A Unified Framework for Examining the Effect of Retirement on Cognitive Performance *

Preliminary Draft

János K. Divényi †
Central European University

September 26, 2014

Abstract

All papers are trying to sell their strategy as the only best. I choose another way to act: instead of throwing out all previous work I put the pieces together to see the broader picture. I draw the attention to potential biases and assess their magnitude. Then, I build a new approach on the lessons learnt from the studies and utilizing the panel structure of the Survey of Health, Ageing and Retirement in Europe (SHARE) I show that if retirement has any adverse effect on cognitive performance it should be really small in magnitude.

*This paper is improved upon my Master thesis at Central European University. I thank Gábor Kézdi for advice, Zsombor Cseres-Gergely, Gábor Rappai and participants of the Hungarian Economics Association/Pécs University of Sciences, Faculty of Business and Economics' 2013 summer workshop for doctoral students and of the Venice International University Summer Institute on Ageing for helpful discussion and comments. The analysis was carried out in Stata with the help of the *ivreg2* package of Baum et al. (2014) and the tables were produced by the *estout* package of (Jann, 2005, 2007). The corresponding codes and the actual version of the paper can be found on GitHub. First draft: May 24, 2012.

†divenyi_janos@ceu-budapest.edu

1 Introduction[1]

In developed countries, increased life expectancy, together with the parallel decline in average retirement age, has increased the average spell of retirement in the last decades (e.g. from 10.5 years in 1970 to 19.8 years in 2007 for Germany[2]). Even if eligibility ages have been raised recently, people often spend 15-20 years of their lives as pensioners, which makes this phase of their life more and more relevant. Beside the individual level, the period of retirement is also of growing importance at the social level as well, because the proportion of retirees is increasing in the aging population. As a natural consequence, various fields of research began to deal with the quality of life of retirees. In this agenda, a particular aspect – namely the cognitive performance of old age individuals – has captured the attention of economists as it highly influences the decisions they make forming their consumption or saving behavior which affects the work of the economy to an increasing extent. Therefore, the age profile of cognitive abilities at the later stages of life is fundamental for many fields from marketing to pension and health policy.

It has been widely documented that individual cognitive performance tends to decline in older ages. According to Schaie (1989) cognitive abilities are relatively stable until the age of 50 but begin to decline afterwards. decline in older ages. However, there is large heterogeneity in the progress of cognitive decay, raising the natural question of what are the driving forces behind the decay and whether there is a way to decelerate it in order to maintain cognitive abilities as long as possible. A popular hypothesis, which is often called as use-it-or- lose-it hypothesis (see for example Rohwedder and Willis, 2010), suggests that the natural decay of cognitive abilities in older ages can be mitigated by intellectually engaging activities. Thus, retirement which goes together with the ceasing of cognitively demanding tasks at work, might accelerate the natural declining process, having a causal effect on cognition. In this respect, the notion of retirement simply refers to not working, and thus incorporates a broader definition than usual (for example, people on disability benefit or who are unemployed could also be regarded as retirees).

Many papers have been investigating recently whether retirement has a causal effect on cognitive abilities in developed countries, yet the results they have delivered are ambiguous. The inconclusive results are most likely due to the difficulty of identification and the resulting variety in the identification strategies.

The main problem is the endogeneity of retirement: a simple comparison of cognitive abilities of retirees and employees is likely to lead to biased estimates, as retirees and employees could hardly be considered as randomly assigned to their groups. One can conveniently argue that the decay of cognitive abilities may induce the individual to retire, that is there is reverse causality

[1] I may need to reconsider this...
[2] Allianz Demographic Pulse, March 2011

going from cognition to retirement, that may result in overestimating the retirement effect on cognition in a simple comparison, even if we control for age. The other part of the confusion might come from the difference in terms for which the effect is identified (the short run effect is likely to differ from the long run one).

In this paper I apply a novel identification strategy which aims to handle the problems which the current literature suffer from. I make use of the first, second and the fourth waves of the Survey of Health, Ageing and Retirement in Europe (SHARE), [3] which collects rich multidisciplinary data about the socio-economic status, health (including cognitive functioning), and other relevant characteristics (like social networks) of people aged 50 or over across 10 developed European countries. I identify the yearly effect of retirement on a long run panel by applying a difference-in-differences approach. Besides accounting for time- invariant individual heterogeneity and controlling for past labor market status, I also handle possible endogeneity by using public policy rules as instrumental variables. To my knowledge, this paper is the first which makes use of a large longitudinal cross-country sample to go after this effect. Contrary to previous findings, my results suggest that retirement does not seem to cause harm for cognition.

2 Model

A general way to model the relationship between cognitive performance and retirement is the following:

$$CP_i = \alpha + f(YR_i; \boldsymbol{\beta}) + u_i \tag{1}$$

$$u_i = \boldsymbol{X}_i^{'}\boldsymbol{\gamma} + \varepsilon_i \tag{2}$$

where CP_i denotes the cognitive performance of individual i, YR_i is the number of years the individual has spent in retirement (i.e. not working). I allow for the cognitive performance to depend upon these years through an arbitrary function f where $\boldsymbol{\beta}$ expresses the marginal effect of one more year spent in retirement. The term u_i contains all factors associated with CP_i except for

[3]This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013(DOI: 10.6103/SHARE.w4.111) or SHARE wave 1 and 2 release 2.6.0, as of November 29 2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1, as of November 24th 2010 (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, No 211909, SHARE-LEAP, No 227822 and SHARE M4, No 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).

YR_i such as age. Equation (2) makes this dependency explicit where $\boldsymbol{X}_i$ is the vector containing these factors.

Clearly, $\mathrm{E}[CP_i|YR_i] = \alpha + f(YR_i; \boldsymbol{\beta}) + \mathrm{E}[u_i|YR_i]$. Assuming that we know f and have a good measure for CP_i the parameter of interest ($\boldsymbol{\beta}$) can be consistently estimated if $\mathrm{E}[u_i|YR_i] = 0$. However, this is hardly the case. There are two sources which make the exogeneity assumption dubious: (1) omitted variable bias and (2) reverse causality.

Omitted variable bias. There are lots of factors which are associated with the cognitive performance and also the years spent in retirement. These are factors in $\boldsymbol{X}_i$ which are correlated with YR_i. The most obvious candidate is age: older individuals are expected to have spent more years in retirement and they also have worse cognitive skills due to age-related decline. Education is also incorporated in $\boldsymbol{X}_i$: worse educated individuals retire earlier and they also have worse cognition. One should take care of these factors when estimating the effect of retirement on cognitive performance. The main challenge here is that we do not know exactly what factors are in $\boldsymbol{X}_i$.

Reverse causality. One can conveniently argue that the decay of cognitive abilities may induce the individual to retire, that is there is reverse causality going from cognition to retirement. That may result in overestimating the retirement effect on cognition in a simple comparison, even if we control for all factors in $\boldsymbol{X}_i$.

Most attempts trying to uncover $\boldsymbol{\beta}$ apply instrumental variables, as they might be able to eliminate both of the problems. Good instrumental variables (let us denote them by the vector $\boldsymbol{Z}_i$) satisfy two requirements: first, they are correlated with the possibly endogenous variable ($\mathrm{Cov}(\boldsymbol{Z}_i, YR_i) \neq 0$), and second, they are related to the cognitive performance only through years of retirement ($\mathrm{E}[u_i|\boldsymbol{Z}_i] = 0$). If these two assumptions hold, both omitted variable bias and reverse causality is overcome.

3 Data

Most papers which are after the effect of interest use the same sources of data provided by three large longitudinal surveys: the Health and Retirement Survey (HRS), the English Longitudinal Survey of Ageing (ELSA) and the Survey of Health, Ageing and Retirement (SHARE).

Aiming to provide a multidisciplinary data about ageing, the United States of America launched the Health and Retirement Survey (HRS) in 1992, and since then the study has collected detailed information about socio-economic status, health (including cognitive functioning), and other relevant characteristics (like social networks) of people aged 50 or over. Respondents of the survey are visited biannually and put through in-depth interviews to collect rich panel micro data about aging population. The English Longitudinal Survey of Aging (ELSA) was designed according to the HRS with its first wave launched in 2002. 2 years later Continental Europe also decided to set up an aging database by establishing the Survey of Health, Ageing and Retirement in Europe (SHARE), a cross-nationally comparable panel database of micro data. SHARE started with 12 countries (Austria, Belgium, Denmark, France, Germany, Greece, Israel, Italy, the Netherlands, Spain, Sweden and Switzerland) in 2004 with wave 1, three countries (the Czech Republic, Ireland and Poland) joined in wave 2, and another four countries (Estonia, Hungary, Portugal and Slovenia), joined in wave 4. The three surveys (HRS, ELSA and SHARE) are carefully harmonized, and thus provide an excellent basis for cross- country investigation of aging population in developed countries.

What makes the surveys appropriate for this particular analysis is that they include a battery of tests about cognitive abilities (memory, verbal fluency and numeracy). The test of memory is done as follows: 10 simple words are read out by the interviewer and the respondent should recall them once immediately after hearing and then at the end of the cognitive functioning module. As a result, both immediate recall and delayed recall scores range from 0 to 10. Often, the two variables are merged to a composite one by adding them up, which is called total word recall. Verbal fluency is tested by asking the respondent to name as many distinct animals as she can within one minute. The length of this list provides a measure for verbal fluency. SHARE also consists of several questions about individual numeracy skills. Respondents who answer the first one correctly get a more difficult one, while those who failed get an easier one. The last question requires the respondent to calculate compound interest. The number of correct answers to these questions provides an objective measure of numeracy ranging from 0 to 4. Finally, there is a test of orientation of four questions which examines whether the respondent is aware of the date of the interview (day, month, year) and the day of the week. This test may be used to detect individuals with serious cognitive problems or progressed dementia.

Various measures of cognitive skills might grab its different aspects as argued in Mazzonna and Peracchi (2012). As most of the papers use the results on memory tests I also focus on that measure for comparison purposes. To have a common unit I use standardized scores to express scales in standard deviation.

Throughout the paper I make use of the first, second and fourth waves of SHARE. The third wave of data collection (SHARELIFE) is omitted, as it is of different nature: it focuses on people's life histories instead of current characteristics.

4 Replications

In this section I replicate the main results of the literature, namely that of Rohwedder and Willis (2010), Mazzonna and Peracchi (2012), Coe et al. (2012) and Bonsang et al. (2012). I put all of these results in my unified framework and show that their differing conclusions actually fit in the broader picture. The ambiguity of their results stems from the differences in their identification strategies that implies that their estimated "effects" of retirement on cognitive performance measure different kinds of things.

The papers differ in three crucial aspect: first, what is their assumption about how retirement should affect cognitive performance (i.e. what is their assumption for f), second, which factors they are controlling for from $\boldsymbol{X}_i$, and third, what is their choice for instrumental variable. Of course, they also differ in the data they use for estimation but considering the goal of uncovering a general relationship this fact does not really matter (as far as the measurements are comparable across the datasets).

The structural equation the papers try to estimate could be summarized as follows:

$$CS_i = \alpha + f(YR_i; \boldsymbol{\beta}) + \boldsymbol{X}_i^{*'}\boldsymbol{\gamma}^* + \tilde{u}_i \tag{3}$$

$$\tilde{u}_i = \tilde{\boldsymbol{X}}_i^{'}\tilde{\boldsymbol{\gamma}} + \varepsilon_i \tag{4}$$

where CS_i is a cognitive score, a measurement of cognitive performance. This formulation helps to differentiate between factors which are controlled for ($\boldsymbol{X}_i^*$) versus factors which remain in the error term ($\tilde{\boldsymbol{X}}_i$). To get a clear causal effect equation (3) is estimated by a 2SLS procedure where the first stage is

$$YR_i = \boldsymbol{Z}_i^{'}\boldsymbol{\pi} + \boldsymbol{X}_i^{*'}\boldsymbol{\rho} + \nu_i \tag{5}$$

For now on let us assume that the cognitive measurements detailed in the previous section describe well the actual cognitive skills. To be more precise, I assume that $CP_i = CS_i + e_i$ where e_i is a classical measurement error in the dependent variable, i.e. $\mathrm{Cov}(e_i, CS_i) = \mathrm{Cov}(e_i, YR_i) = 0$. In this case our estimators remain consistent although less precise. Later on I will ease this assumption to assess how crucial it is.

All of the papers use various public policy rules to instrument retirement (such as pension eligibility rules). Such rules are good candidates for instrument as they vary across country and gender and are strongly correlated with employment status. The crucial question is whether it also satisfies the other assumption, namely the assumption of exogeneity. Using the formal notation the exogeneity assumption can be expressed as $\mathrm{E}\left[\tilde{u}_i | \boldsymbol{Z}_i\right] = 0$. It essentially says that there is no systematic difference related to cognitive performance between an eligible and a non-eligible individual in the sample (once controlling for some other features).

Rohwedder and Willis (2010) provide the first serious attempt to uncover the causal relationship between retirement and individual cognitive performance. Their framework is the simplest one among the papers: they only include a dummy for not working on the right hand side on a restricted sample of people aged between 60 and 64. This is equivalent to estimating the average effect of retirement on cognition conditional on the average duration of retirement the sample, that is assuming that $f(YR_i; \boldsymbol{\beta}) = \tilde{\beta}\mathbf{1}(YR_i > 0)$ where $\tilde{\beta} = \beta\overline{YR_i}$. Beside restricting the sample on a narrow age-range they do not include anything in $\boldsymbol{X}_i^*$. To handle endogeneity they use public pension eligibility rules as instruments: whether the individual is eligible for early or full benefits. See Table A.1 for a summary of the methodologies.

Rohwedder and Willis (2010) estimate their model on the 2004 waves of SHARE, ELSA and HRS, and find that retirement has a large adverse effect on cognition among 60-64 years old, amounting to one-and-a-half standard deviation. Unfortunately, they do not report the average duration of retirement in their sample so I cannot convert this number to yearly average.

Using only the first wave of SHARE (and thus having a much smaller sample than theirs, 4466 versus 8828 observations) I was able to replicate their main findings (see the first column of Table 1). The pattern is the same: retirement seems to decrease cognitive performance. However, my estimation is somewhat smaller, amounting to only 1 standard deviation. Considering that the average duration of retirement in my sample is 6.6 years, it could be translated to an average yearly decline of 0.15 standard deviation.

Table 1: Comparing the methodology of Rohwedder and Willis (2010) by two versions of the instrumental variable: 2SLS estimation

	(1) Rohwedder and Willis (2010)	(2) Mazzonna and Peracchi (2012)
Retired	-1.009^{***}	-0.500^{***}
	(0.14)	(0.13)
Constant	0.736^{***}	0.365^{***}
	(0.10)	(0.097)
Observations	4,462	4,462

Both result are from the second stage estimation of $CS_i = \alpha + \beta R_i + u_i$ where R_i is a retirement dummy ($\mathbf{1}(YR_i > 0)$) which is instrumented by early and normal eligibility dummies. The coefficient of interest in Rohwedder and Willis (2010) is -4.66^{***} on a sample of 8828 observations which amounts to 1.5 standard deviation.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

In order to be able to interpret the previous result as causal effect we should be sure that $\mathrm{E}\left[\tilde{u}_i | \boldsymbol{Z}_i\right] = 0$. Clearly, eligibility rules are not related to unobserved individual idiosyncrasies in cognition, as they generally refer to everyone. So using the instrument indeed helps to handle our problems. However, there are other factors left in $\tilde{u}_i$ which are unlikely to be uncorrelated with the instrument. For example, in most countries eligibility rules differ for males and females: women tend to become eligible earlier. Women also have higher memory scores than men in the same age (for this sample the mean difference amounts to 0.17 standard deviation). Not controlling for gender is likely to lead to underestimated effects as women with better scores are overrepresented in the eligible population. Bingley and Martinello (2013) draw the attention to the fact that people from different countries might differ in their average education as well (e.g. because of different compulsory schooling laws affecting today's pensioners). As different countries also have different eligibility rules, ignoring schooling is also likely to undermine the exogeneity of the instruments. Bingley and Martinello (2013) show that countries with higher eligibility ages also tend to have better educated old age people, and thus the effect of Rohwedder and Willis (2010) is overestimated. The failure of the exogeneity assumption makes the causal interpretation of the results in the first column of Table 1 questionable.

There is an easy way to improve upon the estimation of Rohwedder and Willis (2010). They use eligibility rules that were in effect at the time when the interviews were conducted. Mazzonna and Peracchi (2012), in contrast, use the same eligibility rules but consider that the rules might have had some changes. For each individual they apply the eligibility rules which were in effect for the individual's cohort. This way they have some variation in the rules within country-gender cells. Looking at first stage regression results in Table 2 we can see that both instrumental variables reach the same level of relevance (which is also comparable to that of reached by Rohwedder and Willis (2010)). However, using this IV results in a significant drop in the coefficient of interest (see the second column of Table 1). Introducing within-country- gender variation into the instrumental variable leads to halving of the effect, to 0.075 standard deviation decline per year.

The methodology of Mazzonna and Peracchi (2012) differs from that of Rohwedder and Willis (2010) not only in respect of the instrumental variable. They also assume a different functional form, and control for a different set of features. Instead of using just a retirement dummy (and thus estimating the effect conditional on the average duration of retirement) they enter the number of years spent in retirement linearly in the equation (i.e. they assume that $f(YR_i; \boldsymbol{\beta}) = \beta YR_i$). They control for age in a different manner: instead of restricting the sample to 60-64 years old they control for age linearly in a sample of people aged 50-70. They also control for country dummies and estimate the equation separately for men and women.

Table 3 summarizes the results of moving from the strategy of Rohwedder and Willis (2010) to that of Mazzonna and Peracchi (2012) step by step. Table A.2 shows the corresponding first stage regression results.

Lessons from Table 3 in bullet points:

(1) 0.05 yearly $\sim$ 0.08 average from second column of Table 1

Table 2: Comparing the methodology of Rohwedder and Willis (2010) by two versions of the instrumental variable: first stage

	(1) Rohwedder and Willis (2010)	(2) Mazzonna and Peracchi (2012)
Eligible for early benefits	0.323***	0.246***
	(0.028)	(0.027)
Eligible for full benefits	0.165***	0.185***
	(0.014)	(0.014)
Constant	0.375***	0.439***
	(0.027)	(0.025)
Observations	4,462	4,462
adjusted R^2	0.0644	0.0650

Both result are from the first stage estimation of $CS_i = \alpha + \beta R_i + u_i$ where R_i is a retirement dummy ($\mathbf{1}(YR_i > 0)$) which is instrumented by early and normal eligibility dummies. The corresponding coefficients for early and full benefits in Rohwedder and Willis (2010) are 0.19*** and 0.16**, respectively, with the adjusted R^2 being 0.059 on a sample of 8828 observations.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

(2) Extending the age range does not really matter.

(3) Restricting to those with labor market history makes the effect a bit larger (make the sample of retirees more elite which is OK in this case as the main goal is to assess the effect of moving from working to retirement).

(4) Controlling for age delivers weird results. The effect doubles and the coefficient on age is positive: until retirement age seems to improve cognitive performance, after that it decreases by $\sim$0.14 standard deviation. This could be explained by country differences: as Bingley and Martinello (2013) draws the attention to, eligibility age and schooling is positively correlated (in my sample the correlation is 0.20 and 0.14 for the early and normal eligibility age, respectively). Therefore, comparing two individuals with the same age but differing years after eligibility likely means comparing two individuals from different countries with the older one being from the better educated country. This reasoning justifies the positive age coefficient and underlines the importance of controlling for both age and country.

(5) Controlling for country indeed solves the puzzle of positive age coefficient, changing its sign to the expected negative. However, now the coefficient of interest changes sign and gets positive. The unexpected sign results again from omitted variable bias: gender is not controlled for. As mentioned previously, women perform significantly better on memory

Table 3: Moving from the strategy of Rohwedder and Willis (2010) to that of Mazzonna and Peracchi (2012)

	(1) aged 60-64	(2) aged 50-70	(3) + worked at 50	(4) + age	(5) + country
Years in retirement	-0.051^{***}	-0.053^{***}	-0.083^{***}	-0.176^{***}	0.167^{***}
	(0.0072)	(0.0021)	(0.0029)	(0.016)	(0.035)
Age				0.047^{***}	-0.116^{***}
				(0.0077)	(0.017)
Constant	0.376^{***}	0.359^{***}	0.223^{***}	-2.333^{***}	6.344^{***}
	(0.051)	(0.015)	(0.011)	(0.42)	(0.86)
Country dummies	No	No	No	No	Yes
Observations	4,052	17,448	13,973	13,973	13,973
Weak IV F statistic	118.05	1614.48	6337.18	256.62	55.57

Weak IV F statistic is calculated according to Angrist and Pischke (2008). Stock et al. (2002) suggest that an F below 10 should make us worry about the potential bias in the IV estimation.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

tests (controlling for age) than men (for this sample, they are by 0.28 standard deviation better. Thus, when we control for both age and country, we mainly identify the retirement effect from gender variation. To see that this is really the case, check the results in the first two columns of Table 4 where I also included a control for gender. The positive sign of the coefficient of interest reverse back to the expected negative. The next two columns of the table shows the same result when the numeracy score is used to measure the cognitive skills. Women perform on average by -0.27 standard deviation worse on the numeracy test and correspondingly, we see larger negative effect of years in retirement on numeracy when not controlling for gender.

There is one more puzzle in Table 4. Why is the coefficient on age is positive for numeracy when controlling for country but not for gender? (The same coefficient is negative for TWR.) There is a possible explanation for that. Interestingly, as the sample ages so decreases the share of women (mortality rates would predict the opposite). The coefficient on age is mainly identified on non-eligible population (as for eligible population the age effect is actually the sum of the coefficients on age and years in retirement). As women are better in memory tests, and they are less in relatively older cohort, the composition effect implies a negative coefficient for age. By contrast, the opposite is true for numeracy, composition effect implies a positive coefficient for age as women are worse in numeracy. Controlling for gender eliminates the level differences in

Table 4: Control for gender

	(1) TWR	(2) TWR	(3) numeracy	(4) numeracy
Years in retirement	0.167***	−0.197***	−0.390***	−0.010
	(0.035)	(0.040)	(0.047)	(0.036)
Age	−0.116***	0.059***	0.165***	−0.018
	(0.017)	(0.019)	(0.022)	(0.017)
female		0.291***		−0.303***
		(0.022)		(0.019)
Constant	6.344***	−2.792***	−8.054***	1.478*
	(0.86)	(1.00)	(1.15)	(0.88)
Country dummies	Yes	Yes	Yes	Yes
Observations	13,973	13,973	14,065	14,065
Weak IV F statistic	55.57	40.23	56.13	40.95

Weak IV F statistic is calculated according to Angrist and Pischke (2008). Stock et al. (2002) suggest that an F below 10 should make us worry about the potential bias in the IV estimation.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

cognitive scores by gender, but possibly different rate of decline (i.e. differential retirement effect for men and women) could further complicate the results and make it hard to assess the direction of possible bias.

That is why Mazzonna and Peracchi (2012) estimate the equation separately for men and women. Table 5 show my replication for their strategy for total word recall and numeracy. According to my results, the rate of decline is indeed different in which the relatively better performing gender suffers a larger decline. These estimation is the closest to that of Mazzonna and Peracchi (2012), the only difference is in the use of measurements: they adjust the cognitive scores by the time spent on answering them that I do not. This might be a reason why my estimates do not lie really close to theirs. In columns A of their Table 5 they report a consistent negative effect of retirement amounting to 0.006-0.015 standard deviation per year (I rescale the coefficients to standard deviations using the reported numbers in their Table 3). My results show bigger effects but they are also less precise. However, the gender difference in the rate is decline is the same. I was also able to broadly reproduce the simple OLS results which are available upon request.

There is an additional problem which could contaminate the results. According to previous evidence (Banks and Mazzonna, 2012) education matters in old age cognitive skills even if gained

Table 5: Estimating separately by gender, closest to Mazzonna and Peracchi (2012)

	(1) TWR, men	(2) TWR, women	(3) numeracy, men	(4) numeracy, women
Years in retirement	0.017	−0.043*	−0.061	−0.036
	(0.039)	(0.025)	(0.040)	(0.025)
Age	−0.042**	−0.016	0.009	−0.009
	(0.018)	(0.012)	(0.019)	(0.013)
Constant	2.475***	1.252**	−0.006	0.830
	(0.95)	(0.63)	(0.99)	(0.64)
Country dummies	Yes	Yes	Yes	Yes
Observations	8,017	5,956	8,084	5,981
Weak IV F statistic	32.93	84.10	32.68	83.92

Weak IV F statistic is calculated according to Angrist and Pischke (2008). Stock et al. (2002) suggest that an F below 10 should make us worry about the potential bias in the IV estimation.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

in early stages of life. Today's pensioners are highly affected by the expansion of average schooling: the 50 years old cohort spent on average 2.7 years more in school than the cohort of 70. Mazzonna and Peracchi (2012) try to control for education by including a low-education dummy (they also interact this dummy with the effect) and show that education indeed plays a significant role in explaining the heterogeneity in the levels of cognitive skills (and to a smaller extent in their age-related decline). However, there is also evidence (see for example the PISA surveys) that countries are different in how effective they improve cognitive abilities in childhood. The first PISA survey was performed in 2000, the mean scores of countries in math, science and reading are positively correlated with the eligibility ages. If there is some persistence in the quality of education systems from the time when today's pensioners went to school and that of today, this might introduce a new type of bias in the estimations even if the number of years spent in education is controlled for.

We might also question the assumption that age related cognitive decline and the effect of retirement is the same across countries (previously we only accounted for level differences). Similarly to what we did for gender we could estimate the main equation separately for countries to see whether there are differential effects. Unfortunately, in this case we lose our ability to use our instruments as there is not enough variability in them within country. But we can still see significant differences in the OLS results (see Figure 1). This, again, questions the reliability of the previous identification where we assume that individuals from different countries might be compared once we allow for average level differences.

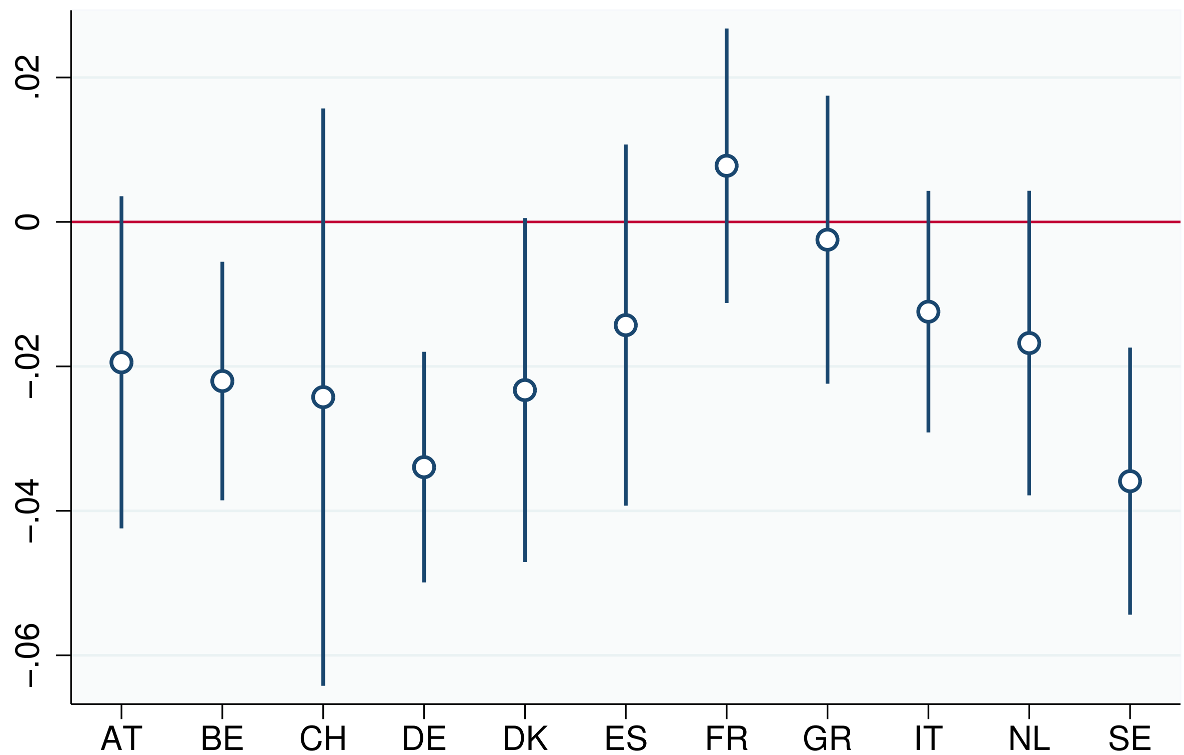

Figure 1: OLS coefficients by country (controlling for age and gender)
Average duration of retirement in the sample (50-70 years, worked at 50) varies a lot as well: above 4 years in Austria, below 1.5 years in Switzerland and Sweden.

Replications of Coe et al. (2012) and Bonsang et al. (2012) comes later...

5 My strategy

I demonstrated in the previous section that getting a clear causal effect in cross-section is really challenging. Some minor improvements upon the strategy of Rohwedder and Willis (2010) resulted in a drop of the estimated effect from 0.22 standard deviation per year to at least a magnitude less. But still, there are plenty of factors left in u_i^* which are likely to bias our results (to a yet unknown extent).

Moving to panel identification could solve a lot of issues. It makes possible to compare the cognitive scores of the exact same individual instead of assuming that an another one is a good subject for comparison. Mazzonna and Peracchi (2012) argue that panel data is inappropriate for identifying the causal effect of retirement on cognition because of two reasons. First, as in each wave the exact same cognitive exercises are performed, the participants may remember their answers from previous waves. However, when we compare the cognitive path of employed retired persons, we should only worry about a case when learning is systematically different for the two groups for which I do not see a reason. Second, there is considerable attrition in the sample which might also bias the result. To see whether this really cause a problem, I will estimate the same relationship on different samples which are likely to suffer from different attrition rates and check whether they are different.

Figure 2 shows the average paths of cognitive score of the individuals who were present in all waves (and satisfy the previous restrictions such as being 50-70 years old at the beginning and worked at age 50), by working history for each country. To illustrate, working history of 110 indicates a person who worked during the first two waves but left between wave 2 and 4. The first thing to note is that no clear pattern arises regarding the effect of retirement. There are some signs of selection and learning, but it does not seem like retirement would have any clear-cut effect on cognitive scores.

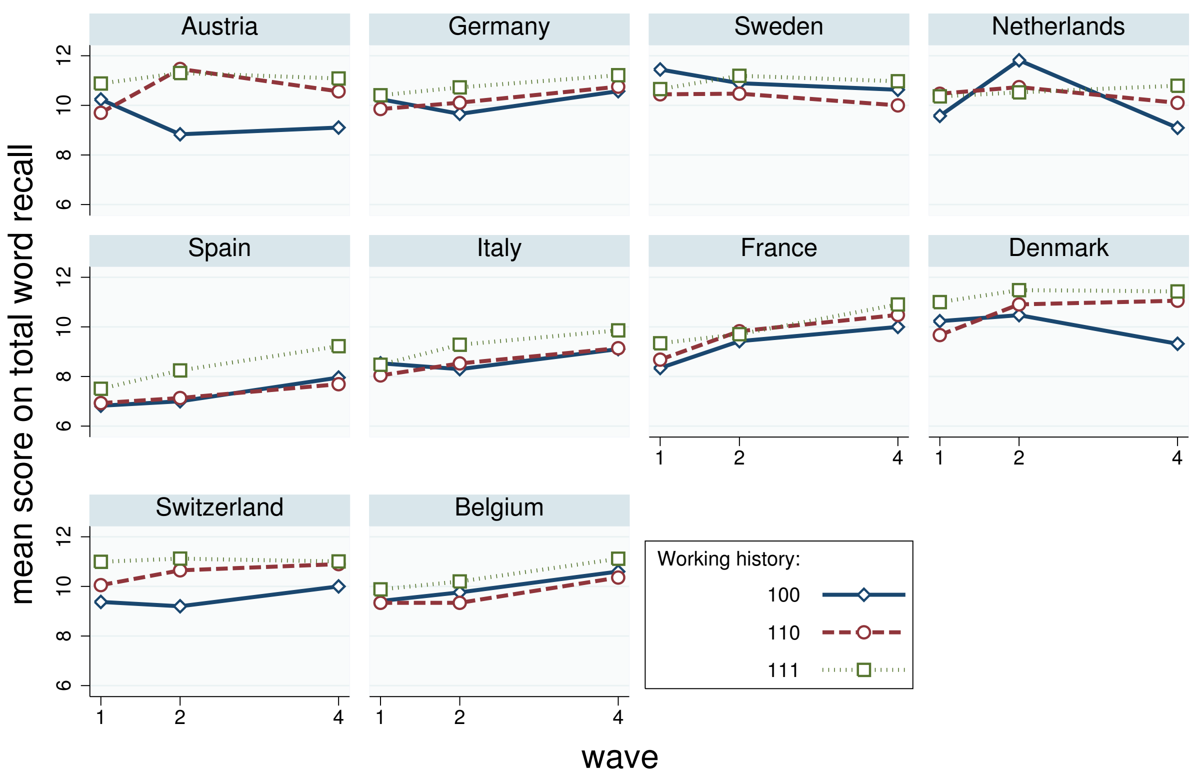

Figure 2: Pattern of cognitive scores across waves by working history
100: worked only in first wave, 110: worked in first two waves, etc.

Using longitudinal data one can control for all time-invariant individual heterogeneity like gender, amount and quality of education, etc. Formally, this means to include an individual specific term a_i in $\boldsymbol{X}_i^*$ in equation (3) as is in equation (6) and then estimate the differenced version of that (7):

$$CS_i = \alpha + \beta YR_i + a_i + \tilde{u}_i \tag{6}$$

$$\Delta CS_i = \alpha^* + \beta \Delta YR_i + \Delta \tilde{u}_i \tag{7}$$

The previous controls (country, education, gender) fall out as they (typically) do not vary within individual in older ages. Additionally, we might possibly control for differential effects by including gender and country in equation (7). There is one peculiarity regarding this data: the time elapsed between different interviews is not the same for all individual (e.g. it ranges from 11 to 40 months between the first two waves). Therefor, the amount of ageing is not the same for each individual which implies the need to include a control for that in the regressions.

The first three columns of table 6 shows the 2SLS results for various estimations for changes in total word recall score between wave 1 and 2 (again, years in retirement is instrumented by years after the eligibility ages). It seems that one more year in retirement decreases the cognitive score by 0.02 standard deviation point. This magnitude corresponds to that of Table 5, the closest replication of Mazzonna and Peracchi (2012). This seems to be robust for allowing differential effect by gender and country. Table 7 and 8 shows the same results but for different periods (for score changes between waves 2 and 4 and between waves 1 and 4). They also seem to be in line with the previous estimation showing an additional yearly decline of 0.02-0.04 standard deviation.

However, there is a reasonable scenario we have not considered yet. If natural age-related cognitive decline is concave instead of being linear then controlling only for a linear age trend might attribute the larger decline in older ages to the effect of retirement. Unfortunately, we are unable to allow for differential age effect and still use our instruments (there is not enough variability within country, gender and age). It still makes sense to run simple OLS regressions to see the difference. If we compare the third and fourth columns of the tables we can see that OLS results are slightly smaller in absolute value than those of 2SLS (although we were afraid of negative bias in them). If we include age (column 5), the estimated effect drops further, to about 0.01 standard deviation yearly. This estimation might suffer from selection but as we eliminated all time-invariant individual heterogeneity it only matters if the rate of decline is different as well. Given the similarity of 2SLS and OLS estimation of columns 3 and 4 I do not think it is a serious problem. Additionally, if any, this would bias the results in the negative direction.

Tables A.3-A.5 show the same estimations for numeracy. Again, we see that the adverse effect of retirement is at most 0.01 standard deviation yearly. The magnitude of this effect is comparable to that of Mazzonna and Peracchi (2012) but being a result of a more reliable identification strategy. Providing its really small size I do not think one should worry about it.

Table 6: Panel estimation: change in total word recall score between wave 1 and 2

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	−0.020*	−0.020	−0.019	−0.016	−0.007
	(0.012)	(0.012)	(0.013)	(0.016)	(0.025)
Years elapsed	−0.001	0.000	0.080*	0.047	0.045
	(0.026)	(0.026)	(0.041)	(0.058)	(0.058)
Female		0.029	0.023	0.006	0.004
		(0.021)	(0.021)	(0.029)	(0.029)
Age at first wave					−0.002
					(0.0042)
Constant	0.023	0.008	−0.086	0.005	0.125
	(0.061)	(0.062)	(0.11)	(0.16)	(0.29)
Country dummies	No	No	Yes	Yes	Yes
Observations	8,631	8,631	8,631	4,704	4,704
Weak IV F statistic	4331.52	4409.99	4384.67		

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

6 The way ahead

Investigate the following topics:

- Dumbledore hypothesis (by Stine-Morrow, 2007) "it is our choices... that show what we truly are, far more than our abilities" (Rowling, 1999) –> no empirics
- Physical inactivity on Cognitive Performance (Aichberger et al., 2010) –> no causal effect, only comparison by controlling for some potential confounders
- nice review about results on enrichment effects can be found in Hertzog et al. (2009)

My hypothesis:

- Retirement in itself has essentially no effect
- Lifestyle changes often happening together the transition from work to retirement might matter – but is heterogeneous across agents: some may retire from a stressful job to get more time to relax and care for grandchildren, others may lose the meaning of their lives

Table 7: Panel estimation: change in total word recall score between wave 2 and 4

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	-0.030^{***}	-0.030^{***}	-0.047^{***}	-0.030^{***}	0.015
	(0.0077)	(0.0077)	(0.0080)	(0.011)	(0.018)
Years elapsed	0.051	0.052	−0.022	−0.101	−0.098
	(0.045)	(0.045)	(0.056)	(0.075)	(0.075)
Female		−0.015	−0.006	0.009	−0.003
		(0.025)	(0.025)	(0.034)	(0.034)
Age at second wave					-0.017^{***}
					(0.0055)
Constant	−0.161	−0.158	0.008	0.200	1.148^{**}
	(0.19)	(0.19)	(0.25)	(0.33)	(0.45)
Country dummies	No	No	Yes	Yes	Yes
Observations	6,781	6,781	6,781	3,775	3,775
Weak IV F statistic	4447.80	4472.96	4221.74		

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

- Goal: try to uncover these patterns (clear the simple comparisons from selection)
- Fear: large drops in sample sizes, lose power...

7 Conclusion

Two main lessons:

1. Identifying a clear causal effect is *really* hard (ideally, we would like to compare individuals from the same cohort and country, with the same age and education, one of them randomly assigned to be retired for a period of time while the other being working... mission impossible).
2. According to my analysis there is most likely no real effect.

Table 8: Panel estimation: change in total word recall score between wave 1 and 4

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	−0.024***	−0.024***	−0.038***	−0.038***	−0.025**
	(0.0056)	(0.0056)	(0.0059)	(0.0075)	(0.012)
Years elapsed	−0.257***	−0.258***	−0.087	−0.108	−0.116
	(0.035)	(0.035)	(0.054)	(0.076)	(0.076)
Female		0.028	0.041	0.079**	0.072*
		(0.028)	(0.028)	(0.039)	(0.039)
Age at first wave					−0.008
					(0.0060)
Constant	1.783***	1.775***	0.587	0.584	1.053*
	(0.23)	(0.23)	(0.37)	(0.52)	(0.64)
Country dummies	No	No	Yes	Yes	Yes
Observations	5,146	5,146	5,146	2,839	2,839
Weak IV F statistic	3551.55	3601.08	3392.05		

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Appendix

Table A.1: Comparing the methodologies of the literature

	$f(YR_i; \boldsymbol{\beta})$	$\boldsymbol{X}_i^*$	$\boldsymbol{Z}_i$
Rohwedder and Willis (2010)	$\tilde{\beta}\mathbf{1}(YR_i > 0)$	-	normal end early eligibility dummies (no variation within country-gender cells)
Mazzonna and Peracchi (2012)	βYR_i	age, country dummies, gender	normal and early eligibility dummies (some variation within country-gender cells)
Coe et al. (2012)	later...	later...	later...
Bonsang et al. (2012)	later...	later...	later...

Using a simple retirement dummy instead of years in retirement is equivalent to estimating the average effect conditional on the average time spent in retirement in the sample, i.e. $\tilde{\beta} = \beta\overline{YR_i}$.

Table A.2: Moving from the strategy of Rohwedder and Willis (2010) to that of Mazzonna and Peracchi (2012): First stage

	(1) aged 60-64	(2) aged 50-70	(3) + worked at 50	(4) + age	(5) + country
Years after early eligibility	0.303***	0.016	0.184***	0.127***	0.019
	(0.071)	(0.033)	(0.012)	(0.013)	(0.024)
Years after normal eligibility	0.580***	0.581***	0.264***	0.138***	0.156***
	(0.079)	(0.033)	(0.012)	(0.015)	(0.018)
Age				0.209***	0.299***
				(0.013)	(0.023)
Constant	6.437***	8.310***	3.629***	−9.241***	−13.363***
	(0.38)	(0.18)	(0.067)	(0.79)	(1.36)
Country dummies	No	No	No	No	Yes
Observations	4,052	17,448	13,973	13,973	13,973
adjusted R^2	0.0546	0.1561	0.4756	0.4854	0.5038

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.3: Panel estimation: change in numeracy score between wave 1 and 2

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	−0.025**	−0.026**	−0.024*	−0.011	0.012
	(0.012)	(0.012)	(0.013)	(0.016)	(0.025)
Years elapsed	0.013	0.014	0.004	−0.027	−0.033
	(0.026)	(0.026)	(0.041)	(0.058)	(0.058)
Female		0.015	0.018	0.030	0.026
		(0.021)	(0.021)	(0.029)	(0.029)
Age at first wave					−0.005
					(0.0043)
Constant	−0.006	−0.013	0.039	0.083	0.392
	(0.060)	(0.061)	(0.11)	(0.16)	(0.29)
Country dummies	No	No	Yes	Yes	Yes
Observations	8,665	8,665	8,665	4,721	4,721
Weak IV F statistic	4351.43	4428.95	4404.51		

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.4: Panel estimation: change in numeracy score between wave 2 and 4

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	−0.003	−0.003	−0.005**	−0.007***	−0.008*
	(0.0018)	(0.0018)	(0.0019)	(0.0027)	(0.0043)
Years elapsed	0.031***	0.031***	0.022	0.006	0.005
	(0.011)	(0.011)	(0.013)	(0.017)	(0.017)
Female		−0.010*	−0.011*	−0.019**	−0.019**
		(0.0058)	(0.0058)	(0.0078)	(0.0079)
Age at second wave					0.000
					(0.0013)
Constant	−0.124***	−0.122***	−0.035	0.027	0.014
	(0.044)	(0.044)	(0.058)	(0.077)	(0.10)
Country dummies	No	No	Yes	Yes	Yes
Observations	6,788	6,788	6,788	3,784	3,784
Weak IV F statistic	4432.90	4455.97	4211.08		

Comes later
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.5: Panel estimation: change in numeracy score between wave 1 and 4

	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) OLS	(5) OLS
Years in retirement	−0.007	−0.007	−0.007	−0.003	0.002
	(0.0051)	(0.0051)	(0.0054)	(0.0069)	(0.011)
Years elapsed	−0.026	−0.026	0.029	−0.037	−0.040
	(0.032)	(0.032)	(0.050)	(0.070)	(0.070)
Female		−0.015	−0.015	−0.036	−0.038
		(0.026)	(0.026)	(0.036)	(0.036)
Age at first wave					−0.003
					(0.0056)
Constant	0.196	0.200	−0.141	0.310	0.464
	(0.21)	(0.21)	(0.34)	(0.48)	(0.59)
Country dummies	No	No	Yes	Yes	Yes
Observations	5,161	5,161	5,161	2,849	2,849
Weak IV F statistic	3552.34	3601.76	3394.82		

Comes later...
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.6: From Rohwedder and Willis (2010) to Mazzonna and Peracchi (2012), only for presentation

	(1)	(2)	(3)	(4)
Years in retirement	−0.083***	−0.176***	0.167***	−0.197***
	(0.0029)	(0.016)	(0.035)	(0.040)
Age		0.047***	−0.116***	0.059***
		(0.0077)	(0.017)	(0.019)
Female				0.291***
				(0.022)
Constant	0.223***	−2.333***	6.344***	−2.792***
	(0.011)	(0.42)	(0.86)	(1.00)
Country dummies	No	No	Yes	Yes
Observations	13,973	13,973	13,973	13,973
Weak IV F statistic	6337.18	256.62	55.57	40.23

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

References

Aichberger, M.C., M.a. Busch, F.M. Reischies, a. Ströhle, a. Heinz, and M.a. Rapp, "Effect of Physical Inactivity on Cognitive Performance after 2.5 Years of Follow-Up," *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, March 2010, *23* (1), 7–15.

Angrist, Joshua D and Jörn-Steffen Pischke, "Mostly harmless econometrics: An empiricist's companion," *An empiricist's companion*, 2008, p. 392.

Banks, James and Fabrizio Mazzonna, "The Effect of Education on Old Age Cognitive Abilities: Evidence from a Regression Discontinuity Design.," *Economic journal (London, England)*, May 2012, *122* (560), 418–448.

Baum, Christopher F, Mark E Schaffer, and Steven Stillman, "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation," July 2014.

Bingley, Paul and Alessandro Martinello, "Mental retirement and schooling," *European Economic Review*, October 2013, *63*, 292–298.

Bonsang, Eric, Stéphane Adam, and Sergio Perelman, "Does retirement affect cognitive functioning?," *Journal of Health Economics*, May 2012, *31* (3), 490–501.

Coe, Norma B, Hans-Martin von Gaudecker, Maarten Lindeboom, and Jürgen Maurer, "The effect of retirement on cognitive functioning.," *Health Economics*, August 2012, *21* (8), 913–27.

Hertzog, Christopher, Arthur F Kramer, Robert S Wilson, and Ulman Lindenberger, "Enrichment Effects on Adult Cognitive Development," *Psychological Science*, 2009, *9* (1), 1–65.

Jann, Ben, "Making regression tables from stored estimates," *The Stata Journal*, 2005, *5* (3), 288–308.

_ , "Making regression tables simplified," *The Stata Journal*, 2007, *7* (2), 227–244.

Mazzonna, Fabrizio and Franco Peracchi, "Ageing, cognitive abilities and retirement," *European Economic Review*, May 2012, *56* (4), 691–710.

Rohwedder, Susann and Robert J Willis, "Mental Retirement.," *The Journal of Economic Perspectives*, January 2010, *24* (1), 119–138.

Rowling, Joanne K, *Harry Potter and the Chamber of Secrets*, New York: Scholastic, Inc., 1999.

Schaie, K Warner, "The hazards of cognitive aging.," *The Gerontologist*, August 1989, *29* (4), 484–93.

Stine-Morrow, Elizabeth A L, "The Dumbledore Hypothesis of Cognitive Aging," *Current Directions in Psychological Science*, December 2007, *16* (6), 295–299.

Stock, James H, Jonathan H Wright, and Motohiro Yogo, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 2002, *20*, 518–529.