Unhealthy Retirement?

Fabrizio Mazzonna
Franco Peracchi

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 jobs characterized by a high level of physical burden.

I. 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

---

Fabrizio Mazzona is an assistant professor of economics at the Università della Svizzera Italiana (USI) and a fellow of the Munich Center for the Economics of Aging (MEA). Franco Peracchi is a professor of econometrics at the University of Rome Tor Vergata, a professor of the practice at Georgetown University, and a fellow of the Einaudi Institute for Economics and Finance (EIEF) in Rome, Italy. *We thank two anonymous referees, Mary Ann Bronson, Janet Currie, Angus Deaton, Peter Eibich, Osea Giuntella, and seminar participants at Princeton, DIW Berlin, Georgetown University and IIPF Conference 2014 for valuable comments. Angus Deaton, Peter Eibich, Osea Giuntella, and seminar participants at Princeton, DIW Berlin, Georgetown University and IIPF Conference 2014 for valuable comments. The paper was written while the second author was visiting the Center for Health and Wellbeing at Princeton University. The authors use data from SHARE Waves 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 programs. 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). The data used in this article can be obtained beginning December 2017 through November 2020 from Franco Peracchi, Department of Economics and Finance, via Columbia 2, 00133 Rome, Italy, franco.peracchi@uniroma2.it.*
[Submitted September 2014; accepted July 2015]; doi:10.3368/jhr.52.1.0914-6627R1
ISSN 0022-166X E-ISSN 1548-8004 © 2017 by the Board of Regents of the University of Wisconsin System
Supplementary materials are freely available online at: http://uwpress.wisc.edu/journals/journals/jhr-supplementary.html

increasing the average retirement age of the work force. 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 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, van Kippersluis, and van Doorslaer 2013 for a review).

On the other hand, there is a view that retirement may be harmful. This may happen if a lack of purpose 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 actually can 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 also can be predicted if retirement reduces social interactions. The social capital literature (see 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 (2014) 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 work force but a positive effect for people working in strenuous jobs.

Empirically, it is difficult to provide causal evidence of the effect of retirement. As 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 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 U.K. and U.S. 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 U.S. old-age surveys show evidence of a negative effect of retirement on cognitive abilities (Rohwedder and Willis 2010; Bonsang, Adam and Perelman 2012; 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 United States.

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-collar 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 multidimensional (e.g., physical and psychosocial). 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 psychosocial 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 work force, 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 II describes the data used for this study. Section III presents our empirical model and discusses a number of issues that complicate the identification of the causal effect of retirement on cognitive abilities. Section IV presents our main results as well as a large battery of robustness checks. Finally, Section V offers some conclusions.

II. 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 biannual 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 socioeconomic status, health, social, and family networks for nationally representative samples of elderly people in the participating countries.

A. 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 and Jürges 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 III.)

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 8,163 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.

1. In Greece, the short fieldwork period in the first wave (roughly two 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.

Table 1
Sample Size by Country and Gender

	Country	Men	Women
AT	Austria	341	328
BE	Belgium	757	495
CH	Switzerland	196	160
DE	Germany	507	413
DK	Denmark	382	348
ES	Spain	316	152
FR	France	530	492
IT	Italy	491	283
NL	Netherlands	487	270
SE	Sweden	570	645
Total		4,577	3,586

B. 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 considered 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 health perception and response style. (See 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) 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, heart 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:

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

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 c_{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. Because roughly 80 percent 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 two 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 (such as depression, anxiety, and suicidality). Our indicator of depression takes value 1 for values of Euro-D lower than 4 and value 0 otherwise. This particular cutoff point was validated in the EURODEP studies against a variety of clinically relevant indicators (Dewey and Prince 2005). 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, in the Online Appendix B we replicate 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, Folstein, and McHugh 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 Study of Ageing (ELSA).

The memory test consists of verbal registration and recall of a list of ten words, carried out twice. The first time is immediately after the encoding phase (immediate recall), while the second time is some five 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.[2] 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; Cawley, Heckman, and Vytlacil 2001). Table 2 presents our PCA results. The first principal component explains almost 60 percent 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 reasonably can rely on a single index of cognitive

2. We trim values above 45, which represents 0.5 percent of the distribution. These values are obviously implausible because they mean that these respondents were able to name almost one animal per second.

Table 2
Principal Component Analysis 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	0.693	0.573
Explained variance	0.578	0.231	0.190

ability. As for health, the Online 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 percent over our two-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 results is partly due to retesting effects, because the battery of cognitive tests is exactly the same in the two waves. Retesting effects may affect results if they differ for employed and retired people. To investigate this issue, we use the information from Wave 2 distinguishing between people from the longitudinal sample, who were exposed to these tests for the second time, and people from the refreshment sample, who were exposed to these tests for the first time.[3] 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.

C. 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

3. We regress our cognitive index on a dummy for belonging to the refreshment sample, interacted with either a dummy for being retired or a dummy for having less than college education. Both regressions also control for age, education, gender, employment status, and country fixed effects. Results are presented in Table B.8 of the Online Appendix.

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

Table 3
Descriptive Statistics

Variable	Wave 1 Mean	Wave 2 Mean	Difference
Health index	0.805	0.791	–0.015***
	(0.002)	(0.002)	
Cognitive index	–0.057	0.057	0.114***
	(0.014)	(0.015)	

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, Sakata and McKenzie 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 percent of all two-digit codes, 94.8 percent of the 3-digit codes, and 78.5 percent of the four-digit codes). From these JEMs, Kroll (2011) derives two indexes, a "physical job index" and a "psychosocial 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 psychosocial 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 psychosocially 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 psychosocial 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, (See Section IV 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

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 psychosocial index for robustness checks.

Table 4
Percentage of Respondents in Physically or Psychosocially Demanding Jobs by Country

	External indexes		
	Physical	Psycho-social	Internal 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
Correlations		40.4	72.9
			39.9

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 the Online Appendix B, we show how we deal with these two issues. Because the results obtained using this internal index are very similar to those presented in Section IV, 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 psychosocially 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 psychosocially demanding (external index of psychosocial 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 psychosocial burden, whose correlation with the two indexes of physical burden (external and internal) is only about 40 percent. This suggests that the physical and psychosocial burden, although correlated across jobs, do not coincide and represent distinct aspects of a job.

III. Model Specification and Estimation

The empirical strategy 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 attainment. The main drawback is panel attrition, which implies a loss of roughly 30 percent of the initial sample. This issue will be discussed in more detail in Section III.C.

Another important difference is that we now use a less restrictive empirical specification of the effect of retirement on health and cognition.

A. Model Specification

As mentioned in the Introduction, previous literature modeled 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.

Following this argument, our baseline specification (Model A) is:

$$H_{it} = \alpha_i + \beta_1 Age_{it} + \beta_2 Retired_{it} + \beta_3 DistR_{it} + \beta_4 D_i + \beta_5 X_{it} + U_{it},$$

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, $DistR_{it} = max\{Age_{it} - R_{it}\}$ is the number of years spent in retirement (equal to zero if the individual is not yet retired), D_i is a set of binary time-invariant indicators for educational attainment and the country of residence (with Belgium as the reference country), X_{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, so it need not correspond to recipiency 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) that 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 also should 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 difference (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 still may bias our estimates. More importantly, taking first differences does not eliminate the bias due to potential reverse causality from health to retirement (such as a health shock that 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 III.C discusses other identification issues that are relevant for our empirical strategy.

Finally, it is important to notice that, because 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* 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, the coefficient on *DistR* simply would be identified as the deviation from the mean value of first differences in health and cognition of respondents already retired in Wave 1.

B. 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.[8] 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. So, our IV strategy picks up the effects on health and cognition of variations in retirement status and distance from retirement induced by changes in eligibility. Whether these legislated changes are expected or not is an issue, because expected changes may affect health investment or cognitive repair activities and

8. A simple alternative is to consider the reduced form relationship and look if health or cognition change in a discernible way around the pension eligibility age, as suggested by Bound and Waidmann (2007). With a single country and only one eligibility age, this strategy is easy to implement and very transparent. With a single country but two eligibility ages, the effects of one of them may be confounded by the presence of the other. The problem becomes even more complicated in our multicountry setting with country-specific early and normal eligibility ages.

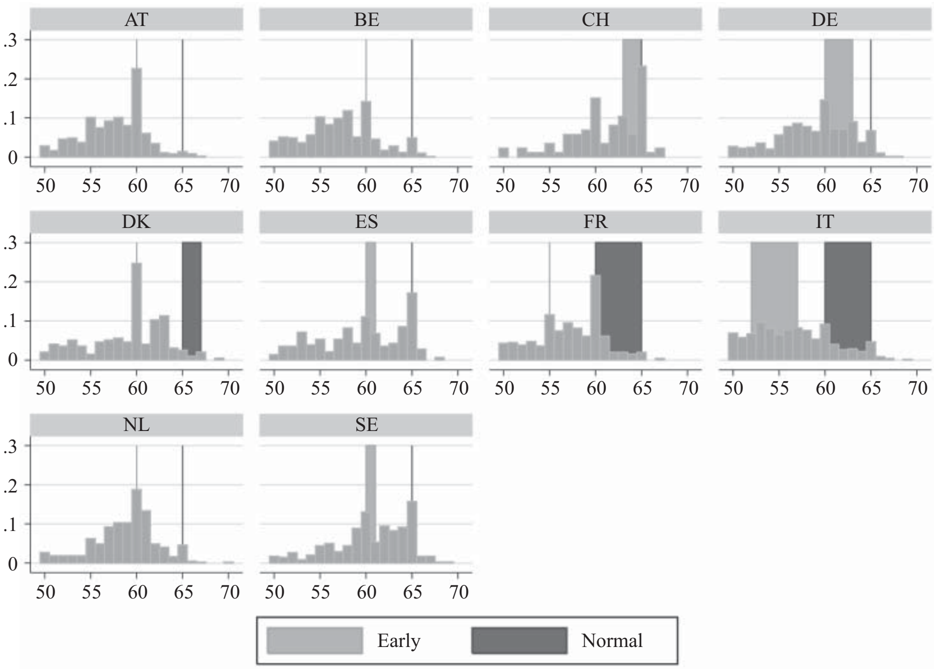

Figure 1
Distribution of Retirement Ages and Changes in Early and Normal Pension Eligibility Ages. Males.

therefore health and cognition. Bound and Waidmann (2007) argue that these effects are small and, if anything, an anticipation effect should downward bias our estimated effects of retirement.

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 1990s. We refer to the Online 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

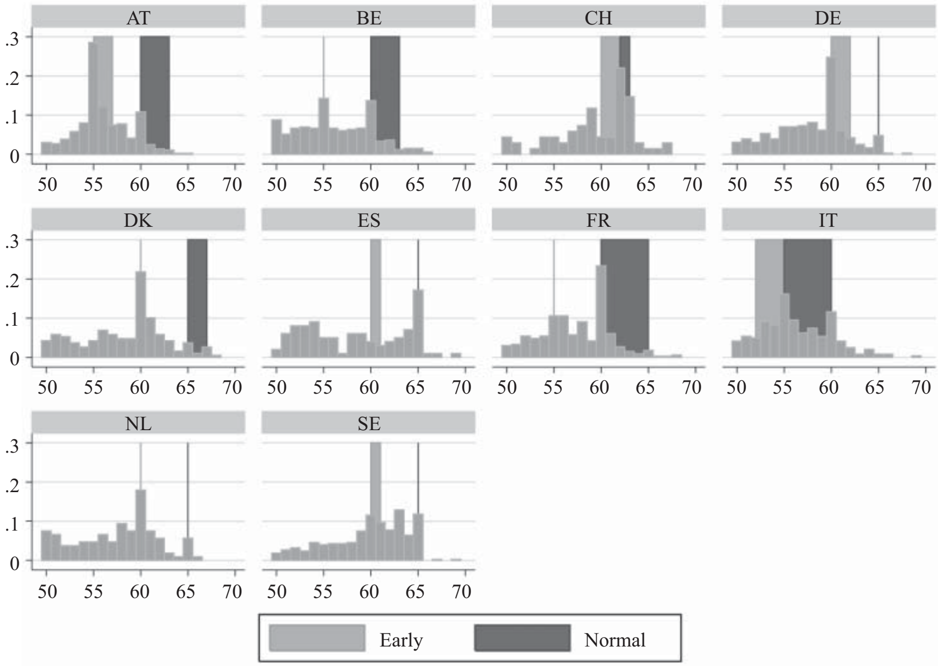

Figure 2
Distribution of Retirement Ages and Changes in Early and Normal Pension Eligibility Ages. Females.

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 (in other words, $DistE_{it} = max\{Age_{it} - E_i\}$ and $DistN_{it} = max\{Age_{it} - N_i\}$. Because we have two endogenous regressors, our model is over-identified, which allows us to test the exogeneity of the instruments through a Sargan-Hansen J test of the over-identifying restrictions.

Instruments relevance is the other fundamental issue to take into account because weak instruments may affect the finite sample properties of the IV estimates and bias them in the direction of OLS (Bound, Jaeger, and Baker 1995). In our context—characterized by two endogenous variables and four instruments—the appropriate diagnostic tool to test for the presence of weak instrument is the Kleibergen-Paap F statistic (Kleibergen and Paap 2006). The usual approach in the literature is to conclude that the instruments are not weak if these test statistics exceed the critical values tabulated by Stock and Yogo (2005) regarding the relative OLS bias and the maximal test size. For this reason, at the bottom of each table presented in Section IV.B we show the value of this statistic (along with the value of Sargan-Hansen test of the over-identifying restrictions).

C. 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. Because education is an important determinant of health and cognition in later life (see for example Mazzonna 2014), 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.[9]

Our FD strategy also may 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 worse health conditions and depend on retirement status. However, if attrition is mainly by characteristics of the respondents that do not change between the two waves, then the FD estimator is unbiased. As a robustness check, we compare our results to those obtained using the Inverse Probability Weighting approach of Fitzgerald, Gottschalk, and Moffitt (1998), which allows attrition to be nonrandom and to depend on individual characteristics observed in the first wave[10], including health, cognitive outcomes, and fieldwork characteristics such as the interviewer's age, gender, and education. (See Section IV.C 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 is convincingly argued by Coe and Zamarro (2011) using the same data. Second, specifying a higher-order polynomial in age also would 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 IV.C, the results are very similar to those obtained from our baseline specification.

Concerns also may 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

9. In the case of cohort heterogeneity we also can exploit age variation for individual belonging to the same cohort since we take into account the exact interview date.
10. As argued by Wooldridge (2010), this assumption is reasonable in short panels.

IV.C 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 experience 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.

IV. 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 III. We also present the results of a number of checks of the robustness of our estimation strategy.

A. 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 III.B, 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 for *DistR* and the bottom panel 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.

B. 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). The number of individuals is slightly smaller than in Table 5 because of the presence of item nonresponse on single health or cognitive questions. At the bottom

Table 5

First-differences Estimates from the First-Stage Regression for DistR = max{0, Age − R} (Top Panel) and Retired (Bottom Panel) by Gender and Type of Job (Low vs. High Physical Burden)

	All		Men	Women	Low	High
	A	B	B	B	B	B
DistR						
DistE	0.458***	0.462***	0.458***	0.488***	0.442***	0.477***
	(0.045)	(0.038)	(0.043)	(0.045)	(0.045)	(0.040)
DistN	0.416***	0.407***	0.400***	0.417***	0.431***	0.394***
	(0.044)	(0.038)	(0.041)	(0.049)	(0.045)	(0.041)
EligE	0.346***	0.358***	0.479***	0.206**	0.315***	0.405***
	(0.099)	(0.082)	(0.111)	(0.083)	(0.094)	(0.089)
*Elig*N	0.358***	0.345***	0.438***	0.191*	0.322***	0.385***
	(0.122)	(0.099)	(0.125)	(0.110)	(0.118)	(0.101)
N	8,163	8,163	4,577	3,586	4,263	3,624
R^2	0.661	0.670	0.683	0.659	0.661	0.679
F^a	1110.5***	779.2***	560.2***	640.0***	555.6***	558.4***
Retired						
DistE	0.041***	0.042***	0.035***	0.059***	0.050***	0.035***
	(0.007)	(0.007)	(0.009)	(0.010)	(0.010)	(0.008)
DistN	−0.045***	−0.046***	−0.042***	−0.058***	−0.053***	−0.037***
	(0.007)	(0.007)	(0.009)	(0.011)	(0.010)	(0.008)
EligE	0.114***	0.115***	0.126***	0.101***	0.087***	0.148***
	(0.022)	(0.022)	(0.026)	(0.030)	(0.024)	(0.028)
EligN	0.115***	0.113***	0.103***	0.120***	0.121***	0.116***
	(0.031)	(0.031)	(0.032)	(0.046)	(0.034)	(0.035)
N	8,163	8,163	4,577	3,586	4,263	3,624
R^2	0.124	0.126	0.119	0.144	0.132	0.127
F^a	27.64***	29.60***	20.13***	19.24***	20.78***	21.74***

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.

[a]*F*-test on the excluded instruments.

of each table we report the values of the Kleibergen-Paap *F*-statistic for weak instruments and the Sargan-Hansen *J*-statistic for the validity of the over-identifying restrictions. The latter is asymptotically distributed as a χ^2 with 2 degrees of freedom under the null hypothesis that the over-identifying restrictions are valid. The null is only rejected test in the case of women's health (and only at the 10 percent level), thus lending

Table 6
Effects of Retirement by Gender (2SLS-FD)

	All		Men	Women
	A	B	B	B
Predicted Good Health				
DistR	–0.009***	–0.007***	–0.010***	–0.003
	(0.002)	(0.002)	(0.003)	(0.004)
Retired	0.038	0.042	0.051	0.033
	(0.030)	(0.028)	(0.045)	(0.037)
N	8,007	8,006	4,479	3,528
Kleibergen-Paap *F*	76.587	78.820	39.694	42.367
Sargan-Hansen *J*	2.597	2.418	2.529	5.246*
Cognitive Score				
DistR	–0.056***	–0.060***	–0.069***	–0.057***
	(0.014)	(0.014)	(0.018)	(0.020)
Retired	0.121	0.107	0.191	0.075
	(0.170)	(0.153)	(0.210)	(0.205)
N	7,888	7,888	4,412	3,476
Kleibergen-Paap *F*	72.814	74.851	36.819	41.418
Sargan-Hansen *J*	0.383	0.452	0.577	2.084

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.

support to our IV strategy. We also can test the hypothesis of weak instruments by comparing the value of the Kleibergen-Paap F-statistic with the critical values tabulated by Stock and Yogo (2005). In particular, with two endogenous variables and four instruments, the critical value for a maximum relative bias of 5 percent relative to OLS is 11.04, while the critical value for size distortion greater than 10 percent is 16.87, far below the values of our statistic.

In the pooled estimates, retirement has a clear negative effect on health and cognitive abilities. In line with the findings in 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 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 percent 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 percent of a standard deviation. 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 the Online Appendix show that these results are substantially unaffected when we use the original health and cognitive measures. On the other hand, replicating our analysis using the psychosocial index shows little evidence heterogeneity across occupational groups. (Results are available upon request.)

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.[11] 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 reestimate 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 percent of the probability of reporting good health (11 percent with respect to the mean value) and about half of a standard deviation 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 last includes most of the jobs in mining, extraction, construction, and manufacturing. The results confirm our findings from the first two columns of the table. In particular, we find negative effects of retirement on health and cognition, mainly among respondents employed in very low physically demanding jobs, and positive effects of retirement among those in highly physically demanding jobs. However, because we the sample size is reduced by roughly one-half, some of the estimated coefficients are no longer statistically significant.

In Table 8 we show our results splitting the sample by gender and occupation to show that the results by occupation are not driven by the gender composition of the different occupational groups. Obviously, the reduction in the sample size affects the power of our estimation strategy. Even though standard errors increase, point estimates are similar

11. It is important to remind that the women sample is highly selected due to the lower labor force attachment.

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)
Predicted good health					
DistR	–0.006*	–0.007**	–0.006	–0.008**	–0.006
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Retired	–0.011	0.092**	–0.025	0.055	0.098*
	(0.035)	(0.043)	(0.037)	(0.044)	(0.054)
N	4,193	3,546	2,251	2,801	2,687
Kleibergen-Paap *F*	41.807	36.455	23.953	27.643	28.626
Sargan-Hansen *J*	2.504	0.578	0.142	3.109	1.854
Cognitive index					
DistR	–0.060***	–0.057***	–.074***	–0.057**	–0.047**
	(0.018)	(0.020)	(0.021)	(0.023)	(0.023)
Retired	–0.122	0.521**	–0.037	–0.175	0.632**
	(0.219)	(0.235)	(0.245)	(0.274)	(0.274)
N	4,137	3,485	2,228	2,765	2,629
Kleibergen-Paap *F*	39.489	34.832	22.330	26.704	27.169
Sargan-Hansen *J*	0.520	0.393	0.059	1.551	1.212

Notes: The specification corresponds to Model B in 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.

to those reported in Table 7 with more evidence of heterogeneity across occupational groups for men. This is understandable as there are very few women in the most physically demanding occupations (Categories 8 to 10).

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.

C. 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 IV 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

Table 8
Effects of Retirement by Physical Burden of the Job and Gender (2SLS-FD)

	Men		Women	
	Low (1–5)	High (6–10)	Low (1–5)	High (6–10)
Predicted Good Health				
DistR	–0.011**	–0.008	0.001	–0.004
	(0.004)	(0.005)	(0.004)	(0.006)
Retired	0.006	0.134*	–0.017	0.042
	(0.044)	(0.081)	(0.046)	(0.051)
N	2,331	1,962	1,862	1,584
Kleibergen-Paap *F*	25.26	15.04	19.44	22.85
Sargan-Hansen *J*	2.215	4.603	0.726	5.001*
Cognitive index				
DistR	–0.069***	–0.069***	–0.049*	–0.053*
	(0.027)	(0.026)	(0.026)	(0.031)
Retired	0.039	0.648*	–0.068	0.399
	(0.292)	(0.360)	(0.330)	(0.271)
N	2,301	1,926	1,836	1,559
Kleibergen-Paap *F*	22.67	14.50	19.61	21.69
Sargan-Hansen *J*	2.115	0.489	6.79**	1.045

Notes: The specification corresponds to Model B in 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.

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 2.1 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 already mentioned, the results of our analysis 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 IV is correlated with our instruments, the eligibility ages for early and normal retirement. In Table 2.5 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 career 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. As expected, people with lower education attainment are concentrated in the high burden group. This may raise the concern that the real source of heterogeneity in our data is education, through its impact on occupational choices. For this reason, we replicate our analysis using education instead of job type. The results presented in Tables B.6 and B.7 of the Online Appendix confirm that the main source of heterogeneity is type of occupation, not education. In particular, Table B.7 focuses on people with less than college education and shows that, even within this group, there is evidence of heterogeneity across jobs.

As discussed in Section III.C, we deal with panel attrition by reestimating our baseline model using the approach of Fitzgerald, Gottschalk, and Moffitt (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 2.4 we compare the unweighted and the weighted estimates for depression and memory.[12] The table shows that the IPW approach does not lead to substantially different results. If anything, the precision of the estimates increases despite the fact that some observations are lost due to item nonresponse on the variables employed to construct the weights (mainly fieldwork characteristics). Because our weights take into account important observable information, such as fieldwork characteristics and baseline health and cognitive status, we conclude it is unlikely that unobservable factors driving the attrition process 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 reestimated 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.

12. Results for the other outcomes are similar and are available upon request.

V. Conclusions

In this paper we estimate the causal effect of retirement on health and cognitive abilities using data from ten European countries. Unlike previous papers (Rohwedder and Willis 2010; Coe and Zamarro 2011; Bonsang, Adam and Perelman 2012), we use panel data and take the distance from retirement into account. We also exploit within- and cross-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 suggest that the average effects of retirement on health and cognitive abilities are negative. Further, these negative effects become larger as the number of years spent in retirement increases. However, we also find evidence of substantial heterogeneity depending on previous occupation. In particular, the negative effect of retirement disappears when we focus on people who worked 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 policymakers 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 heterogeneity of the effect across job types suggests that the design of pension reforms also should take care of the relatively small fraction of workers in very physically demanding occupations—for whom we find evidence of a positive effect of retirement.

Some European countries, such as Italy, have recently increased the eligibility age for normal retirement but allow people in very strenuous jobs (e.g., mining) to retire earlier. Similarly in France, the 2010 pension reform includes early retirement scheme for workers in strenuous jobs. Our results lend support to this kind of policies. On the other hand, one cannot ignore the recent macro literature (see Manuelli, Seshadri, and Shin 2012), which warns against the potentially distortive effects that these policies may have on human capital accumulation and occupational choices.

References

Bingley, Paul, and Alessandro Martinello. 2013. "Mental Retirement and Schooling." *European Economic Review* 63:292–98.

Bonsang. Eric, Stéphane Adam, and Sergio Perelman. 2012. "Does Retirement Affect Cognitive Functioning?" *Journal of Health Economics* 31(3):490–501.

Bound, John, David Jaeger, and Regina Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90(430):443–50.

Bound, John, Michael Schoenbaum, Todd Stinebrickner, and Timothy Waidmann. 1999. "The Dynamic Effects of Health on the Labor Force Transitions of Older Workers." *Labour Economics* 6(2):179–202.

Bound, John, and Timothy Waidmann. 2007. "Estimating the Health Effects of Retirement." Working Paper WP 2007-168. Ann Arbor, Mich.: University of Michigan Retirement Research Center.

Börsch-Supan, Axel. 2013. "Myths, Scientific Evidence and Economic Policy in an Aging World." *Journal of the Economics of Ageing* 1–2:3–15.

Börsch-Supan, Axel, and Hendrik Jürges. 2005. *The Survey of Health, Aging, and Retirement in Europe. Methodology.* Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

Börsch-Supan, Axel, and Morten Schuth. 2014. "Early Retirement, Mental Health and Social Networks." In *Discoveries in the Economics of Aging*, ed. David Wise, 225–50. Chicago: University of Chicago Press.

Case, Anne, and Angus Deaton. 2005. "Broken Down by Work and Sex: How Our Health Declines." In *Analyses in the Economics of Aging*, ed. David Wise, 185–212. Chicago: University of Chicago Press.

Case, Anne, and Christina Paxson. 2005. "Sex Differences in Morbidity and Mortality." *Demography* 42(2):189–214.

Cawley, John, James Heckman, and Edward Vytlacil. 2001. "Three Observations on Wages and Measured Cognitive Ability." *Labour Economics* 8(4):419–52.

Charles, Kerwin. 2004. "Is Retirement Depressing? Labor Force Inactivity and Psychological Well-Being in Later Life." *Research in Labor Economics* 23:269–99.

Coe, Norma, Hans-Martin von Gaudecker, Maarten Lindeboom, and Jürgen Maurer. 2012. "The Effect of Retirement on Cognitive Functioning." *Health Economics* 21(8):913–27.

Coe, Norma, and Gemma Zamarro. 2011. "Retirement Effects on Health in Europe." *Journal of Health Economics* 30(1): 77–86.

d'Hombres, Beatrice, Lorenzo Rocco, Marc Suhrcke, and Martin McKee. 2010. "Does Social Capital Determine Health? Evidence from Eight Transition Countries." *Health Economics* 19 (1): 56–74.

Dewey, Michael, and Martin Prince. 2005. Cognitive Function." In *Health, Ageing and Retirement in Europe: First Results from the Survey of Health, Ageing and Retirement in Europe*, ed. Axel Börsch-Supan et al., 118–125. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33 (2):251–99.

Folstein, Marshal, Susan Folstein, and Paul McHugh. 1975. "'Mini-Mental State': A Practical Method for Grading the Cognitive State of Patients for the Clinician." *Journal of Psychiatric Research* 12(3):189–98.

Glass, Thomas, Carlos Mendes de Leon, Richard Marottoli, and Lisa Berkman. 1999. "Population Based Study of Social and Productive Activities as Predictor of Survival Among Elderly Americans." *British Medical Journal* 319:478.

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

Gruber, Jonathan, and David Wise. 2004. *Social Security Programs and Retirement around the World: Micro-Estimation*. Chicago: University of Chicago Press.

Herzog, Regula, and Robert Wallace. (1997), "Measures of Cognitive Functioning in the AHEAD Study." *Journals of Gerontology: Series B* 52(Special Issue):37–48.

Johnston, David, and Wang-Sheng Lee. 2009. "Retiring to the Good Life? The Short-Term Effects of Retirement on Health." *Economics Letters* 103(1):8–11.

Jürges, Hendrik. 2007. "True Health vs. Response Styles: Exploring Cross-Country Differences in Self-Reported Health." *Health Economics* 16(2):163–78.

Kajitani, Shinya, Kei Sakata, and Colin McKenzie. 2013. "Occupation, Retirement and Cognitive Functioning." Joint Research Center for Panel Studies Discussion Paper 2012-002. Keio, Japan: Keio University.

Kleibergen, Frank, and Richard Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition", *Journal of Econometrics* 133(1):97–126.

Kroll, Lars. 2011. "Construction and Validation of a General Index for Job Demands in Occupations Based on ISCO-88 and KldB-92." *Methoden—Daten—Analysen* 5(1):63–90.

Manuelli, Rodolfo, Ananth Seshadri, and Yongseok Shin. 2012. "Lifetime Labor Supply and Human Capital Investment." Working Paper 2012-004A. St. Louis, Mo.: Federal Reserve Bank of St. Louis.

Mazzonna, Fabrizio. 2014. "The Long Lasting Effects of Education on Old Age Health: Evidence of Gender Differences." *Social Science and Medicine* 101(1):129–38.

Mazzonna, Fabrizio, and Franco Peracchi. 2012. "Ageing, Cognitive Abilities and Retirement." *European Economic Review* 56(4):691–710.

Neuman, Kevin. 2008. "Quit Your Job and Get Healthier? The Effect of Retirement on Health." *Journal of Labor Research* 29(2):177–201.

Peracchi, Franco, and Claudio Rossetti. 2012. "Heterogeneity in Health Responses and Anchoring Vignettes." *Empirical Economics* 42(2):513–38.

Ravesteijn, Bastian, Hans van Kippersluis, and Eddy van Doorslaer. 2013. "The Contribution of Occupation to Health Inequality." *Research on Economic Inequality* 21:311–32.

Rohwedder, Susan, and Robert Willis. 2010. "Mental Retirement." *Journal of Economic Perspectives* 24(1):1–20.

Santi, Irene, Lars Kroll, Andreas Dietz, Heiko Becher, and Heribert Ramroth. 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.

Stock, James, and Motohiro Yogo. 2005 "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. Donald Andrews and James Stock, 80–108. New York: Cambridge University Press.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*, second edition. Cambridge, Mass.: MIT Press, 2010.

Zantinge, Else, Mathijs van den Berg, Henriëtte Smit, and Susan Picavet. 2014. "Retirement and a Healthy Lifestyle: Opportunity or Pitfall? A Narrative Review of the Literature." *European Journal of Public Health* 24(3):433–39.