from eight countries up to 2014.[3] Our study expands the work of Nishimura, Oikawa,

[3] The aforementioned data source pertains to the Health and Retirement Study (HRS), the Survey of Health, Ageing and Retirement in Europe (SHARE), the English Longitudinal Study on Ageing (ELSA), the Japanese Study of Aging and Retirement (JSTAR), and the Korean Longitudinal Study of Aging (KLoSA).

and Motegi (2018) in several ways. Foremost, they investigated each country separately, leaving unsolved the issue of cross-country variations in the health implications of retirement. Contrarily, we tested the effect heterogeneity by countries by including interaction terms between retirement and the characteristics of country, such as region, income level, and the percentage of the older population. We confirmed that the effect of retirement was homogeneous across countries and thus showed pooled estimates from data of 35 countries. Moreover, given that many countries started raising their SPA around 2015 (Organisation for Economic Co-operation and Development 2021), this study applies more recent data from a larger number of countries. Finally, we explored not only health outcomes but also health behaviors including physical inactivity, smoking, and binge drinking, which enabled us to reveal the underlying mechanism that contributes to heterogeneous effects on health outcomes.

The remainder of the paper is organized as follows: Section 2 outlines the data used in this study; Section 3 presents the empirical model; Section 4 reports the results; and Section 5 discusses the results and concludes the paper.

2. Data

2.1. Harmonized Data

We implemented the harmonized datasets of the Health and Retirement Study (HRS) and its sister surveys provided by the Gateway to Global Aging Data project (Lee, Phillips, and Wilkens 2021), which is "a free public resource designed to facilitate cross-national and longitudinal studies on aging." It includes data on various individual characteristics regarding demographics, health, health services use, work/employment, economic status, and family structure/social network.[4]

[4] GATEWAY TO GLOBAL AGING DATA. https://g2aging.org/ (Accessed: January 21, 2023)

Data from the Irish Longitudinal Study on Ageing, the Longitudinal Aging Study in India, and the Malaysia Ageing and Retirement Study were not used because they were harmonized only for one wave. Given this, our datasets comprised of waves 1, 2, and 4 through 8 (2004–2019) of the Survey of Health, Ageing and Retirement in Europe (SHARE)[5]; waves 1 through 9 (2002–2018) of the English Longitudinal Study on Ageing (ELSA); waves 1 through 5 (2005–2012) of the Costa Rican Longevity and Healthy Aging Study (CRELES); waves 1 through 5 (2001–2018) of the Mexican Health and Aging Study (MHAS); waves 1 through 14 (1992–2018) of the HRS; waves 1 through 4 (2011–2018) of the China Health and Retirement Longitudinal Study (CHARLS); waves 1 through 3 (2007–2011) of the Japanese Study of Aging and Retirement (JSTAR); and waves 1 through 7 (2006–2018) of the Korean Longitudinal Study of Aging (KLoSA). Finally, the CRELES included a cohort interviewed in waves 1 through 3 (2005–2009) and another cohort interviewed in waves 4 through 5 (2010–2012). The surveys were all designed to represent the national older population except for the JSTAR, which randomly recruited participants from 10 specific municipalities. The same individuals were reviewed biennially; however, the MHAS and CHARLS conducted interviews triennially since 2012 and 2015, respectively.

Originally, the harmonized data involved 1,912,071 observations from 276,930 unique individuals who were surveyed in multiple timings. Then, our inclusion and exclusion criteria were applied. First, we included 609,422 observations from 205,022 unique individuals aged 50 to 70 years, whose timings could be affected by the SPA of each country.[6] Second, regarding the

---

[5] SHARE was conducted in 29 countries, namely Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, England, Estonia, Finland, France, Germany, Greece, Hungary, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and Switzerland. The third wave of the SHARE, referred to as SHARELIFE, featured a distinct questionnaire from the preceding waves.

[6] The CRELES conducted interviews with individuals aged ≥ 60 years in waves 1 through 3, and with individuals aged 55 through 65 years in waves 4 and 5. Whereas the HRS interviewed adults aged ≥ 51 years.

CHARLS, 39,682 observations from rural residents were not included because China had different pension systems in rural and urban areas; therefore, rural residents were not affected by the SPA (Lei and Liu 2018). Third, we excluded 118,100 observations that corresponded to individuals who were not working for reasons other than retirement, such as being unemployed, disabled, or a homemaker. Fourth, we did not include 355 observations because of missing values for explanatory variables necessary for our analysis. Finally, 54,381 individuals who were observed only once in the survey were excluded from analyses because maintaining them in a FE model could underestimate standard errors (Correia 2015). Thus, at the baseline, our study consisted of 396,904 observations from 106,927 individuals in 35 countries with a mean follow-up period of 6.7 years (Table 1). Notably, the number of observations varied across regressions, as those with missing values, differed across outcomes.

Table 1: Cohort characteristics of the surveys

Survey	Country	Interview years	No. of unique individuals	Mean follow-up years	Mean no. of interviews	% of men
SHARE	Austria	2004, 2006, 2011, 2013, 2015, 2017, 2019	2,877	5.4	3.3	46.2
	Belgium	2004, 2006, 2011, 2013, 2015, 2017, 2019	4,118	5.7	3.3	51.8
	Bulgaria	2017, 2019	377	2.0	2.0	43.0
	Croatia	2015, 2017, 2019	1,119	2.8	2.4	49.5
	Cyprus	2017, 2019	124	2.0	2.0	42.7
	Czech Republic	2006, 2011, 2013, 2015, 2017, 2019	3,827	5.8	3.4	41.4
	Denmark	2004, 2006, 2011, 2013, 2015, 2017, 2019	3,031	6.6	3.5	48.2
ELSA	England	2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018	9,895	7.7	4.5	47.6
SHARE	Estonia	2011, 2013, 2015, 2017, 2019	3,662	4.8	3.2	42.3
	Finland	2017, 2019	550	2.0	2.0	47.3
	France	2004, 2006, 2011, 2013, 2015, 2017, 2019	3,540	6.4	3.4	47.1

	Germany	2004, 2006, 2011, 2013, 2015, 2017, 2019	3,437	5.3	3.2	49.7
	Greece	2004, 2006, 2015, 2017, 2019	2,187	6.5	2.7	60.5
	Hungary	2011, 2017, 2019	788	6.4	2.3	41.1
	Israel	2004, 2006, 2013, 2015, 2017, 2019	1,447	8.2	3.3	46.9
	Italy	2004, 2006, 2011, 2013, 2015, 2017, 2019	3,026	5.9	3.2	56.1
	Latvia	2017, 2019	303	2.0	2.0	41.6
	Lithuania	2017, 2019	528	2.0	2.0	37.1
	Luxembourg	2013, 2015, 2017, 2019	841	3.9	2.8	54.5
	Malta	2017, 2019	239	2.0	2.0	72.8
	Netherlands	2004, 2006, 2011, 2013, 2019	1,862	6.4	2.7	55.6
	Poland	2006, 2011, 2015, 2017, 2019	1,700	5.6	2.7	41.4
	Portugal	2011, 2015, 2017	761	4.6	2.3	50.5
	Romania	2017, 2019	560	2.0	2.0	45.5
	Slovakia	2017, 2019	665	2.0	2.0	47.7
	Slovenia	2011, 2013, 2015, 2017, 2019	2,531	4.5	3.1	44.6
	Spain	2004, 2006, 2011, 2013, 2015, 2017, 2019	2,731	5.3	3.0	59.6
	Sweden	2004, 2006, 2011, 2013, 2015, 2017, 2019	3,151	6.2	3.2	45.1
	Switzerland	2004, 2006, 2011, 2013, 2015, 2017, 2019	2,178	6.4	3.6	48.8
CRELES	Costa Rica	2005, 2007, 2009, 2010, 2012	1,244	2.3	2.1	76.9
MHAS	Mexico	2001, 2003, 2012, 2015, 2018	8,148	6.8	2.7	66.7
HRS	United States	1992, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018	25,753	9.2	5.2	46.9
CHARLS	China	2011, 2013, 2015, 2018	2,819	4.8	2.9	54.1
JSTAR	Japan	2007, 2009, 2011	1,775	3.0	2.5	64.8
KLoSA	South Korea	2006, 2008, 2010, 2012, 2014, 2016, 2018	5,133	7.0	4.2	53.1
Overall			106,927	6.7	3.7	50.5

Note: SHARE stands for the Survey of Health, Ageing and Retirement in Europe; ELSA stands for the English Longitudinal Study on Ageing; CRELES stands for the Costa Rican Longevity and Healthy Aging Study; MHAS stands for the Mexican Health and Aging Study; HRS stands for the Health and Retirement Study; CHARLS stands for the China Health and Retirement Longitudinal Study; JSTAR stands for the Japanese Study of Aging and Retirement; KLoSA stands for the Korean Longitudinal Study of Aging.

2.2. Retirement and State Pension Age

The retirement statuses of survey participants were determined using the harmonized variable of self-reported labor force status in Appendix A and described by Zamarro and Lee (2012). Individuals who self-identified as retired during the interview, regardless of their working status (i.e., including those who were "partly retired"), were included in the retired group for comparison with workers, as defined in previous literature (Bianchini and Borella 2016; Atalay, Barrett, and Staneva 2019), and outlined in Appendix B. Other studies defined retirement as not working (Coe and Zamarro 2011; Bonsang, Adam, and Perelman 2012; Bingley and Martinello 2013). In light of this alternative definition, individuals who identified as retired but were still engaged in paid work were classified by combining labor force status and employment engagement variables, and subsequently excluded from the analysis in a robustness check.

To address potential endogeneity in retirement decisions, we employed the SPA as an IV for retirement. In some countries, early pensions are granted under specific circumstances, such as reduced benefits or sufficient social security contributions. Thus, we employed the joint instruments of the early retirement age (ERA) and the official retirement age (ORA) to predict retirement. A binary ERA variable indicated whether participants had attained the earliest age of eligibility for reduced pensions or full pensions with certain conditions. Similarly, a binary ORA variable indicated whether participants had attained the age of entitlement to minimum guaranteed pensions or full pensions without any requirements. In countries where early pensions are not available, the ERA variable was set to zero for all participants. We obtained ERA, ORA, and their modifications during the study period from "Social Security Programs Throughout the World" (United States Social Security Administration, 2020), "Pensions at a Glance" (Organisation for

Economic Co-operation and Development 2021), and websites of the national authorities (as listed in Appendix C). To provide a descriptive overview of retirement patterns, we created a graph depicting age and the corresponding retirement rate by country, which demonstrated changes in the retirement rate around the SPA (as illustrated in Appendix D).

2.3. Outcome Measures

As a measure of cognitive function, we focused on episodic memory involving a neurocognitive system that is responsible for recollecting past experiences. Episodic memory constitutes an appropriate measure to assess the impact of retirement because it exhibits a decline with advancing age (Tulving 2002) and can capture the preliminary stages of cognitive impairments. Moreover, it is less subjective to floor and ceiling effects, given a wide range of scores (Bonsang et al. 2012). In accordance with the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery (Morris et al. 1989), a list of common words was verbally presented to the participants, following which they were immediately asked to recollect as many words from the list as possible. After approximately 5 min, the participants were requested to, once again, recall the words from the list. The episodic memory score was calculated by adding the number of words remembered during both the immediate and delayed recalls. Typically, most surveys included a list of ten words, thereby offering a score range from 0 to 20. As shown in Appendix Figure E.1, the scores appeared to be normally distributed for surveys featuring a 10-word list. However, the number of words varied across surveys and waves. Waves 1 through 2 of the HRS comprised 20 words on the list, while the MHAS consisted of 8 words, and the CRELES and the KLoSA contained 3 words (Shih, Lee, and Das 2012). Thus, to enable comparison, we standardized the scores to z-scores[7] for each

[7] The z-score is a score converted so that the mean is 0 and the standard deviation is 1, which makes values with different units of measurement (such as outcome measures in the harmonized data) comparable.

survey. Additional analysis, using only surveys with a 10-word list, was performed for robustness.

The assessment of physical function in our study was based on the individual's capability to carry out activities of daily living (ADL) and instrumental activities of daily living (IADL). A combination of ADL and IADL items is known to accurately predict physical limitations (Roehrig, Hoeffkin, and Pientka 2007). To facilitate comparability, we selected eight activities that participants were capable of performing. These activities included four ADL items (bathing, eating, getting in and out of bed, and using the toilet) and four IADL items (managing money,[8] taking medications, shopping for groceries, and preparing meals). As wave 1 of the HRS did not include questions about the capability of toilet use and four IADL items, the score was recorded as missing. The responses to the eight items displayed consistency (Cronbach's $\alpha = 0.79$). As shown in Appendix Figure E.2, most participants were capable of performing all eight activities. Hence, we categorized the participants into two groups: those who were independent in all activities and those who were not. Subsequently, we created a binary variable indicating 1 for those who were fully independent and 0 for otherwise.

Self-rated health was measured using a 5-point Likert scale (1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent). In some prior studies, self-rated health was dichotomized into a binary variable indicating good or poor health (Behncke 2012; Coe and Lindeboom 2008; Hessel 2016; Johnston and Lee 2009; Messe and Wolff 2019; Neuman 2008; Rose 2020; Zhu 2016). However, as it appeared normally distributed (as illustrated in Appendix Figure E.3), this study standardized self-rated health to z-scores for each survey and treated it as a continuous variable, similar to other studies (Calvo, Sarkisian, and Tamborini 2013; Gorry, Gorry, and Slavov 2018).

[8] The JSTAR contained three distinct queries concerning financial management, namely paying bills, withdrawing money from the bank, and filling out a pension document. To ensure consistency with other surveys, we used the first inquiry to create the variable for the Japanese participants.

Finally, to explore the underlying mechanisms linking retirement to the primary outcomes, we also examined the impact of retirement on physical inactivity, smoking, and binge drinking as these factors have been identified as potential risk factors for cognitive and physical impairments (Agahi et al. 2018; Maurage et al. 2012; Moore, Endo, and Carter 2003; Okusaga et al. 2013; Sato et al. 2021). Those who engaged in vigorous or moderate physical activity less than once per week were considered physically inactive individuals. [9] The smoking status of participants was categorized into current smokers and non-smokers.[10] Binge drinking was defined as consuming five or more drinks per day for men and four or more for women, in accordance with the definition provided by Centers for Disease Control and Prevention (2022).[11] These three variables were converted into binary categories.

2.4. Covariates

The estimation model utilized in this study incorporated adjustments for covariates including age, age squared (divided by 10), and marital status. Age was centered at the mean in regression models for ease of interpretation. During each interview, participants were asked to report their marital

---

[9] In the case of South Korea, we relied solely on the variable of pertaining to vigorous physical activity since the KLoSA did not inquire about the frequency of moderate physical activity. For wave 7 of SHARE, only individuals who had also participated in wave 3 were asked about the frequency of physical activity. Thus, all observations from Bulgaria, Cyprus, Finland, Latvia, Lithuania, Malta, Romania, and Slovakia were excluded from the analysis because those only had one observation. In waves 1 through 3 of CHARLS, only half of the participants were queried about physical activity, and certain observations were excluded owing to question incompatibility. Specifically, the MHAS asked whether participant engaged in vigorous physical activity three or more times per week; in contrast, the JSTAR asked for the number of minutes of exercise on weekdays and weekends.

[10] In wave 6 of SHARE, individuals who had previously been interviewed were not queried about their smoking status. In wave 7, only new participants and those who previously reported smoking in wave 3 were asked about their current smoking status. Consequently, all the observations from Bulgaria, Cyprus, Finland, Latvia, Lithuania, Malta, Romania, and Slovakia were excluded from the analysis owing to a single observation.

[11] Waves 1 and 6–8 of SHARE, wave 1 of ELSA, waves 1 and 2 of HRS, and all waves of CRELES and CHARLS did not provide the information on the number of drinks per day. Thus, all observations from Bulgaria, Croatia, Cyprus, Finland, Greece, Hungary, Latvia, Lithuania, Luxembourg, Malta, Portugal, Romania, and Slovakia were excluded from the analysis since these individuals had only one observation. Additionally, the question on drinking habits was not asked in wave 2 in some study sites of JSATR.

status, with the response options being: married, partnered, separated, divorced, widowed, or never married. We coded 1 for those who were married or partnered and 0 otherwise.

To access potential effect heterogeneity across various demographic and occupational characteristics, we included interaction terms between retirement and sex, educational levels, pre-retirement job characteristics, and country characteristics in our statistical models. Educational attainment was classified into three groups using the 1997 International Standard Classification of Education codes—less than upper secondary education as low, upper secondary and vocational training as middle, and tertiary education as high. We also investigated whether retirement from physically demanding jobs and jobs with low control modified the impact of retirement on the outcomes. During the surveys, participants who were currently employed were asked to rate their agreement with statements regarding the physical demands[12] and control of their job[13] using a four-point Likert scale that included "strongly disagree," "disagree," "agree," or "strongly agree." Participants who responded with "agree" or "strongly agree" at least once during the interview were considered to have experience in physical labor and low-control jobs, respectively. Participants who had never engaged in paid work during the study period were excluded from the models with job characteristics interactions. Additionally, we performed interaction tests across country characteristics, grouping regions into Europe (including Israel), America (Costa Rica, Mexico, and the United States), and Asia (China, Japan, and South Korea); classifying countries as high-income (all European countries, United States, Japan, and Korea) or low-middle income (Bulgaria, Romania, Costa Rica, Mexico, and China) based on Gross National Income per capita as defined by the World Bank[14]; and considering a country to be an aged society if the percentage

[12] The question was not included in wave 1 of the ELSA, CRELES, MHAS, and CHARLS.
[13] The question was not included in wave 1 of the ELSA; CRELES; MHAS; and waves 1 through 7 of HRS, CHARLS, JSTAR, and KLoSA.
[14] Although Poland did not meet the threshold for high-income countries in 2006, we categorized it as a high-income

of the population aged 65 years and older[15] exceeded 14%, as defined by the Organisation for Economic Co-operation and Development and World Health Organization (2020).

3. Empirical Model

We investigated the impact of retirement on the outcomes using linear probability models estimated by the FEIV with the two-stage least squares procedure. In the first stage, the probability of retirement was predicted as follows:

$$R_{ijt} = \beta_1 age_{ijt} + \beta_2 age_{ijt}^2 + \beta_3 ERA_{ijt} + \beta_4 ERA * age_{ijt} + \beta_5 ERA * age_{ijt}^2 + \beta_6 ORA_{ijt} + \beta_7 ORA * age_{ijt} + \beta_8 ORA * age_{ijt}^2 + \beta_9 mstat_{ijt} + \alpha_i + u_j + \lambda_t + u_j * \lambda_t + \varepsilon_{ijt}$$

where R_{ijt} denotes whether individual *i* residing in country *j* was retired in interviewed year *t*. ERA_{ijt} and ORA_{ijt} are instruments indicating whether the participants reached the ERA and the ORA. We include age, age squared, and their interactions with ERA_{ijt} and ORA_{ijt}. $mstat_{ijt}$ denotes whether the participants were married. α_i, u_j, and λ_t are FEs for individuals, countries, and survey years, respectively. The model also includes interaction between country and year FEs. ε_{ijt} is an error term. Subsequently, the second stage was estimated using the following equation:

$$Y_{ijt} = \gamma_1 \hat{R}_{ijt} + \gamma_2 age_{ijt} + \gamma_3 age_{ijt}^2 + \gamma_4 mstat_{ijt} + \alpha_i + u_j + \lambda_t + u_j * \lambda_t + \nu_{ijt}$$

country through the study period.

[15] We obtained the percentage of the population aged 65 and older from "World Development Indicators" published by the World Bank (2022).

where Y_{ijt} is health outcomes and health behaviors as risk factors, $\hat{R}_{ijt}$ is the predicted probability of retirement from the first-stage estimation, and ν_{ijt} is an error term. In all analyses, we estimated robust standard errors clustering for individual, country, year, and interactions between country and year.

The FEIV model has several advantages with respect to pooling data from different countries and estimating the causal effect of retirement on outcomes. The FEs of individuals and countries provide controls for both observable and unobservable time-invariant factors, such as genetic predisposition and educational attainment, as well as institutional and cultural differences among countries. Additionally, the interaction between FEs and countries represents heterogeneous time trends across countries. Moreover, we applied the IV method using the ERA and ORA as instruments to mitigate the endogeneity of retirement. To be valid, IVs must meet two conditions, namely (i) the relevance condition (the IV is associated with treatment, i.e., retirement) and (ii) the exclusion restriction (the IV has no association with potential outcomes under different values of treatment). Furthermore, the assumption of monotonicity (i.e., the IV does not have conflicting effects on treatment in any individual) enables us to interpret the point estimate of $\hat{\gamma}_1$ as a local average treatment effect (LATE) among "compliers" (i.e., individuals who would retire upon reaching the SPA).

4. Results

4.1. Descriptive Statistics

Table 2 presents the descriptive statistics by labor force status for a large sample of 396,904 observations from 106,927 individuals, which consisted of 217,166 (54.7%) with a working status and 179,738 (45.3%) individuals with a retired status. At first glance, Table 2 indicates that retirement seems to deteriorate all health statuses, although it would be beneficial for health behaviors as risk factors. Notably, retirees were found to be older and less likely to be men, married, and highly educated. Finally, retirees are less likely to have experienced physical labor and a job with low control than workers.

Table 2: Descriptive statistics of observations by labor force status (396,904 obs. from 106,927 individuals)

Variables	Working (obs.=217,166)			Retired (obs.=179,738)			Difference
	Obs.	Mean	SD	Obs.	Mean	SD	
Outcome variables							
Health status							
Cognitive function (z-score)	204,541	0.172	0.952	172,735	-0.0595	0.990	0.232***
Physical independence	194,608	0.959	0.198	168,365	0.883	0.322	0.076***
Self-rated health (z-score)	209,867	0.263	0.913	174,764	-0.105	0.986	0.368***
Health behavior as risk factors							
Physical inactivity	142,764	0.187	0.390	130,060	0.196	0.397	-0.009***
Smoking	183,993	0.204	0.403	140,526	0.183	0.387	0.021***
Binge drinking	138,350	0.103	0.304	103,861	0.062	0.241	0.041***
Covariates							
Age	217,166	57.940	4.694	179,738	64.217	4.253	-6.278***
Married	217,166	0.799	0.401	179,738	0.759	0.428	0.040***
Potential effect of heterogeneity							
Men	217,166	0.522	0.500	179,738	0.467	0.499	0.055***
Education							
Low	199,341	0.254	0.435	172,376	0.308	0.462	-0.054***
Middle	199,341	0.468	0.499	172,376	0.484	0.500	-0.016***
High	199,341	0.278	0.448	172,376	0.207	0.405	0.071***
Physically demanding job	181,723	0.574	0.494	77,187	0.518	0.500	0.056***
Low control job	121,760	0.355	0.479	44,475	0.323	0.468	0.032***

Note: Obs and SD denote the number of observations and standard deviation, respectively. Unpaired t-tests were performed.
*** p<0.01, ** p<0.05, * p<0.1.

As previously mentioned, we only included individuals who were followed up with at least twice, which may introduce selection bias caused by attrition. To assess this potential impact, we compared characteristics between individuals who were followed up and those who were lost in the follow-up (Appendix F). Consequently, we confirmed almost no differences in characteristics between these two groups; however, we observed that those who were lost in the follow-up were on average 0.84 years older and had poorer self-rated health by 0.11 points than those who were followed-up. Thus, exercising causion is necessary to interpret our results of age and subjective health.

4.2. Overall Regression Results

Tables 3 and 4 present comprehensive analyses of the effects of retirement on health outcomes and health behaviors as risk factors, respectively. Regressions were adjusted for age, age squared, and marital status. In Table 3, the point estimates obtained from pooled ordinary least squares (OLS) and FE models demonstrate that retirement has negative effects on health outcomes. However, the FEIV estimates reveal positive effects on health outcomes, with the difference between these models being the treatment of the decision to retire as an endogenous variable. These findings suggest that the inconsistencies in the effects of retirement on health outcomes in prior studies may partially be attributed to variations in estimation methods. Moreover, the FEIV estimates in our study do not show the paradoxical evidence observed in previous investigations, which indicate that retirement enhances self-rated health but has detrimental impacts on cognitive and physical functions. Specifically, our results indicate that retirement significantly improves the z-score of cognitive function by 0.048 standard deviation (SD), the likelihood of physical independence by 2.7% point, and the z-score of self-rated health by 0.144 SD. To elucidate the mechanisms underlying these results, we further evaluate the effect of retirement on health behaviors as risk factors. Table 4 shows that retirees exhibit a 3.0% point reduction in physical inactivity compared with workers in the FEIV model. The associations of retirement with smoking and binge drinking are not statistically significant in the FEIV models.

Table 3. The effects of retirement on health outcomes

	Cognitive function			Physical independence			Self-rated health		
	Pooled OLS	FE	FEIV	Pooled OLS	FE	FEIV	Pooled OLS	FE	FEIV
Retirement	-0.104***	-0.002	0.048**	-0.087***	-0.023***	0.027***	-0.326***	-0.055***	0.144***
	(0.022)	(0.005)	(0.021)	(0.011)	(0.003)	(0.006)	(0.021)	(0.008)	(0.019)
Individual FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Country FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Year FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Country x Year FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
IV	NO	NO	YES	NO	NO	YES	NO	NO	YES
Observations	377,276	377,276	377,276	362,973	362,973	362,973	384,631	384,631	384,631
Adjusted R^2	0.023	0.478		0.027	0.418		0.041	0.564	
Kleibergen-Paap F			2230.315			2222.211			2184.288
Hansen J			0.684			0.313			2.510

Note: OLS, FE, and FEIV denote ordinary least squares, fixed effect, and fixed effect with instrumental variable, respectively. All regressions are adjusted for age, age squared, and marital status. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. The effects of retirement on health behaviors

	Physical inactivity			Smoking			Binge drinking		
	Pooled OLS	FE	FEIV	Pooled OLS	FE	FEIV	Pooled OLS	FE	FEIV
Retirement	-0.004	-0.016***	-0.030***	0.020**	-0.011***	-0.006	-0.026***	-0.004**	0.011
	(0.024)	(0.005)	(0.010)	(0.008)	(0.002)	(0.007)	(0.009)	(0.002)	(0.010)
Individual FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Country FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Year FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Country x Year FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
IV	NO	NO	YES	NO	NO	YES	NO	NO	YES
Observations	272,824	272,824	272,824	324,519	324,519	324,519	242,211	242,211	242,211
Adjusted R^2	0.002	0.418		0.012	0.762		0.008	0.448	
Kleibergen-Paap F			1855.314			1592.798			788.561
Hansen J			2.471			3.627*			0.640

Note: OLS, FE, and FEIV denote ordinary least squares, fixed effect, and fixed effect with instrumental variable, respectively. All regressions are adjusted for age, age squared, and marital status. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

For all FEIV models, the Kleibergen-Paap Wald F statistics (Kleibergen and Paap 2006) indicated a strong correlation between IVs and retirement. In addition, the over-identification tests (Hansen 1982) did not reject the null hypotheses at the 5% significance level that the instruments were uncorrelated with residuals, which indicates that the IVs are plausible and satisfy the requirements for being valid. The first stage estimates of FEIV models are presented in Appendix G. It shows that reaching ERA and ORA had significantly positive effects on the probability of retirement, which suggests that raising the ERA or ORA would delay the timing of retirement. The effect of ORA was more pronounced than that of ERA. The probability of retirement increased with age; conversely, the negative interaction terms of ERA and ORA with age indicate that the slope slowed down once ERA and ORA were reached.

4.3. FEIV Models Incorporating Interactions

To determine the extent of heterogeneity, interaction terms between retirement and several demographic, socio-economic, and other contextual factors were included in the FEIV, such as sex (see Appendix H), educational levels (see Appendix I), pre-retirement job characteristics including physically demanding job (see Appendix J) and a job with low control (see Appendix K), region (see Appendix L), country income levels (see Appendix M), and population aging rates (see Appendix N). In summary, most of interaction terms failed to achieve statistical significance, indicating that retirement has a homogenous impact on health outcomes and behaviors across these characteristics. An encouraging development is that the homogenous outcomes observed across diverse country attributes, including economic status and demographic trends related to aging, added support to the authenticity of our investigation, which entailed aggregating data from numerous countries. However, there was a heterogeneous effect on cognitive function, self-rated health, and smoking across sex, as demonstrated in Appendix H. Compared with women, retirement was found to be less likely to enhance the z-score of cognitive function by 0.088 SD and that of self-rated health by 0.071 SD, while it was more likely to increase smoking by a 4.8% point among men. Nevertheless, it should be noted that the over-identification test did not meet the requirement for self-rated health and physical inactivity at the 5% significance level.

4.4. Stratified Analysis by Sex

Given the notable interactions between retirement and sex in the preceding section, we performed a stratified analysis based on sex. The results are presented in Table 5. Among men, we found no significant association, except for self-rated health; male retirees demonstrated a 0.100 SD increase in self-rated health. Conversely, female retirees showed a 0.100 SD increase in cognitive function, a 3.8%-point increase in physical independence, and a 0.193 SD increase in self-rated health with respect to health outcomes. Moreover, female retirees curtailed their physical inactivity by 4.3%-points and smoking by 1.9% points with respect to health behaviors. Based on the results, the mechanism of retirement to enhance health behavior and, consequently, health outcomes, would be more lucid for women than men.

Table 5. Stratified FEIV models for the effect of retirement on health outcomes and behaviors by sex

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
			Men			
Retirement	-0.005	0.015	0.100***	-0.017	0.015	0.026
	(0.032)	(0.009)	(0.030)	(0.015)	(0.011)	(0.017)
Observations	183,386	175,722	190,480	131,204	162,583	119,539
Kleibergen-Paap F	925.970	912.002	911.613	745.062	643.583	335.596
Hansen J	1.568	2.927*	1.006	0.115	2.118	0.098
			Women			
Retirement	0.100***	0.038***	0.193***	-0.043***	-0.019**	-0.006
	(0.027)	(0.008)	(0.024)	(0.013)	(0.008)	(0.011)
Observations	193,890	187,251	194,151	141,620	161,932	122,672
Kleibergen-Paap F	1333.937	1344.962	1300.021	1142.730	977.660	481.215
Hansen J	0.000	0.493	2.049	4.886**	0.002	0.711

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.5. Robustness Check

We performed additional analyses to ascertain the robustness of our findings. First, we excluded participants who claimed to be retired but were still working, or partly retired, thereby considering the alternative definition of retirement i.e., fully retired. Although the point estimates were similar to the main results, the association between retirement and cognitive function was not statistically significant (as presented in Appendix O).[16]

Second, given that many participants were retired throughout the study period, we restricted the sample to those who reported being engaged in paid work at least once in the interviews (as demonstrated in Model 1 in Appendix P). From Model 1, we excluded those who were self-employed (as shown in Model 2) and further eliminated those who engaged in a part-time job[17] (as demonstrated in Model 3). The results were almost consistent with the main findings even though the associations of retirement with cognitive function and physical inactivity were no longer statistically significant in Model 3.

Third, we narrowed the age range of participants between 52 and 68 years old. While the association between retirement and cognitive function became statistically insignificant, other outcomes were consistent with the main result (as presented in Appendix Q).

Fourth, we present the outcomes of country-by-country analyses in Appendix R, which revealed that certain countries, namely Greece, Latvia, Malta, Portugal, Romania, Costa Rica, Japan, and South Korea, had weak IVs with F statistics below the Stock-Yogo's critical value of 10% maximal relative bias (Stock and Yogo 2002). These were excluded from the analysis. The results obtained after exclusions (presented in Appendix S) were similar to the main findings.

[16] Nishimura, Oikawa, and Motegi (2018) also demonstrated that the robustness of retirement effects estimates across different definitions of retirement in their replication of previous studies.
[17] We defined part-time work as labor force participation of fewer than 35 hours per week.

Moreover, given that 24.1% of the participants were from the United States, we excluded data from the HRS and confirmed through Appendix T that this exclusion did not affect the outcomes. In addition, countries with no changes in their SPA during the study period, namely Cyprus, Finland, Luxembourg, Sweden, Switzerland, Costa Rica, Mexico, China, and Japan, were excluded from the analysis as they did not have a within-country variation in FEIV estimation, and the results obtained were consistent with the main findings (presented in Appendix U).

Sixth, as mentioned above, the length of a word list in the cognitive function test differed by surveys, and thus, we restricted our analysis to surveys with a 10-word list and investigated the association between retirement and the raw scores of cognitive function scores. The results revealed no clear association among men, while female retirees could recall 0.281 more words than workers, which is in line with the main findings (presented in Appendix V).

Seventh, to mitigate potential bias from missing observations, we adopted multiple imputations using the algorithm of expectation–maximization with bootstrapping (Honaker and King 2010) and created ten imputed datasets.[18] Although the association between retirement and cognitive function was not statistically significant using imputed data, other outcomes were similar to the main results (presented in Appendix W).

Finally, we investigated short-term and long-term retirement effects, as previous studies have suggested that these effects are dynamic (Blake and Garrouste, 2019; Calvo, Sarkisian, and Tamborini 2013; Gorry, Gorry, and Slavov 2018; Mazzonna and Peracchi, 2017). Specifically, we compared each group of retirees; that is, those who retired within five years and those who retired

[18] Imputing missing values using a hierarchical model is a common strategy in longitudinal studies. In this case, the imputation model used a linear time trend to account for changes over time and included several variables to predict missing values, such as sex, age, marital status, working status, retirement status, ERA and ORA status, three outcome variables, three health behavior variables, and country. Assuming missing at random, the imputation model estimates the missing values based on the available data and the variables that are predictive of the missing values.

more than five years ago were compared with those who were still working. Appendix X shows that retirement was associated with improved physical independence and self-rated health in both groups, whereas the associations between retirement and cognitive function were not statistically significant. Furthermore, we observed a significant association between retirement and decreased physical inactivity among those who retired more than five years ago but not among those who retired within five years.

5. Discussion and Conclusions

This is the first study to examine the impact of retirement on cognitive function, physical independence, and self-rated health using harmonized longitudinal data from 35 countries. Our FEIV models revealed that retirement was associated with improved cognitive function and physical independence among women. In both sexes, retirement improved self-rated health, but women indicated a larger effect than men.

The sex difference in the effect of retirement on cognitive function has been reported in previous studies (Atalay, Barrett, and Staneva 2019; Ebeid and Oguzoglu 2023); however, its detrimental effect among men was statistically insignificant in our study. The observed effect size of 0.10 SD among women was not negligible, given that Kraft (2020) proposed to consider the effect of 0.05–0.20 SD on cognition as a medium size.[19] Our stratified analysis of health behaviors showed that retirement was associated with decreased physical inactivity and smoking among women but not among men. Post-retirement health behaviors can be mechanisms through which retirement influences health (Eibich 2015). Given that unhealthy lifestyles including physical

---

[19] The author reviewed 1,942 effect sizes from 747 randomized control trials that evaluated the effect of educational interventions on cognitive skills. He found that the distribution of effect sizes had a median of 0.10 SD and even the 90th percentile was under 0.50 SD.

inactivity and smoking are potential risk factors for cognitive and memory declines (Sabia et al. 2009; Kesse-Guyot et al. 2014; Jia et al. 2023), the sex differences in health behaviors may induce the heterogeneous effect of retirement on cognitive function.

Our findings on physical independence were contrary to non-IV studies that showed negative associations with retirement (Dave, Rashad, and Spasojevic 2008; Stenholm et al. 2014). As we demonstrated, the effect of retirement flipped to be positive after adopting the IV, which suggested that the previous studies could not fully address its endogeneity. This study was consistent with more recent studies (Szabó et al. 2019; van Zon et al. 2016). For example, adopting FEIV models, Nishimura, Oikawa, and Motegi (2018) showed the positive impact of retirement on ADL in England using data from the ELSA between 2002 and 2014, as well as in the United States, Germany, Switzerland, Japan, and South Korea. Similar to cognitive function, the estimates of physical independence were consistent with the associations of retirement with physical inactivity and smoking in favor of women.

Similar to many previous studies (Coe and Zamarro 2011; Eibich 2015; Gorry, Gorry, and Slavov 2018; Hessel 2016; Johnston and Lee 2009; Messe and Wolff 2019; Neuman, 2008; Nishimura, Oikawa, and Motegi 2018; Rose 2020; Zhu 2016), we found improved self-rated health among retirees compared with full-time workers. There are several possible pathways that link retirement to better self-rated health. Retirement provides opportunities for individuals to engage in health-promoting activities, including exercise, healthy eating, and adequate sleep (Barnett, van Sluijs, and Ogilvie 2012; Kämpfen and Maurer 2016; Myllyntausta et al. 2018; Helldán et al. 2012). We observed that women exhibited a greater effect size on self-rated health than men, which may reflect their healthier lifestyles after retirement (i.e., decreased physical inactivity and smoking). Furthermore, retirees tend to spend more time in social activities that promote mental and physical well-being (Bogaard, Henkens, and Kalmijn 2014; Kobayashi et al. 2022). Relief from job strain can be another mechanism explaining the relationship between retirement and improved self-rated health (Lerner et al. 1994).

There were several limitations in this study. First, certain discrepancies across surveys were recognized despite the fact that field experts harmonized data (Shih, Lee, and Das 2012; Zamarro and Lee 2012; Lee, Phillips, and Wilkens 2021). Even though these discrepancies could lead to estimations being biased, some potential biases could be eliminated by including the country FEs. Second, measurement errors could occur because most of the measures were self-reported. Nonetheless, the performance of outcome measures has been validated (Morris et al. 1989; Welsh et al. 1994; Roehrig et al. 2007; Bond et al. 2006; Brenowitz et al. 2014; Idler and Benyamini 1997). In addition, the straightforward inquiry into retirement status ensured face validity to measure individuals' recognition, which could induce behavioral adjustments (Eibich 2015). Third, further studies are needed to determine the mechanism linking retirement to improved outcomes. Our study demonstrated sex differences in physical inactivity and smoking after retirement, which was consistent with the heterogeneous effect of retirement on health by sex. Nonetheless, mediation analysis is necessary to confirm the mechanism. In addition, there may be other factors such as sleep, diet, and social participation that were not provided in the harmonized data.

This study suggests that retirement benefits health, especially for women. While we observed sex heterogeneity, the effects of retirement on the outcomes appeared constant across different educational levels, pre-retirement job characteristics, and countries. Notably, increasing the SPA delays retirement timing and might dampen citizens' health. Nevertheless, promoting healthy behaviors such as engaging in physical activity and refraining from smoking can offset the potentially detrimental effects of delayed retirement and contribute to realizing healthy aging.

References

Agahi, Neda, Stefan Fors, Johan Fritzell, and Benjamin A Shaw. 2018. "Smoking and Physical Inactivity as Predictors of Mobility Impairment During Late Life: Exploring Differential Vulnerability Across Education Level in Sweden." *The Journals of Gerontology: Series B* 73 (4): 675–83. https://doi.org/10.1093/geronb/gbw090.

Atalay, Kadir, Garry F. Barrett, and Anita Staneva. 2019. "The Effect of Retirement on Elderly Cognitive Functioning." *Journal of Health Economics* 66 (July): 37–53. https://doi.org/10.1016/j.jhealeco.2019.04.006.

Barnett, Inka, Esther M. F. van Sluijs, and David Ogilvie. 2012. "Physical Activity and Transitioning to Retirement: A Systematic Review." *American Journal of Preventive Medicine* 43 (3): 329–36. https://doi.org/10.1016/j.amepre.2012.05.026.

Behncke, Stefanie. 2012. "Does Retirement Trigger Ill Health?" *Health Economics* 21 (3): 282–300. https://doi.org/10.1002/hec.1712.

Bianchini, Laura, and Margherita Borella. 2016. "Retirement and Memory in Europe." *Ageing & Society* 36: 1434–58. https://doi.org/10.1017/S0144686X15000434.

Bingley, Paul, and Alessandro Martinello. 2013. "Mental Retirement and Schooling." *European Economic Review* 63 (October): 292–98. https://doi.org/10.1016/j.euroecorev.2013.01.004.

Blake, Hélène, and Clémentine Garrouste. 2019. "Collateral Effects of a Pension Reform in France." *Annals of Economics and Statistics*, no. 133: 57–86. https://doi.org/10.15609/annaeconstat2009.133.0057.

Bogaard, Levi Van Den, Kène Henkens, and Matthijs Kalmijn. 2014. "So Now What? Effects of

Retirement on Civic Engagement.” *Ageing & Society* 34 (7): 1170–92. https://doi.org/10.1017/S0144686X13000019.

Bond, John, Heather O. Dickinson, Fiona Matthews, Carol Jagger, Carol Brayne, and MRC CFAS. 2006. “Self-Rated Health Status as a Predictor of Death, Functional and Cognitive Impairment: A Longitudinal Cohort Study.” *European Journal of Ageing* 3 (4): 193–206. https://doi.org/10.1007/s10433-006-0039-8.

Bonsang, Eric, Stéphane Adam, and Sergio Perelman. 2012. “Does Retirement Affect Cognitive Functioning?” *Journal of Health Economics* 31 (3): 490–501. https://doi.org/10.1016/j.jhealeco.2012.03.005.

Brenowitz, Willa D., Rebecca A. Hubbard, Paul K. Crane, Shelly L. Gray, Oleg Zaslavsky, and Eric B. Larson. 2014. “Longitudinal Associations between Self-Rated Health and Performance-Based Physical Function in a Population-Based Cohort of Older Adults.” *PLOS ONE* 9 (11): e111761. https://doi.org/10.1371/journal.pone.0111761.

Calvo, Esteban, Natalia Sarkisian, and Christopher R. Tamborini. 2013. “Causal Effects of Retirement Timing on Subjective Physical and Emotional Health.” *The Journals of Gerontology: Series B* 68 (1): 73–84. https://doi.org/10.1093/geronb/gbs097.

Centers for Disease Control and Prevention. 2022. “Binge Drinking.” Centers for Disease Control and Prevention. January 6, 2022. https://www.cdc.gov/alcohol/fact-sheets/binge-drinking.htm.

Coe, Norma B., Hans-Martin von Gaudecker, Maarten Lindeboom, and Jürgen Maurer. 2012. “The Effect of Retirement on Cognitive Functioning.” *Health Economics* 21 (8): 913–27. https://doi.org/10.1002/hec.1771.

Coe, Norma B., and Maarten Lindeboom. 2008. “Does Retirement Kill You? Evidence from Early

Retirement Windows." *IZA Discussion Paper* No. 3817 (November). https://doi.org/10.2139/ssrn.1298260.

Coe, Norma B., and Gema Zamarro. 2011. "Retirement Effects on Health in Europe." *Journal of Health Economics* 30 (1): 77–86. https://doi.org/10.1016/j.jhealeco.2010.11.002.

Correia, Sergio. 2015. "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix." November 2015. http://scorreia.com/research/singletons.pdf.

Dave, Dhaval, Inas Rashad, and Jasmina Spasojevic. 2008. "The Effects of Retirement on Physical and Mental Health Outcomes." *Southern Economic Journal* 75 (2): 497–523. https://doi.org/10.1002/j.2325-8012.2008.tb00916.x.

Ebeid, Mohamed, and Umut Oguzoglu. 2023. "Short-Term Effect of Retirement on Health: Evidence from Nonparametric Fuzzy Regression Discontinuity Design." *Health Economics*, March. https://doi.org/10.1002/hec.4669.

Eibich, Peter. 2015. "Understanding the Effect of Retirement on Health: Mechanisms and Heterogeneity." *Journal of Health Economics* 43 (September): 1–12. https://doi.org/10.1016/j.jhealeco.2015.05.001.

Garrouste, Clémentine, and Elsa Perdrix. 2022. "Is There a Consensus on the Health Consequences of Retirement? A Literature Review." *Journal of Economic Surveys* 36 (4): 841–79. https://doi.org/10.1111/joes.12466.

Gorry, Aspen, Devon Gorry, and Sita Nataraj Slavov. 2018. "Does Retirement Improve Health and Life Satisfaction?" *Health Economics* 27 (12): 2067–86. https://doi.org/10.1002/hec.3821.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.

Helldán, Anni, Tea Lallukka, Ossi Rahkonen, and Eero Lahelma. 2012. "Changes in Healthy Food

Habits after Transition to Old Age Retirement.” *European Journal of Public Health* 22 (4): 582–86. https://doi.org/10.1093/eurpub/ckr060.

Hessel, Philipp. 2016. “Does Retirement (Really) Lead to Worse Health among European Men and Women across All Educational Levels?” *Social Science & Medicine (1982)* 151 (February): 19–26. https://doi.org/10.1016/j.socscimed.2015.12.018.

Honaker, James, and Gary King. 2010. “What to Do About Missing Values in Time Series Cross-Section Data.” *American Journal of Political Science* 54 (3): 561–81.

Idler, E. L., and Y. Benyamini. 1997. “Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies.” *Journal of Health and Social Behavior* 38 (1): 21–37.

Jia, Jianping, Tan Zhao, Zhaojun Liu, Yumei Liang, Fangyu Li, Yan Li, Wenying Liu, et al. 2023. “Association between Healthy Lifestyle and Memory Decline in Older Adults: 10 Year, Population Based, Prospective Cohort Study.” *BMJ* 380 (January): e072691. https://doi.org/10.1136/bmj-2022-072691.

Johnston, David W., and Wang-Sheng Lee. 2009. “Retiring to the Good Life? The Short-Term Effects of Retirement on Health.” *Economics Letters* 103 (1): 8–11. https://doi.org/10.1016/j.econlet.2009.01.015.

Kämpfen, Fabrice, and Jürgen Maurer. 2016. “Time to Burn (Calories)? The Impact of Retirement on Physical Activity among Mature Americans.” *Journal of Health Economics* 45 (January): 91–102. https://doi.org/10.1016/j.jhealeco.2015.12.001.

Kesse-Guyot, Emmanuelle, Valentina A. Andreeva, Camille Lassale, Serge Hercberg, and Pilar Galan. 2014. “Clustering of Midlife Lifestyle Behaviors and Subsequent Cognitive Function: A Longitudinal Study.” *American Journal of Public Health* 104 (11): e170-177. https://doi.org/10.2105/AJPH.2014.302121.

Kobayashi, Erika, Ikuko Sugawara, Taro Fukaya, Shohei Okamoto, and Jersey Liang. 2022. "Retirement and Social Activities in Japan: Does Age Moderate the Association?" *Research on Aging* 44 (2): 144–55. https://doi.org/10.1177/01640275211005185.

Kraft, Matthew A. 2020. "Interpreting Effect Sizes of Education Interventions." *Educational Researcher* 49 (4): 241–53. https://doi.org/10.3102/0013189X20912798.

Lee, Jinkook, Drystan Phillips, Jenny Wilkens, and Gateway to Global Aging Data Team. 2021. "Gateway to Global Aging Data: Resources for Cross-National Comparisons of Family, Social Environment, and Healthy Aging." *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 76 (Suppl 1): S5–16. https://doi.org/10.1093/geronb/gbab050.

Lei, Xiaoyan, and Hong Liu. 2018. "Gender Difference in the Impact of Retirement on Cognitive Abilities: Evidence from Urban China." *Journal of Comparative Economics* 46 (4): 1425–46. https://doi.org/10.1016/j.jce.2018.01.005.

Lerner, D J, S Levine, S Malspeis, and R B D'Agostino. 1994. "Job Strain and Health-Related Quality of Life in a National Sample." *American Journal of Public Health* 84 (10): 1580–85. https://doi.org/10.2105/AJPH.84.10.1580.

Maurage, P., F. Joassin, A. Speth, J. Modave, P. Philippot, and S. Campanella. 2012. "Cerebral Effects of Binge Drinking: Respective Influences of Global Alcohol Intake and Consumption Pattern." *Clinical Neurophysiology* 123 (5): 892–901. https://doi.org/10.1016/j.clinph.2011.09.018.

Mazzonna, Fabrizio, and Franco Peracchi. 2017. "Unhealthy Retirement?" *Journal of Human Resources* 52 (1): 128–51. https://doi.org/10.3368/jhr.52.1.0914-6627R1.

Messe, Pierre-Jean, and François-Charles Wolff. 2019. "The Short-Term Effects of Retirement on

Health within Couples: Evidence from France." *Social Science & Medicine (1982)* 221 (January): 27–39. https://doi.org/10.1016/j.socscimed.2018.12.008.

Moore, Alison A., Justin O. Endo, and M. Kallin Carter. 2003. "Is There a Relationship Between Excessive Drinking and Functional Impairment in Older Persons?" *Journal of the American Geriatrics Society* 51 (1): 44–49. https://doi.org/10.1034/j.1601-5215.2002.51008.x.

Morris, J. C., A. Heyman, R. C. Mohs, J. P. Hughes, G. van Belle, G. Fillenbaum, E. D. Mellits, and C. Clark. 1989. "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and Neuropsychological Assessment of Alzheimer's Disease." *Neurology* 39 (9): 1159–65. https://doi.org/10.1212/wnl.39.9.1159.

Myllyntausta, Saana, Paula Salo, Erkki Kronholm, Jaana Pentti, Mika Kivimäki, Jussi Vahtera, and Sari Stenholm. 2018. "Changes in Sleep Difficulties During the Transition to Statutory Retirement." *Sleep* 41 (1): zsx182. https://doi.org/10.1093/sleep/zsx182.

Neuman, Kevin. 2008. "Quit Your Job and Get Healthier? The Effect of Retirement on Health." *Journal of Labor Research* 29 (2): 177–201. https://doi.org/10.1007/s12122-007-9036-8.

Nishimura, Yoshinori, Masato Oikawa, and Hiroyuki Motegi. 2018. "What Explains the Difference in the Effect of Retirement on Health? Evidence from Global Aging Data." *Journal of Economic Surveys* 32 (3): 792–847. https://doi.org/10.1111/joes.12215.

Okusaga, Olaoluwa, Marlene C. W. Stewart, Isabella Butcher, Ian Deary, F. Gerry R. Fowkes, and Jackie F. Price. 2013. "Smoking, Hypercholesterolaemia and Hypertension as Risk Factors for Cognitive Impairment in Older Adults." *Age and Ageing* 42 (3): 306–11. https://doi.org/10.1093/ageing/afs193.

Organisation for Economic Co-operation and Development. 2021. *Pensions at a Glance 2021:*

OECD and G20 Indicators. Paris: OECD Publishing. https://doi.org/10.1787/ca401ebd-en.

Organisation for Economic Co-operation and Development and World Health Organization. 2020. *Health at a Glance: Asia/Pacific 2020: Measuring Progress Towards Universal Health Coverage*. Paris: OECD Publishing. https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-asia-pacific-2020_26b007cd-en.

Roehrig, Bernd, Klaus Hoeffken, Ludger Pientka, and Ulrich Wedding. 2007. "How Many and Which Items of Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) Are Necessary for Screening." *Critical Reviews in Oncology/Hematology* 62 (2): 164–71. https://doi.org/10.1016/j.critrevonc.2006.10.001.

Romero Starke, Karla, Andreas Seidler, Janice Hegewald, Anna Klimova, and Katie Palmer. 2019. "Retirement and Decline in Episodic Memory: Analysis from a Prospective Study of Adults in England." *International Journal of Epidemiology* 48 (6): 1925–36. https://doi.org/10.1093/ije/dyz135.

Rose, Liam. 2020. "Retirement and Health: Evidence from England." *Journal of Health Economics* 73 (September): 102352. https://doi.org/10.1016/j.jhealeco.2020.102352.

Sabia, Séverine, Hermann Nabi, Mika Kivimaki, Martin J. Shipley, Michael G. Marmot, and Archana Singh-Manoux. 2009. "Health Behaviors from Early to Late Midlife as Predictors of Cognitive Function: The Whitehall II Study." *American Journal of Epidemiology* 170 (4): 428–37. https://doi.org/10.1093/aje/kwp161.

Sato, Koryu, Naoki Kondo, Masamichi Hanazato, Taishi Tsuji, and Katsunori Kondo. 2021. "Potential Causal Effect of Physical Activity on Reducing the Risk of Dementia: A 6-Year Cohort Study from the Japan Gerontological Evaluation Study." *The International Journal of Behavioral Nutrition and Physical Activity* 18 (1): 140. https://doi.org/10.1186/s12966-

021-01212-w.

Shih, Regina A., Jinkook Lee, and Lopamudra Das. 2012. "Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Cognition." RAND Corporation. https://www.rand.org/pubs/working_papers/WR861z7.html.

Stenholm, Sari, Hugo Westerlund, Paula Salo, Martin Hyde, Jaana Pentti, Jenny Head, Mika Kivimäki, and Jussi Vahtera. 2014. "Age-Related Trajectories of Physical Functioning in Work and Retirement: The Role of Sociodemographic Factors, Lifestyle and Disease." *Journal of Epidemiology and Community Health* 68 (6): 503–9. https://doi.org/10.1136/jech-2013-203555.

Stock, James H., and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear IV Regression." Working Paper 284. Technical Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/t0284.

Szabó, Ágnes, Joanne Allen, Christine Stephens, and Fiona Alpass. 2019. "Is Retirement Associated with Physical Health Benefits? A Longitudinal Investigation with Older New Zealanders." *Age and Ageing* 48 (2): 267–72. https://doi.org/10.1093/ageing/afy176.

the United States Social Security Administration. 2020. "Social Security Programs Throughout the World." March 2020. https://www.ssa.gov/policy/docs/progdesc/ssptw/index.html.

the World Bank. 2022. "World Development Indicators." The World Bank. May 25, 2022. https://databank.worldbank.org/source/world-development-indicators/.

Tulving, Endel. 2002. "Episodic Memory: From Mind to Brain." *Annual Review of Psychology* 53 (1): 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114.

Welsh, K. A., N. Butters, R. C. Mohs, D. Beekly, S. Edland, G. Fillenbaum, and A. Heyman. 1994. "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A

Normative Study of the Neuropsychological Battery." *Neurology* 44 (4): 609–14. https://doi.org/10.1212/wnl.44.4.609.

Zamarro, Gema, and Jinkook Lee. 2012. "Harmonization of Cross-National Studies of Aging to the Health and Retirement Study: Employment and Retirement Measures." RAND Corporation. https://www.rand.org/pubs/working_papers/WR861z4.html.

Zhu, Rong. 2016. "Retirement and Its Consequences for Women's Health in Australia." *Social Science & Medicine (1982)* 163 (August): 117–25. https://doi.org/10.1016/j.socscimed.2016.04.003.

Zon, Sander K. R. van, Ute Bültmann, Sijmen A. Reijneveld, and Carlos F. Mendes de Leon. 2016. "Functional Health Decline before and after Retirement: A Longitudinal Analysis of the Health and Retirement Study." *Social Science & Medicine (1982)* 170 (December): 26–34. https://doi.org/10.1016/j.socscimed.2016.10.002.

Appendixes

A. Measurement of labor force status

1. SHARE

SHARE asks participants, “In general, how would you describe your current situation?” They then choose the best description of their current labor force status from a list of options: 1) retired, 2) employed or self-employed (including working for a family business), 3) unemployed and looking for work, 4) permanently sick or disabled, 5) homemaker, or 6) other (renter, living off own property, student, or doing voluntary work). The harmonized variable was constructed based on responses to this direct question.

2. ELSA

ELSA asks participants, “Which of these, would you say, best describes your situation?” They then choose the best description of their current labor force status from a list of options: 1) employed, 2) self-employed, 3) unemployed, 4) partly retired, 5) retired, 6) permanently sick or disabled, or 7) looking after home or family. The harmonized variable was constructed based on responses to this direct question.

3. CRELES

CRELES asks participants whether they have ever had a job for which they received payment in money or kind. If the respondent answered yes, they were asked what they did during most of the last week: worked, worked with a family business, did not work but had a job, looking for work, did household chores, or did not work. Participants who were not working were then asked if they had not worked: less than 2 years, more than 2 years, or had never worked.

If the participant answered that they worked, helped with a family business, or did not work last week but had a job, the harmonized variable was set to “working.” If the participant reported that they were looking for work, it was set to “unemployed.” If the participant reported that they were doing household chores, it was set to “doing household chores.” If the participant reported that they had worked in the past but were not currently working, it was set to “retired.” If the participant reported that they had never worked, it was set to “never worked.”

4. MHAS

MHAS asks participants whether they have worked or are currently working. In waves 2–4, it also asks the main reason why they were not currently working: dedicated to household chores, retired, old age, sick or temporarily disabled, unable to work for the rest of life, and does not have customers or cannot find work.

If the participant reported that they were currently working, the harmonized variable was set to “working.” If the participant is currently looking for work or does not work but “does not have customers or cannot find work,” it is set to “unemployed.” If the participant mentioned retirement, regardless of current work, it was set to “retired.” If the participant is “sick or temporarily disabled” or “unable to work for rest of life,” it is set to “disabled.” Otherwise, the variable is set to “not in the labor force.” The question asking the reason for not working is not included in wave 1; thus, the harmonized variable has an integrated category indicating “unemployed, retired, or disabled” in wave 1. If the present study, we treated those in the integrated category in wave 1 as retirees only if they were categorized as “retired” in wave 2; otherwise, they were not included in the analyses.

5. HRS

Participants in HRS provided information on their labor force status at several time points in an interview. First, HRS asks the participants to select all applicable options from a list that includes 1) working now, 2) unemployed and looking for work, 3) temporarily laid off, on sick or other leave, 4) disabled, 5) retired, 6) homemaker, or 7) other (specify). It also asks them whether they are currently working for payment, the usual number of hours per week if applicable, and whether they consider themselves partly retired, completely retired, or not retired.

If the participant reports working full-time (i.e., working 35+ hours per week or 36+ weeks per year), the harmonized variable is set to "working full-time." If the participant is working part-time and does not mention retirement, it is set to "working part-time." If the participant is working part-time and mentions retirement, it is set to "partly retired." If the participant is not working but is looking for a job, it is set to "unemployed." If the participant is not looking for a job and there is any mention of retirement, it is set to "retired." If retirement is not mentioned and disabled employment status is given, it is set to "disabled." Otherwise, the variable is set to "not in the labor force."

6. CHARLS

CHARLS asks participants whether they engaged in agricultural work for more than 10 days in the past year, worked for at least 1 hour last week if not engaged in agricultural work, were temporarily laid-off or on sick or other leave, worked for at least a few months, whether they were homemakers, completed retirement procedures, and currently retired (including early retirement or internal retirement).

If the participants report working for other farmers, the harmonized variable is set to "agricultural employed." If the participants reported working for their household, it is set to "agricultural self-employed." If the participants describe their non-agricultural job as employment, it is set to "non-agricultural employed." If the participants describe their non-agricultural job as self-employment, it is set to "non-agricultural self-employed." If the participants describe their non-agricultural job as an unpaid family business, it is set to "non-agricultural unpaid family business." If the participants report not currently working but had worked for at least 3 months and have searched for a job in the past month, it is set to "unemployed." If the participants declare to have completed retirement procedures or describe themselves as retired, it is set to "retired." If the participants reported never worked, it was set to "never worked."

7. JSTAR

JSTAR asks participants whether they are currently employed, looking for a job, or intend to look for work in the future. If they were neither a worker nor a job seeker, they were asked about their current status with the following response options: 1) retired, 2) keep house, 3) receive medical care, 4) other, 5) do not know, and 6) refused to answer.

If the participant reports working full-time or working 35+ hours per week and 36+ weeks per year, the harmonized variable is set to "working full-time." If the participant reports working part-time or less than 35 hours per week or 36 weeks per year, it is set to "working part-time." If the participant reports currently working as an owner of an independent business or having a side job at home, it is set to "self-employed."

If the participant reports not working but is looking for a job and there is no mention of retirement, it is set to "unemployed." If the participant reports looking for a part-time job and mentions retirement, it is set to "partly retired." If the participant is not working and not looking for work, and there is any mention of retirement, it is set to "retired." If retirement is not mentioned and disabled employment status is given, it is set to "disabled." If neither retirement nor disability is mentioned, but a homemaker situation is given, it is set to "not in the labor force."

8. KLoSA

KLoSA asks participants whether they are currently working or looking for a job. If they are neither a worker nor a job seeker, they are asked about their retirement status with the following response options: 1) worked before but currently retired, 2) worked before and intended to work in the future but currently not looking for a job, and 3) never had a job before.

If the participant is employed by another person or company for payment, the harmonized variable is set to "employed full-time" or "employed part-time," based on the working classification the participant gave for the job. If the participant report being self-employed, it is set to "self-employed." If the participant is employed and reports working without payment for family more than 18 hours per week, it is set to "help with family 18 hours or more per week." If a non-working participant is looking for work and reports being able to work if offered a job and then confirms that they have done something to find work in the last 4 weeks, it is set to "unemployed." If a non-working participant is not looking for work and reports being retired, it is set to "retired." If the participant reports being retired but later mentions working for payment or looking for paid work, it is set to "partly retired." If a non-working participant is looking for work but then reports not being able to accept work or looking for work due to poor health or a disability, it is set to "disabled." Otherwise, it is set to "not in labor force."

B. Summary of the Harmonized Variable of Labor Force Status

Retirement status was determined based on the harmonized variable of labor force status (RwLBRF). The table below summarizes the codes of labor force status for each survey and how we treated them.

This study	SHARE	ELSA	CRELES	MHAS	HRS	CHARLS (urban residents only)	JSTAR	KloSA
Included as those "working"	1. employed or self employed	1. employed	1. working	1. working	1. working full-time	1. agricultural employed	1. working full-time	1. working full-time
		2. self-employed			2. working part-time	2. agricultural self-employed	2. working part-time	2. working part-time
						3. non-agricultural employed	8. self-employed	3. self-employed
						4. non-agricultural self-employed		4. help with family 18 hours or more per week
						5. non-agricultural unpaid family business		
Included as those being "retired"	5. retired	4. partly retired	4. retired	3. retired	4. partly retired	7. retired	4. partly retired	6. partly retired
		5. retired			5. retired		5. retired	7. retired
Excluded from analyses	3. unemployed	3. unemployed	2. unemployed	2. unemployed	3. unemployed	6. unemployed	3. unemployed	5. unemployed
	6. permanently sick or disabled	6. disabled	3. doing household chores	4. disabled	6. disabled	8. never worked	6. disabled	8. disabled
	8. homemaker	7. looking after home or family	5. never worked	5. not in labor force	7. not in labor force		7. not in labor force	9. not in labor force

C. Early and official retirement age for each country

Country	Year	Men		Women	
		ERA	ORA	ERA	ORA
Austria[a]	2018	NA	65	NA	60
Belgium[b]	2018	63	65	63	65
Bulgaria[c]	2018	63.08	64.08	60.17	61.17
Croatia[d]	2018	60	65	57	62
Cyprus	2018	63	65	63	65
Czech Republic[e]	2018	60	63.16	59.66	62.66
Denmark[f]	2018	NA	65	NA	65
England[g]	2018	NA	65	NA	65
Estonia[h]	2018	60.5	63.5	60.5	63.5
Finland	2018	63	65	63	65
France[i]	2018	62	67	62	67
Germany[j]	2018	63	65.58	63	65.58
Greece[k]	2018	62	67	62	67
Hungary[l]	2018	NA	63.5	NA	63.5
Israel[m]	2018	NA	67	NA	62
Italy[n]	2018	62	66.58	62	66.58
Latvia[o]	2018	61.25	63.25	61.25	63.25
Lithuania[p]	2018	58.67	63.67	57.33	62.33
Luxembourg	2018	57	65	57	65
Malta[q]	2018	61	62	61	62
Netherlands[r]	2018	NA	66	NA	66
Poland[s]	2018	NA	65	NA	60
Portugal[t]	2018	60	66.33	60	66.33
Romania[u]	2018	60	65	55.92	60.92
Slovakia[v]	2018	60.42	62.42	60.42	62.42
Slovenia[w]	2018	60	65	59.67	64
Spain[x]	2018	61.5	65.5	61.5	65.5
Sweden[y]	2018	61	65	61	65
Switzerland	2018	63	65	62	64
Costa Rica	2013	62	65	60	65
Mexico	2018	60	65	60	65
United States[z]	2018	62	66	62	66
China[aa]	2016	NA	60	NA	50
Japan	2012	60	65	60	65
South Korea[ab]	2018	57	61	57	61

Source: The United States Social Security Administration "Social Security Programs Throughout the World"; Organisation for Economic Co-operation and Development "Pensions at a Glance"; websites of the authorities of each country.

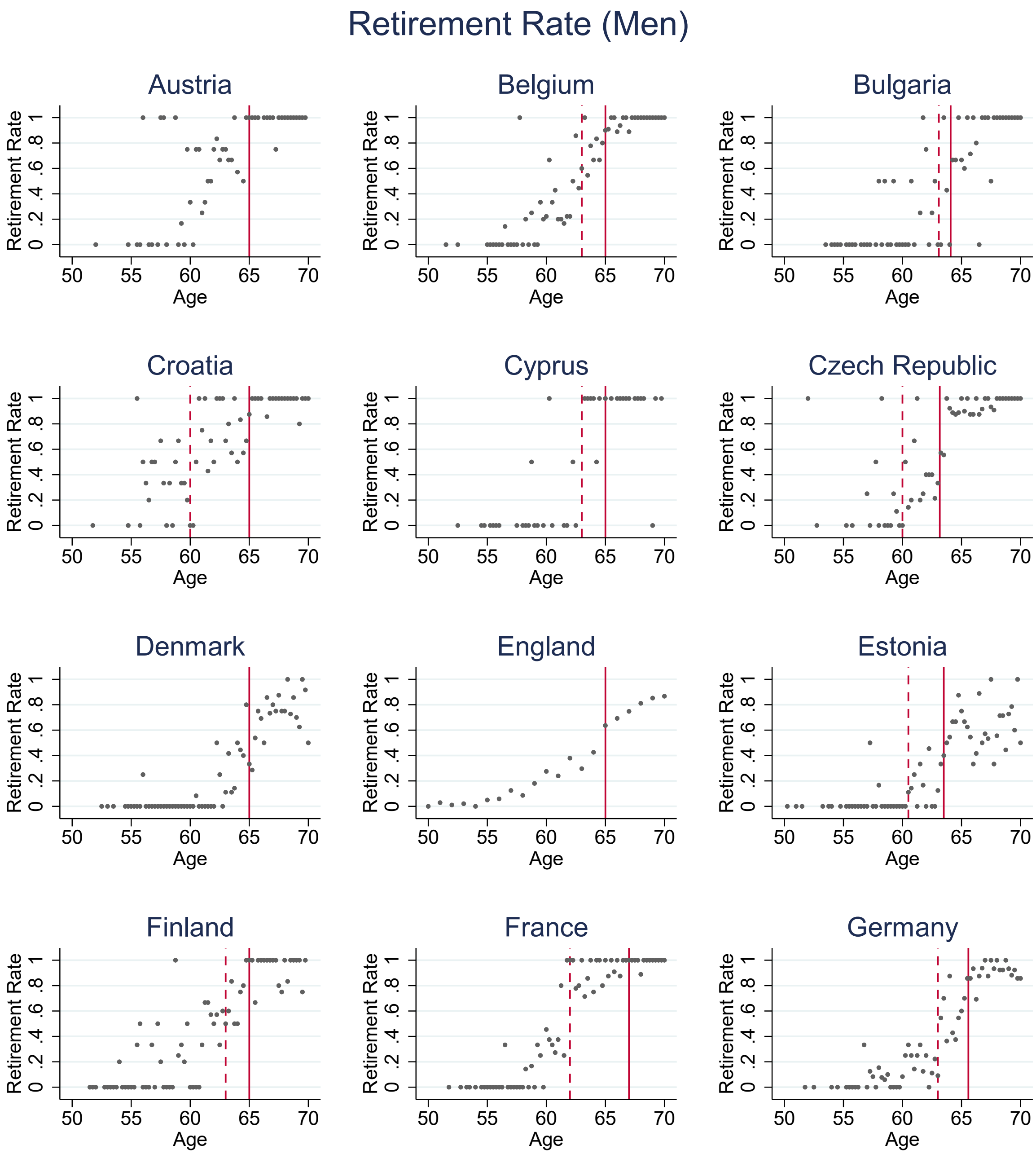
Note: ERA, ORA, and NA denote early retirement age, official retirement age, and not applicable, respectively.

[a] ERA was 61.5 for men and 56.5 for women in 2004 and gradually increased to be phased out in 2017.

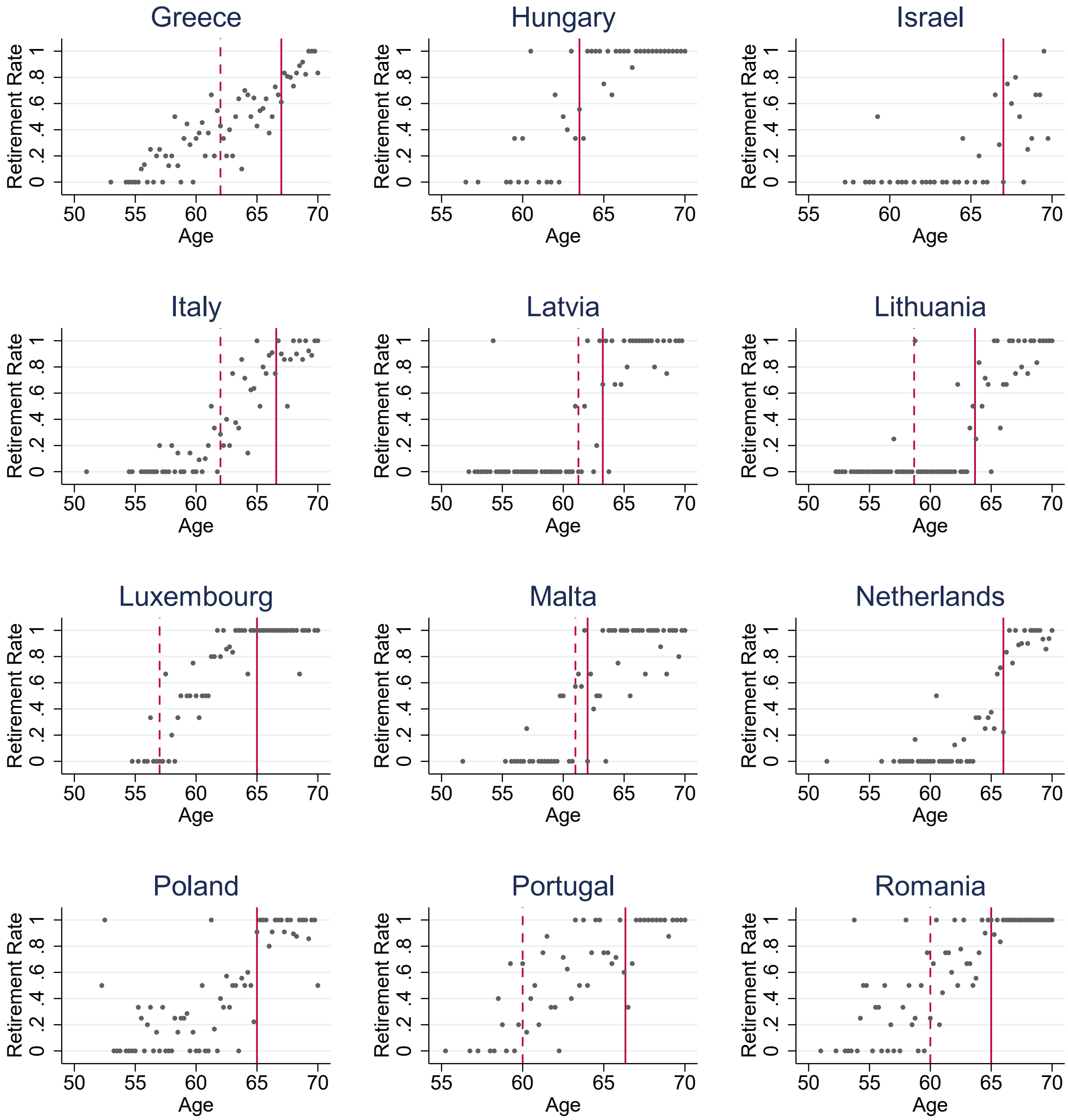
[b] ERA gradually increased from age 60 to 63 from 2013 to 2018.
[c] ORA is gradually increasing from age 63 to 65 by 2029 for men and from age 60 to 65 by 2037 for women. Early retirement is possible up to one year prior to the ORA.
[d] ERA and ORA for women are gradually increasing from age 55 to 60 and 60 to 65 by 2030, respectively.
[e] ORA is gradually increasing from age 60 to 65 for men and 57 to 65 for women without children by 2030.
[f] ORA was 67 for those who reached age 60 before 1 July 1999. ORA is gradually increasing to age 67 from 2019 to 2022.
[g] ORA for women gradually increased from age 60 to 65 from 2010 to 2018.
[h] ORA is gradually increasing from age 63 to 65 from 2017 to 2026. Early retirement is possible up to three years prior to the ORA.
[i] ERA is increasing from age 60 to 62, and ORA from 65 to 67, depending on the year of birth.
[j] ERA and ORA are gradually increasing from age 63 to 65 and 65 to 67 by 2029, respectively.
[k] ERA increased from age 60 to 62 for men and 55 to 62 for women, and ORA from 65 to 67 for men and 60 to 67 for women in 2013.
[l] ORA is gradually increasing from age 62 to 65 by 2022.
[m] ORA is increasing from age 65 to 67 for men and 60 to 62 for women, depending on the year of birth.
[n] ERA gradually increased from age 57 to 62, and ORA from 65 to 67 for men and 60 to 67 for women in 2019.
[o] ORA is gradually increasing from age 62 to 65 from 2013 to 2025. Early retirement is possible up to two years prior to the ORA.
[p] ORA is gradually increasing to age 65 by 2026. Early retirement is possible up to five years prior to the ORA.
[q] ORA is gradually increasing to age 65, depending on the year of birth.
[r] ORA is gradually increasing to age 67 by 2024, depending on the year of birth.
[s] ORA increased from age 65 to 65.58 for men and 60 to 60.58 for women from 2012 to 2015 but returned to age 65 and 60 in 2017.
[t] ERA increased from age 55 to 60 in 2015, and ORA is gradually increasing to age 66.5 by 2021.
[u] ORA for women is gradually increasing to age 63 by 2030. Early retirement is possible up to five years prior to the ORA.
[v] ORA is gradually increasing from age 62 based on increases in life expectancy from 2016. Early retirement is possible up to two years prior to the ORA.
[w] ERA (with 40 years of contribution) gradually increased from age 58 to 60 in 2018 for men and 2019 for women. ORA (with 20 years of contribution) gradually increased from age 63 to 65 from 2012 to 2016 for men and 61 to 65 from 2012 to 2020 for women.
[x] ORA is gradually increasing from age 65 to 67 from 2012 to 2027. Early retirement is possible up to four years prior to the ORA in the case of involuntary unemployment.
[y] The earning-related national pension and guarantee pension benefits are available from ages 61 and 65, respectively.
[z] ORA gradually increased from age 65 to 66, depending on the year of birth.
[aa] ORA of 50 is for non-professional salaried women.
[ab] ERA is gradually increasing from age 55 to 60 from 2012 to 2029, and ORA from 60 to 65 from 2012 to 2034.

D. Retirement rate by country

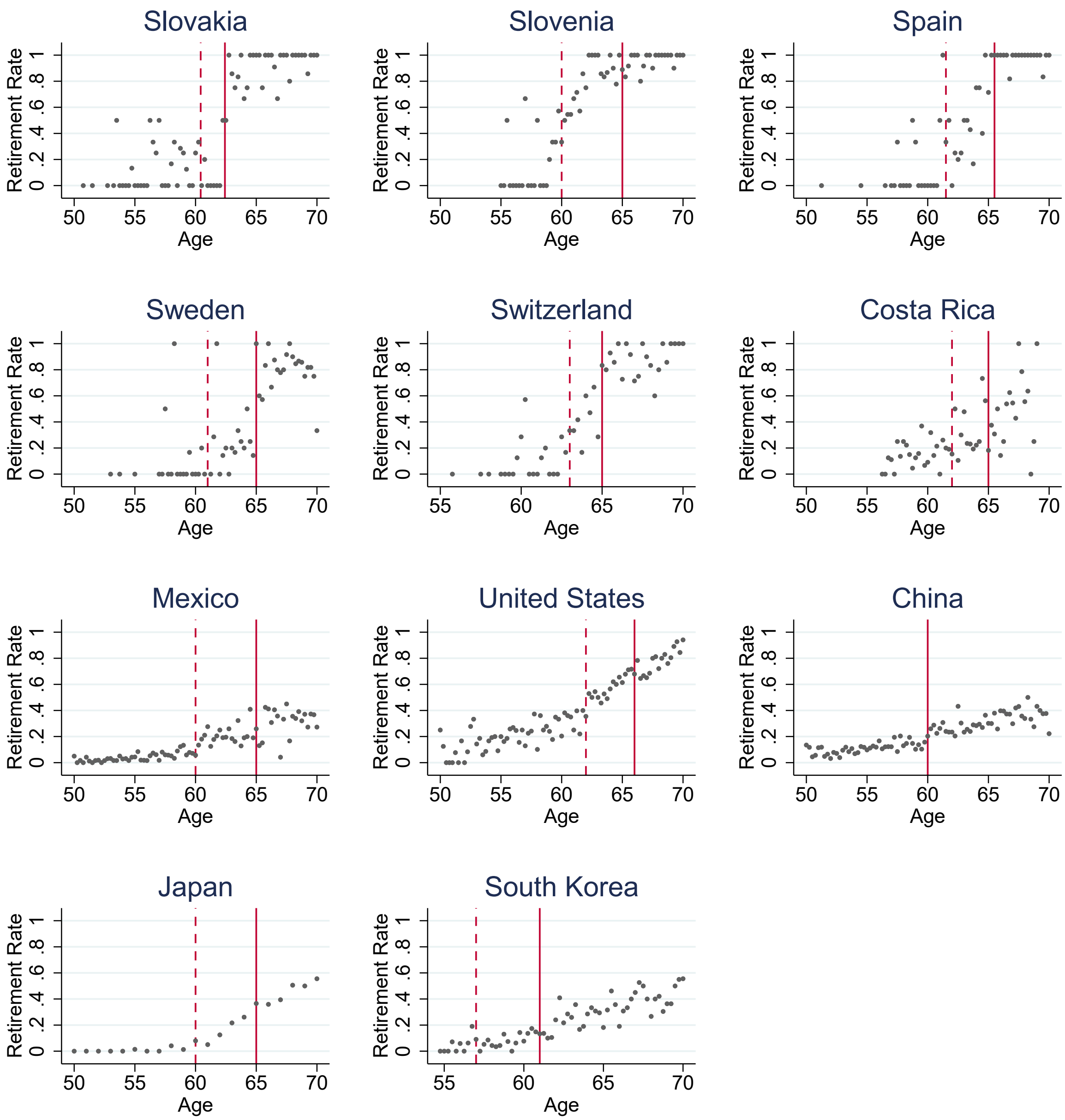
The latest surveys were utilized to illustrate the figures. Each individual dot represents the average retirement rate for each 3-month intervals, as monthly age data was unavailable in England and Japan. The retirement rate is calculated by dividing the number of retirees by the sum of retirees and non-retired individuals who are not working due to reasons other than retirement (such as being unemployed, disabled, or homemaker], with the exclusion of the latter. The dashed red line denotes the ERA, while the solid red line represents the ORA that corresponds to the survey year.



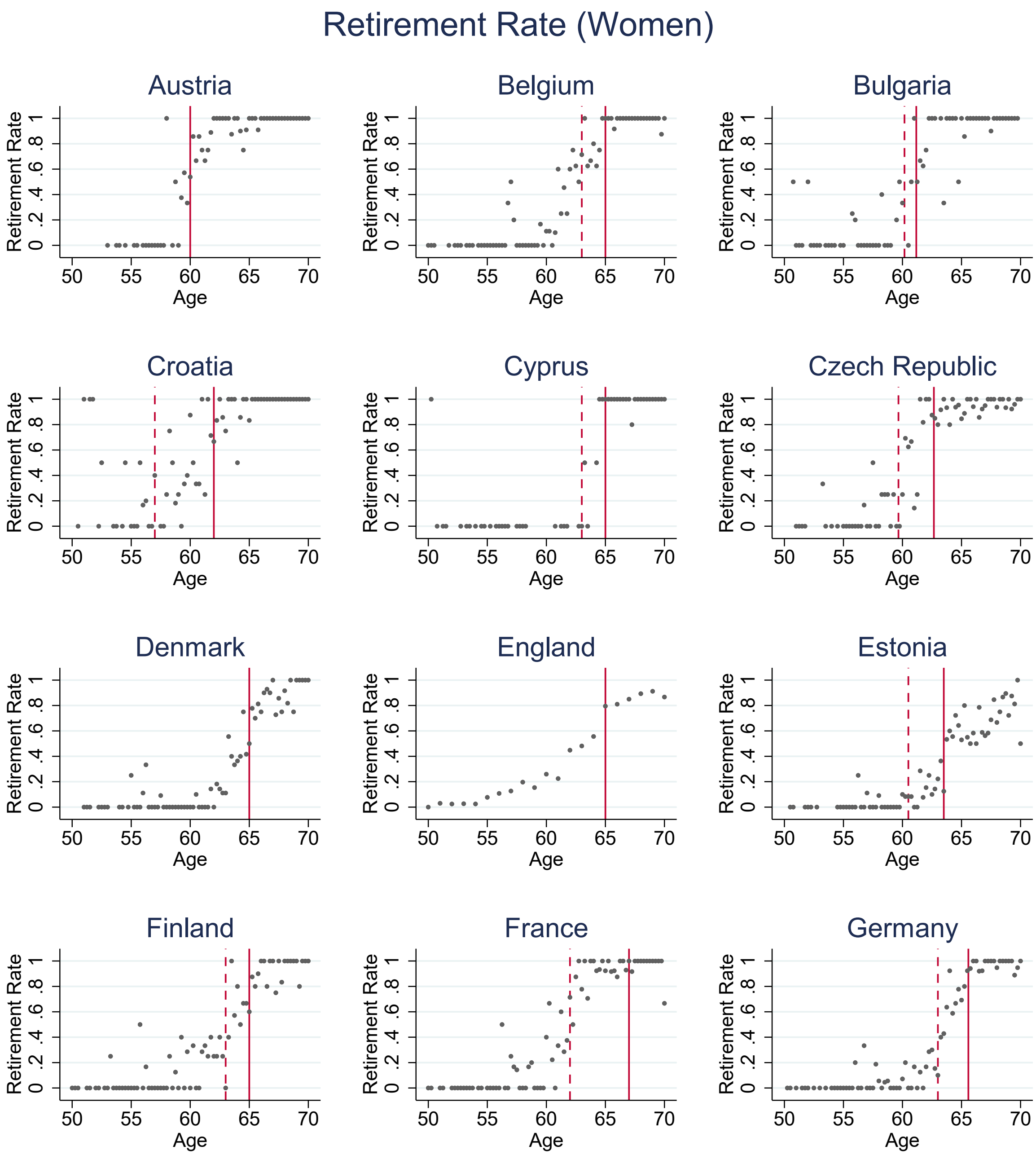
Retirement Rate (Men, cont.)

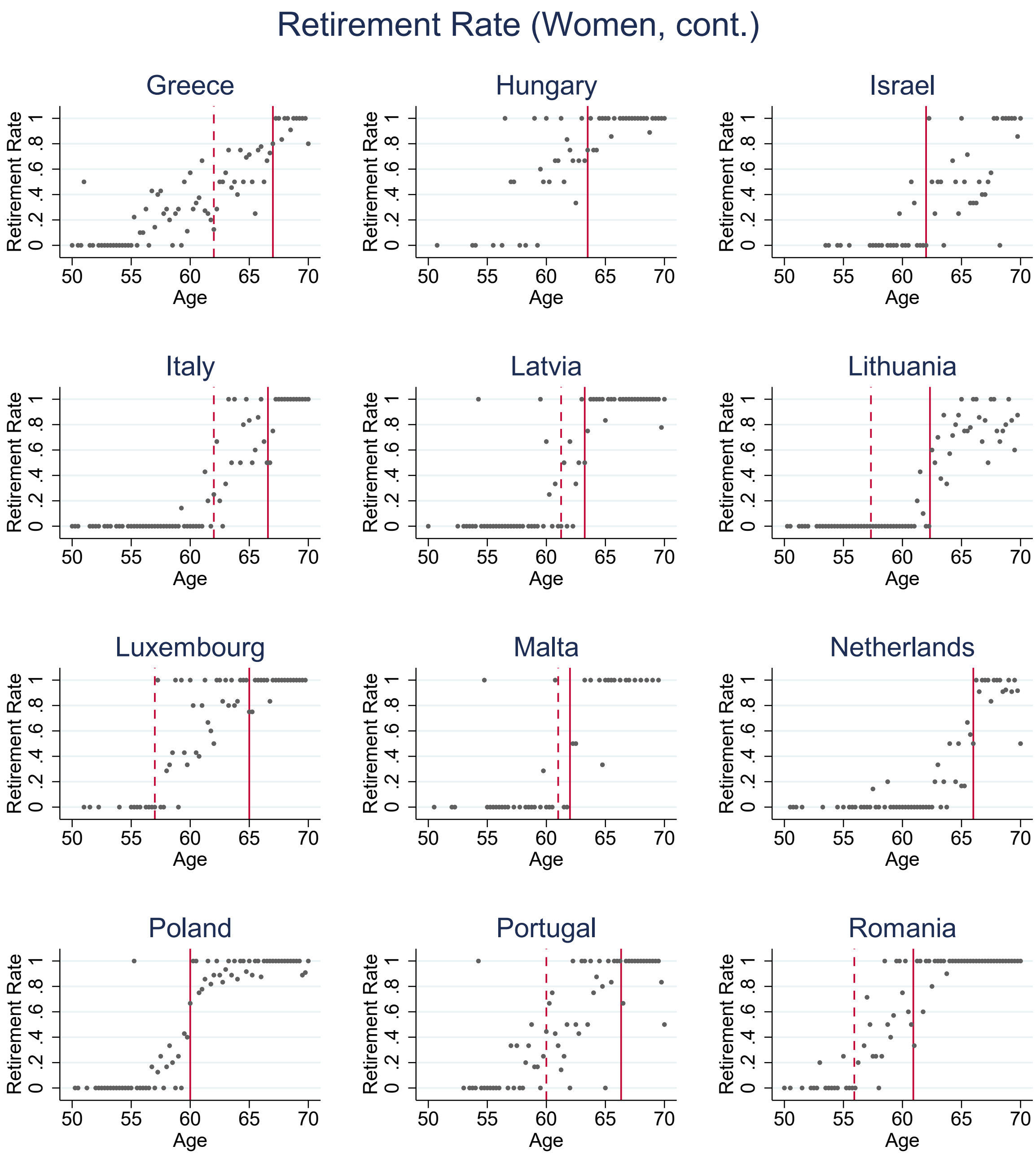


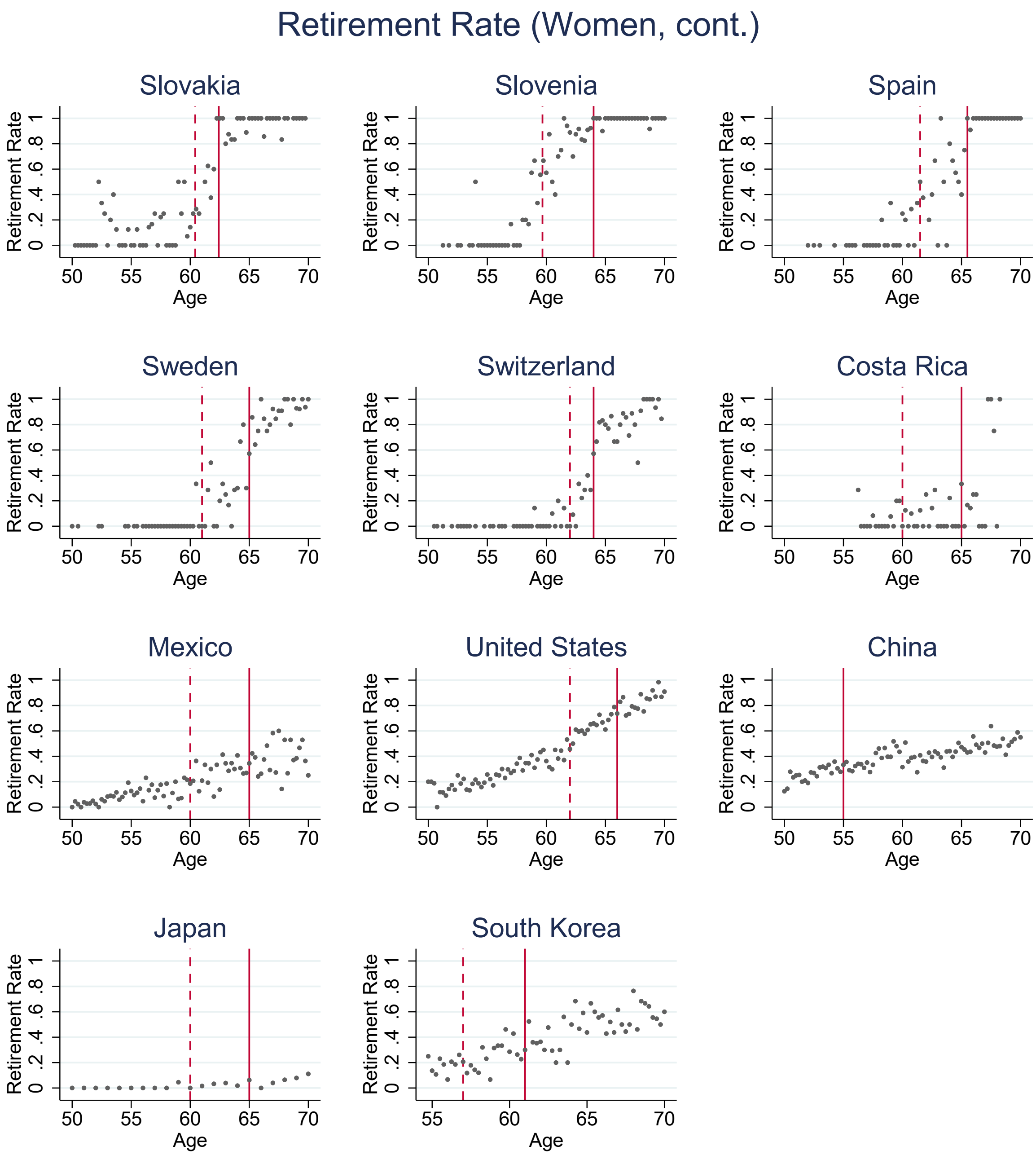
Retirement Rate (Men, cont.)



Appendix Figure D.1. Men's retirement rate by country

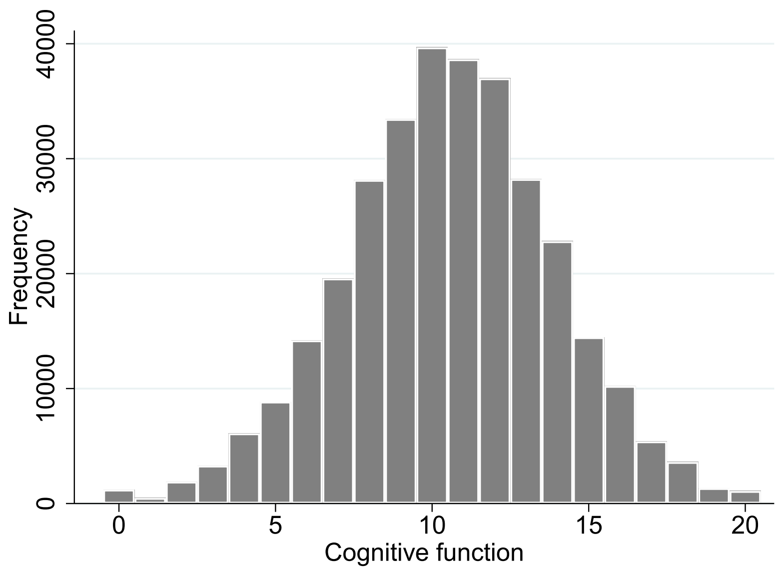

Retirement Rate (Women)
Austria
Belgium
Bulgaria
Croatia
Cyprus
Czech Republic
Denmark
England
Estonia
Finland
France
Germany
Retirement Rate
0 .2 .4 .6 .8 1
50 55 60 65 70
Age


Retirement Rate (Women, cont.)
Greece
Hungary
Israel
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Poland
Portugal
Romania
Retirement Rate
0 .2 .4 .6 .8 1
Age
50 55 60 65 70

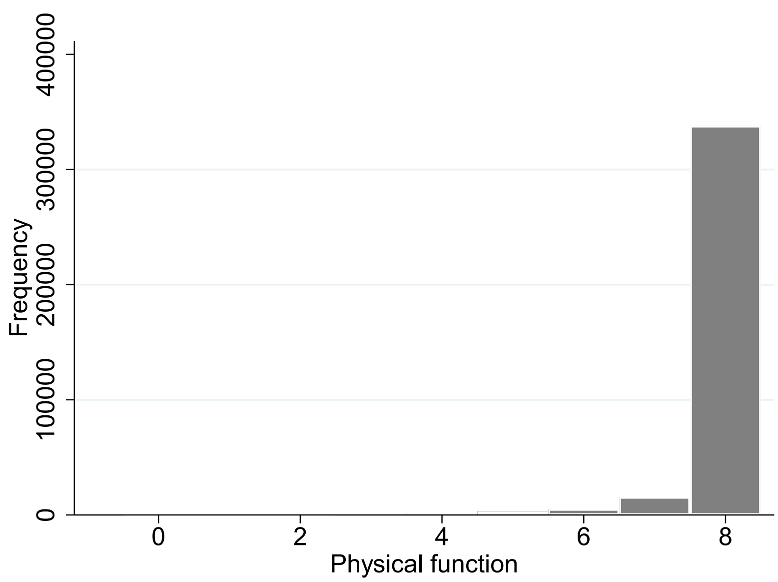


Appendix Figure D.2. Women's retirement rate by country

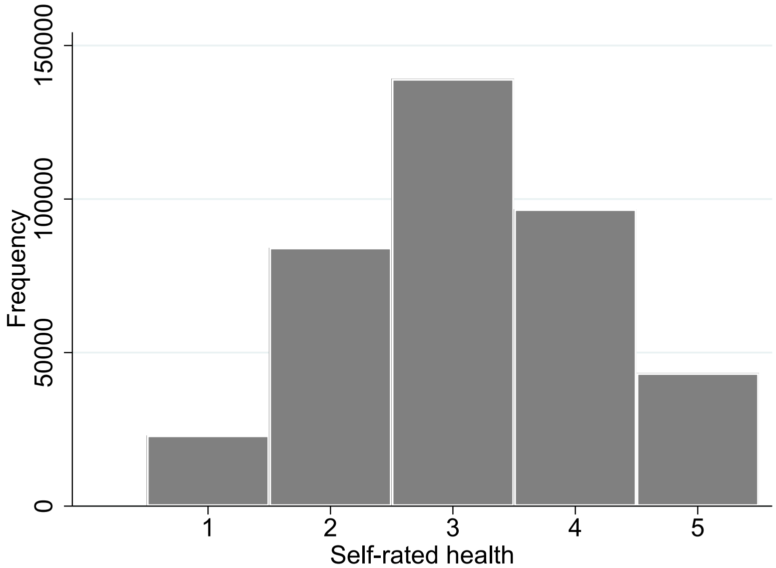
E. Distributions of health outcome measures



Appendix Figure E.1. Distribution of cognitive function score (only for surveys using a ten-word list)



Appendix Figure E.2. Distribution of physical function score



Appendix Figure E.3. Distribution of self-rated health score

F. Comparison of characteristics between participants followed up and those lost to follow-up

Characteristics in the previous interview	Lost to follow-up		Followed up		Standardized difference
	Mean (Obs.)	SD (%)	Mean (Obs.)	SD (%)	
Retired	0.487	0.500	0.441	0.497	0.092
Outcome variables					
Health status					
Cognitive function (z-score)	-0.011	1.019	0.063	0.972	-0.074
Physical independence	0.918	0.275	0.925	0.264	-0.026
Self-rated health (z-score)	0.003	1.007	0.112	0.961	-0.110
Health behavior as risk factors					
Physical inactivity	0.180	0.384	0.190	0.392	-0.026
Smoking	0.226	0.418	0.196	0.397	0.072
Binge drinking	0.085	0.278	0.086	0.280	-0.003
Covariates					
Age	61.28	5.406	60.44	5.509	0.154
Married	0.783	0.412	0.783	0.413	0.001
Potential effect of heterogeneity					
Men	0.525	0.499	0.494	0.500	0.061
Education					0.098
Low	(13,700)	(31.7)	(79,318)	(27.4)	
Middle	(19,175)	(44.3)	(139,831)	(48.2)	
High	(10,407)	(24.0)	(70,693)	(24.4)	
Physically demanding job	0.537	0.499	0.559	0.497	-0.043
Low control job	0.354	0.478	0.342	0.474	0.025

Note: Obs and SD denote the number of observations and standard deviation, respectively. The scores of cognitive function and self-rated health are standardized. In general, a standardized difference less than 0.1 indicates a well balance between the two groups.

G. First stage estimates of adjusted FEIV models

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Age	0.017***	0.018***	0.016***	0.021***	0.015***	0.013***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Age^2	0.027***	0.027***	0.026***	0.029***	0.027***	0.023***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Married	0.007**	0.003	0.005*	-0.001	0.005	0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
ERA	0.086***	0.083***	0.084***	0.087***	0.083***	0.064***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
ERA x Age	-0.006***	-0.007***	-0.006***	-0.008***	-0.008***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ERA x Age^2	-0.005***	-0.004***	-0.004***	-0.002**	-0.002**	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ORA	0.180***	0.182***	0.175***	0.180***	0.165***	0.123***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
ORA x Age	-0.010***	-0.009***	-0.009***	-0.004***	-0.009***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
ORA x Age^2	-0.035***	-0.036***	-0.035***	-0.042***	-0.034***	-0.033***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	377,276	362,973	384,631	272,824	324,519	242,211

Note: FEIV, ERA, and ORA denote fixed effects instrumental variable, early retirement age, and official retirement age, respectively. Age squared was divided by 10 for ease of interpretation. All regressions are adjusted for fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

H. FEIV models with an interaction of sex

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.096***	0.034***	0.184***	-0.040***	-0.024***	-0.002
	(0.026)	(0.008)	(0.024)	(0.013)	(0.008)	(0.011)
Retirement x Men	-0.088**	-0.015	-0.071*	0.020	0.048***	0.024
	(0.041)	(0.012)	(0.037)	(0.019)	(0.013)	(0.020)
Observations	377,276	362,973	384,631	272,824	324,519	242,211
Kleibergen-Paap F	505.009	499.381	497.442	402.296	357.436	208.600
Hansen J	0.850	1.149	3.897**	4.070**	3.175*	0.417

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and sex, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

I. FEIV models with interactions of education

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.050*	0.020**	0.153***	-0.028**	-0.010	0.004
	(0.027)	(0.008)	(0.024)	(0.013)	(0.009)	(0.011)
Retirement x Low education	-0.025	-0.002	-0.043	0.013	-0.009	-0.012
	(0.045)	(0.015)	(0.043)	(0.022)	(0.015)	(0.022)
Retirement x High education	0.032	0.009	0.025	-0.018	-0.004	0.018
	(0.054)	(0.013)	(0.047)	(0.022)	(0.016)	(0.025)
Observations	353,219	337,831	360,025	247,837	299,372	218,016
Kleibergen-Paap F	147.951	153.754	144.239	132.078	97.432	48.761
Hansen J	2.628	2.729*	5.220**	10.999***	9.587***	4.423**

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and education, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

J. FEIV models with an interaction of a physically demanding job

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.068*	0.036***	0.154***	-0.042**	-0.015	0.019
	(0.037)	(0.009)	(0.032)	(0.017)	(0.011)	(0.014)
Retirement x Physically demanding	-0.062	-0.015	-0.025	0.039	0.022	-0.023
	(0.050)	(0.014)	(0.045)	(0.024)	(0.015)	(0.020)
Observations	247,236	233,908	254,372	194,225	223,630	174,681
Kleibergen-Paap F	361.632	328.579	375.916	316.143	440.371	257.944
Hansen J	0.308	1.245	6.694**	1.243	3.286*	1.359

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and engagement in a physically demanding job, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

K. FEIV models with an interaction of a low control job

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.054	0.003	0.149***	-0.044***	-0.017	0.009
	(0.037)	(0.009)	(0.033)	(0.015)	(0.012)	(0.018)
Retirement x Low control	-0.030	0.022	0.072	0.025	0.039**	-0.022
	(0.056)	(0.014)	(0.051)	(0.023)	(0.018)	(0.028)
Observations	161,552	161,559	161,606	144,311	131,504	97,343
Kleibergen-Paap F	261.945	268.037	254.415	267.031	177.343	103.537
Hansen J	1.125	0.341	13.919***	1.130	1.980	7.049***

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and engagement in a low control job, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

L. FEIV models with interactions of region

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.055***	0.019***	0.138***	-0.036***	-0.010	0.006
	(0.020)	(0.005)	(0.019)	(0.008)	(0.007)	(0.013)
Retirement x America	-0.022	0.025	-0.021	0.006	0.004	0.002
	(0.051)	(0.019)	(0.043)	(0.037)	(0.014)	(0.017)
Retirement x Asia	0.129	-0.034	-0.184	0.019	-0.058	-0.018
	(0.194)	(0.058)	(0.300)	(0.130)	(0.051)	(0.172)
Observations	377,276	362,973	384,631	272,824	324,519	242,211
Kleibergen-Paap F	8.540	9.443	4.151	6.001	10.023	2.706
Hansen J	5.166**	0.963	10.413***	6.725**	2.719*	0.960

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and regions, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

M. FEIV models with an interaction of country income level

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.047**	0.030***	0.151***	-0.029***	-0.006	0.013
	(0.021)	(0.006)	(0.019)	(0.010)	(0.007)	(0.010)
Retirement x LMIC	-0.198	-0.043	-0.258	0.139	-0.008	0.114
	(0.164)	(0.059)	(0.236)	(0.147)	(0.047)	(0.097)
Observations	377,276	362,973	384,631	272,824	324,519	242,211
Kleibergen-Paap F	16.236	16.814	9.357	5.596	17.492	7.590
Hansen J	1.606	0.413	2.850*	3.977**	7.636***	0.493

Note: FEIV and LMIC denote fixed effect with instrumental variable and low-middle income countries, respectively. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and country income level, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

N. FEIV models with an interaction of aged society

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.032	0.043**	0.096**	0.009	0.004	0.002
	(0.049)	(0.018)	(0.042)	(0.044)	(0.013)	(0.012)
Retirement x Aged society	0.018	-0.022	0.042	-0.046	-0.012	0.010
	(0.053)	(0.019)	(0.046)	(0.045)	(0.015)	(0.018)
Observations	377,276	362,973	384,631	272,824	324,519	242,211
Kleibergen-Paap F	184.996	142.173	191.637	69.539	186.734	152.033
Hansen J	0.494	2.260	5.451**	2.057	3.310*	1.555

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for covariates (age, age squared, and marital status), interactions between covariates and aged society, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

O. FEIV models for the effect of full-retirement on outcomes

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Full retirement	0.037	0.035***	0.150***	-0.025*	-0.003	0.014
	(0.025)	(0.007)	(0.023)	(0.013)	(0.008)	(0.012)
Observations	349,739	336,718	356,752	249,571	299,763	223,520
Kleibergen-Paap F	1681.882	1688.270	1634.389	1302.014	1147.875	586.978
Hansen J	0.090	0.048	3.954**	3.671*	1.484	1.540

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

P. FEIV models according to pre-retirement employment status

	Cognitive function			Physical independence			Self-rated health		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Retirement	0.042*	0.059**	0.056	0.020***	0.019***	0.030***	0.135***	0.151***	0.121***
	(0.024)	(0.024)	(0.035)	(0.007)	(0.007)	(0.010)	(0.022)	(0.022)	(0.032)
Observations	292,796	213,426	121,608	280,106	204,848	116,951	299,649	217,932	125,624
Kleibergen-Paap F	1443.132	1392.887	715.787	1424.532	1375.721	702.718	1409.601	1355.670	698.016
Hansen J	0.000	0.002	0.288	0.546	1.260	0.187	3.716*	3.822*	4.519**
	Physical inactivity			Smoking			Binge drinking		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Retirement	-0.017	-0.025**	-0.014	-0.002	-0.001	-0.000	0.012	0.012	0.011
	(0.011)	(0.012)	(0.017)	(0.008)	(0.008)	(0.011)	(0.011)	(0.010)	(0.015)
Observations	209,740	155,283	86,794	261,011	185,434	104,796	200,446	140,680	80,202
Kleibergen-Paap F	1188.162	1113.025	565.052	1096.946	1057.489	584.311	620.373	610.673	336.670
Hansen J	0.681	4.346**	0.503	3.047*	1.879	6.004**	0.021	0.015	0.329

Note: FEIV denotes fixed effect with instrumental variable. Model 1 restricted the participants to those who answered that they were in paid work at least once in the interviews; Model 2 additionally excluded those who were self-employed from Model1; Model 3 additionally excluded those who experienced a part-time job from Model 2. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Q. FEIV models for people aged 52-68

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	-0.004	0.016**	0.154***	-0.033**	0.003	0.022
	(0.026)	(0.008)	(0.024)	(0.013)	(0.009)	(0.014)
Observations	316,949	303,868	323,716	226,667	271,785	203,917
Kleibergen-Paap F	1343.216	1345.309	1321.374	1080.231	926.187	469.269
Hansen J	0.132	0.025	0.992	1.506	4.614**	0.069

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

R. Country-by-country FEIV models for the effect of retirement on outcomes

R.1. Country-by-country FEIV models for the effect of retirement on cognitive function

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	9,028	0.148 (0.091)	118.716
Belgium	13,415	0.071 (0.069)	239.358
Bulgaria	752	0.934 (0.357)***	11.900
Croatia	2,596	0.081 (0.334)	16.431
Cyprus	248	-1.097 (0.295)***	15.848
Czech Republic	12,565	-0.022 (0.051)	411.344
Denmark	10,614	0.129 (0.084)	308.888
England	42,988	0.025 (0.048)	729.130
Estonia	10,999	0.159 (0.142)	48.371
Finland	1,058	0.314 (0.183)*	21.825
France	11,874	0.015 (0.051)	440.918
Germany	10,745	0.024 (0.064)	306.879
Greece	5,971	-0.129 (0.339)	9.785
Hungary	1,802	0.059 (0.197)	74.571
Israel	4,468	0.231 (0.183)	51.575
Italy	9,477	-0.027 (0.117)	94.900
Latvia	596	0.026 (0.330)	8.082
Lithuania	1,054	-0.896 (0.303)***	16.359
Luxembourg	2,240	0.031 (0.168)	49.293
Malta	466	-0.411 (0.387)	10.630
Netherlands	4,902	0.226 (0.102)**	278.824
Poland	4,505	0.213 (0.188)	68.037
Portugal	1,634	0.339 (0.510)	7.044
Romania	1,120	0.024 (0.820)	3.568
Slovakia	1,328	0.232 (0.241)	15.314
Slovenia	7,561	0.174 (0.131)	77.866
Spain	7,913	0.050 (0.087)	195.794
Sweden	10,168	0.110 (0.064)*	259.288
Switzerland	7,833	-0.008 (0.067)	177.932
Costa Rica	2,577	-1.686 (1.764)	0.717
Mexico	20,058	-0.070 (0.287)	15.625
United States	125,283	0.020 (0.046)	378.917
China	7,044	-0.539 (0.203)***	29.214
Japan	1,760	0.311 (1.012)	2.718
South Korea	20,634	0.537 (0.420)	7.818

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

R.2. Country-by-country FEIV models for the effect of retirement on physical independence

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	9,402	0.010 (0.021)	124.056
Belgium	13,481	0.031 (0.018)*	242.973
Bulgaria	738	-0.244 (0.085)***	12.543
Croatia	2,620	0.049 (0.085)	16.425
Cyprus	244	0.093 (0.051)*	16.346
Czech Republic	12,866	-0.005 (0.013)	412.966
Denmark	10,670	0.017 (0.017)	310.583
England	44,684	0.029 (0.014)**	765.277
Estonia	11,707	0.126 (0.040)***	53.531
Finland	1,094	-0.093 (0.045)**	21.660
France	12,034	0.004 (0.011)	447.183
Germany	10,836	0.035 (0.017)**	307.756
Greece	5,991	0.004 (0.057)	10.021
Hungary	1,814	0.105 (0.051)**	79.989
Israel	4,701	-0.028 (0.044)	55.968
Italy	9,645	0.011 (0.023)	97.435
Latvia	600	-0.061 (0.078)	7.702
Lithuania	1,046	0.032 (0.063)	16.707
Luxembourg	2,342	-0.047 (0.026)*	54.277
Malta	472	0.031 (0.089)	10.549
Netherlands	4,937	-0.044 (0.025)*	282.648
Poland	4,549	0.043 (0.055)	67.549
Portugal	1,755	-0.024 (0.135)	7.667
Romania	1,092	-0.098 (0.267)	3.560
Slovakia	1,308	0.179 (0.079)**	14.457
Slovenia	7,767	0.032 (0.027)	83.140
Spain	8,160	-0.020 (0.018)	202.179
Sweden	10,230	-0.003 (0.013)	260.822
Switzerland	7,880	0.007 (0.012)	179.354
Costa Rica	2,599	0.545 (0.453)	1.142
Mexico	18,245	-0.029 (0.087)	14.812
United States	105,591	0.043 (0.018)**	298.573
China	7,963	-0.064 (0.071)	35.715
Japan	2,384	0.285 (0.155)*	5.565
South Korea	21,526	0.126 (0.102)	7.247

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

R.3. Country-by-country FEIV models for the effect of retirement on self-rated health

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	9,416	0.346 (0.088)***	123.818
Belgium	13,502	0.234 (0.068)***	242.790
Bulgaria	754	-0.152 (0.251)	12.556
Croatia	2,634	0.138 (0.303)	16.413
Cyprus	248	0.263 (0.276)	15.848
Czech Republic	12,879	0.091 (0.047)*	412.274
Denmark	10,678	0.236 (0.077)***	310.363
England	37,816	0.028 (0.045)	604.670
Estonia	11,747	0.345 (0.123)***	54.254
Finland	1,098	0.619 (0.213)***	21.663
France	12,051	0.093 (0.050)*	447.339
Germany	10,880	0.311 (0.061)***	308.773
Greece	5,990	-0.139 (0.281)	9.981
Hungary	1,838	0.306 (0.186)	78.811
Israel	4,688	0.050 (0.151)	55.869
Italy	9,666	0.016 (0.113)	97.374
Latvia	606	-0.016 (0.289)	8.361
Lithuania	1,056	-0.193 (0.294)	16.378
Luxembourg	2,355	-0.087 (0.155)	55.232
Malta	478	0.463 (0.340)	10.548
Netherlands	4,938	0.037 (0.094)	282.518
Poland	4,562	-0.057 (0.175)	69.196
Portugal	1,753	0.389 (0.451)	7.632
Romania	1,120	0.064 (0.778)	3.568
Slovakia	1,330	0.091 (0.225)	15.383
Slovenia	7,783	0.320 (0.128)**	81.565
Spain	8,163	0.088 (0.086)	201.757
Sweden	10,236	0.065 (0.057)	260.253
Switzerland	7,888	0.151 (0.065)**	179.021
Costa Rica	2,625	0.267 (1.013)	1.133
Mexico	20,093	0.018 (0.310)	15.746
United States	133,109	0.127 (0.038)***	397.264
China	4,701	-0.342 (0.328)	10.867
Japan	4,423	-0.899 (1.199)	1.816
South Korea	21,527	-0.303 (0.419)	7.252

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

R.4. Country-by-country FEIV models for the effect of retirement on physical inactivity

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	8,039	0.011 (0.037)	98.916
Belgium	11,333	0.017 (0.025)	200.816
Croatia	858	-0.015 (0.094)	13.153
Czech Republic	11,029	-0.018 (0.020)	342.321
Denmark	9,222	-0.016 (0.022)	279.523
England	44,647	-0.079 (0.020)***	765.176
Estonia	9,009	-0.044 (0.054)	43.081
France	10,725	-0.033 (0.019)*	408.578
Germany	9,061	-0.050 (0.021)**	262.482
Greece	5,403	-0.030 (0.092)	11.179
Hungary	616	-0.203 (0.122)*	36.201
Israel	3,782	-0.072 (0.076)	42.819
Italy	7,937	-0.102 (0.056)*	86.290
Luxembourg	1,624	-0.044 (0.051)	46.627
Netherlands	4,938	-0.046 (0.028)*	282.518
Poland	3,070	-0.133 (0.115)	34.864
Portugal	1,134	-0.617 (0.390)	3.455
Slovenia	5,438	0.015 (0.041)	68.322
Spain	6,613	-0.092 (0.037)**	162.698
Sweden	9,150	-0.003 (0.017)	221.178
Switzerland	7,015	0.015 (0.026)	159.139
Costa Rica	1,870	0.598 (0.574)	1.379
United States	75,515	-0.027 (0.036)	171.749
China	3,269	0.224 (0.157)	18.299
South Korea	21,527	0.154 (0.222)	7.252

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

R.5. Country-by-country FEIV models for the effect of retirement on smoking

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	6,190	-0.101 (0.037)***	85.269
Belgium	8,196	-0.082 (0.025)***	138.372
Croatia	856	-0.085 (0.086)	13.146
Czech Republic	7,600	0.007 (0.020)	315.464
Denmark	6,313	-0.063 (0.027)**	230.762
England	44,416	0.008 (0.011)	760.243
Estonia	6,380	-0.048 (0.049)	29.340
France	8,548	0.005 (0.017)	324.711
Germany	5,623	0.024 (0.025)	163.204
Greece	4,295	-0.008 (0.132)	7.958
Hungary	616	0.206 (0.098)**	36.201
Israel	2,777	-0.028 (0.067)	32.812
Italy	5,750	-0.041 (0.040)	80.315
Luxembourg	728	-0.068 (0.079)	44.489
Netherlands	4,938	-0.012 (0.027)	282.518
Poland	2,115	-0.086 (0.103)	21.048
Slovenia	3,668	0.052 (0.037)	71.991
Spain	3,929	0.003 (0.039)	108.252
Sweden	6,325	-0.011 (0.025)	154.875
Switzerland	5,349	-0.030 (0.030)	95.367
Costa Rica	2,623	0.386 (0.247)	1.135
Mexico	21,848	0.100 (0.091)	15.535
United States	132,462	-0.009 (0.012)	395.455
China	7,379	-0.052 (0.053)	37.260
Japan	4,065	0.286 (0.296)	2.207
South Korea	21,526	-0.144 (0.122)	7.231

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

R.6. Country-by-country FEIV models for the effect of retirement on binge drinking

Country	Observations	Coef. (SE)	Kleibergen-Paap F
Austria	4,492	0.004 (0.040)	42.544
Belgium	4,804	-0.032 (0.051)	64.264
Czech Republic	5,706	0.065 (0.040)	137.946
Denmark	3,506	-0.002 (0.040)	144.161
England	33,054	-0.002 (0.022)	526.304
Estonia	4,740	0.032 (0.151)	8.853
France	5,250	0.042 (0.031)	100.022
Germany	1,899	-0.038 (0.041)	43.505
Israel	1,340	-0.028 (0.023)	31.520
Italy	3,194	0.019 (0.063)	28.817
Netherlands	2,908	-0.002 (0.050)	129.470
Poland	1,192	-0.123 (0.224)	8.670
Slovenia	1,998	0.141 (0.078)*	10.959
Spain	2,232	0.019 (0.054)	40.283
Sweden	2,798	-0.033 (0.035)	86.837
Switzerland	3,754	0.004 (0.038)	47.289
Mexico	21,627	0.126 (0.097)	15.175
United States	113,912	0.007 (0.009)	341.350
Japan	2,278	-0.347 (0.511)	0.805
South Korea	21,527	-0.022 (0.177)	7.252

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual and year. Robust standard errors clustering at individual and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

S. FEIV models excluding countries with weak IVs

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.051**	0.024***	0.143***	-0.038***	-0.009	0.005
	(0.020)	(0.006)	(0.018)	(0.009)	(0.006)	(0.009)
Observations	342,518	326,554	346,109	242,890	292,006	218,406
Kleibergen-Paap F	2411.781	2433.463	2375.399	2098.785	1775.578	901.990
Hansen J	0.053	0.279	0.855	3.806*	2.036	2.955*

Note: FEIV denotes fixed effect with instrumental variable. Greece, Latvia, Malta, Portugal, Romania, Costa Rica, Japan, and South Korea are excluded from analysis. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

T. FEIV models excluding data from the United States

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.051**	0.019***	0.132***	-0.029***	-0.008	0.015
	(0.022)	(0.006)	(0.021)	(0.009)	(0.008)	(0.017)
Observations	251,993	257,382	251,522	197,309	192,057	128,299
Kleibergen-Paap F	2082.328	2120.431	2052.623	1844.518	1382.119	498.713
Hansen J	1.716	0.773	9.416***	0.221	4.273**	1.415

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

U. FEIV models excluding countries without changing the SPA

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.052**	0.032***	0.149***	-0.030***	-0.004	0.013
	(0.022)	(0.006)	(0.020)	(0.011)	(0.007)	(0.010)
Observations	324,290	309,992	330,964	249,896	276,202	211,754
Kleibergen-Paap F	2028.669	2009.672	1979.637	1644.936	1486.158	763.661
Hansen J	0.558	0.060	2.041	3.547*	1.387	0.569

Note: FEIV and SPA denote fixed effect with instrumental variable and state pension age, respectively. Cyprus, Finland, Luxembourg, Sweden, Switzerland, Costa Rica, Mexico, China, and Japan are excluded from analysis. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

V. Stratified FEIV models by sex for the effect of retirement on the raw scores of cognitive function

	Men	Women
Retirement	-0.089	0.281***
	(0.101)	(0.087)
Observations	147,516	168,730
Kleibergen-Paap F	1027.651	1422.153
Hansen J	1.782	0.006

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

W. FEIV models with multiple imputation

	Cognitive function	Physical independence	Self-rated health	Physical inactivity	Smoking	Binge drinking
Retirement	0.033	0.032***	0.141***	-0.025**	-0.012	0.014
	(0.022)	(0.006)	(0.020)	(0.010)	(0.009)	(0.010)
Observations	402164	402164	402164	317028	402164	253572
Kleibergen-Paap F	2282.936	2282.936	2282.936	2218.316	2282.936	844.731
Hansen J	0.530	0.142	3.451*	0.478	3.646*	0.982

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

X. Subgroup FEIV models by retirement duration

	Cognitive function		Physical independence		Self-rated health	
	<5 years	≥5 years	<5 years	≥5 years	<5 years	≥5 years
Retirement	0.026	0.041	0.017**	0.046***	0.156***	0.121***
	(0.026)	(0.036)	(0.007)	(0.010)	(0.024)	(0.032)
Observations	277,378	300,186	264,810	288,898	284,966	307,455
Kleibergen-Paap F	1183.118	982.391	1151.145	1032.190	1169.870	960.784
Hansen J	0.072	0.158	1.233	0.779	3.395*	11.931***

	Physical inactivity		Smoking		Binge drinking	
	<5 years	≥5 years	<5 years	≥5 years	<5 years	≥5 years
Retirement	0.002	-0.041**	0.010	-0.001	0.026*	0.016
	(0.014)	(0.019)	(0.009)	(0.013)	(0.013)	(0.021)
Observations	190,124	208,611	239,323	256,064	180,880	191,665
Kleibergen-Paap F	741.853	740.301	775.172	534.881	437.062	247.862
Hansen J	0.190	4.007**	2.582	4.672**	0.009	0.271

Note: FEIV denotes fixed effect with instrumental variable. All regressions are adjusted for age, age squared, marital status, and fixed effects of individual, country, year, and interactions between country and year. Robust standard errors clustering at individual, country, year, and interactions between country and year are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.