EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 09/14
September 2014

Unhealthy Retirement?

by

Fabrizio Mazzonna
(University of Lugano and MEA)
Franco Peracchi
(Università di Roma "Tor Vergata" and EIEF)

Unhealthy retirement?*

Fabrizio Mazzonna
University of Lugano and MEA

Franco Peracchi
University of Rome Tor Vergata and EIEF

September 15, 2014

Abstract

We investigate the causal effect of retirement on health and cognitive abilities by exploiting the panel dimension of the first two waves of the Survey of Health Ageing and Retirement in Europe (SHARE) and the variation between and within European countries in old age retirement rules. We show evidence of substantial heterogeneity in the effect of retirement across occupational groups. In particular, we find that retirement increases the age-related decline of health and cognitive abilities for most workers. On the other hand, we find evidence of a positive immediate effect of retirement for those employed in highly physically demanding jobs.

Keywords: Aging; cognitive abilities; retirement; occupation; SHARE.

JEL codes: C26, I14, J14, J24, J26.

* Corresponding author: Franco Peracchi, Department of Economics and Finance, via Columbia 2, 00133 Rome, Italy (`franco.peracchi@uniroma2.it`). We thank Janet Currie, Angus Deaton, Peter Eibich, Osea Giuntella, seminar participants at Princeton (CHW) and DIW Berlin, and participants at the 2014 IIPF Conference for useful comments. This paper uses data from SHARE wave 1 and 2, release 2.5.0 as of November 29, 2013. The SHARE data collection has been primarily funded by the European Commission through the 5th, 6th and 7th framework programmes. Additional funding from the U.S. National Institute on Aging and the German Ministry of Education and Research, as well as from various national sources is gratefully acknowledged (see `www.share-project.org` for a full list of funding institutions).

1 Introduction

Declining fertility, continuing growth in life expectancy and declining labor force participation among older workers have raised serious concerns about the financial stability of social security programs in most developed economies. In order to meet these challenges, many governments have implemented policies aimed at increasing the average retirement age of the workforce. However, there are dissenting opinions not only about the magnitude of the effects of these policies, but also about their sign (see Börsch-Supan 2013 for a discussion).

On the one hand, there is a view that retirement enables individuals to enjoy their leisure time and eliminates work-related stress, with positive spillovers on their mental health and well-being. Retirement may be particularly beneficial for those who work in strenuous and unhealthy occupations. This argument has been strongly supported by labor unions who oppose increases in the retirement age. Recent literature in Economics has indeed shown the presence of negative health effects of working in physically demanding occupations (see Case and Deaton 2005 and Ravesteijn et al. 2013 for a review).

On the other hand, there is a view that retirement may be harmful. This may happen if a lack of purposes in the retiree's life affects individual well-being, mental health and cognitive abilities (Rohwedder and Willis 2010). As argued in our previous study on the effect of retirement on cognition (Mazzonna and Peracchi 2012), this negative effect can actually be predicted using the theoretical framework proposed by Grossman (1972). The main intuition is that retired individuals lose the market incentive to invest in cognitive repair activities, which may lead to an increase in the rate of cognitive decline after retirement. A negative effect on health can be also predicted if retirement reduces social interactions. The social capital literature (see e.g. Glass et al. 1999 and d'Hombres et al. 2010) suggests that the social networks formed at work may buffer from health shocks. To this end, Börsch-Supan and Schuth (2013) argue that at least one third of the decline in cognition after retirement can be attributed to the shrinkage of social networks. More generally, retirement may affect lifestyles, such as drinking and smoking habits, dietary consumption and most importantly physical activity (Zantinge et al. 2014), which in turns may affect individuals' health. For instance, if work is the main form of physical activity for many individuals, then we may expect a negative effect of retirement on health.

In this paper we present evidence of heterogeneity in the effect of retirement on health and cognition across occupations, thus showing that these two apparently opposite views can in fact coexist. In particular, we present evidence of large heterogeneity in the effect of retirement, with a negative effect for a large fraction of the workforce but a positive effect for people working in strenuous jobs.

Empirically, it is difficult to provide causal evidence of the effect of retirement. Being a choice, retirement may be related to bad health, cognitive decline or other unobserved factors (e.g., time preferences). Therefore, even the simple comparison of health status for the same individual before and after retirement may lead to wrong conclusions. Recent studies try to address endogeneity of retirement by exploiting retirement incentives provided by exogenous laws and social security regulations (see for example the country studies in Gruber and Wise 2004), such as between- or within-country variation in eligibility ages for early and normal retirement benefits. The empirical evidence from these studies is mixed, as results are sensitive to the countries analyzed, the identification strategy employed, and the health or cognitive outcome considered. Charles (2004), Neuman (2008) and Johnston and Lee (2009) find a positive effect of retirement on subjective measures of health by exploiting age specific incentives in the UK and US social security regulations. Similar results are reported by Coe and Zamarro (2011), who mainly exploit between-country variation in eligibility ages across European countries. On the contrary, studies based on both European and US old-age surveys show evidence of a negative effect of retirement on cognitive abilities (Rohwedder and Willis 2010, Bonsang et al. 2012, and Mazzonna and Peracchi 2012). The only exception is a paper by Coe et al. (2012) who find no evidence of a causal effect of retirement on cognition in the US.

Most existing literature has the important drawback of regarding retirement as a binary treatment that only causes an immediate one-time shift in the level of health or cognition. This ignores the possibility that the effect of retirement may be cumulative and depend also on the years into retirement. Moreover, many studies (Rohwedder and Willis 2010, Coe and Zamarro 2011, Coe et al. 2012) only rely on cross-country variation in the eligibility ages at one point in time (the time of the interview) as the source of the exogenous variation needed to estimate the causal effect of retirement.

In this paper we present estimates of the causal effect of retirement on health and cognitive functioning that overcome the limitations of the existing literature. First, we account for the endogeneity of retirement by exploiting the panel dimension of the first two waves of the Survey of Health, Ageing and Retirement in Europe (SHARE) and by using an instrumental variable (IV) strategy that accounts for residual time-varying heterogeneity that might confound the estimates of the causal effect of interest. Second, we take into account both the short and the long run effects of retirement by including a control for the years spent in retirement. Third, we exploit the heterogeneity in the effect of retirement across occupational groups. Previous literature (e.g., Coe et al. 2012) tried to exploit this potential source of heterogeneity by distinguishing between blue and white collars jobs, without finding any significant difference. However, the blue/white collar distinction is too crude as it is based on a coarse job categorization, namely the first digit of the ISCO-88 code. Moreover, such distinction cannot capture the fact that the burden of a job may be multi-dimensional (e.g., physical and psycho-social). In this paper, thanks to the availability in the first (2004) wave of SHARE of detailed information on respondents' last job, we are able to associate each occupation to specific levels of physical and psycho-social burden using both internal and external indexes. Our results show evidence of heterogeneity in the effect of retirement across jobs and suggest that the physical dimension is the more relevant. In particular, we find that, for people working in more physically demanding jobs, retirement has an immediate beneficial effect on both mental and physical health (depression and mobility limitation) and on cognitive abilities (memory and fluency). On the contrary, for the rest of the workforce, retirement has negative long-run effects on the age profile of health and cognitive abilities.

The remainder of this paper is organized as follows. Section 2 describes the data used for this study. Section 3 presents our empirical model and discusses a number of issues that complicate the identification of the causal effect of retirement on cognitive abilities. Section 4 presents our main results as well as a large battery of robustness checks. Finally, Section 5 offers some conclusions.

2 Data

In this paper we mainly use data from Release 2 of the first two waves (2004 and 2006) of the Survey of Health, Ageing and Retirement in Europe (SHARE), a multidisciplinary and cross-national bi-annual household panel survey coordinated by the Munich Center for the Economics of Aging (MEA) with the technical support of CentERdata at Tilburg University. The survey collects detailed information on socio-economic status, health, social and family networks for nationally representative samples of elderly people in the participating countries.

2.1 Description of SHARE

SHARE is designed to be cross-nationally comparable and is harmonized with the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA). The baseline (2004) survey covers 11 countries, representing different regions of continental Europe, from Scandinavia (Denmark, Sweden) through Central Europe (Austria, Belgium, France, Germany, the Netherlands, Switzerland) to Mediterranean countries (Greece, Italy, Spain). The target population consists of individuals aged 50+ who speak the official language of each country and do not live abroad or in an institution, plus their spouses or partners irrespective of age. The common questionnaire and interview mode, the effort devoted to translation of the questionnaire into the national languages of each country, and the standardization of fieldwork procedures and interviewing protocols are the most important design tools adopted to ensure cross-country comparability (Börsch-Supan et al. 2005).

It is worth noting that there is a substantial variation in the interview date within and between countries. Interviews in wave 1 took place between March 2004 and November 2005, while interviews in wave 2 took place between September 2006 and September 2007. Unlike the previous literature, we take into account and exploit such temporal variation by constructing precise measures of age and distance from retirement at the time of the interview (further information is provided in Section 3).

We consider all countries that contributed to the 2004 baseline except Greece, which is excluded because of important sample selection issues.[1] Our working sample consists of people aged 50–70 at the time of their first interview, who classified themselves as employed, unemployed or retired, answered the retrospective question on past employment status, reported being in the labor force at age 50, and participated to both wave 1 and 2 of SHARE. These selection criteria, which are meant to avoid excessive noise in the measurement of labor force status by excluding people with weak labor force attachment, result in a balanced panel of 8,226 individuals, each interviewed twice.

Table 1 shows the composition of our working sample by country and gender. It is clear from the table that the exclusion of people out of the labor force at the age of 50 leads to a gender imbalance, especially in Italy, the Netherlands and Spain where female attachment to the labor force attachment is particularly weak. Finally, it is worth noticing that we lose roughly 400 units from this sample when we analyze the heterogeneity in the effect of retirement across occupational groups because of missing or incomplete information on the last job.

2.2 Health and cognitive measures

SHARE contains several measures of health and cognition, which allows us to investigate the effect of retirement on a variety of health and cognitive dimensions.

Self-rated health (SRH) is generally consider a good summary of the overall health of an individual, although it may suffer from substantial reporting heterogeneity, in particular by gender and country, resulting from differences in both health perception and response style (see, e.g., Case and Paxson 2005, Jürges 2007, and Peracchi and Rossetti 2012). In SHARE, respondents are asked to rate their overall health on a five-point scale: very good, good, fair, bad and very bad. In addition to SRH, the survey also includes several other health measures, some objective (e.g., grip strength)

[1] In Greece, the small fieldwork period in the first wave (roughly 2 months because of the beginning of the 2004 Olympic Games) and the use of the telephone directory as the sampling frame cast doubts on the representativeness of the Greek sample.

and some self-reported (e.g., suffering of specific illnesses or diseases such as high blood pressure, high cholesterol, diabetes, chronic lung disease, various types of cancer, hearth attack, or stroke).

To summarize this large amount of health information and to facilitate the presentation of the results, we follow Bound et al. (1999) and employ a single health index constructed by estimating the following model

$$SRH_{it} = \pi' H_{it} + \delta_t + C_i + \epsilon_{it},$$

where SRH_{it} is SRH of individual $i = 1, \ldots, N$ in wave $t = 1, 2,$, H_{it} is a vector of health measures that vary across individuals and over waves, δ_t is a wave dummy, C_i is a set of country dummies, and ϵ_{it} is a regression error. The health measures in H_{it} include grip strength and binary indicators for instrumental activities of daily living, mobility limitations, chronic conditions and depression. As for mobility limitations, respondents were asked whether they had difficulties with various activities because of a health or physical problem. Since roughly 80% of the respondents report no or at most one limitation, our mobility indicator takes value 1 if the respondent reports to suffer from less than 2 mobility limitations and value 0 otherwise. As for depression, SHARE contains a measure based on the Euro-D index, a 12-item depression symptoms scale that considers several dimensions of mental health (depression, anxiety, suicidality, etc.). Our indicator of depression takes value 1 for values of Euro-D lower than 4 and value 0 otherwise.[2] We also include country fixed effects to account for country differences in self-assessment of health. Our procedure is similar to that implemented by Coe and Zamarro (2011), except that we include indicators for each chronic condition instead of a single index, and exclude obesity and physical inactivity because they may reflect differences in lifestyle and behavior.

We estimate the model using ordered probit, separately by gender to account for differences in reporting behavior of men and women. Our health index is just the predicted probability of being at least in good health, a number ranging between 0 and 1. To show that our predicted health index is consistent with the information provided by the original health measures, Appendix B) replicates the estimates reported in the main text using SRH and the original indicators for depression and mobility limitations.

The SHARE cognitive function module contains measures of cognitive abilities based on simple tests of memory, verbal fluency and numeracy. These tests are based on the well-known Mini-Mental State Exam (Folstein et al. 1975), follow a protocol aimed at minimizing the potential influences of the interviewer and the interview process, and are comparable with similar tests implemented in other surveys, in particular the Health and Retirement Study (HRS) and the English Longitudinal

[2] This particular cutoff point was validated in the EURODEP studies against a variety of clinically relevant indicators (Dewey and Prince 2005).

Study of Ageing (ELSA).

The memory test consists of verbal registration and recall of a list of 10 words, carried out twice. The first time is immediately after the encoding phase (immediate recall), while the second time is some 5 minutes later, at the end of the cognitive function module (delayed recall). A general measure of memory is constructed by adding the individual scores in the two tests. The resulting memory variable ranges between 0 and 20. The test of verbal fluency consists of counting how many distinct elements from the animal kingdom the respondent can name in a minute. Our fluency variable is the score in this test, which ranges between 0 and 50.[3] The test of numeracy consists of a few questions involving simple arithmetical calculations based on real life situations. Our numeracy variable is the score in this test, which ranges between 0 and 4.

To summarize the information provided by the three cognitive tests into a single measure of cognitive skills we use principal component analysis (PCA), a statistical method extensively employed in the literature to summarize the information from a large battery of cognitive and non-cognitive tests (e.g., Herzog and Wallace 1997 and Cawley, Heckman and Vytlacil 2001). Table 2 presents our PCA results. The first principal component explains almost 60% of the total variance and is the only one with an eigenvalue greater than 1 (the standard criterion to select the number of principal components) and a positive sign for all factor loadings. This means that the three tests are strongly positively correlated and we can reasonably rely on a single index of cognitive ability. As for health, Appendix B contains the results obtained by replicating all the estimates reported in the main text using the three original cognitive scores.

Table 3 presents the mean of our health and cognitive indexes in wave 1 and 2, along with their mean difference between the two waves. For the probability of reporting good health, the table shows a statistically significant decline of about 1.5% over our 2-year period. More controversial is the descriptive evidence for the cognitive index, as the table shows on average an increase over time. Such apparently puzzling result is partly due to retesting effects, since the battery of cognitive tests is exactly the same in the two waves. Retesting effects may seriously compromise our analysis of the effect of retirement if they differ for employed and retired people. To investigate this issue, we exploit the information from people in the refreshment sample from wave 2, who were exposed to these tests for the first time. As expected, we find evidence of retesting effect for the longitudinal sample when compared with the refreshment sample. However, we do not find evidence of differential retesting effects by labor force status.

[3] We trim values above 45 which represents 0.5% of the distribution. These values are obviously implausible since they mean that these respondents were able to name almost one animal per second.

2.3 Physically demanding occupations

To analyze heterogeneity in the effects of retirement across job types, we divide respondents into two groups: those who are (or were) employed in more physically demanding occupations, and those who are (or were) employed in less physically demanding occupations. The distinction between the two groups is based on detailed information on the last job in which the respondent is employed (or was employed if retired), collected in the first wave of the survey, with jobs classified by the fourth digit of the ISCO-88 classification.[4] Previous literature (e.g., Coe et al. 2012) relies on the classical distinction between blue- and white-collar jobs, which is typically based on the assumed skill level of each occupation using only the information from the first digit of the ISCO-88 code. However, this distinction is too coarse and therefore unable to capture differences in physical burden across occupations in the same group (see Kajitani at al. 2013 for a similar point).

For this reason, we instead rely on a pair of "external" occupational indexes based on the Job Exposure Matrices (JEMs) constructed by Kroll (2011) using a large scale representative survey on working conditions of about 20,000 employees in Germany.[5] These JEMs link almost all the ISCO 88-classified jobs (100% of all 2-digit codes, 94.8% of the 3-digit codes and 78.5% of the 4-digit codes). From these JEMs, Kroll (2011) derives two indexes, a "physical job index" and a "psycho-social index".[6] The first is a measure of the physical burden of a job based on its ergonomic stress and environmental pollution. The second is a measure of its psycho-social burden based on its mental stress, social stress and temporal loads. Both indexes range between 1 and 10, with higher values meaning higher burden. For example, value 1 of the Physical Job Index corresponds to jobs with the lowest physical burden (e.g., draftsmen, bookkeepers, and teachers), while value 10 refers to particularly heavy jobs (e.g., miners, bricklayers, and metal and machinery workers). We classify an occupation as physically or psycho-socially demanding if the corresponding index is larger than 5. The availability of these two indexes is important because it allows us to investigate which job characteristics—physical burden or psycho-social stress—is more closely associated with heterogeneity in the effect of retirement on health.[7] In addition, we exploit the extended range of values of these indexes and distinguish those who work (or used to work) at the two ends of the distribution of physical burden, namely in the least and the most physically demanding occupations

[4] The International Standard Classification of Occupations (ISCO) is an international classification produced by the International Labour Organization (ILO). In particular, ISCO-88 provides a system for classifying and aggregating occupational information obtained from population censuses and other statistical surveys, as well as from administrative records.

[5] The survey is part of the European Working Conditions Survey, which has been conducted regularly since the 1980s in all countries of the European Union.

[6] See Santi et al. (2013) for details on the construction of the two indexes.

[7] In the main text we only use the physical job index and leave the psycho-social index for robustness checks.

(see Section 4 for further details).

To check the robustness of our results, we also construct an "internal" index based on self-evaluation of the level of physical strength required for the job in which the respondent is employed. Construction of this index exploits the information collected during the SHARE interview, where respondents are asked whether they agree with the following statement: "*Your job is physically demanding*". Respondents can choose among the following alternatives: *Strongly agree*, *Agree*, *Disagree*, and *Strongly disagree*. Unfortunately, this question is only asked to those who are currently employed. Further, the answers are likely to be affected by substantial reporting heterogeneity, at least by gender and country. In Appendix C, we show how we deal with these two issues. Since the results obtained using this internal index are very similar to those presented in Section 4, we do not report them here but make them available upon request.

Table 4 presents, separately by country, the percentage of respondents in physically and psycho-socially demanding jobs according to the external index, and the percentage in physically demanding jobs according to our internal index. It also presents the correlation between the three indexes in our sample. The table shows that about 45 percent of SHARE respondents work in jobs that may be classified as physically demanding (external index of physical burden greater than 5), while 53 percent work in jobs that may be classified as psycho-socially demanding (external index of psycho-social burden greater than 5). There is also evidence of heterogeneity across countries, with a much higher fraction of workers in physically demanding jobs in Spain (70 percent) and a much lower fraction in Switzerland (30 percent). We find similar cross-country heterogeneity in our internal index of physical burden, whose correlation with the external index of physical burden exceeds 70 percent. The table shows instead less cross-country heterogeneity in the external index of psycho-social burden, whose correlation with the two indexes of physical burden (external and internal) is only about 40 percent. This suggests that the physical and psycho-social burden, although correlated across jobs, do not coincide and represent distinct aspects of a job.

3 Model specification and estimation

The empirical strategy that we follow in this paper differs in many respects from our earlier work on the effect of retirement on cognition (Mazzonna and Peracchi 2012). The most important difference is that we now exploit the panel dimension of SHARE, which has the important advantage of allowing us to control for time-invariant characteristics of the respondents, such as gender, birth cohort and educational attainments. The main drawback is the non-negligible rate of panel attrition, which implies a loss of roughly 30% of the initial sample. This issue will be discussed in more detail in Section 3.2.

Another important difference is that we now use a less restrictive empirical specification of the effect of retirement on health and cognition. As mentioned in the Introduction, previous literature modelled the effect of retirement on health or cognitive abilities only as a binary treatment, ignoring the possibility that the effect of retirement may depend on retirement duration, i.e., the length of time spent in retirement. In our previous paper, guided by the implications of our theoretical model, we instead specified the effect of retirement as depending solely on its duration. In fact, both effects may play a role. In other words, it can be argued that retirement may cause both an initial shock due to the changed environment—such as an increase or a decrease in depression symptoms at the time of retirement—and a change in the rate of health deterioration after retirement.

Our baseline specification (Model A) has the following form:

$$H_{it} = \alpha_i + \beta_1 Age_{it} + \beta_2 Retired_{it} + \beta_3 DistR_{it} + \beta_4^\top X_i + \beta_5^\top Z_{it} + U_{it},$$

where H_{it} is either the health or the cognitive index of individual i in wave t, α_i is a time-invariant unobservable individual effect, Age_{it} is age of individual i in wave t, $Retired_{it}$ is a binary indicator of retirement, $DistR_{it} = \max\{0, Age_{it} - R_{it}\}$ is the number of years spent in retirement (equal to zero if the individual is not yet retired), X_i is a set of binary time-invariant indicators for educational attainments and the country of residence (with Belgium as the reference country), Z_{it} is a set of time-varying controls, such as marital status, and U_{it} is a regression error potentially correlated with $DistR_{it}$. Retirement status is self-reported and does not necessarily coincide with either the recipience of pension income or a drastic decline in the number of hours of paid work. Notice that, conditional on U_{it}, the total effect of retirement is equal to β_2 plus the effect due to the years spent in retirement, namely $\beta_3 * DistR_{it}$.

We also consider a second specification (Model B) which allows the linear age term to differ across countries. This specification adds to the regressors in Model A a set of interactions between the linear age term and the country indicators. Both models are estimated first on the pooled data and then separately by gender or type of job (low and high physically demanding jobs) to take into account these two important sources of potential heterogeneity.

It is well known that OLS estimates of our baseline model may be biased due to potential reverse causality (people with poor health may decide to retire earlier) or correlation between the retirement choice and unobservable factors included in the regression error. Moreover, other important identification issues should also be taken into account, such as failure of functional form assumptions, panel attrition, and endogeneity of education and occupational choices.

Given the panel dimension of our data, the first differences (FD) estimator solves some of these identification issues because it nets out the effect of all time-invariant sources of heterogeneity.

However, in particular in the case of retirement, time-varying unobservable individual characteristics may still bias our estimates. More importantly, taking first differences does not eliminate the bias due to potential reverse causality from health to retirement (for example a health shock which hits the respondent between waves leading her to retire). In addition, FD estimates are susceptible to attenuation bias due to measurement error in the retirement variables, an issue particularly relevant in survey data. For these reasons, we address the problem using an instrumental variable (IV) strategy described in detail in the next section. Section 3.2 discusses other identification issues that are relevant for our empirical strategy.

Finally, it is important to notice that since we use a FD estimator, we need a sufficient number of individuals to change their labor force status in order to identify the coefficient on the binary indicator *Retired*. In our sample, 758 respondents switch from employment to retirement, while 49 respondents switch from retirement to employment. Moreover, since we exploit the variability of both the interview and the retirement dates, the effect of the distance from retirement ($DistR$) is identified not only through the respondents who are already retired in wave 1, but also through the respondents who retire between the first and the second wave. Indeed, the value of $DistR$ in both waves varies across respondents of the same cohort who retired at the same date (year and month) but were interviewed at different dates. Without taking the exact interview and retirement dates into account, the coefficient on $DistR$ would be simply identified as the deviation from the mean value of first differences in health and cognition of respondents already retired in wave 1.

3.1 Endogeneity of retirement

Following Mazzonna and Peracchi (2012), we address the endogeneity problem by using an IV strategy. Here, however, we take advantage of the panel dimension of SHARE by using a two-stage least squares FD (2SLS-FD) estimator. Our instruments are based on the legislated early and normal ages of eligibility for a public old-age pension, two variables that are arguably exogenous and easily shown to be relevant for the actual retirement age.

Figures 1 and 2 present the histograms of reported retirement age by country, respectively for men and women. The vertical blue and red bars respectively denote the range of eligibility ages for early and normal retirement that are relevant for the cohorts in our sample, the width of each bar measuring the amount of the changes introduced during the period considered. The two figures differ slightly from those presented in Mazzonna and Peracchi (2012), as we made an additional effort at incorporating the many pensions reforms introduced in several European countries during the 1990's. We refer to Appendix A for a discussion of the changes in pension eligibility rules in the SHARE countries that are relevant for the cohorts considered in this paper. The two figures show that eligibility ages differ substantially by country and gender at each point in time, but also changed substantially over time for some country. For instance, in 1994 the early retirement age for males was about 52 years in Italy and 65 years in Switzerland (where early retirement was introduced in 1997, so in 1994 only retirement at the normal age of 65 was possible). Between 1994 and 2001, however, it was increased from about 52 to about 57 years in Italy, but was lowered from 65 to 63 years in Switzerland. For the normal retirement age, the differences across countries and by gender are much smaller, but changes over time have been large for some countries.

Based on our eligibility data, we construct four instruments: two binary indicators of eligibility, respectively for early and normal retirement ($EligE$ and $EligN$), and two variables that measure the distance of the respondent's age at the time of the SHARE interview from the eligibility ages for early and normal retirement ($DistE$ and $DistN$). By analogy with $DistR$, the last two instruments are constructed as the positive part of the difference between the actual age of individual i at time t and the eligibility ages for early and normal retirement that are relevant for individual i (i.e., $DistE_{it} = \max\{0, Age_{it} - E_i\}$ and $DistN_{it} = \max\{0, Age_{it} - N_i\}$). Since we have two endogenous regressors, our model is over-identified, which allows us to test the exogeneity of the instruments through a Sargan-Hansen test of the over-identifying restrictions.

3.2 Other identification issues

As argued by Bingley and Martinello (2013), education may be a source of bias if cross-country differences in retirement ages are positively correlated with cross-country differences in average educational attainment. Since education is an important determinant of health and cognition in later life (see for example Mazzonna 2013), this "would invalidate the use of retirement ages as instruments without appropriate controls" (Bingley and Martinello 2013). It is worth noting that, unlike other papers (e.g., Rohwedder and Willis 2010), our FD strategy allows us to control for any time-invariant determinant of retirement, whether observed or unobserved, thus including education, country fixed effects and cohort heterogeneity.[8]

Our FD strategy may also help control for panel attrition. As already mentioned, the attrition rate between the first two waves of SHARE is substantial. We are fully aware that the resulting selection process may be far from random. In fact, attrition is more likely for people who are in worst health conditions and also depends on retirement status. However, if attrition is mainly explained by characteristics of the respondents that do not change between the two waves,[9] then the FD estimator is unbiased. As a robustness check, we compare our results to those obtained

[8] In the case of cohort heterogeneity we can also exploit age variation for individual belonging to the same cohort since we take into account the exact interview date.

[9] As argued by Wooldridge (2012), this assumption is reasonable in short panels.

using the Inverse Probability Weighting approach of Fitzgerald et al. (1998), which allows attrition to be non-random and to depend on individual characteristics observed in the first wave, including health, cognitive outcomes and fieldwork characteristics such as the interviewer's age, gender and education (see Section 4.3 for further details).

Another potential concern is failure of the assumption of a linear age profile of health and cognitive abilities. There are two main reasons for this assumption. First, our age window is relative short (ages 50–70), so linearity of the relationship between age and health is not unreasonable—as convincingly argued by Coe and Zamarro (2011) using the same data. Second, specifying a higher-order polynomial in age would also require specifying a higher-order polynomial for the time spent in retirement, with the need of finding additional instruments. However, as a robustness check, we also consider a flexible specification of the age profile of health by including a set of age dummies. As discussed in Section 4.3, the results are very similar to those obtained from our baseline specification.

Finally, concerns may also arise because of mobility across jobs. For example, jobs at the end of a working life are likely to be less physically demanding than at the beginning. This may cause problems to our identification strategy if proximity to retirement, and in particular to retirement eligibility, affects an individual's occupational choice. In Section 4.3 we show that our instrument is uncorrelated with the threshold value that we use for the heterogeneity analysis, but there is some evidence of an age gradient in these occupations. However, even in the absence of bias, it is important to recognize that our analysis is confined to the last job, as reported in the first wave of SHARE (we do not have precise information on the respondents' occupation in the second wave). For instance, our analysis cannot establish whether people feel retirement as a relief because of the more recent exposure (the last job) or because the last job is a proxy for a long exposure to physically demanding jobs.

4 Results

In this section we report the result from 2SLS-FD estimation of the effect of retirement on health and cognition using the identification strategy presented in Section 3. We also present the results of a number of checks of the robustness of our estimation strategy.

4.1 First-stage results

Table 5 shows the results from the first-stage regression of our two endogenous variables ($DistR$ and $Retired$) on the exogenous regressors and the excluded instruments. As discussed in Section 3.1, we use four instruments: two binary indicators of eligibility for early and normal retirement ($EligE$

and *EligN*), and the two variables that measure the distance from retirement (*DistE* and *DistN*). The table is divided in two panels: the top panel (Panel A) for *DistR* and the bottom panel (Panel B) for *Retired*. Each panel shows the results for the whole sample and separated by gender and the level of physical burden (low and high). In the case of gender and physical burden we only show the results from Model B. The table also shows the sample size (N), the regression R^2 and the F-test statistic for the joint significance of the excluded instruments.

Our results confirm that eligibility rules are important determinants of retirement decisions. For both genders and job types, and for both models, all instruments are strong predictors of our two endogenous variables. However, in the case of *Retired*, the effect of the distance from the eligibility age for normal retirement appears to be negative. Put differently, conditional on age, the eligibility indicators and *DistE*, the longer is the distance from the eligibility age for normal retirement, the lower is the probability that an individual will retire. This somewhat puzzling result is partly a consequence of our sample selection criterion that restricts the sample to people aged 50–70. Finally, notice that our estimates are unaffected by the introduction, in Model B, of a country-specific linear trend in age.

4.2 Second-stage results

Table 6 shows the estimated coefficients on *DistR* and *Retired* for our health and cognitive indexes, in the whole sample (both Models A and B) and separately by gender (only Model B). At the bottom of each table we report the value of the Sargan-Hansen J-statistic for testing the validity of the over-identifying restrictions, which is asymptotically distributed as a χ^2 with 2 degrees of freedom under the null hypothesis that the over-identifying restrictions are valid. Except for case of women's health (but only at the 10% level), the test never rejects the null hypothesis, thus lending support to our IV strategy.

In the pooled estimates, retirement has quite clear negative effect on both health and cognitive abilities. Consistent with Mazzonna and Peracchi (2012), this effect seems to be due to the years spent in retirement (*DistR*), not to the immediate short-run effect measured by the dummy for being retired (*Retired*). However, the coefficient on *Retired* is very imprecisely estimated. We will show below that this is not due to a problem of power (see the value of the F-test reported at the bottom of the table), but to the large heterogeneity in the effect of interest across occupational groups.

In the case of health, each year into retirement decreases by almost 1% the probability of reporting a good health status. In the case of cognition, consistently with our previous cross-sectional work, each year in retirement decreases cognitive abilities by about 6% standard deviations. Results are also robust to the inclusion of a country-specific age trends (Model B). In the third and fourth columns, we look for the presence of gender heterogeneity. In the case of health, the negative effect on $DistR$ in the pooled sample appears to be largely driven by men. In the case of cognitive abilities, instead we do not find evidence of substantial gender heterogeneity.

Tables B.2 and B.3 in Appendix B show that these results are substantially unaffected when we use three original health measures (SRH, mobility limitation and depression) and the three original cognitive scores.

Taken at face value, the results presented so far would suggest a clear-cut answer to our research question: retirement increases the age-related rate of decline of health and cognition, more for men than for women.[10] However, the large standard errors on the estimates of the short-run effect of retirement suggest caution and invite further investigations.

Thus, our next step is to investigate whether the evidence in Table 6 is homogeneous across job types. For this reason, in Table 7 we re-estimate our baseline model separately by type of jobs, distinguished by the value of our external index of physical strength. In the first two columns, we split the sample in two (almost equally sized) groups: *Low* (index values from 1 to 5) and *High* (index values from 6 to 10). Our results indicate that the immediate effect of retirement (the coefficient on the indicator *Retired*) is positive and statistically significant for those employed in more physically demanding jobs, corresponding to an increase by more than 9% of the probability of reporting good health (11% with respect to the mean value) and about half standard deviations for the cognitive index. On the other hand, for people employed in less physically demanding jobs, the coefficient is negative although not statistically significant. This helps explain why, in the whole sample, the estimated coefficients on the retirement indicator are characterized by very large standard errors. Although the effects of the years spent in retirement is negative for both groups, the immediate positive effects of retirement (*Retired*) is so large for people employed in strenuous jobs that the overall effects of retirement remains positive for at least 10 years.

To further explore the possibility of heterogeneity across jobs, in the last three columns we split the sample in three groups: *Very low* (index values from 1 to 3), *Median* (index values from 4 to 6) and *Very high* (index values from 7 to 10). The first category includes mainly managers, business professionals and most office workers, while the sat includes most of the jobs in mining, extraction, construction and manufacturing. To avoid weak instrument problems, we do not consider a further split by gender, at least in the main text. However, as described in Section 4.3, replicating the analysis by gender gives very similar results, in particular for the male subsample. The results

[10] It is important to remind once more, that the women sample is highly selected given their very low labor force attachment.

substantially confirm our findings from the first two columns of the table. In particular, we find negative effects of retirement on health and cognition (the coefficients on both the retirement indicator and the time spent in retirement) mainly among respondents employed in very low physically demanding jobs, and positive effects among those in highly physically demanding jobs. Obviously, since we reduce the sample size, the power of the instruments does not always allow us to estimate statistically significant coefficients. However, the sign and the size of the coefficient are consistent across outcomes.

Overall, our results are in line with those in Johnston and Lee (2009), who find positive short-term effect of retirement on individuals' mental health. The fact that they found positive effect of retirement for all workers can be explained by the use of a different estimation strategy—regression discontinuity design—that only allows the evaluation of the short-term effects of retirement, not its long term effects captured in our study by $DistR$, the years spent in retirement.

As for the effects by gender, Tables B.4 and B.5 in Appendix B present the results obtained by replicating our estimates using the original health and cognitive measures. In the case of the health measures, the results for depression and mobility limitations are fully consistent with those reported so far, while for SRH we find less evidence of heterogeneity across occupational groups. In the case of the cognitive measures, the results are also qualitatively and quantitatively similar to those reported in the main text in particular for memory and fluency.

Finally, we replicate our analysis using the psycho-social index but we do not find evidence of similar heterogeneity across occupational groups (results are available upon request).

4.3 Robustness checks

This section presents the results of a number of checks of the robustness of our estimation strategy.

A first set of robustness checks involves testing our baseline model against alternative specifications. In Section 4 we specified the age profile of health and cognition as linear, and distinguished between an immediate effect of retirement captured by the binary retirement indicator and a cumulative effect captured by the number of years spent in retirement. As an alternative, we now consider a more flexible specification of the age profile of health and cognition by including a set of age dummies (ages 50–55, 56–60, 61–65, and 66+). The results in Table B.1 show that the results of the two alternative specifications are quantitative and qualitatively similar. We also considered different specifications of the effect of the time spent in retirement. In particular, we considered replacing the linear specification with various alternatives, such as a quadratic or a cubic and one that includes a set of dummies for the years in retirement. Although in principle appealing, these alternatives are not feasible because they require additional and specific instruments, which dramatically affects the power of the 2SLS strategy.

As discussed in Section 3.2, the analysis of the heterogeneity of the effect of retirement across job types may be affected by the fact that people at the end of their working career might move to less physically demanding jobs. Specifically, our analysis might be biased if the selection into the occupational groups we use in Section 4 is correlated with our instruments, the eligibility ages for early and normal retirement. In Table B.7 we show that this does not occurs. In particular, the first column shows the result of a probit model for the probability of being in the high physically demanding group according to the external index. The table shows that our two instruments do not significantly affect the probability of being in the high group. It also shows that age and education help predict the probability of belonging to the high-burden group. In the case of age, the coefficient is negative and statistically significant but small in magnitude. This might suggest that people at end of their carrier move toward less physically demanding jobs. However, since this analysis is conducted on a cross-section, it might also capture cohort differences that disappear in the FD estimates.

Even more important is the strong correlation with educational attainments. As expected, people with lower education attainments (high school dropouts) are concentrated in the high burden group. Such result may raise the concern that the real source of heterogeneity in our data is education that leads to different occupation choices. For this reason, we replicate our analysis using education instead of job type. The results are qualitative similar only in the case of health, although more noisy (results are available upon request).

As discussed in Section 3.2, we deal with panel attrition by re-estimating our baseline model using the approach of Fitzgerald et al. (1998). This approach is based on the assumption that all determinants of attrition can be controlled for (selection on observables) and exploits the panel dimension of the data. Specifically, we estimate a probit model for the probability of participation in the second wave of SHARE by conditioning on the value of variables observed in the first wave, including lagged values of the dependent variables and information about the interviewer's characteristics (age, gender and education). The results—available upon request—indicate that people with lower cognitive abilities or in poor health, who are employed or are interviewed by an older interviewer are less likely to participate to the second wave. We then use the inverse of the fitted probability to construct the weights that we use in our main equation. In Table B.6 we compare the unweighted and the weighted estimates for depression and memory.[11] The table shows that the IPW approach does not lead to significantly different results. If anything, the precision of the estimates actually increases, despite the fact that we end up losing some of the observations

[11] Results for the other outcomes are similar and are available upon request.

in the original sample due to item nonresponse on some of the variables employed in constructing the weights (mainly fieldwork characteristics). Of course, by this we do not want to say that attrition can safely be ignored. However, considering that our weights take into account important observable information, such as fieldwork characteristics and baseline health and cognitive status, we believe that is really unlikely that there are unobservable factors driving the attrition process that may substantially change our results.

Another set of robustness checks looks for evidence of country heterogeneity. Our main concern here is that the results reported so far may reflect the influence of a small set of countries. Thus, we re-estimated our baseline model separately by region, distinguishing between Mediterranean countries (Greece, Italy and Spain), Continental European countries (Austria, Belgium, France, Germany, the Netherlands and Switzerland) and Scandinavian countries (Denmark and Sweden). The results—available upon request—show evidence of some heterogeneity across regions depending on the outcome of interest. However, we find no evidence of systematic differences across regions. Most importantly, heterogeneity across job types also holds when we focus on individual regions.

An additional concern regards the possibility that the heterogeneity in the effect of retirement may reflect gender differences in the probability of being employed in physically demanding jobs. In particular, there is a very low share of women employed in the high physical burden group and most of them are employed in jobs that are at the margin with the low category (index value of 6 or 7). However, in the light of our results, it would be sufficient to look at Table 6 to see that the heterogeneity we found by gender is completely different from that we found by job type. At any rate, as a robustness check we re-estimate our baseline separately by gender and job type (results are available upon request). It can be shown that the positive effect of retirement is larger among men in high burden jobs, but in this case the standard error is large due to the small sample size.

Finally, in Appendix C we describe how we construct the internal index of physical burden that we use as robustness check. Given the high correlation between the internal and the external index, the results using the internal index are fully consistent with those presented in the main text using the external one.

5 Conclusions

In this paper we estimate the causal effect of retirement on health and cognitive abilities using data from 10 European countries. Unlike previous papers (Rohwedder and Willis 2010, Coe and Zamarro 2011, Bonsang et al. 2012), we use panel data and we take the distance from retirement into account. We also exploit both cross-country and within-country variation in early and normal retirement ages as the key source of identification.

Consistently with our previous work (Mazzonna and Peracchi 2012), our results for the population as a whole suggest a negative effect of retirement on health and cognitive abilities. Further, these negative effects increase with the number of years spent in retirement. However, we also find evidence of substantial heterogeneity across occupations. In particular, the negative effect of retirement disappears when we focus on people who work or used to work in more physically demanding occupations. For these people, retirement has an immediate beneficial effect on both health and cognitive abilities.

Our results are particularly relevant for policy makers who, especially in Europe, worry about the effects of raising the retirement age as a way of improving the financial stability of social security programs. In fact, they provide further evidence of a negative effect of retirement on health and cognition for most of the population. On the other hand, the large heterogeneity of the effects of retirement across job types suggests that the design of pension reforms should also take care of the relatively small fraction of workers in very physically demanding occupation—for whom we find evidence of a positive effect of retirement.

References

Angelini V., Brugiavini A., and Weber G. (2009), "Ageing and unused capacity in Europe: Is there an early retirement trap?" *Economic Policy*, 24: 463–508.

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

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

Bound J., Schoenbaum M., Stinebrickner T., and Waidmann T. (1999) "The dynamic effects of health on the labor force transitions of older workers". *Labour Economics*, 6: 179–202.

Börsch-Supan A., and Jürges H. (2005), *The Survey of Health, Aging, and Retirement in Europe. Methodology.* Mannheim Research Institute for the Economics of Aging (MEA).

Börsch-Supan A. (2013), "Myths, scientific evidence and economic policy in an aging world." *The Journal of the Economics of Ageing*, 1: 3–15.

Börsch-Supan A., and Schuth M. (2013), "Early Retirement, Mental Health and Social Networks." In Wise D.A. (ed.), *Discoveries in the Economics of Aging*, University of Chicago Press.

Case A., and Deaton A. (2005), "Broken down by work and sex: How our health declines." In Wise D.A. (ed.), *Analyses in the Economics of Aging*, University of Chicago Press.

Case A., and Paxson C.H. (2008), "sex differences in morbidity and mortality." *Demography*, 42: 189–214.

Cawley J., Heckman J., and Vytlacil E. (2001), "Three observation on wages and measured cognitive ability." *Labour Economics*, 8: 419–452.

Charles K.K. (2004), "Is retirement depressing? Labor force inactivity and psychological well-being in later life." *Research in Labor Economics*, 23: 269–299.

Coe N.B., von Gaudecker H.M., Lindeboom M., and Maurer J. (2012), "The effect of retirement on cognitive functioning." *Health Economics*, 21: 913–927.

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

d'Hombres, B., Rocco L., Suhrcke M., and McKee M. (2010), "Does social capital determine health? Evidence from eight transition countries." *Health Economics*, 19: 56–74.

Dewey M. E., and Prince M.J. (2005), "Cognitive function." In Börsch-Supan A., Brugiavini A., Jürges H., Mackenbach J., et al. (eds.), *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe.* Mannheim Research Institute for the Economics of Aging (MEA).

Fitzgerald J., Gottschalk P., and Moffitt R. (1998), "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics". *Journal of Human Resources*, 33: 251–299.

Folstein M.F., Folstein S.E., and McHugh P.R. (1975), "Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician." *Journal of Psychiatric Research*, 12: 189–198.

Glass T.A., Mendes de Leon C., Marottoli R.A., and Berkman L.F (1999), "Population based study of social and productive activities as predictor of survival among elderly Americans." *British Medical Journal*, 319: 478–483.

Glymour M.M. (2007), “When bad genes look good: APOE-e4, cognitive decline, and diagnostic thresholds.” *American Journal of Epidemiology*, 165: 1239–1246.

Grossman M. (1972), “On the concept of health capital and the demand for health.” *The Journal of Political Economy*, 80: 223–255.

Gruber J., and Wise D.A. (2004), *Social Security Programs and Retirement around the World: Micro-Estimation*. University of Chicago Press, Chicago.

Herzog, A.R., and Wallace, R.B. (1997), “Measures of cognitive functioning in the AHEAD study”. *Journals of Gerontology Series B*, 52: 37–48.

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

Jürges H. (2007), ”True health vs. response styles: Exploring cross-country differences in self-reported health.” *Health Economics*, 16: 163–178.

Kajitani S., Sakata K., and McKenzie C. (2013), “Occupation, retirement and cognitive functioning.” Keio University, Joint Research Center for Panel Studies DP 2012-002.

Kasen S., Cohen P. Chen, H., and Castille D. (2003), “Depression in Adult Women: Age Changes and Cohort Effects.” *American Journal of Public Health*, 93(12): 2061–2066.

Kroll L.E. (2011), “Construction and validation of a general index for job demands in occupations based on ISCO-88 and KldB-92.” *Methoden – Daten – Analysen*, 5: 63–90.

Mazzonna F. (2013), “The long lasting effects of education on old age health: Evidence of gender differences.” *Social Science and Medicine*, 101: 129–138.

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

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

Peracchi F., and Rossetti C. (2012), ”Heterogeneiy in health responses and anchoring vignettes.” *Empirical Economics*, 42: 513–538.

Ravesteijn B., van Kippersluis H., and van Doorslaer E. (2013), “The contribution of occupation to health inequality.” *Research on Economic Inequality*, 21: 311–332.

Rohwedder S., and Willis R. (2010), “Mental retirement.” *Journal of Economic Perspectives*, 24: 1–20.

Santi I., Kroll L.E., Dietz A., Becher H., and Ramroth H. (2013), “Occupation and educational inequalities in laryngeal cancer: The use of a job index.” *BMC Public Health*, 13:1080. Available from `doi:10.1186/1471-2458-13-1080`.

Zantinge E.M., van den Berg M., Smit H.A., and Picavet H.S. (2014), “Retirement and a healthy lifestyle: Opportunity or pitfall? A narrative review of the literature.” *European Journal of Public Health* 24: 433–9.

Table 1: Sample size by country and gender.

	Country	Men	Women
AT	Austria	344	329
BE	Belgium	765	501
CH	Switzerland	204	162
DE	Germany	508	413
DK	Denmark	385	349
ES	Spain	317	152
FR	France	531	496
IT	Italy	491	284
NL	Netherlands	491	273
SE	Sweden	580	654
Total		4614	3612

Table 2: Principal component analysis (PCA) for cognitive tests.

	Components		
Variable	1st	2nd	3rd
Memory	0.593	-0.400	0.700
Fluency	0.595	-0.369	-0.714
Numeracy	0.543	0.840	0.019
Eigenvalue	1.734	.693	.573
Explained variance	.578	.231	.190

Table 3: Descriptive statistics.

Variable	Wave 1 mean	Wave 2 mean	Diff.
Health index	.805	.791	-.0148 ***
	(0.002)	(0.002)	
Cognitive index	-.057	.057	.114 ***
	(0.014)	(0.015)	

Table 4: Percentage of respondents in physically or psycho-socially demanding jobs by country.

	External indexes		Internal
	Physical	Psycho-social	index
AT	51.9	58.3	53.5
BE	36.5	47.4	41.8
CH	30.0	49.1	37.1
DE	44.3	52.6	48.0
DK	44.4	51.6	50.5
ES	70.0	59.5	76.3
FR	40.7	46.5	41.9
IT	61.9	55.9	59.0
NL	43.4	60.7	47.6
SE	41.8	53.8	43.8
Total	45.2	52.8	47.7
Correlation		40.4	72.9
			39.9

Table 5: First-differences estimates from the first-stage regression for $DistR = \max\{0, Age - R\}$ (Panel A) and *Retired* (Panel B) by gender type of job (low vs. high physical burden).

Panel A	All		Men	Women	Low	High
Model:	A	B	B	B	B	B
DistE	.457 ***	.461 ***	.458 ***	.487 ***	.440 ***	.475 ***
	(.045)	(.038)	(.043)	(.045)	(.045)	(.040)
DistN	.417 ***	.407 ***	.401 ***	.418 ***	.432 ***	.395 ***
	(.044)	(.038)	(.041)	(.049)	(.045)	(.041)
EligE	.340 ***	.352 ***	.470 ***	.205 **	.308 ***	.404 ***
	(.099)	(.082)	(.112)	(.083)	(.094)	(.089)
EligN	.355 ***	.342 ***	.439 ***	.182 *	.312 ***	.389 ***
	(.121)	(.098)	(.125)	(.106)	(.115)	(.101)
N	16452	16452	9228	7224	8582	7314
R^2	.660	.668	.682	.656	.659	.678
$F^\dagger$	1110.51 ***	779.23 ***	560.22 ***	640.02 ***	555.57 ***	558.36 ***

Panel B	All		Men	Women	Low	High
Model:	A	B	B	B	B	B
DistE	.041 ***	.042 ***	.034 ***	.059 ***	.050 ***	.035 ***
	(.007)	(.007)	(.009)	(.010)	(.010)	(.008)
DistN	-.046 ***	-.047 ***	-.042 ***	-.058 ***	-.053 ***	-.038 ***
	(.007)	(.007)	(.009)	(.011)	(.010)	(.008)
EligE	.116 ***	.118 ***	.129 ***	.102 ***	.089 ***	.149 ***
	(.023)	(.023)	(.027)	(.030)	(.024)	(.029)
EligN	.116 ***	.114 ***	.103 ***	.123 ***	.124 ***	.115 ***
	(.031)	(.031)	(.032)	(.046)	(.034)	(.035)
N	16452	16452	9228	7224	8582	7314
R^2	.125	.127	.120	.146	.133	.127
$F^\dagger$	27.64 ***	29.60 ***	20.13 ***	19.24 ***	21.06 ***	21.45 ***

Notes: The table reports the result from the first-stage regressions for *DistR* and *Retired* using the FD estimator. Model A also includes a linear age term and a binary indicator for marital status. Model B adds country-specific age trends. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level.
† F-test on the excluded instruments.

Table 6: Effects of retirement by gender (2SLS-FD).

	All		Men	Women
	A	B	B	B
Panel A: Predicted good health				
DistR	-.009 ***	-.007 ***	-.010 ***	-.003
	(.002)	(.002)	(.003)	(.004)
Retired	.035	.041	.049	.033
	(.029)	(.028)	(.045)	(.037)
J-test	2.809	2.436	2.536	5.401 *
Panel B: Cognitive score				
DistR	-.056 ***	-.060 ***	-.069 ***	-.057 ***
	(.014)	(.014)	(.018)	(.020)
Retired	.125	.106	.191	.076
	(.170)	(.152)	(.207)	(.206)
J-test	.370	.455	.632	2.076
$F^{\dagger}$	77.72	78.55	40.27	41.79
N	16452	16452	16452	16452

Notes: Model A also includes a linear age term and a binary indicator for marital status. Model B adds country-specific age trends. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are robust to clustering at the country and cohort level.
The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Table 7: Effects of retirement by physical burden of the job (2SLS-FD).

	Low (1–5)	High (6–10)	Very low (1–3)	Median (4–6)	Very high (7–10)
Panel A: Predicted good health					
DistR	-.006 *	-.007 *	-.006	-.008 **	-.005
	(.003)	(.004)	(.004)	(.004)	(.004)
Retired	-.011	.090 **	-.029	.055	.095 *
	(.035)	(.043)	(.037)	(.044)	(.053)
J-test	2.640	.574	.216	3.104	1.713
Panel B: Cognitive index					
DistR	-.053 ***	-.054 ***	-.075 ***	-.058 **	-.046 **
	(.019)	(.020)	(.022)	(.023)	(.023)
Retired	-.075	.505 **	-.035	-.182	622 **
	(.237)	(.239)	(.244)	(.275)	(.271)
J-test	1.142	.297	.062	1.574	1.261
$F^{\dagger}$	41.39	36.67	24.07	27.35	28.89
N	8384	7092	4502	5600	5374

Notes: The table reports results from 2SLS-FD estimates by job type (according to the external index of physical burden), as in Model B of Table 6. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level.
The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Figure 1: Early and normal eligibility ages for pension benefits (males).

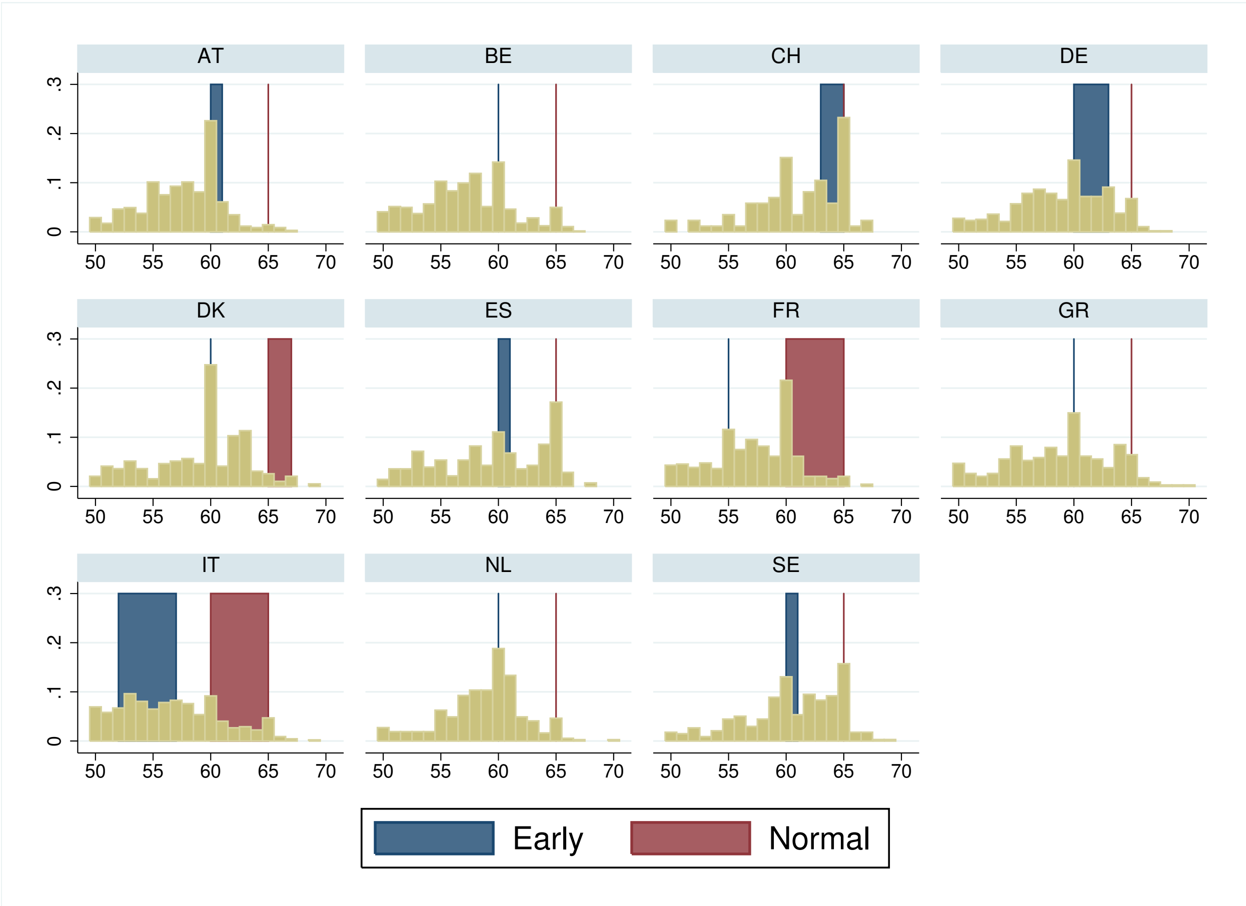
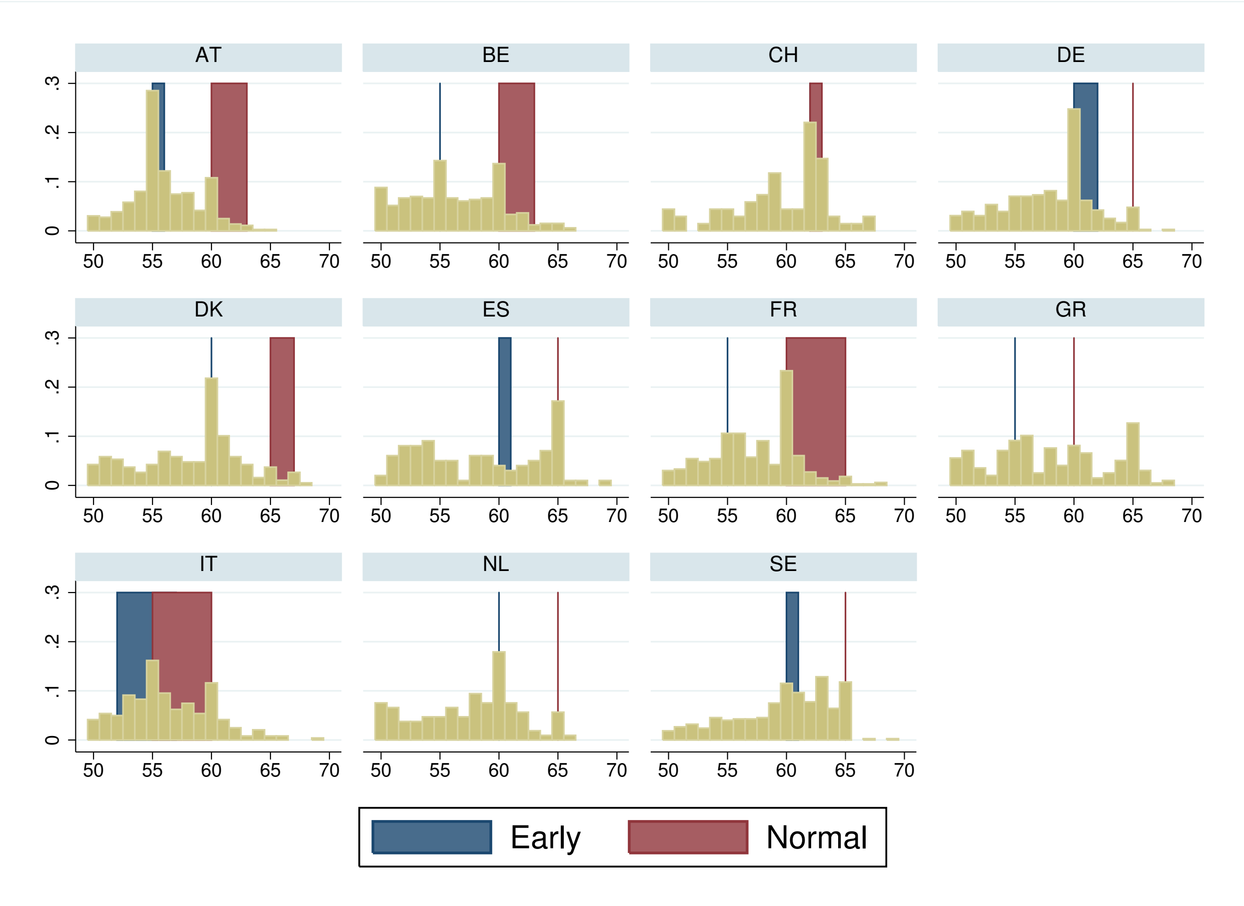


Figure 2: Early and normal eligibility ages for pension benefits (females).



A Pension eligibility rules in the SHARE countries

The eligibility ages reported in our paper are the results of the country specific retirement rules that involve the cohorts in the study (born in 1934–1954) until 2004 (the first wave of SHARE). Actually, in most of these countries eligibility for early retirement also depends on work experience through the number of years of contributions. Since we do not have this information, we consider the minimum age at which an individual may have access to early retirement. For changes in the legislations, we consider the first cohort "potentially" affected (FCA). For instance, in Belgium normal retirement age for women has been gradually raised from 60 to 63: age 61 in 1997, to age 62 in 2000, and age 63 in 2003. Thus, the FCA was the 1937 cohort in 1997, the 1939 cohort in 2000, and the 1941 cohort in 2003.

Our strategy cannot perfectly predict individuals' behaviors for several reasons. First, there are additional routes to early retirement in each country (e.g. long-term unemployment, disability, plant disclosure or redundant workers). In this case, we consider the retirement program that affects the largest fraction of respondents, avoiding to consider individual characteristics that might be endogenously correlated with respondents' health and cognition. Second, during the 90's several countries introduced changes in the legislation that removes adverse incentive effects by moving from a "pay as you go" system to less generous and more actuarially fair schemes. However, as in the case of Denmark, Germany and the Netherlands, these changes did not affect the legal early and normal retirement age. Nevertheless, these retirement ages are able to predict the most important jumps in the retirement age distribution reported in Figure 1 and 2.

The starting source of information on early and normal ages of eligibility for public old-age pensions in the SHARE countries is the Mutual Information System on Social Protection (MISSOC) database and the Social Security Administration (SSA) website. The MISSOC collects information on social protection for the member states of the European Union and other countries, including Switzerland. The SSA website highlights the principal features of social security programs in more than 170 countries every 2 years. These initial sources were supplemented with information from Gruber and Wise (1999, 2004, 2007), Angelini et al. (2009) and several other country specific auxiliary data sources. Below we report the statutory old age and early retirement ages used in this paper for each country. The retirement age reported are slightly different from Mazzonna and Peracchi (2012). In particular, we improve our information about past pension reforms, increasing the within country variability.

Austria

The statutory old retirement age is 65 for men and 60 for women.

The early retirement age was 60 for men and 55 for women. The 2000 and 2003 pension reforms gradually increased the eligibility age of 2 years (step-wise increase of 2 month each quarter) for men born after September 1940 and for women born after September 1945 (see Staubli and Zweimueller 2012).

Belgium

The statutory old age retirement is 65 for males. For women it was 60 but it has been gradually raised to 61 in 1997 (FCA 1937), 62 in 2000 (FCA 1939), and 63 in 2003 (FCA 1941).

The early retirement age is 55 for women and 60 for men.

Denmark

The statutory old age retirement is 67 for both men and women up to 2003. From 2004 it has been decreased to 65 (1999 reform).

The early retirement age is 60 for both men and women.

France

In France the statutory old retirement was 65 but it was lowered to 60 in 1983 for both men and women.

The early retirement age is 55 for some categories and 60 for others (we use 55 since it affected the largest share of workers).

Germany

The statutory old age retirement is, from 1961, 65 for both men and women.

In the case of early retirement age it is possible from 1973 to have a pension with full benefits at 63 years of age (with 35 years of contribution) for men, and 60 years of age (with 15 years of contribution) for women. However, public retirement insurance pays pensions without adjustment for employees from age 60 if certain conditions are met (e.g. unemployed, part-time employed and workers who cannot be appropriately employed). As shown by Berkel and Börsch-Supan (2004), the interpretation of the laws was particularly generous, so 60 became the actual early retirement age (see also the peak in Figure 1 in particular in the East Germany. In addition, retirement before 60 was also possible mainly using unemployment compensation. With two consecutive reforms (1992 and 2001), the system was simplified and the generosity of the system dramatically decreased. In particular, after a transitional period the early retirement age was set to 63 for everybody. For women, the step-wise increase took place in 2001 reaching 62 in 2004 and 63 in 2006. For men, the step-wise increase was heterogeneous across the categories aforementioned and affected the cohorts in this paper only marginally.

Italy

The statutory old age retirement was 60 (65 in the public sector) for men and 55 (60 in the public sector) for women from 1961 to 1993. Several consecutive reforms (1992, 1995 and 1998) increased the statutory old age retirement to 65 for men and 60 for women with step-wise increments from 1994 to 2000.

From 1965 to 1995, early retirement was possible at any age with 35 years of contributions (25 in the public sector) for both men and women; From 1996 to 2004 it was stepwise increased up to 57 for both the private and public sector.

The Netherlands

The statutory old age retirement is 65 both for men and women.

At the end of the eighties the eligibility age for many early retirement schemes (including disability and unemployment) was 60 or 61 for the majority of the employees without actuarial adjustment for both men and women. Starting from the mid-1990s, several financial parameters of the early retirement schemes were changed (e.g. the contribution requirements to obtain the maximum benefit) See also Euwals, Van Vuuren and Wolthoff 2010.

Spain

The statutory old age retirement is 65 both for men and women.

From 1983 to 1993, early retirement was possible at age 60 for both men and women; From 1994 to 2001, it was raised at 61 for both men and women; From 2002, early retirement is possible with 61 years of age and 30 years of contributions for both men and women.

Sweden

The statutory old age retirement was 67 for both men and women from 1961 to 1994. From 1995, it was changed to 65 for both men and women.

From 1961 to 1962 there was no early retirement. From 1963 to 1997, it was allowed at age 60 for both men and women; From 1998, it was raised to 61 for both men and women.

Switzerland

The statutory old age retirement is 65 for men and 62 for women from 1975. In 2001, women's retirement age was raised to 63.

Only since 1997 old-age insurance pensions can be claimed prior to the legal retirement age (1991 reform). In particular, men were allowed to retire at 64 from 1997 to 2001, and at 63 from 2001. For women early retirement was allowed at age 62 only from 2001.

Additional references for retirement ages

Berkel B., and Börsch-Supan A. (2004), "Pension reform in Germany: The impact on retirement decision." *FinanzArchiv: Public Finance Analysis*, 60(3): 393-421.

Euwals R., van Vuuren D., and Wolthoff R. (2010), "Early retirement behaviour in the Netherlands: Evidence from a policy reform." *De Economist* 158(3): 209–236.

Gruber J., and Wise D.A. (1999), *Social Security Programs and Retirement around the World.* University of Chicago Press, Chicago.

Gruber J., and Wise D.A. (2007), *Social Security Programs and Retirement around the World. Fiscal implications of the reform.* University of Chicago Press, Chicago.

Hanel B., and Riphahn R.T. (2006), "Financial incentives and the timing of retirement: evidence from Switzerland." IZA Discussion Paper No. 2492.

Jousten A., Léfebvre M., Perelman S., and Pestieau P. (2010), "The Effects of Early Retirement on Youth Unemployment: The Case of Belgium". In Gruber J. and Wise D.A. *Social Security Programs and Retirement around the World: The Relationship to Youth Employment.* University of Chicago Press, Chicago.

Social Security Administration (SSA) (2012). "Social Security Programs Throughout the World." http://www.ssa.gov/policy/docs/progdesc/ssptw/

Staubli S., and Zweimueller J. (2012), "Does raising the early retirement age increase employment of older workers?" *Journal of Public Economics*, 108: 17–32.

B Effects of retirement on the original measures of health and cognitive ability

Table B.1: Effects of retirement on health by physical burden of the job (specification with age dummies).

	Low (1-5)	High (6-10)	Very low (1–3)	Median (4–6)	Very high (7–10)
Panel A: Predicted good health					
DistR	-.010***	-.007*	-.009**	-.011**	-.008
	(.004)	(.004)	(.004)	(.004)	(.005)
Retired	-.033	.090*	-.032	.019	.109
	(.040)	(.053)	(.042)	(.053)	(.067)
J -test	3.176	1.070	.586	2.690	2.270
Panel B: Cognitive index					
DistR	-.037*	-.028	-.069***	-.026	-.012
	(.021)	(.025)	(.024)	(.025)	(.029)
Retired	-.021	.663**	-.079	.008	.816**
	(.297)	(.293)	(.294)	(.378)	(.336)
J -test	3.176	1.070	.586	2.690	2.270
$F^{\dagger}$	27.43	23.79	16.70	16.76	19.65
N	8384	7092	4502	5600	5374

Notes: The model also includes 4 age dummies, a binary indicator for marital status and country-specific age trends. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are robust to clustering at the country and cohort level. The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Table B.2: Effects of retirement on health by gender (2SLS-FD).

	All		Men	Women
	A	B	B	B
Not depressed				
DistR	-.010 *	-.010 *	-.020 ***	.006
	(.005)	(.005)	(.007)	(.008)
Retired	.021	.031	.064	-.015
	(.074)	(.071)	(.085)	(.107)
J-test	.281	.668	2.301	.664
No mobility limitations				
DistR	-.015 ***	-.014 ***	-.017 **	-.007
	(.005)	(.005)	(.007)	(.008)
Retired	.092	.095	.168 *	.033
	(.067)	(.064)	(.089)	(.099)
J-test	3.578	3.917	1.584	2.855
Good health				
DistR	-.015 **	-.018 ***	-.024 ***	-.006
	(.006)	(.005)	(.007)	(.007)
Retired	.046	.054	.120	-.007
	(.079)	(.065)	(.095)	(.075)
J-test	4.856 *	6.559	7.127	3.173
$F^{\dagger}$	78.54	79.18	40.32	42.51
N	16452	16452	9228	7224

Notes: Model A also includes a linear age term and a binary indicator for marital status. Model B adds country-specific age trends. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level.

The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.

† F-test on the excluded instruments.

Table B.3: Effects of retirement on cognitive ability by gender (2SLS-FD).

	All		Men	Women
	A	B	B	B
Memory				
DistR	-.121 ***	-.141 ***	-.140 ***	-.163 **
	(.046)	(.044)	(.054)	(.069)
Retired	.269	.190	.373	.283
	(.482)	(.449)	(.632)	(.640)
J-test	1.325	1.292	.741	5.775 *
Fluency				
DistR	-.245 ***	-.273 ***	-.290 ***	-.264 **
	(.081)	(.079)	(.108)	(.109)
Retired	.513	.363	1.279	-.891
	(1.115)	(.964)	(1.458)	(1.211)
J-test	.614	1.811	3.843	1.369
Numeracy				
DistR	-.025 **	-.022 *	-.037 **	-.009
	(.012)	(.012)	(.015)	(.019)
Retired	.101	.125	.118	.173
	(.135)	(.132)	(.182)	(.185)
J-test	.803	1.332	2.461	1.654
$F^{\dagger}$	78.54	79.18	40.32	42.51
N	16452	16452	9228	7224

Notes: Model A also includes a linear age term and a binary indicator for marital status. Model B adds country-specific age trends. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level.
The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Table B.4: Effects of retirement on health by physical burden of the job (2SLS, Model B).

	Low (1–5)	High (6–10)	Very low (1–3)	Median (4–6)	Very high (7–10)
Not depressed					
DistR	-.006	-.011	-.019 **	.006	-.013
	(.007)	(.009)	(.010)	(.008)	(.010)
Retired	-.179 **	.226 **	-.142	-.076	.265 **
	(.080)	(.100)	(.093)	(.120)	(.106)
J-test	.152	1.087	6.342	2.469	.464
No mobility limitations					
DistR	-.013 *	-.015 *	-.016 *	-.016 *	-.011
	(.007)	(.008)	(.009)	(.009)	(.009)
Retired	.000	.205 **	-.147	.243 **	.184
	(.079)	(.099)	(.092)	(.107)	(.118)
J-test	3.171	1.467	2.304	2.190	3.012
Good health					
DistR	-.020 ***	-.009	-.025 ***	-.011	-.019 **
	(.007)	(.009)	(.009)	(.009)	(.009)
Retired	.051	.058	-.063	.057	.122
	(.087)	(.115)	(.096)	(.128)	(.108)
J-test	1.948	4.592	2.633	4.208	3.478
$F^{\dagger}$	40.84	37.73	23.94	27.77	29.34
N	8522	7238	4570	5710	5480

Notes: The table reports results from 2SLS-FD estimates by job type (according to the external index of physical burden), as in Model B of Table 6.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are robust to clustering at the country and cohort level. The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Table B.5: Effects of retirement on cognitive abilities by physical burden of the job (2SLS, Model B): Low (1–5) vs. high (6–10).

	Low (1–5)	High (6–10)	Very low (1–3)	Median (4–6)	Very high (7–10)
Memory					
DistR	-.156 ***	-.120 *	-.151 **	-.120 *	-.157 **
	(.060)	(.061)	(.074)	(.066)	(.070)
Retired	-.492	1.394 **	-.456	-.904	2.124 ***
	(.564)	(.710)	(.637)	(.804)	(.824)
J-test	1.401	1.032	1.045	3.744	1.109
Fluency					
DistR	-.310 ***	-.169	-.313 **	-.266 **	-.159
	(.113)	(.106)	(.134)	(.129)	(.123)
Retired	-.755	1.972	-.251	-1.590	2.578 *
	(1.463)	(1.226)	(1.695)	(1.699)	(1.483)
J-test	1.637	.229	1.803	2.585	1.346
Numeracy					
DistR	-.020	-.035 *	-.047 **	-.020	-.009
	(.015)	(.018)	(.022)	(.022)	(.021)
Retired	.037	.205	.047	.272	.150
	(.209)	(.237)	(.233)	(.274)	(.263)
	40.636	36.058	23.786	26.880	28.468
$F^\dagger$	40.84	37.73	23.94	27.77	29.34
N	8522	7238	4570	5710	5480

Notes: The table reports results from 2SLS-FD estimates by job type (according to the external index of physical burden), as in Model B of Table 6.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are robust to clustering at the country and cohort level. The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1.
† F-test on the excluded instruments.

Table B.6: Unweighted vs. weighted estimates of the effects of retirement on predicted good health and cognitive index by physical burden of the job.

	Low (1-5)		High (6-10)	
	Unweighted	Weighted	Unweighted	Weighted
Panel A: Predicted good health				
DistR	-.005 *	-.005 *	-.007 *	-.007 *
	(.003)	(.003)	(.004)	(.004)
Retired	-.018	-.028	.091 **	.096 **
	(.035)	(.034)	(.041)	(.043)
Panel B: Cognitive index				
DistR	-.059 ***	-.068 ***	-.056 ***	-.062 ***
	(.018)	(.020)	(.020)	(.022)
Retired	-.136	-.101	.503 **	.526 **
	(.219)	(.225)	(.227)	(.227)
N	8294	8294	7044	7044
$F^{\dagger}$	41.351	40.610	37.132	35.508

Notes: In this table we compare unweighted and weighted (IPW) estimates of the effect of retirement on depression and Memory by job physical burden using 2SLS-FD. The table reports results from 2SLS-FD estimates by job type (according to the external index of physical burden), as in Model B of Table 6.
Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level. The asymptotic null distribution of the Sargan-Hansen J statistic for testing the overidentifying restrictions is χ^2_1 .
† F-test on the excluded instruments.

Table B.7: Probit for the probability of being employed in high physically demanding jobs.

Age	-.010 **
	(.004)
DistE	.000
	(.004)
DistN	.006
	(.004)
EligE	.013
	(.019)
EligN	.030
	(.021)
Low educ	.357 ***
	(.011)
N	7893

The model also includes country fixed effects, country specific linear age trends and controls for marital status.
Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Standard errors are robust to clustering at the country and cohort level.

C Internal index of physical burden.

We construct our index by using the response from those who are currently employed to estimate the following linear model

$$PB_{ij} = \alpha + \beta A_i + \delta_j O_{ij} + U_{ij}, \tag{1}$$

where PB_{ij} is the self-reported level of physical burden (that ranges between 1 and 4) of the occupation of individual i working in occupation j, A_i is the age of the respondent, O_{ij} is a vector of binary indicators for each occupation recorded in SHARE based on the second digit of the ISCO-88 classification, and U_{ij} is a regression error with the usual properties. To control for differences by gender and country, we also include among the regressors in model (1) an indicator for being a female and a set of indicators for the country of residence. Our index of physical burden is just $\widehat{\delta}_j O_{ij}$, where $\widehat{\delta}_j$ is the OLS estimate of the vector of coefficients δ_j in model (1). One advantage of this approach is that we can construct the index also for those who are retired using information on their last job. Based on the value of our index, we separate respondents into two groups: those in less physically demanding jobs and those in more physically demanding jobs. We use the threshold value of 2, corresponding to "agreement" with the statement reported above. The results using this internal index are largely consistent with those reported in the main text.