Journal of Health Economics 56 (2017) 113–125

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Full length article

Retirement and cognitive decline. A longitudinal analysis using SHARE data

Martina Celidoni, Chiara Dal Bianco, Guglielmo Weber*

Department of Economics and Management, University of Padua, Via del Santo 33, 35123 Padua, Italy

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form
29 November 2016
Accepted 6 September 2017
Available online 23 September 2017

JEL Classifications:
I12
I1
J26
J1
J24

Keywords:
Ageing
Cognition
Retirement
Instrumental variable estimation
SHARE

ABSTRACT

We show that a new measure of cognitive decline, that can be computed in longitudinal surveys where respondents perform the same recall memory tests over the years, is highly predictive of the onset of dementia. Using SHARE data, we investigate the way retirement affects cognitive decline over time controlling for age, education and other confounding factors. We find that retirement has a long-term detrimental effect on cognition for individuals who retire at the statutory eligibility age. It plays instead a protective role for those who retire on an early retirement scheme.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Population ageing in Europe and other developed countries challenges the sustainability of the health care and long-term care systems. One of the key reasons individuals require long-term care in old age is cognitive decline, leading to dementia when it interferes with independent functioning (American Psychiatric Association, 2000). According to a recent study about the United States (Hurd et al., 2013), dementia affects a large and growing number of older adults and represents a substantial financial burden for the society with estimated costs similar to those related to heart disease and cancer.

Cognitive abilities later in life have been widely investigated in epidemiology and gerontology (Dixon et al., 2004; Schaie, 1994): the related literature documents a decline of cognition at older ages with strong effects on fluid abilities such as memory when recalling specific past events (Peterson et al., 2002; Bäckman et al., 2005).[1] Individual heterogeneity in cognition levels and changes with age are likely to be associated with individuals' engagement in mentally stimulating activities (Salthouse, 2006; Maguire et al., 2000).

Particular attention has been devoted (since the seminal paper by Adam et al., 2007) to the effects of retirement on cognition since this transition marks a change in individuals' life-style. According to Rohwedder and Willis (2010), retirees are engaged in less mental exercise than workers: the latter are exposed to environments that are more cognitively challenging and stimulating compared to the non-work condition (the so-called "unengaged lifestyle hypothesis"). If they are right, the spate of recent pension reforms increasing retirement age would also reduce long-term care expenditure (Dave et al., 2008; Bonsang et al., 2012).

When assessing the role of retirement on cognition, endogeneity issues have to be taken into account. There could be a reverse

* Corresponding author.
E-mail addresses: martina.celidoni@unipd.it (M. Celidoni), chiara.dalbianco@unipd.it (C. Dal Bianco), guglielmo.weber@unipd.it (G. Weber).

[1] Episodic memory is traditionally considered an information processing system that receives and stores information about temporally dated episodes or events, retains various aspects of this information and, upon instructions, transmits specific retained information (Tulving, 1972).

http://dx.doi.org/10.1016/j.jhealeco.2017.09.003
0167-6296/© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

causal link – individuals who experienced a bad health shock retire as soon as possible (see Insler, 2014). Also, there is likely to be a selection problem (Coe and Zamarro, 2011): people self-select into retirement based on their gains from retirement – those with the most physically demanding jobs, or who enjoy their jobs the least, retire earlier to relieve themselves of the daily strain.

In this paper we use individual panel data from a host of European countries to investigate the relation between cognitive decline and years from retirement, controlling for age, education and other confounding factors. We estimate the causal effect of retirement on cognitive decline by using eligibility ages for early retirement and statutory (old-age) pension in several European countries over time as instruments for retirement and retirement-related variables (such as years from retirement). The variability across individuals in public pension eligibility reflects not only age, but also gender, country of residence, time of retirement and years of pension contributions, and this ensures that the instruments we construct are informative. The existence of two different types of pension eligibility criteria is a key feature of our data that allows us to investigate whether retirement has heterogeneous effects on cognition.

We contribute to the literature in several dimensions. First, we exploit the longitudinal dimension of SHARE data.[2] Second, we adopt a sharp measure of cognitive decline, based on a 20% drop in the number of words recalled between waves and show that it predicts well the onset of dementia in a commonly used US data set. Most importantly, we investigate and document heterogeneity in retirement effects related to the existence of early and statutory retirement ages.

In a preliminary analysis that does not address the possibility of heterogenous effects, we find that retirement status *per-se* has a negative effect on cognitive decline, but years in retirement has a positive effect, after controlling for age and education, in line with Bonsang et al. (2012). Our evidence therefore supports both the claim that retirement improves cognition in the short-run, and is detrimental for cognition in the long-run.

However, our analysis also shows that there are significant differences in estimated parameters when we focus on two distinct groups of retirees: those who retire as soon as possible (that is, at the time when they become eligible for an early retirement pension) and those who instead retire as late as possible (at the time when they qualify for an old-age pension – when in most cases retirement from the job is mandatory or employment protection ceases to operate). For the former, we find that retirement has positive short-term and no negative long-term effects on cognition; for the latter, retirement has strong negative cumulative effects. We document to what extent early and late retirees differ in terms of observable characteristics: we find that early retirees tend to be men in low-skill jobs. We also find that late retirees tend to report more often satisfaction with the salary and the freedom connected to their job.

We find that our results are robust to the definition of the dependent variable, to the way we control for age and to non-random attrition in the panel data we use. The key results of our robustness analysis are shown in the paper; more details and further analyses are instead available in the online appendix.

The paper is organized as follows. Section 2 presents the data; the validation analysis is described in Section 3 and the empirical strategy in Section 4. Section 5 comments the results. Section 6 presents robustness checks and Section 7 concludes.

2. The data

In our empirical analysis we use data drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE)[3] which collects information on health, socio-economic status and social and family networks. The SHARE target population are individuals aged fifty or over who speak the official language(s) of their country, plus their partner regardless of age. The baseline study, which took place in 2004, involved a balanced representation of the various regions in Europe, ranging from Scandinavia (Denmark and Sweden) through Central Europe (Austria, France, Germany, Switzerland, Belgium, and the Netherlands) to the Mediterranean (Spain, Italy and Greece). To this first set of 11 countries several others have been added in the following waves.

2.1. Sample selection

In our study we restrict the sample of analysis to respondents, aged 50 or over, taking part in the first wave, or those interviewed for the first time in the second wave (refreshment sample).

Among these, we keep only individuals re-interviewed both in the third wave, called SHARELIFE given its retrospective nature, and in the fourth one. Since we are interested in studying the effect of retirement on cognition, we select respondents who were working or retired from work in the baseline (i.e. the first or the second wave depending on when respondents entered the sample).[4] For the most part we pool males and females.

While Rohwedder and Willis (2010), Mazzonna and Peracchi (2012) and Bonsang et al. (2012) define an individual as retired if he or she reports not to be working, for the purpose of our analysis we define a respondent as retired if he or she declares to be retired from work and has work experience higher than or equal

[2] Almost all papers on European data use cross sections rather than the available longitudinal information. To our knowledge, the only exceptions are Bianchini and Borella (2014) and Mazzonna and Peracchi (2014), that present fixed effect estimates on SHARE data. However, fixed effect estimation controlling for age relies on transitions into retirement for identification, and is therefore unsuitable to estimate the cumulative effect of retirement in a short panel. For this reason, we prefer to control for the lagged endogenous variable, instead (Imbens and Wooldridge, 2009).

[3] This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013 or SHARE wave 1 and 2 release 2.5.0, as of May 24th 2011 or SHARELIFE release 1, as of November 24th 2010. This paper uses data from the generated Job Episodes Panel (DOI:10.6103/SHARE.jep.200), see Brugiavini et al. (2013) and Antonova et al. (2014) for methodological details. The Job Episodes Panel release 2.0.0 is based on SHARE Waves 1 and 2 (release 2.5.0, May 24th 2011) and SHARELIFE (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N211909, SHARE-LEAP, N227822 and SHARE M4, N 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) 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.

[4] The number of released interviews with the cognitive functioning module of waves 1, 2 or 4 is 85,783. We drop 33,298 individuals belonging to countries that did not participate in all four waves. We further drop 29,848 individuals who participated only to one regular wave and 1058 individuals born after 1956. We drop 2660 individuals who are neither retired nor employed/self-employed or who do not provide information about their employment status. We are left with 18919 individuals: for 13,636 of these we have information from the Job Episode panel (see Brugiavini et al., 2013; Antonova et al., 2014). We then drop individuals (1734) for whom we do not know the retirement year, who retired before 40 or after 70, or whose job experience is lower than 15 years. We thus keep 7426 individuals who participated to waves 1 and 4 and 2513 individuals who were part of the wave 2 refreshment sample and participated to wave 4. For 224 of these individuals we were not able to compute our cognitive decline measure, since they did not do the word recall test in at least two waves. Finally, we drop 603 observations due to missing information about covariates.

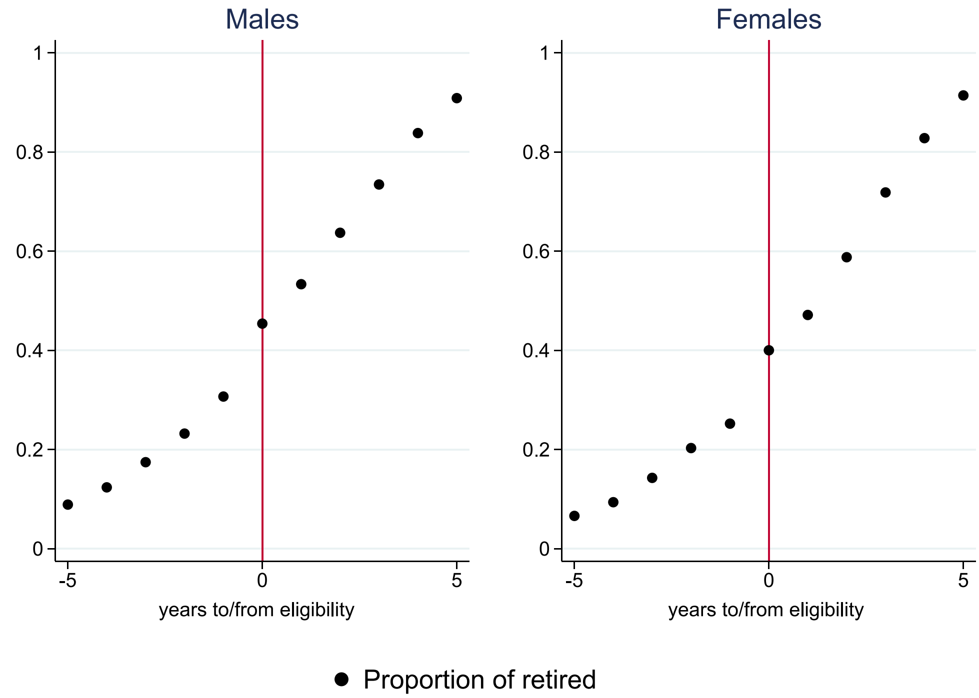

Fig. 1. Early retirement - all countries.

to 15 years. By doing this, we avoid to include unemployed or disabled individuals among the retirees so that we can strictly focus on the consequences of retirement from work on cognitive abilities, reducing the influence of long inactivity periods on cognition for other reasons.

Coe and Zamarro (2011) observe that there are individuals still working but who declare themselves as retired simply because they left their *career job*. In our sample 95% of individuals that in baseline declare to be retired did not do any work for pay in the previous four weeks. Even if we reclassify that 5% of individuals as not retired, results do not change.

Further we exclude proxy interviews because in those cases individuals do not perform cognitive tests and we do not consider interviewees with missing values in tests' scores in at least one of the two measurement occasions (i.e. baseline and wave 4). The final sample is a balanced, two-period panel for 8932 individuals (6583 observed in waves 1 and 4, 2349 in waves 2 and 4).[5]

2.2. Covariates

In addition to retirement, another key variable in our analysis is years spent in retirement. This variable is computed on the basis of the year when the SHARELIFE respondent reports to have retired from work. When this information is missing or the respondent retired between waves 3 and 4, we use a comparable question from wave 4 questionnaire. In Figs. 1 and 2 we show the proportions of individuals who self-report being retired by gender as a function of the years from/to the eligibility age separately for early and statutory retirement (in online Appendix C we describe the eligibility criteria used for each country, and show the proportions by country – see Figs. C.1–C.4). It is interesting to notice that there are sizeable increases in the proportion of retired at both eligibility rates in most countries. In Table 1 we report descriptive statistics by country for our set of retirement variables (retired, transition into retirement and years in retirement) and eligibility age for early and statutory retirement.

In our baseline specification we include education, expressed according to the International Standard Classification of Education (ISCED), the logarithm of age, a gender dummy and nine country dummies (Germany is the reference country). We include in the model a dummy, *low cognition*, that takes value one if the baseline memory score is lower than the median value by wave, country and gender, as a 20% decrease is less likely to be observed if the initial value is already low. Finally, we have also a control for the wave 2 refreshment sample (to take into account the shorter time distance from wave 4) and *less repetitions* which identifies individuals who performed the test only twice (because they enter the sample in wave 2 or because they were interviewed in wave 1 and wave 4 but not in wave 2). See Table A.1 in the online Appendix for summary statistics.

3. Outcome variable: definition and validation

Data about cognitive abilities are collected in each regular wave of SHARE[6]: a series of brief tests are included in the CAPI questionnaire. Among them, only verbal fluency and verbal learning tests are performed in all three waves.[7] It is worth noting that, since cognitive decline is a multidimensional phenomenon, each test usually measures a different aspect of cognition.

In agreement with the economics literature, we focus on memory scores (based on a modified version of the Rey's Auditory Verbal Learning Test-RAVLT). In this ten-word-list learning test the respondent is asked to learn a list of ten common words and recall them immediately (immediate recall or first trial) as well as after an interference period (delayed recall or second trial), roughly 5 minutes later. As in Rohwedder and Willis (2010) and

[5] Our final sample includes the following countries: Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Switzerland and Belgium.

[6] We refer to waves 1, 2 and 4 as regular.

[7] The verbal learning test has the same technical features in all three waves, the only exception is that words used in the fourth wave of SHARE are different from those used in the previous waves, for details see Malter and Boersch-Supan (2013).

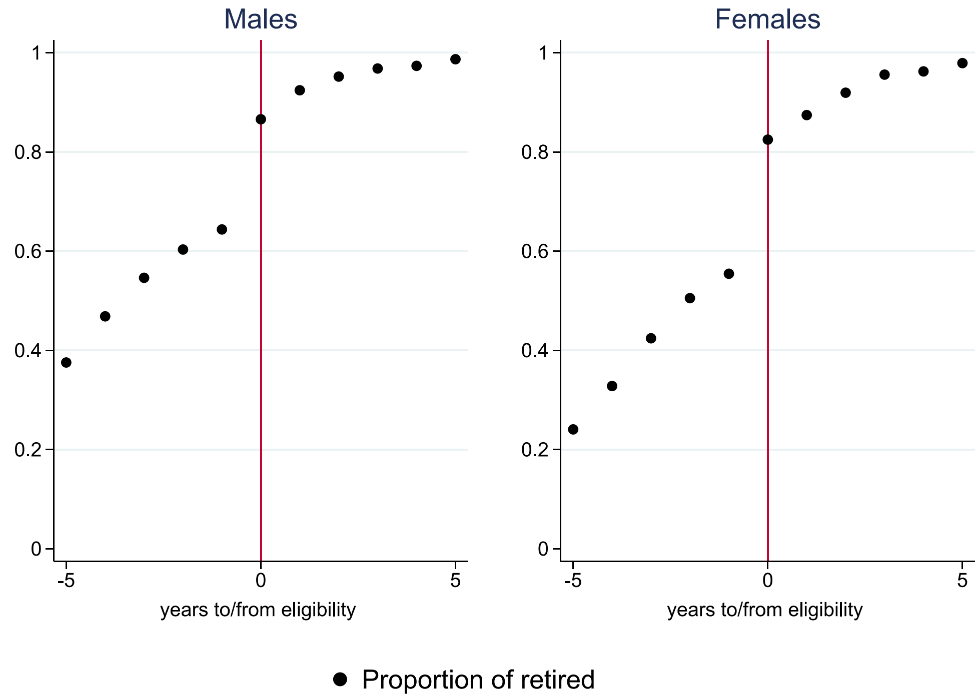

Fig. 2. Statutory retirement - all countries.

Table 1
Summary statistics by country.

	Nr. of obs #	Age Mean	Retired %	Transitions into retirement %	Years in retirement (if retired) Mean	Early ret. age Males Mean	Early ret. age Females Mean	Statutory ret. age Males Mean	Statutory ret. age Females Mean
Austria	345	63	0.73	0.17	8.3	61	55	65	60
Germany	842	64	0.60	0.17	7.6	61	61	65	65
Sweden	1064	64	0.53	0.22	6.8	61	61	65	65
Netherlands	788	62	0.49	0.21	5.9	62	62	65	65
Spain	568	65	0.64	0.14	8.3	61	61	65	65
Italy	1113	64	0.73	0.14	9.8	54	54	63	58
France	1121	63	0.60	0.19	9.4	57	57	60	60
Denmark	1168	64	0.50	0.18	6.5	60	60	66	66
Switzerland	688	64	0.47	0.17	7.5	63	62	65	63
Belgium	1235	64	0.63	0.18	8.8	60	60	65	62

Bonsang et al. (2012), we measure cognitive abilities as the sum of words remembered in the immediate and delayed recalls with a score ranging from 0 to 20. This test is preferred to verbal fluency (whereby respondents list as many animal names as possible within a short period of time) because memory is particularly affected by ageing and, in addition, it does not suffer from floor and ceiling effects (Bonsang et al., 2012). On the basis of the memory score, we shall compute our main outcome of interest, “high decrease”.

The standard negative association between cognitive abilities and age, that most papers on cognition refer to, is confirmed in Fig. 3: looking at the cross-sectional variability in memory scores in SHARE it is possible to notice how the total number of words recalled decreases almost linearly with age. This is the kind of relation that most previous studies about the effect of retirement on cognition have noted, further highlighting a drop around retirement ages. To better understand cognitive decline, however, the longitudinal information should be exploited as following the same individuals over time allows us to look at changes over time.

We base our measure of cognitive decline on the percentage change in words recalled between waves. Formally, if $score_{i,t}$ denotes the number of words recalled in both immediate and delayed test, we define $y*_{i,t}$, the percentage change in memory score, as follows:

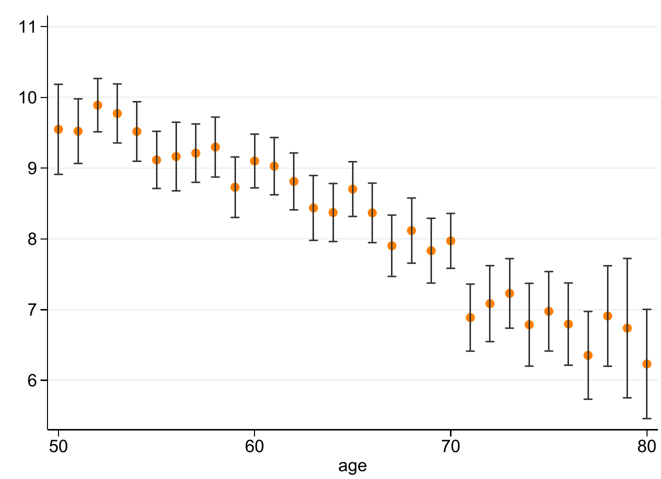

Fig. 3. Average Memory Score by age.

$$y*_{i,t} = (score_{i,t} - score_{i,t-1})/score_{i,t-1} \tag{1}$$

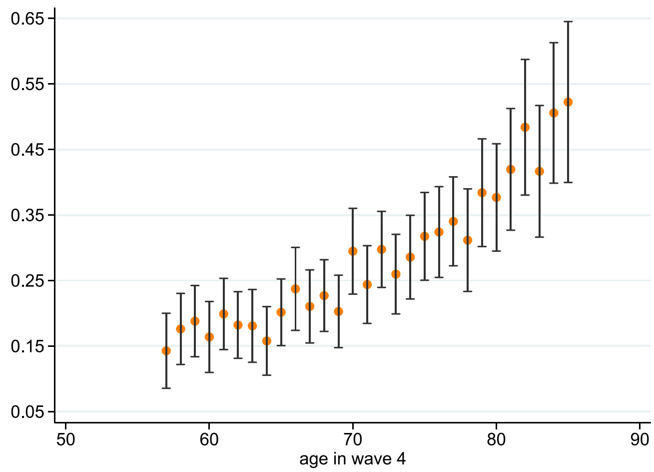

Fig. 4. Percentage of High Decrease in Cognitive Abilities, by age in wave 4.

We further define our sharper measure of high cognitive decline, $y_{i,t}$, as a dummy variable that takes value 1 if $y^*_{i,t}$ is lower than −0.2 and zero otherwise.[8] This is more appropriate than a simple measure of the change in words recalled in presence of re-testing effects, typically found in longitudinal data: respondents tend to improve their performance in memory tests across waves, particularly the second time they are interviewed. Dal Bianco et al. (2013) present descriptive evidence on this issue using SHARE data and argue that high decreases in words recalled are more informative about actual declines, as opposed to straight changes in the score.

In the literature there is no standard threshold to discriminate different levels of cognitive declines. However, as Table 2 reveals, by focusing on drops higher than 20% we select those changes that are in the bottom quarter of the distribution of memory score variations between waves. This is a fairly stable result: in Table 2 we show the distribution of percentage changes in memory scores respectively between wave 1 (2004) and wave 4 (2011) and, for the refreshment sample of wave 2, between wave 2 (2006) and wave 4 (2011). We also show the same statistics for a restricted sample where very old individuals are excluded. In all cases the 25th percentile corresponds to falls between 20% and 23%.[9] In Fig. 4 we show the percentage of high decrease by age in wave 4.

Focusing on high decreases rather than any decrease should also help in reducing measurement issues, since mild decreases might not reveal true cognitive deterioration but only measurement error. However, as we shall see in Section 6, our analysis is robust to changes in the definition of the outcome variable.

To understand if our measure of cognitive decline can be considered symptomatic of a pathological impairment related to dementia, we use data drawn from HRS where, for a sub-sample of individuals aged 70+ (ADAMS), we have both a memory test, similar to that proposed in SHARE, with immediate and delayed recall, and a clinical assessment for dementia. We are especially interested in cognitive declines between waves that we will compare with the clinical assessment of dementia provided by a nurse and a neuropsychology technician specifically trained in data collection for dementia evaluation. Exploiting the correlation between the clinical assessment of dementia and the high decline measure, we see that high decline in cognition is a good measure as in the 70% of cases it corresponds to a clinical assessment of the pathology. To validate our measure, we also provide in online Appendix B results on Receiver Operating Characteristic (ROC) curve that corroborates our chosen 20% threshold.

4. Empirical strategy

The aim of the empirical analysis is to estimate the role of years spent in retirement on cognitive decline accounting for re-testing effects, i.e. looking at the within-person drop in cognition over time as a function of the years spent in retirement, controlling for age and other covariates and taking into account the endogenous nature of retirement.

We adopt the following linear specification for $y_{i,t}$, the dummy variable that takes value 1 if the percentage change in words recalled has fallen by 20% or more across waves:

$$y_{i,t} = \beta_1 Retired_{i,t-1} + \beta_2 fromWtoR_{i,t} + \beta_3 \log(1 + yearsinR_{i,t-1}) + \mathbf{X}^{T}{}_{i,t-1}\beta_4 + \varepsilon_i \quad (2)$$

where we assume that the probability of observing a decline in cognitive abilities depends on retirement status in t − 1, $Retired_{i,t-1}$, on whether we observe the transition from work to retirement between waves, $fromWtoR_{i,t}$ and on the logarithm of years spent in retirement, $\log(1+yearsinR_{i,t-1})$. We include in the model also a vector of covariates, $\mathbf{X}^{T}{}_{i,t-1}$ as described above (that includes age, gender and education).[10]

The first retirement-related variable captures the status effect of retirement on cognitive decline. For instance, individuals who are retired from work may be more often depressed, and this may lead to faster loss of cognition over time, or instead feel relieved, and this may lead them to engage in stimulating activities. The second retirement-related variable instead captures the immediate effects of retiring from work – that may be beneficial if work had become a psychological or physical burden ("honeymoon effect"). The third and last retirement-related variable captures the progressive loss of fluid memory induced by a less engaged life style, for a given age. One might expect this loss to be zero at the beginning of the retirement period, and to build up over the years.

The advantage of the linear probability model specification is that we can easily account for the potential endogeneity of the retirement decision. For this, we need instruments that are both relevant, i.e. directly related to retirement decisions, and exogenous – that have an effect on cognition only through their impact on retirement. As by now standard in the literature (following Battistin et al., 2009), retirement decisions are instrumented by legislated ages of eligibility for early retirement and old-age pension. Differently from other studies, that adopted the same instrumental variables strategy (Rohwedder and Willis, 2010), we exploit not only the cross-country variability in eligibility ages, but also variations over time as in Angelini et al. (2009). As Mazzonna and Peracchi (2012) observe, in fact, SHARE data offer substantial within-country variability in eligibility rules arising from the pension reforms of the 1990s, which contributes to the European heterogeneity of pension entitlements. Pooling the data from all countries and all periods of retirement, we know (from Fig. 1) that the proportion of males and females who are retired is a steeply ascending function of the years to/from eligibility to an early retirement pension (with an upward

[8] We stress that we exclude proxy interviews because individuals in those cases did not perform cognitive tests; we also delete interviews with missing values in tests scores in either baseline or wave 4. This selection could affect our estimates, since there is a high probability of not observing a cognitive drop for individuals cognitively impaired that did a proxy interview or did not participate in wave 4 due to poor health conditions. Our results do not change if we include in the cognitive decline group individuals who performed the memory test in baseline but did not in wave 4.

[9] In Table 2, we notice that for some percentiles, individuals observed between wave 2 and 4 exhibit a larger drop than those observed in wave 1 and 4. This might be explained by selection (attrition) issues, that we investigate in the robustness section, or by re-testing effects.

[10] Bingley and Martinello, 2013, show that omitting education can seriously bias the estimated effect of retirement on cognition, because low-education individuals, who perform poorly in cognitive tests, tend to retire earlier.

Table 2
Percentage change in memory score between waves.

Waves	Percentiles 5	10	25	50	75	90	95
			Whole sample				
W1-W4	−0.615	−0.429	−0.214	0.000	0.300	0.714	1.000
W2-W4	−0.600	−0.444	−0.231	0.000	0.250	0.625	1.000
			Sample: Age 50–80				
W1-W4	−0.584	−0.417	−0.200	0.000	0.300	0.714	1.000
W2-W4	−0.571	−0.429	−0.222	0.000	0.273	0.625	1.000

jump at eligibility for both), and the same is true for statutory retirement (when there are more pronounced upward jumps at eligibility for both men and women) as shown in Fig. 2. This strong association between years to/from eligibility to either type of pension and years to/from retirement confirms that the standard identification strategy that instruments retirement with eligibility can be extended to years from retirement (and to the variable $\log(1 + yearsinR_{i,t-1})$ in equation (2)).

We use the following instruments for $Retired_{i,t-1}$, $fromWtoR_{i,t}$ and log $(1 + yearsinR_{i,t-1})$: two dummy variables that take value 1 if the individual is eligible for an early or statutory (old age) retirement pension, two dummies that equal 1 if we observe the transitions from not being eligible to being eligible between waves and two variables indicating the logarithm of years since eligibility for the two types of retirement. Therefore in our two-stage least squares (TSLS) specification, we have three endogenous variables and six instruments. As we show in the next section, the instruments are relevant but the over-identification tests are rejected at the 10% level (and this may suggest heterogeneity between early and statutory retirement).

Differently from Coe and Zamarro (2011) and Rohwedder and Willis (2010) that consider only the binary treatment of retirement, we also take into account the cumulative role of years spent in retirement. As argued in Bonsang et al. (2012), Coe et al. (2012) and Mazzonna and Peracchi (2012), in fact, the effect of retirement may not be instantaneous. According to Atchley (1976, 1982), individuals might experience, right after retirement, a so-called *honeymoon phase* in which they can engage in different activities that were set aside because of work-related constraints. This engagement in desired activities may attenuate the negative effects of retirement on cognition. We might also expect that changes in activities would translate only progressively into changes in cognitive abilities. If this is true, considering in the empirical model only the retirement status could provide just a partial description of changes in cognition and be uninformative for policy purposes. There could be, in fact, a cumulative effect of years spent in retirement: the longer the period of time since the individual retired the more likely that he or she experiences a high decrease in cognition for any given age. As Bonsang et al. (2012) argue, if the impact is cumulative, there could be gains in terms of lower long-term care expenditures coming from policies that increased retirement eligibility ages, since they might delay the appearance of cognitive impairment at older ages.

5. Estimation results

In this section we present our estimation results when the dependent variable is the cognitive decline measure described above.

In Table 3, we report our baseline specification results. Column (1) replicates what has been done in previous studies – it presents TSLS (or GIVE) estimates when the additional instruments are the set of six variables based upon eligibility for Early Retirement (ER) as well as Statutory Retirement (SR). In column (1) there are more additional instruments (six) than endogenous explanatory variables (three) – the model is over-identified. Column (2) presents TSLS (or IV) estimates when the additional instruments are based upon eligibility for Early Retirement, while column (3) presents TSLS (or IV) estimates when the additional instruments are based upon eligibility for Statutory Retirement. In columns (2) and (3) there are as many additional instruments (three) as endogenous explanatory variables – the models are just-identified. Estimated standard errors are robust to clustering at the country, gender and cohort level. Column (4) presents OLS estimates – that are useful for comparison with the IV estimates shown in columns (1)–(3).

Focusing on the TSLS estimates shown in column (1), we can see that the dummy indicator "being retired" is significant with negative sign, the indicator for transiting from "work to retirement" has a negative but insignificant effect while the variable (logarithm of) "1 + years spent in retirement" is significant (at the 10% level) with a positive sign. The positive sign on the key variable of interest (years in retirement) is consistent with the notion that a long period in retirement is associated to an increased probability of experiencing a large decline. Given that we control for (log) age (that has a positive, significant effect), we can infer that retiring early accelerates cognitive decline (the longer you have been retired for a given age, the earlier you must have retired). Other controls include low cognition in baseline (that has a strong negative effect), gender (women are less likely to suffer from cognitive decline),[11] education, a set of country dummies, a dummy that controls for the shorter period in the panel for those who were first interviewed in wave 2 and a dummy for individuals who have taken the memory test less than three times (either because they entered the sample in wave 2 or because they entered in wave 1 but skipped the test in wave 2).

At the bottom of Table 3 we report the Sanderson and Windmeijer (2016) version of Angrist-Pischke weak instruments F-tests, the Kleibergen-Paap Wald rank test and the Sargan-Hansen test of the over-identifying restrictions (Tables D.1–D.3 in Appendix report the first stage regressions). On the basis of the test results, we conclude that the instruments used are not weak, but the over-identification restrictions are rejected at the 10% significance level (the p-value of the test in fact is 0.054). We discuss in the sequel a possible reason for this rejection. We also report an endogeneity test, that rejects the null hypothesis that all three retirement-related variables can be treated as exogenous in estimation.

The evidence provided so far confirms previous findings from HRS and SHARE that working appears to have a protective role on cognition in the long run – the later individuals retire the less likely they are to develop cognitive decline, given age, gender and education. To arrive at this conclusion, we exploited (presumably) exogenous variability across time and space of pension eligibility rules. To clarify, we estimated the causal effect on cognition of

[11] It should be remembered that women in our analysis are those who worked or were retired from work in baseline, and this might be a positively selected sample (home makers are excluded). However, it is also possible that women are less likely to experience memory losses because of their more engaged life style (Steeves et al., 2015).

Table 3
Cognitive decline – OLS and TSLS estimates (baseline model).

	(1) TSLS ER + SR	(2) TSLS ER	(3) TSLS SR	(4) OLS
retired	−0.144***	−0.167***	−0.086**	−0.057***
	(0.030)	(0.032)	(0.040)	(0.017)
fromWtoR	−0.055	−0.180**	0.045	−0.020
	(0.042)	(0.078)	(0.055)	(0.013)
log(1+yearsinR)	0.041*	−0.031	0.104***	0.031***
	(0.025)	(0.049)	(0.035)	(0.009)
log(age)	0.817***	1.296***	0.298	0.679***
	(0.174)	(0.310)	(0.275)	(0.064)
Observations	8932	8932	8932	8932
R-squared	0.082	0.069	0.077	0.084
Adj R-squared	0.080	0.067	0.075	0.082
Sargan-Hansen (p-value)	0.054			
F-test (p-value)				
retirement variables	0.000	0.000	0.000	0.000
Angrist and Pischke, first-stage F stat (Weak identification test)				
retired	379.998	890.06	223.422	
fromWtoR	79.225	58.16	116.798	
log (1+yearsinR)	75.603	51.23	135.627	
Kleibergen-Paap Wald rk F statistic	F(4,744)	F(1,744)	F(1,744)	
	47.434	16.203	35.362	
Endogeneity test (p-value)	0.0008	0.0002	0.0057	

Notes: Standard errors are robust to clustering at the country, gender and cohort level. The Angrist and Pischke F statistics are computed as in Sanderson and Windmeijer (2016). Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

retirement for those individuals who were either induced or forced to retire by these rules. If the effect was the same across the population, we could extend this result to all retirees. If however, the effect varies across individuals, what we estimate is a weighted average effect for those who are close to the eligibility thresholds (Local Average Treatment Effect – LATE).

So far we have grouped together individuals who retire as soon as possible (as soon as they qualify for an early retirement pension) and individuals who instead retire as late as possible (when they reach normal or old age retirement pension eligibility, that is normally associated with mandatory retirement from the job). If the effects of retirement on cognition are heterogeneous, the estimates we presented therefore are a weighted average of causal effects for instrument-specific compliers where the weights depend on the strength of each instrument in the first stage (see Angrist and Pischke, 2009).[12]

We might expect retirement to have different effects for such widely different groups of individuals, and this is consistent with the rejection (albeit at the 10% level) of the over-identifying restrictions (Sargan-Hansen) test.

In columns (2) and (3) of Table 3 we separate these two groups by using one set of instruments at the time: early retirement pension eligibility on the one hand, statutory retirement age eligibility on the other.

Column (2) of Table 3 shows that for individuals who retire as soon as possible (via the ER route – ER compliers) retirement is unambiguously beneficial: as in column (1), being retired significantly reduces the probability of experiencing a high decline in cognitive abilities. Also, transiting into retirement has a negative effect, that is statistically significant. However, in this case years into retirement do not have the positive, significant effect that we have seen in column (1). At the bottom of the table we report the Sanderson and Windmeijer (2016) version of Angrist-Pischke weak instruments F-tests – that are all well above ten – and the Kleibergen-Paap Wald rank test. Given that the number of instruments equals the number of endogenous explanatory variables (just identification) we cannot compute the Sargan-Hansen test. It is worth recalling that in this just-identified case weak instruments are less of a concern, because the IV estimator is median-unbiased. We also report the p-value of the endogeneity test, that very strongly rejects the null hypothesis that all three retirement-related variables can be treated as exogenous in estimation.

Column (3) of Table 3 shows that for individuals who retire as late as possible (via the SR route – SR compliers) retirement is beneficial in the short-run, but detrimental in the long-run. The coefficient on the retirement status dummy is negative and significant, as in all previous columns. The effect on cognitive decline of transiting from work to retirement is positive, but insignificant. But the striking result is the very large, highly significant positive cumulative effect on cognitive decline (the point estimate on the years in retirement variable is 0.104 with a standard error of 0.035). The instruments in this case are much stronger compared to column (2), but the effect of age is less precisely estimated. Finally, the endogeneity test rejects the null, albeit less strongly than in columns (1) and (2).

In Table 4 we address the issue of whether the parameter estimates reported in Columns (2) and (3) are significantly different. To compute the standard errors of the difference we follow a bootstrap approach. We draw 1000 bootstrap samples stratified by country, cohort and gender. Table 4 shows that the differences between the estimated coefficients on all three retirement-related variables are statistically different from zero.

To summarize the substantive implications of the parameter estimates shown in columns (1)–(3) of Table 3 we report a graph (see Fig. 5) where we show in a concise and more effective way how the probability of cognitive decline changes over time as a function of years since retirement. The continuous line corresponds to

[12] In a simple case where there are two instruments, z_{1i} and z_{2i}, if the population first stage fitted values for TSLS are given by $D_i = \gamma_{11} z_{1i} + \gamma_{12} z_{2i}$, then the TSLS coefficient to be estimated can be expressed as follows $\rho = w\,\rho_1 + (1-w)\,\rho_2$, where ρ_1 and ρ_2 are the instrument-specific LATE using z_1 and z_2 respectively, and $w = \gamma_{11} cov(D_i, z_{1i}) / [\gamma_{11} cov(D_i, z_{1i}) + \gamma_{12} cov(D_i, z_{2i})]$.

Table 4
Bootstrap estimates – Difference between Early retirement (ER) and Statutory retirement (SR) coefficients – (baseline model).

	Coef.	Std. Err.	lower bound 95% C.I.	upper bound 95% C.I.
retired SR	−0.086	0.040	−0.161	−0.002
retired ER	−0.167	0.034	−0.230	−0.093
Difference	0.081	0.035	0.011	0.149
fromWtoR SR	0.045	0.057	−0.069	0.156
fromWtoR ER	−0.180	0.074	−0.330	−0.042
Difference	0.224	0.092	0.052	0.410
log(1 + yearsinR) SR	0.104	0.034	0.039	0.172
log(1 + yearsinR) ER	−0.031	0.047	−0.132	0.055
Difference	0.135	0.057	0.024	0.245

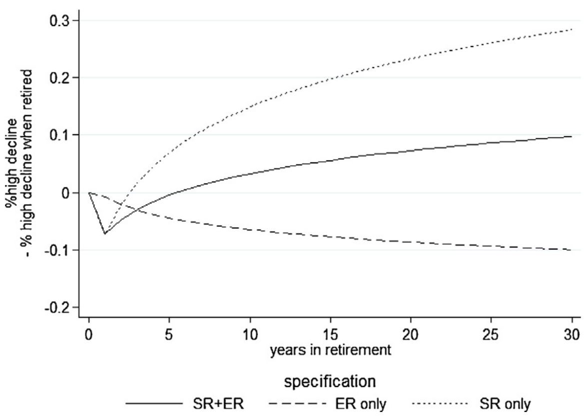

Fig. 5. Estimated effect of years in retirement on the probability of cognitive decline.

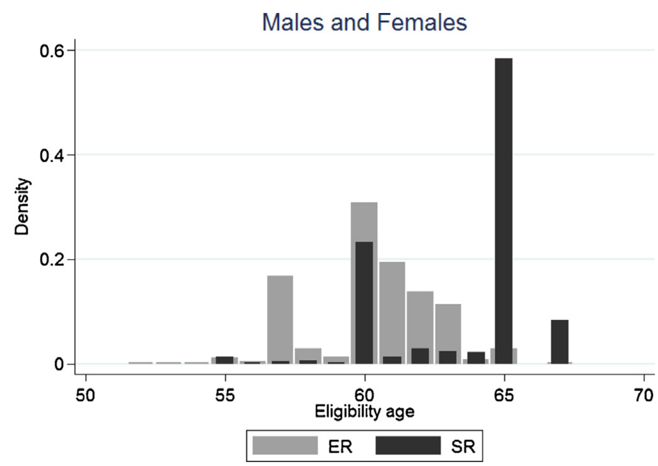

Fig. 6. Histograms of eligibility ages by type of retirement, males and females.

column (1), that pools ER and SR compliers together. The dashed line corresponds to column (2) (ER compliers) and the dotted line corresponds to column (3) (SR compliers).

Fig. 5 clearly shows that retirement is at first beneficial in all three cases, but then a very different picture obtains for ER compliers (whose risk of cognitive decline keeps falling) and for SR compliers (whose risk is instead steadily increasing). The estimates that pool ER and SR compliers together imply a slowly increasing risk – but hide the very different time pattern for the two groups.

A concern one might have in separating the two groups of compliers is that they are fundamentally different, for instance in terms of age. Of course for each respondent early retirement eligibility age is by construction lower or equal to statutory retirement age (in the 292 cases where this is not true, we set early retirement age equal to statutory retirement age). However, retirement eligibility varies across countries and over time – this ensures that there is common support in age for the instruments, as shown graphically in Fig. 6, that displays the histogram of eligibility ages by type of retirement for all respondents (similar evidence by gender is shown in the online Appendix, Figs. D.1 and D.2). We see that eligibility ages for early and statutory retirement overlap, and this is prima facie evidence of common support of the two sets of instruments.

Even though the (common support) conditions are met under which one could in principle pool the two sets of instruments, it seems likely that ER and SR compliers differ in many respects, some of which may be related to observable characteristics. One way to document these differences, allowing for the endogenous nature of retirement, is to run IV regressions of equations similar to Eq. (2) where the dependent variable is an indicator for some potentially relevant, time-invariant observable characteristics.

The observable characteristics we have considered are education (the dependent variable is a dummy taking value 1 if the respondent's education is above the country- and gender-specific median level of education), blue collar worker or low-skill worker, depression symptoms before the age of 50, gender and self-

Table 5
IV-regressions of indicators on retirement status, age, country and gender – with bootstrap standard errors.

	Coef.	Std. Err.	95% confidence interval	
			lower bound	upper bound
Low-skill job				
retired SR	−0.041	0.030	−0.095	0.019
retired ER	0.052	0.033	−0.009	0.119
Difference	−0.093	0.032	−0.153	−0.026
Female				
retired SR	0.034	0.032	−0.031	0.095
retired ER	−0.123	0.035	−0.191	−0.060
Difference	0.158	0.034	0.093	0.225
Freedom				
retired SR	0.069	0.030	0.007	0.125
retired ER	−0.008	0.034	−0.074	0.057
Difference	0.077	0.033	0.009	0.139
Salary				
retired SR	0.023	−0.007	0.084	0.036
retired ER	0.024	−0.066	0.028	−0.020
Difference	0.023	0.011	0.100	0.056

assessed work quality indicators. The explanatory variables include a retirement dummy, that is instrumented in turn by early retirement or statutory retirement eligibility, as well as age (results are robust to the functional form for age), plus country (and gender where applicable) dummies.

Table 5 shows estimation and test results – standard errors and test statistics are bootstrapped. In many cases we do not find significant differences. But we do find that individuals who either currently work in a low-skill job or used to work in a low-skill (main) job are more likely to retire using the ER route as opposed to the SR route and the same applies to men. We also regressed indicators of work quality on the same set of explanatory variables.[13] To be precise, we defined a 0–1 indicator based on whether the respondent strongly agrees with each quality statement. We find significant differences between ER and SR routes only for two domains: freedom and salary. Individuals reporting satisfaction with the freedom or salary of their current/main job are more likely to have taken the SR route.

To summarize: the ER compliers are more likely to be men in low-skill jobs who are not particularly happy with their salary and freedom on the job.

6. Robustness analysis

In this section, we present estimation results for some of the many robustness analyses that we performed. We focus here on three main issues: definition of the dependent variable; specification of age effects and effects of non-random attrition.[14]

6.1. Definition of the dependent variable

In Section 5 we have shown parameter estimates for the specification (Eq. (2)) where the dependent variable is a dichotomous outcome variable that takes value 1 if the total number of words recalled has fallen by more than 20% across waves. In Section 3 we motivated our choice.

To check if our results are robust to changes in the dependent variable we changed the threshold in our "high decline variable" from 20% to 10% in one case, and to 30% in another case. We also directly used as dependent variable the change in words recalled ("raw score") and found slightly less precise estimates, but very much in line.

Table 6 reports estimation and test results – column (1) corresponds to our baseline definition (20% drop), already shown in Table 3. The remaining columns correspond instead to different definitions of the dependent variable: column (2) reports results for a 10% drop, column (3) for a 30% drop and column (4) for (minus) the change in the raw score ($-\Delta$ raw score).

The top panel in Table 6 shows key parameter estimates and test statistics for the case where we use both sets of instruments (ER and SR). Parameter estimates have the same signs across all columns, even though significance is not always retained. It is worth stressing that the key Sargan-Hansen over-identifying restrictions test has p-values lower than 5% for the two alternative "drop" measures (10% and 30%). Its p-value is close to 20%, instead, when we use (minus) the change in words recalled.

The middle panel (based on ER instruments) shows the same pattern of signs and significance levels for all key parameters as in column (2) of Table 3.

The bottom panel (based on SR instruments) shows that the parameter of interest (the coefficient on log years in retirement) always has a positive value and is significant at the 5% (or even 1%) level in the first three columns, but only at the 10% in the final column. In the second and fourth column the estimated coefficient on the retirement dummy is not significant, even though the sign is the same across columns.

[13] For a discussion of these variables see Siegrist (1996), Siegrist et al. (2006), and Dal Bianco et al. (2015).

[14] We investigated the role of depression as a pathway to cognitive impairment – depression following retirement may be responsible for cognitive decline in later years (DeGrip et al., 2012). To understand whether this is driving our results, we first included among controls indicators for at least four among EURO-D casenesses in baseline and changes between waves, and second we excluded individuals who have ever been depressed from our estimation sample. In both cases our baseline results hold (see Table D.4).

We conclude from this that our results are robust to our choice of dependent variable.

6.2. Specification of age effects

In this paper, we investigate how years in retirement affect cognitive decline. However, answering this question poses a serious challenge, as we have to disentangle the effect of years since retirement from the effect of age.

Following Bonsang et al. (2012), our specification includes the logarithm of years spent in retirement (plus 1). The effects of age are controlled for by adding the logarithm of age as an additional regressor. To show that our results are really robust to the specification of the age trend, we shall now consider what happens when we add second, third and fourth order polynomials of age.

In Table 7 we address this issue. The top panel presents estimation and test results for the case where we use both ER and SR instruments. Thus column (1) in Table 7 reproduces column (1) of Table 3. Columns (2)–(5) show what happens when an age polynomial of the first to the fourth order is introduced. We see that age is significant in column (2), and that the age terms are jointly significant in columns (3) to (5) (an F-test is reported at the bottom of the table). In columns (2) and (3) the point estimates on $\log(1 + yearsinR)$ are significant (at the 10% level) and of comparable size to column (1); also the coefficients on the other retirement variables are quite close. Point estimates are less stable in columns (4) and columns (5), but in all specifications the p-value of the Sargan-Hansen test is in the 5%–15% range. We therefore conclude that our key result is not driven by the way we control for age.

The middle panel of Table 7 reports estimation and test results when we use the ER set of instruments. The age effects are always significant – the point estimates and the significance levels of the parameters on the retirement variables are quite similar in all five specifications.

The bottom panel of Table 7 presents estimation and test results for the case where we use SR instruments. In this case age effects tend to be insignificant – except in the case of a third order age polynomial. The negative and significant coefficient on the retirement dummy is observed in the first two columns – the coefficient becomes insignificant and very small when we control for higher order polynomials in age. However, the positive and significant sign on the key variable of interest (the logarithm of years in retirement) is confirmed in all specifications, even though the coefficient is smaller in the last two columns (and its t-ratio falls somewhat below the 5% threshold).

We can thus conclude that our key results are robust to much richer specifications of the age effects.[15]

6.3. Non-random attrition

Last, but by no means least, we address the point of non-random attrition and selection. Mazzonna and Peracchi (2012) and Zamarro et al. (2008) notice that panel attrition might be a problem because people in poor health and with low cognitive abilities are more likely to exit the panel, and this may lead to invalid inference. In the remainder of this section and in the online Appendix we show that this is not an issue in our case.

If there is non-random attrition, we would expect differences in memory score at baseline between those staying in the panel and those dropping out. In Table E.1 we show that this is true only (or mostly) for wave 1 respondents (see online Appendix for a detailed

[15] In a separate exercise, we show that estimation and test results are also robust to a flexible specification of the effect of years in retirement (see Table D.5 of the online Appendix).

Table 6
Cognitive decline – Robustness analysis – Different outcome variables.

	(1) Drop 20%	(2) Drop 10%	(3) Drop 30%	(4) -Δ(raw score)
		ER + SR		
retired	−0.144***	−0.095***	−0.137***	−0.619***
	(0.030)	(0.034)	(0.027)	(0.218)
fromWtoR	−0.055	−0.031	−0.093***	−0.547*
	(0.042)	(0.047)	(0.036)	(0.317)
log(1 + yearsinR)	0.041*	0.056**	0.020	0.063
	(0.025)	(0.027)	(0.021)	(0.173)
log(age)	0.817***	0.600***	0.776***	7.032***
	(0.174)	(0.199)	(0.148)	(1.266)
Sargan-Hansen (p-value)	0.0537	0.0157	0.0255	0.196
F-test (p-value) retirement variables	0.000	0.000	0.000	0.005
Kleibergen-Paap Wald rk F statistic	47.434	47.434	47.434	47.434
		ER		
retired	−0.167***	−0.119***	−0.160***	−0.760***
	(0.032)	(0.037)	(0.028)	(0.239)
fromWtoR	−0.180**	−0.195**	−0.171***	−1.458**
	(0.078)	(0.088)	(0.060)	(0.610)
log(1 + yearsinR)	−0.031	−0.045	−0.018	−0.473
	(0.049)	(0.054)	(0.038)	(0.378)
log(age)	1.296***	1.255***	1.050***	10.491***
	(0.310)	(0.348)	(0.246)	(2.404)
F-test (p-value) retirement variables	0.000	0.001	0.000	0.000
		SR		
Retired	−0.086**	−0.024	−0.088***	−0.282
	(0.040)	(0.048)	(0.033)	(0.306)
fromWtoR	0.045	0.091	−0.017	0.105
	(0.055)	(0.064)	(0.047)	(0.448)
log(1 + yearsinR)	0.104***	0.139***	0.065**	0.448*
	(0.035)	(0.040)	(0.029)	(0.268)
log(age)	0.298	−0.077	0.389*	3.913*
	(0.275)	(0.323)	(0.225)	(2.138)
F-test (p-value) retirement variables	0.000	0.000	0.000	0.013

Notes: Standard errors are robust to clustering at the country, gender and cohort level. The Angrist and Pischke F statistics are computed as in Sanderson and Windmeijer (2016). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

description). Similar conclusions can be drawn when looking at the Kolmogorov-Smirnov test for equality of memory score distributions between individuals observed and not observed in wave 4 (Table E.2).

Given the evidence of sample selection affecting (wave 1) respondents in our longitudinal sample, we next show (in Table E.3) the results of estimating the model on the full sample of wave 1 and wave 2 baseline respondents correcting for endogenous selectivity by the standard two-step Heckman procedure (see Wooldrige, 2002). We exploit information from a final section of the questionnaire that should be completed by the interviewer, known as IV section, providing information on how the interview proceeded and on the general surroundings (living area and type of building). In particular, we use information on the average time of completion of the IV section to generate additional exogenous variables that determine selection. This set of variables includes the length of the IV section[16] (*length_iv_m*), its square and a dummy variable indicating whether this information is missing.[17] The average time that interviewers need to complete the IV section should capture interviewers' characteristics such as the extra burden in terms of interview length or characteristics which are unobserved but might play a key role in gaining survey cooperation (Korbmacher and Schroeder, 2013; Krosnick, 1991). It is reasonable to assume that interviewer specific information, captured by the IV module, does not affect respondents' cognitive impairment. As De Luca and Peracchi (2012) observe, these variables are external to the individuals under investigation and are not under their control – they are therefore expected to be irrelevant for the performance in the memory tests.

We include among covariates of the selection equation demographic controls (age, gender, education), country dummies, a wave 2 indicator, the set of instruments we use for retirement and the length of IV section variables described above that we will exclude from the cognitive decline analysis. Focusing on column (1) of Table E.3, we can see that the IV module length variables are highly significant.

In column (2) of Table E.3, we report TSLS estimates for the effect of retirement on cognitive decline where the Inverse of the Mills Ratio is included – this is the same specification as column (1) of Table 3, except for the (insignificant) variable "less repetitions" that is not defined for those dropping out of the panel. We report in column (3), to enhance comparability, TSLS estimates for the same specification, where we do not control for selection. The Mills Ratio coefficient is insignificantly different from zero and parameter estimates in columns (2) and (3) are very similar: being retired has a strong, negative effect; transiting from work to retirement has

[16] The variable is defined at the interviewer level (average time of completion in seconds).

[17] This information can be computed from the so-called keystroke files. In those text files, every time a key is pressed on the keyboard of the laptop, this is registered and stored by the software. Missing information is due to missing interviewer id or too few completed interviews (less than five) that would not provide a meaningful value for the interviewer time of completion.

Table 7
Cognitive decline – TSLS estimates – Different age specifications.

	(1) Log (Age)	(2)	(3) Age polynomial:	(4)	(5)
		Degree: 1	Degree: 2	Degree: 3	Degree: 4
	ER + SR				
retired	−0.144***	−0.104***	−0.175***	−0.116*	−0.108
	(0.030)	(0.027)	(0.060)	(0.065)	(0.070)
fromWtoR	−0.055	−0.027	−0.072	−0.008	−0.007
	(0.042)	(0.037)	(0.051)	(0.061)	(0.061)
log(1+yearsinR)	0.041*	0.043*	0.041*	0.023	0.025
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
log(age)	0.817***				
	(0.174)				
age		0.011***	0.037**	−0.194*	0.070
		(0.002)	(0.018)	(0.102)	(0.696)
$age^2/10$			−0.002	0.032**	−0.027
			(0.001)	(0.015)	(0.155)
$age^3/100$				−0.002**	0.004
				(0.001)	(0.015)
$age^4/1000$					−0.000
					(0.001)
Sargan-Hansen (p-value)	0.054	0.0536	0.059	0.154	0.110
F-test (p-value)					
retirement variables	0.000	0.000218	0.011	0.102	0.338
age			0.000	0.000	0.000
Angrist and Pischke, first-stage F stat (Weak identification test)	F(4,744)	F(4,744)	F(4,744)	F(4,744)	F(4,744)
retired	379.998	683.4	102.780	67.044	78.33
fromWtoR	79.225	96.85	59.320	46.151	44.15
log (1+yearsinR)	75.603	87.66	78.710	83.254	82.26
Kleibergen-Paap Wald rk F statistic	F(4,744)	F(4,744)	F(4,744)	F(4,744)	F(4,744)
	47.434	52.98	34.250	28.659	28.84
	ER				
retired	−0.167***	−0.105***	−0.223***	−0.175**	−0.156*
	(0.032)	(0.029)	(0.071)	(0.086)	(0.090)
fromWtoR	−0.180**	−0.124*	−0.219**	−0.160*	−0.186*
	(0.078)	(0.066)	(0.085)	(0.097)	(0.103)
log(1+yearsinR)	−0.031	−0.021	−0.036	−0.043	−0.048
	(0.049)	(0.047)	(0.050)	(0.049)	(0.052)
log(age)	1.296***				
	(0.310)				
age		0.017***	0.061***	−0.115	0.868
		(0.004)	(0.022)	(0.124)	(0.836)
$age^2/10$			−0.003**	0.023	−0.195
			(0.001)	(0.017)	(0.186)
$age^3/100$				−0.001	0.020
				(0.001)	(0.018)
$age^4/1000$					−0.001
					(0.001)
F-test (p-value)					
retirement variables	0.000	0.000	0.009	0.068	0.165
age			0.000	0.000	0.000
Angrist and Pischke, first-stage F stat (Weak identification test)					
retired	890.06	374.703	210.328	130.626	143.782
fromWtoR	58.16	77.112	47.217	39.132	33.704
log (1 + yearsinR)	51.23	62.927	54.591	59.594	47.486
Kleibergen-Paap Wald rk F statistic	F(1,744)	F(1,744)	F(1,744)	F(1,744)	F(1,744)
	16.203	19.222	14.659	12.065	11.177
	SR				
retired	−0.086**	−0.074**	−0.048	−0.023	0.001
	(0.040)	(0.031)	(0.081)	(0.082)	(0.098)
fromWtoR	0.045	0.049	0.063	0.097	0.100
	(0.055)	(0.048)	(0.070)	(0.074)	(0.074)
log(1 + yearsinR)	0.104***	0.099***	0.100***	0.060*	0.064*
	(0.035)	(0.034)	(0.035)	(0.035)	(0.035)
log(age)	0.298				
	(0.275)				
age		0.005	−0.004	−0.257**	0.224
		(0.004)	(0.024)	(0.114)	(0.854)

Table 7 (*Continued*)

	(1) Log (Age)	(2) Age polynomial:	(3)	(4)	(5)
		Degree: 1	Degree: 2	Degree: 3	Degree: 4
$age^2/10$			0.001	0.039**	−0.069
			(0.002)	(0.017)	(0.190)
$age^3/100$				−0.002**	0.009
				(0.001)	(0.018)
$age^4/1000$					−0.000
					(0.001)
F-test (p-value)					
retirement variables	0.000	0.000	0.010	0.100	0.190
age			0.299	0.028	0.092
Angrist and Pischke, first-stage F stat (Weak identification test)					
retired	223.422	718.98	82.884	77.789	76.033
fromWtoR	116.798	153.73	71.623	78.733	73.784
log (1 + yearsinR)	135.627	183.31	140.433	175.768	136.015
Kleibergen-Paap Wald rk F statistic	F(1,744)	F(1,744)	F(1,744)	F(1,744)	F(1,744)
	35.362	48.104	20.446	22.502	20.163

Notes: Standard errors are robust to clustering at the country, gender and cohort level. The Angrist and Pischke F statistics are computed as in Sanderson and Windmeijer (2016). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

an insignificant, negative effect; the logarithm of years in retirement has a positive and (10%) statistically significant effect on the probability of experiencing a large decline in cognition.

In Table E.4 and E.5 we report estimates controlling for sample selection when using separately early and statutory retirement instruments respectively. In neither case is the coefficient on the Inverse Mill's Ratio significantly different from zero. Parameter estimates of interest also in this case are very much in line with estimates shown in the previous section (Table 3 columns (2) and (3)).

7. Conclusions

In this paper we have used a new measure of cognitive decline, based on a 20% drop in words recalled between waves, that can be computed in standard surveys where recall memory tests are administered to the same individuals over the years and is highly predictive of the onset of dementia. Using a small sample of individuals aged 70 or more who took part in the US Aging, Demographics and Memory Study and were later medically assessed for dementia, we showed that the 20% decline measure correctly classifies 70% of individuals according to their later dementia status.

Using SHARE data and a plausible identification strategy that exploits gender, country and time variability in pension eligibility to instrument retirement, we have investigated the causal effect of retirement on cognitive decline and confirmed previous findings on the protective role of work on cognition. We have further explored whether for individuals who retire as soon as possible (early retirees) and individuals who retire as late as possible (statutory retirees) the effect of retirement on cognition differs. Consistently with the rejection of the over-identification test, we indeed find heterogeneous effects for the two groups. For early retirees retirement has beneficial effects on cognition, whereas for late retirees it has a detrimental effect in the long-run (that gets worse over time).

We have documented to what extent early and late retirees differ in terms of observable characteristics: we find that early retirees are more likely to be men in low-skill jobs and late retirees tend to report more often satisfaction with the salary and the freedom connected to their job.

While these differences are significant and help explain the different retirement age chosen by or imposed to older European workers, further research is needed to characterize the mechanisms driving the heterogeneity we find in the link between retirement and cognitive decline.

Acknowledgement

This paper was presented at the EUGMS 2013 conference, the 4th SHARE user conference (Liège, November 2013), the Netspar International Workshop (Venice, June 2014), VIU Summer Institute (Venice, June 2015), the 2015 NBER SI on Aging and the SHARE wave 7 meeting (Vilnius, March 2016). We are grateful for comments and suggestions made by participants, as well as by two referees and the editor, and for in-depth discussions with Marco Bertoni. This research was financed by the UE, SHARE-M4 (N261982) , the NIA (R01 AG052527) and by the University of Padua (POPA_EHR). An earlier version of this paper was circulated under the title: "Early retirement and cognitive decline. A longitudinal analysis using SHARE data".

Online Appendices A-E Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jhealeco.2017.09.003.

References

Adam, S., Bonsang, E., Germain, S., Perelman, S., 2007. Retirement and Cognitive Reserve: a Stochastic Frontier Approach Applied to Survey Data, Tech. Rep., HEC-ULg, CREPP Working Papers 2007/04.
American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington DC http://www.psychiatry.org/practice/dsm.
Angelini, V., Brugiavini, A., Weber, G., 2009. Ageing and unused capacity in Europe: is there an early retirement trap? Econ. Policy 24, 463–508.
Angrist, J., Pischke, J., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
Antonova, L., Aranda, L., Pasini, G., Trevisan, E., 2014. Migration, Family History and Pension: The Second Release of the SHARE Job Episodes Panel. SHARE Working Paper 18–2014.
Atchley, R.C., 1976. The Sociology of Retirement. Halsted Press.
Atchley, R.C., 1982. Retirement as a sociological institution. Ann. Rev. Sociol. 8, 263–287.
Bäckman, L., Jones, S., Berger, A.K., Laukka, E.J., Small, B.J., 2005. Cognitive impairment in preclinical Alzeimer's disease: a meta-analysis. Neuropsychology 19, 520–531.
Battistin, E., Brugiavini, A., Rettore, E., Weber, G., 2009. The retirement consumption puzzle: evidence from a regression discontinuity approach. Am. Econ. Rev. 99, 2209–2226.

Bianchini, L., Borella, M., 2014. Cognitive Functioning and Retirement in Europe CeRP Working Papers 139.
Bingley, P., Martinello, A., 2013. Mental retirement and schooling. Eur. Econ. Rev. 63, 292–298.
Bonsang, E., Adam, S., Perelman, S., 2012. Does retirement affect cognitive functioning? J. Health Econ. 31, 490–501.
Brugiavini, A., Cavapozzi, D., Pasini, G., Trevisan, E., 2013. Working Life Histories from SHARELIFE: a Retrospective Panel SHARE Working Paper 11–2013.
Coe, N.B., Zamarro, G., 2011. Retirement effects on health in Europe. J. Health Econ. 30, 77–86.
Coe, N.B., von Gaudecker, H.M., Lindeboom, M., Maurer, J., 2012. The effect of retirement on cognitive functioning. Health Econ. 21 (August (8)), 913–927.
Dal Bianco, C., Garrouste, C., Paccagnella, O., 2013. Early-life circumstances and cognitive functioning dynamics in later life. In: Boersch-Supan, H.L.A., Brandt, M., Weber, G. (Eds.), Active Ageing and Solidarity Between Generations in Europe: First Results from SHARE After the Economic Crisis. De Gruyter, Berlin, pp. 209–224.
Dal Bianco, C., Trevisan, E., Webe, r.G., 2015. 'I want to break free'. The role of working conditions on retirement expectations and decisions. Eur. J. Ageing 12, 17–28, http://dx.doi.org/10.1007/s10433-014-0326-8.
Dave, D., Rashad, I., Spasojevic, J., 2008. The effects of retirement on physical and mental health outcomes. South. Econ. J. 75, 497–523.
De Luca, G., Peracchi, F., 2012. Estimating Engel curves under unit and item nonresponse. J. Appl. Economet. 27, 1076–1099.
DeGrip, A., Lindeboom, M., Montizaan, R.M., 2012. Shattered dreams: the effects of changing the pension system late in the game. Econ. J. 122, 1–25.
Dixon, R., Backman, L., Nilsson, L.G., 2004. New Frontiers in Cognitive Aging. Oxford University Press.
Hurd, M.D., Martorell, P., Delavande, A., Mullen, K.J., Langa, K.M., 2013. Monetary costs of dementia in the United States. N. Engl. J. Med. 368, 1326–1334.
Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47, 5–86.
Insler, M., 2014. The health consequences of retirement. J. Hum. Resour. 49 (1), 195–233, University of Wisconsin press.
Korbmacher, J., Schroeder, M., 2013. Consent when linking survey data with administrative records: the role of the interviewer. Surv. Res. Methods 7, 115–131.
Krosnick, J., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl. Cognit. Psychol. 5, 213–236.
Maguire, E., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J., Frith, C.D., 2000. Navigation-related structural change in the hippocampi of taxi drivers, Tech rep. Proceedings of the National Academy of Sciences of the United States of America 97, 4398–4403.
Malter, F., Boersch-Supan, A., 2013. Share Wave 4: Innovations & Methodology Tech. Rep. MEA, Max Planck Institute for Social Law and Social Policy, Munich.
Mazzonna, F., Peracchi, F., 2012. Ageing, cognitive abilities and retirement. Eur. Econ. Rev. 56, 691–710.
Mazzonna, F., Peracchi, F., 2014. Unhealthy Retirement? EIEF Working Paper 09/14.
Peterson, M., Kramer, A., Colcombe, A., 2002. Contextual guidance of attention in younger and older adults, Tech. rep. In: 2002 Cognitive Aging Conference, Atlanta, GA.
Rohwedder, S., Willis, R.J., 2010. Mental retirement. J. Econ. Perspect. 24, 119–138.
Salthouse, T.A., 2006. Mental exercise and mental aging. Perspect. Psychol. Sci. 1, 68–87.
Sanderson, E., Windmeijer, F., 2016. A weak instrument F-test in linear IV models with multiple endogenous variables. J. Economet. 190, 212–221.
Schaie, T.A., 1994. The course of adult intellectual development. Am. Psychol. 49, 304–313.
Siegrist, J., Wahrendorf, M., von dem Knesebeck, O., Jürges, H., Börsch-Supan, A., 2006. Quality of work, well-being, and intended early retirement of older employees – baseline results from the SHARE study. Eur. J. Public Health 17, 62–68.
Siegrist, J., 1996. Adverse health effects of high-effort/low-reward conditions. J. Occup. Health Psychol. 1, 27–41.
Steeves, J.A., Murphy, R.A., Zipunnikov, V., Strath, S.J., Harris, T.B., 2015. Women workers and women at home are equally inactive: NHANES 2003–2006. Med. Sci. Sports Exerc. 47, 1635–1642.
Tulving, E., 1972. In: Tulving, E., Donaldson, W. (Eds.), Episodic and Semantic Memory, in Organization of Memory. Academic Press, New York.
Wooldrige, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge MA.
Zamarro, G., Meijer, E., Fernandes, M., et al., 2008. Mental health and cognitive ability. In: Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A. (Eds.), First Results from the Survey of Health, Ageing and Retirement in Europe (2004–2007). Starting the Longitudinal Dimension. Mannheim Research Institute for the Economics of Aging (MEA), Mannheim, pp. 2004–2007.