Aging, cognitive abilities and education in Europe*

Giuseppe De Luca
ISFOL

Fabrizio Mazzonna
Tor Vergata University

Franco Peracchi
Tor Vergata University and EIEF

December 2008

Abstract

This paper investigates the relationship between aging, cognitive abilities and social, behavioral and environmental factors using the first two waves of the Survey on Health, Aging and Retirement in Europe (SHARE), a longitudinal survey that offers the possibility of comparing several European countries using nationally representative samples of the population aged 50+. After controlling for sample attrition, we find that education plays a fundamental role in explaining heterogeneity in the level of cognitive abilities and in delaying their age-related decline. The role of childhood environment, proxied here by body height and parents' occupation, is also important. Other factors accounted for by the literature, such as social and physical activities, predict a delay in the process of neurodegeneration. Finally, we find large and systematic differences in cognitive function across European countries, with people living in Mediterranean countries showing lower test scores in all cognitive domains, even after controlling for observable individual characteristics.

Keywords: Aging, cognitive abilities, education, panel data, panel attrition, SHARE.

JEL codes: J14, J24, C23.

* Preliminary and incomplete. Corresponding author: Fabrizio Mazzonna, Faculty of Economics, Tor Vergata University, via Columbia 2, 00133 Rome, Italy (fabrizio.mazzonna@uniroma2.it). We thank Anne Case, Chris Paxson and Frank Vella for insightful discussions.

1 Introduction

In many countries, aging is one of the great social and economic challenges of the 21st century. In Europe, for example, the ratio of persons aged over 65 as a percentage of the working age population (the old-age dependency ratio) is expected to increase from its current levels of 25.4 percent to 53.5 percent in 2060 (Eurostat 2008). This increase of the old-age dependency ratio is likely to place a heavy financial burden on the pension and health-care systems of most European countries.

A fundamental issue related to the aging process is the decline of mental health and cognitive abilities. Declining cognitive functioning may have a number of direct economic consequences. First, it may lower the level of human capital of the labor force. Second, being associated with declining individual ability to work, it may induce early retirement. Third, it may affect individual ability to process information, which is a crucial element for the appropriate formulation of retirement and saving plans (Park 1999; Banks *et al.* 2006; Christelis, Jappelli and Padula 2006). The decline in cognitive functioning is also a predictor of later limitations in Instrumental Activities of Daily Living (IADL), physical impairments, dementia and Alzheimer's disease. The economic costs of these pathologies is substantial. For example, the review paper by Jönsson and Berr (2005) estimates that the annual cost of long-term care due to dementia ranges between Euro 9,000 and 16,000 Euro per patient. Multiplying these figures by a dementia prevalence rate of about 5.4 percent for the European population aged 60+ (Ferri *et al.* 2005), one obtains an average annual cost of long-term care for people aged 60+ that ranges between Euro 486 and 864.

Despite its importance, the age-related process of neurodegeneration is complex and its determinants are still not well understood. Longitudinal studies by Schaie (1986, 1989) show that cognitive functions are relatively stable until the fifth decade of life. After this period, the process of decline becomes apparent and the incidence of cognitive impairments increases sharply with age. At all ages, however, there is large variation across individuals in the level of cognitive performance.

A useful conceptual framework, due to Stern (2002, 2003), is that individuals have different levels of cognitive reserve, and that higher levels allow them to prevent or slow down the organic process of neurodegeneration associated with aging. Individual heterogeneity in cognitive performances may reflect both genetic differences in the level of cognitive reserve and other life experience events (individual choices or exogenous shocks) that may affect the initial cognitive endowment and the rate of age-related decline. As shown by several empirical studies, important factors in this process are education (Le Carret 2003; Leibovici *et al.* 1996), occupational choices (Adam *et al.* 2006), leisure activities (Scarmeas 2003), home environment and parental influences in childhood (Kaplan *et al.* 2001; Knudsen *et al.* 2006; Cunha & Heckman 2007; Case and Paxson 2008) and adolescence (Richards *e al.* 2004), physical exercise (Cervilla *et al.* 2000; Maraldi and Pahor 2006), social activities (Trouton *et al.* 2006), and lifestyles and chronic diseases like hypertension or heart disease (Cervilla *et al.* 1999; Meyer *et al.* 1999).

The available empirical literature shares some common limitations. First, results are usually based on small sample sizes and cross country comparisons are lacking. The few existing longitudinal studies (Schaie 1986, 1989; Leibovici *et al.* 1996; Cervilla *et al.* 1999; Richards *et al.* 2004) do not account for sample selection due to panel attrition, a potentially serious problem in samples of older people. Other studies focus their attention only on pathological aspects or single risk factors (Cervilla *et al.* 1999; Trouton et al. 2006).

A second limitation of this literature is its unidimensional approach (see for example Adam *et al.* 2006). In fact, cognitive ability may be defined as "the capacity to perform higher mental processes of reasoning, remembering, understanding, and problem solving" (Bernstein *et al.* 2006). This definition suggests that cognitive ability is a multi-dimensional concept encompassing a variety of different skills which are probably involved in the aging process in different ways.

A conceptualization of cognitive functioning that involves the concept of modularity in the mental process is the theory of fluid and crystallized abilities introduced by Horn and Cattel (1967) and Salthouse (1985). According to these theories, one can distinguish between two types of abilities. The first type, "fluid abilities", consists of the basic mechanisms of processing information which are closely related to biological and physical factors. One important aspect of these abilities is the speed with which many operations can be executed (Salthouse 1996). The second type, "crystallized abilities", consists of the knowledge acquired during the life with education and other life experiences. Unlike fluid abilities, which are subject to a clear decline as people get older, crystallized abilities tend to be maintained at older ages and are subject to a lower rate of age-related decline. As argued by Salthouse (1985), dimensions of cognitive functioning such as orientation, memory, fluency and numeracy, are generally based on different combinations of fluid and crystallized abilities. This consideration suggests that accounting for the different dimensions of cognitive functioning may be important for the analysis of the organic process of neurodegeneration associated with aging.

The aim of this paper is to investigate the relationship between cognitive performance and aging in Europe, while accounting for a variety of social, behavioral and environmental factors. In particular, we would like to understand which factors help explain the heterogeneity within and between countries in the age-related decline in cognitive functions.

This kind of analysis, new in the economic literature, is made possible by the recent availability of the first two waves (2004, 2006) of the Survey of Health, Aging, and Retirement in Europe (SHARE), a large household panel which contains data on the individual life circumstances of about 30,000 individuals aged 50+ in eleven European countries. In particular, SHARE provides measures of cognitive function based on simple tests of orientation, memory, verbal fluency and numeracy. We focus our attention on cognitive performance in the first wave and on differences in test scores between the two waves. This approach allow us to asses the role of a number of predictors of cognitive ability and of its age-related decline. As with most household panel surveys, attrition is a key problem that we must address, because it may seriously affect the quality of the inference that can be made from the data.

Our results indicate that education plays a fundamental role in explaining heterogeneity in the level of cognitive abilities and in delaying their age-related decline. Not less important, however, is the role of childhood environment, proxied here by body height and parents' occupation. Also other factors accounted for by the literature, such as social and physical activities, predict a delay in the process of neurodegeneration. Finally, we find large and systematic differences in cognitive function across European countries, with people living in Mediterranean countries showing lower test scores in all cognitive domains, even after controlling for observable individual characteristics.

The remainder of this paper is organized as follows. Section 2 describes the data used for this study. Section 3 provides preliminary evidence and examines the attrition problem in the longitudinal sample. Section ref4 presents our main results. Finally Section 5 offers some conclusions.

2 Data

2.1 SHARE

Our data are from the first two waves (2004, 2006) of the Survey of Health, Ageing, and Retirement in Europe (SHARE), a multidisciplinary and cross-national household panel survey coordinated by the Mannheim Research Institute for the Economics of Aging (MEA). We use release 2.0.1 of wave 1 and release 1.0.1 of wave 2. The target population of SHARE consists of individuals aged 50+ and their co-resident partners irrespective of age. The survey has been designed to ensure comparability with the U.S. Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA), two panel surveys with very similar structure.

We use data from the eleven countries that contributed to the 2004 baseline study. They represent different regions of Europe, from Scandinavia (Denmark, Sweden) through Central Europe (Austria, Belgium, France, Germany, Netherlands, Switzerland) to Mediterranean countries (Greece, Italy, Spain).

The SHARE main questionnaire consists of 20 modules, covering demographics, physical and mental health, behavioral risks, health care, employment and pensions, grip strength and walking speed, children, social support, housing, consumption, household income, assets, financial transfers, social and physical activities, and expectations. It also contains measures of cognitive function based on simple tests of orientation in time, memory, verbal fluency and numeracy. All data are collected by face-to-face, computer-aided personal interviews (CAPI), supplemented by a self-completion paper-and-pencil questionnaire. For a detailed description of sampling procedures, questionnaire contents and other detailed information, see Börsch-Supan *et al.* (2005).

Interviews are carried using Blaise, a software for computer-assisted interviewing and survey processing. The Blaise audit trail files, also known as "keystroke files", offer an unprecedented amount of detail on individual interviews because they record for each interview each single keystroke made by the interviewer together with the exact time when it was made. The keystroke files were converted by CentERdata to Stata files containing the time spent on each question asked in the survey. These files enable one not only to compute the duration of each module (as with conventional time-stamps), but also to analyze the time interviewers spend on each single question. This information may be used as a diagnostic tool, to identify possible problems with the respondents or whether the interviewers have not followed the protocol. We use the time spent on each question in a novel way, namely as a measure of processing speed, a second dimension of cognitive abilities evaluation. We refer to the next paragraph for more detail on the use of the keystroke files for this purpose.

We restrict our analysis to individuals aged 50–80 who participated in the first wave. We do not consider people aged 80+ because this sample presents more missing values and is entirely representative of the 80+ population (Börsch-Supan *et al.*, 2005). We consider the main sample and exclude the vignette sample because it was added only in eight countries and its data collection period varied considerably across countries.

We use two working samples, a cross-sectional and a longitudinal one. The cross-sectional sample is obtained by selecting from the first wave all observations with nonmissing values on all the variables of interest. We use this sample for the cross-sectional analysis of heterogeneity in cognitive test scores. The longitudinal sample consists of individuals with completed interviews in both waves. We use this sample to analyze differences in cognitive test scores between the two waves. Table 1 shows the composition of our two working samples by country and gender. As the table makes clear, sample attrition is a key issue: about 8,000 individuals (over one third of the original sample) are lost between the first and the second wave.

2.2 Measures of cognitive ability

Bernstein *et al.* (2006) define cognitive ability as "the capacity to perform higher mental processes of reasoning, remembering, understanding, and problem solving". This definition suggests that cognitive ability is a multidimensional concept which consists of a set of different domains, the most important of which are orientation, memory, executive function (planning, sequencing) and language (Dewey and Prince, 2005). SHARE measures cognitive ability using simple tests of orientation in time, memory, verbal fluency (a test of executive function) and numeracy (arithmetical calculations). These abilities are tested after the "Household demographics", the "Demographics and networks", the "Physical health" and the "Behavioral risks" modules. Thee tests are comparable with similar tests implemented in the HRS and ELSA. An important drawback of SHARE is that exactly the same cognitive tests were repeated in the first two waves of the survey. Repeated exposure to the same tests may cause learning effects, improving the respondents' performance over time.

The test format adopted by SHARE is based on the Telephone Interview of Cognitive Status-Modified (TICS-M) test (Brandt *et al.* 1993), a test format for the assessment of cognitive function that can be administered in person or by telephone and is highly correlated with the Mini-Mental State Exam (MMSE) (Folstein *et al.* 1975), a screening tool frequently used by health-care providers to assess overall brain function. While the MMSE is limited by a ceiling effect, and therefore is relatively insensitive to early evidence of cognitive impairment (de Jager *et al.* 2003), the TICS-M test allows more discrimination in the range of cognitive performance because of its higher item difficulty (for instance 10-word recall, instead of 3-word as in the MMSE).

As for the domains of cognitive ability, the test of orientation in time consists of four questions about the interview date (day, month, year) and day of the week. Although this is a simple and effective test of memory, it is not particular useful for our purposes because it shows very little variability across respondents (almost 90% of the sample answered correctly all four questions, 65% of the errors are about date-day of the month, the residual are equally distributed).

The test of memory consists of verbal registration and recall of a list of 10 items. The respondent hears the list only once. The test is carried out immediately following the encoding phase, and then again after the fluency and numeracy questions. A similar test of memory is used in the HRS and ELSA. The main difference is that SHARE employs only one list, while the other two surveys employ up to four to avoid learning effects, both across individuals during the same household interview and over time for the same individual. As in Zamarro *et al.* (2008), we compute memory scores for this task by considering only the number of target words recalled in the immediate recall phase (score ranging from 0 to 10) to to avoid problems of comparability due to differences in the number and nature of the questions between the immediate and the delayed recall phase.

Verbal fluency is assessed by counting how many distinct elements from a particular category the respondent can name within a specific time interval. The category used in SHARE is members of the animal kingdom, real or mythical, except repetitions or proper nouns. The time interval is one minute.

Numeracy is assessed by asking a few questions that involve simple calculations based on real life situations. Respondents who correctly answer the first question are asked a more difficult one, while those who make a mistake are asked an easier one. The last question is about basic financial knowledge. The resulting "raw" total score ranges from 0 to 4. In addition to the "raw" score, we make use of the keystroke files, and in particular of the time spent to answer by the respondent on each single question. Two arguments may be given to justify this choice. First, as shown by Salthouse (1985), aging is associated with a decrease in the speed at which many cognitive operations can be executed. So using the keystroke files we are also able to capture this characteristic of cognitive deterioration. Second, because it takes only five values, the amount of variability in the "raw" score is low. The time length from the last correct answer may be used to correct for these effects.

Our procedure for including time length is as follow. We first drop observations for which the recorded time spent to provide the last correct answer is too long (more than five minutes). Except for the respondents with score equal to 0 (those who did not answer any question correctly), we then divide the respondents according to the value of their "raw" score (from 1 to 4). For each score group, we divide the respondents according to the quartile distribution of the time spent to provide the last correct answer, generating four additional subgroups for each score group. In this way, we obtain a new score that ranges between 0 and 16.

For more details on the tests of cognitive ability in SHARE see Appendix A.

2.3 Predictors of cognitive abilities

Education is a key predictor of cognitive performance. Because of the specificity of educational institutions in the various countries, the information on education collected by SHARE must be made comparable across countries. To allow comparability, the educational variables is generated by asking a local expert to map the SHARE categories into the ISCED-97 code. In this paper, we use the highest level of education completed, distinguishing between three different aggregated categories: primary education, secondary education, and tertiary education.

Activity status is based on self-reported current economic status of the respondent: employed or self employed (the reference category), unemployed, retired, doing housework, or sick.

The information at the household level includes household size, presence of children in the household, gross household income and residential location (urban or rural). Gross annual household income, in Euro, is adjusted for purchasing power parity and then divided by household size to obtain a real per-capita amount.

As for social activities, we distinguish between doing either voluntary work or caring for a sick or disabled adult, providing help to family, friends and neighbors, attending educational courses, religion or politics.

We control for physical and mental health, in particular for the presence of cardiovascular diseases that may lead to cognitive impairment (Cervilla *et al.* 1999; Meyer *et al.* 1999), by exploiting the self-reported information on diagnosed chronic conditions. We aggregate this information into six categories: high blood pressure, high blood cholesterol, stroke or cerebral vascular disease, diabetes, chronic lung disease, cancer or malignant tumor, and other conditions. We also control for the presence of mobility limitations, physical inactivity, or depression. The latter is measured using the EUROD scale, which considers different symptoms or indicators of depression.

We would like to also control for childhood environment because of its influence on cognitive skill formation (Heckman and Cuhna 2007). Unfortunately, SHARE provides little information on this. The only direct information is the last occupation of the parents of the respondent, classified according to the ISCO88 code. Because of the link between childhood environment and adult height, indirect information is also provided by self-reported body height.

We refer to Appendix A for more details on the variables that we use.

3 Descriptive statistics

In this section we present a preliminary descriptive analysis of our two working samples. The cross-sectional sample is used in Section 3.1 for the cross-sectional analysis of heterogeneity in cognitive test scores. The longitudinal sample is used in Section 3.2 to analyze the differences in cognitive test scores between the two waves. In both cases, we make use of the cross-sectional survey weights provided by SHARE, which are meant to make the distribution of the sample by gender and age in each country and each wave the same as the distribution of the target population.

3.1 The cross-sectional sample

Figure 1 shows the age-profiles of the average test score in the three domains that we consider (recall, fluency and numeracy), separately for men and women. These age-profiles have been obtained by smoothing the original average scores by age and sex using a 3-year centered moving average. The figure shows clear evidence of a decline in average test scores with age. This decline may reflect both the effect of aging and cohort effects. Due to the cross-sectional nature of the data, one cannot distinguish between these two different effects. The figure also shows substantial differences by gender in the outcomes of the various tests. In the recall test, women tend to do better than men, especially at younger ages. By contrast, in the fluency test, men tend to do better than women, especially at older ages. As for numeracy, men tend to score higher than women at all ages. In the case of fluency and recall, the cross-sectional age-profile is steeper for women than for men. In the case of numeracy, it does not vary much by gender except for the level.

Figure 2 shows, for each cognitive domain, the age-profile of average test scores (again, a 3-year centered moving average) by education level. This figure is consistent with the hypothesis that education is a determinant factor of the heterogeneity in cognitive function at older age (Le Carret 2003, Leibovici *et al.* 1996). It is clear from the figure that higher levels of education correspond to better scores in all cognitive tests at all ages. Further, the differences by education level are mainly in the intercept, not in the rate of decline with age. The largest gap is between people who have not completed secondary education and high school graduates. Education seems to play a particularly important role in the numeracy test, where the differences by education level are more marked. By contrast, the differences seems to be much smaller in the case of recall.

Our descriptive analysis strongly supports the view that cognitive ability is multidimensional. It also suggests that the distinction between fluid and crystallized abilities may be relevant. For example, Salthouse (1985) argues that the age-related decline in recall is primarily a function of the age-related decline in "fluid intelligence", most closely linked to the biological and physical processes of aging. By contrast, numeracy appears to be based more on formal education and informal experience ("crystallized intelligence"), and shows less clear evidence of decline at older ages. The behavior of fluency, on the other hand, seems to suggest a combination of fluid and crystallized intelligence.

Figure 3 shows average test scores by macro-region (again, a 3-year centered moving average). Our macro-regions correspond to the classical geographical aggregation into Scandinavia (Denmark and Sweden), Central Europe (Austria, Belgium, France, Germany, Netherlands and Switzerland) and Mediterranean countries (Greece, Italy, Spain). For all cognitive domains, the figure shows large differences in average test scores between Mediterranean countries and the other European countries in SHARE. On the other hand, the differences between Scandinavian and Central European countries are much less marked, especially in the case of numeracy. Theses difference could be due to cross-country differences in the composition of the sample, for example by education, or other observable characteristics.

3.2 The longitudinal sample

The evidence from the longitudinal sample is less clear-cut. The aggregate information in Table 3 shows that mean test scores are higher in the second wave (two years later) for all three measured cognitive skills, a result also noticed by Zamarro *et al.* (2008). This result, which apparently contrasts with the hypothesis of an age-related cognitive decline, could be due to three different factors: composition effects, possible learning effects due to the fact that the cognitive questions are exactly the same in both waves, and nonrandom attrition that selects individual with lower test scores out of the panel.

To control for possible composition effects, Figure 4 plots the age-profile of mean score differences between the two waves, separately for men and women. Because of sampling noise, we smooth the original average scores by age and sex using a 5-year centered moving average. The figure shows that mean score differences are positive at younger ages, but negative at older ages. Further, the age when mean differences become negative differs by cognitive skill. In particular, mean score differences are negative after about age 67 for fluency but only later, after about age 75, for recall and numeracy .

As for learning effects, we cannot exclude the possibility that repeated exposure to the same test may improve performance over time. This learning effect could be due to mere experience with the type of task, or learning with the specific question. If the first effect cannot easily be controlled, the second effect could partially be controlled (as done in ELSA and HRS but not in SHARE) using different but equivalent lists for the memory test.

3.3 Attrition

It is evident from Table 1 that attrition rates vary substantially by country, ranging from about 16% in Greece to about 50% in Germany. Further, in all countries except Spain, attrition rates are higher for men than for women.

Attrition could be due either to the individual decision to leave the panel or to health-related problems, such as death or institutionalization. As for the role of mortality, Table 2 compares average attrition rates by gender and age in SHARE with average mortality rates by gender and age from the life tables provided by Eurostat. We average using weights proportional to the population size of each country. The comparison of the two sets of rates shows that mortality is an important determinant of attrition only at very old ages. Before age 70, mortality cannot explain more than 10 per cent of attrition, and other factors must therefore be considered. One possibility are factors that affect the individual cost of continuing panel participation. In addition to socio-demographic characteristics on the interviewee, these factors may include socio-demographic characteristics of the interviewer and characteristics of the data collection.

The missing-data process operating through attrition could depend on both observable and unobservable factors. This distinction is important, because it leads to different ways of controlling for attrition. Following Rubin (1976), we distinguish between three attrition mechanisms: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). The NMAR assumption allows attrition to depend on unobservable factors, which is the typical case considered in the econometric literature on sample selection models (see Vella 1998 for a review). In this case, ignoring attrition or relying on the MAR assumption may lead to invalid inference about population parameters. The concern about NMAR attrition is especially important here, because those who are lost may more often be those with low cognitive skills.

To better understand which factors help predict nonresponse, and to provide a simple test of the MCAR hypothesis, we estimate a simple probit model for the probability of response in the second wave. The model uses as predictors the available set of socio-demographic variables, health related variables, and cognitive tests scores in the first wave. We also include sex, age and education of the interviewer, and the respondent's willingness to participate to the survey, as assessed by the interviewer. Asymptotic standard errors are robust to the presence of heteroskedasticity and intra-household clustering effects.

The estimation results, presented in Table 4, clearly reject the MCAR hypothesis as people in poor health and with poor cognitive abilities are more likely to drop out from the panel. These results, also noted by Zamarro *et al.* (2008)), confirm the findings in the health economics literature (Contoyannis, Jones and Rice 2004; and Jones, Koolman and Rice 2005) of an inverse relationship between attrition and health. On the other hand, socio-economic factors seem to help little in predicting attrition. Two other sets of factors also seem to matter. One is factors related to the interview process, such as the respondent's willingness to answer (as assessed by the interviewer) and the socio-demographic characteristics of the interviewers. In particular, the probability of attrition has a quadratic relationship with the interviewer's age (initially negative but positive at older age), and it is negative related with interviewers with a college degree. The other is test scores in the first wave. In particular, the probability of attrition appears to be strongly negatively related to fluency and recall scores, while its relationship with numeracy is weaker. These results suggest the need of avoiding strong assumptions, such as exogeneity of attrition with respect to cognitive and health outcomes.

We would like to stress the large degree of variation across countries. Other things being equal, all countries except Greece, Belgium and possibly Denmark report significantly higher levels of attrition than Italy (the reference country). The difference in attrition rates are especially sizeable for Germany and Spain.

4 Empirical results

In this section we use the two available waves of SHARE to investigate both the variability of test scores in the first (2004) wave and the differences in test scores over the two-year period between the first and the second (2006) wave.

4.1 Cross-sectional heterogeneity

Since all our outcome variables are categorical in nature, an ordered probit model would be the most appropriate. However, because the number of categories is fairly large, this model and the simpler and more intuitive linear regression model tend to produce very similar results. For this reason we only report the evidence from the latter model. To facilitate comparisons, we standardize all test scores by subtracting off their mean and dividing by their standard deviation.

After some experimentation, we include in our preferred specifications the same set of regressors, namely socio-demographic variables (a third-order polynomial in age and indicators for educational attainments, working status, and marital status), household characteristics (household size, the log of per-capita household income, and indicators for the presence of children and urban location), physical and mental health variables (log of body height and body weight, and indicators for diabetes, lung disease, stroke, high blood pressure, cancer, physical inactivity, mobility limitation, and the Euro-D depression scale), interview characteristics (indicators for the willingness to participate to the survey), information on social activities (indicators for the different activities) and on parents' last occupation (indicators for a white-collar father and for a blue- or a white-collar mother), plus a full set of country indicators. The reference category for each model is a person aged 65, Italian, with a high-school degree, retired, married, in a two-person household with no children present, located in a small town, with per-capita household income of 20,000 Euro, a body height of 169 cm, a body weight of 74 Kg, without chronic conditions, not involved in any social activity, with a blue-collar father and a homemaker mother.

Table 5 reports the estimated coefficients for each of our three cognitive domains, separately for men and women. Asymptotic standard errors are robust to the presence of heteroskedasticity and intra-household clustering effects.

The effects of age and education are statistically significant and have the expected sign, namely negative for age and positive for education. These results confirm the findings in the literature stressing the key role of education in explaining heterogeneity in the level of cognitive function in old ages (Le Carret 2003; Leibovici *et al.* 1996). A closer inspection of the estimated coefficients shows that numeracy scores seem to be less negatively affected by aging but more sensitive to education. By contrast, scores in the recall tests seem to be the most affected by aging. These results are consistent with the multi-dimensional view of cognitive ability, and with the theories that suggest a distinction between fluid and crystallized abilities (Horn and Cattel 1967; Salthouse 1999).

As for the other covariates, occupational status, activity, and health conditions appear to be important predictors of individual differences in test scores. Body height also appears to be a strong predictor of cognitive scores, consistently with similar findings presented by Case and Paxson (2008) using HRS data. Another strong predictor is the information on the parents' last occupation, in particular whether the father was a white collar. The role played by body height and by parents' occupation is in line with the recent literature that considers the individual's childhood environment as a critical factor in explaining cognitive skill formation (Knudsen *et al.* 2006; Cunha and Heckman 2007).

Not less important, however, are the country effects. These could reflect institutional effects or other effects not included in our set of covariates. In particular, the cross-country differences shown in Section 3.1 between Mediterranean countries and other Europeans countries seems to persist even after controlling for our broad set of covariates.

Separate estimates by country, reported in Appendix B, confirm the presence of substantial cross-country heterogeneity in the effects of age and education for all the cognitive domains.

4.2 Differences in test scores and attrition

Using a linear regression model for the differences in test scores may be criticized on two grounds. First, differences of categorical variables are still categorical. The number of categories, however, is not small (from 20 in the case of recall to nearly 100 in the case of fluency), which again justifies treating test score differences as a continuous variable. Second, and most importantly, panel attrition may be NMAR.

To control for selection due to NMAR attrition, we adopt the parametric 2-step procedure originally proposed by Heckman (1979), based on the normality assumption. The first step of the procedure involves estimating the probability of continued panel participation. This is what we already did in Section 3.3. The second step corrects for selection due to NMAR attrition by inserting into the linear model an additional regressor (the "inverse Mills ratio") generated using the estimates from the first step.

The normality assumption is strong, but could be weakened using a number of semi-parametric alternatives. We do not pursue this here, since we are more concerned with identification of the coefficients on the covariates that may help explain the differences in test scores. Usually, strong identification of these parameters requires exclusion restrictions, namely forcing some of the variables that enter the model for panel attrition out of the model for the conditional mean of the score differences. These exclusion restrictions should be valid, that is, it should be plausible to assume that the excluded variables affect only the probability of panel attrition (Vella 1998). In general, finding a valid set of exclusion restrictions is not easy. Following Nicoletti and Peracchi (2005), we argue that a valid set consists of variables describing the characteristics of the data collection process in the first wave (sex, age and education of the interviewer) since they help predict panel attrition in the second wave but are unlikely to have causal effects on the outcome of interest.

In addition to the inverse Mills ratio, our second-step regression includes as covariates socio-demographic variables (a third-order polynomial in age and indicators for educational attainments, working status, and marital status), household characteristics (household size, log of per-capita household income and indicators for the presence of children and urban location), physical and mental health variables (log of body height and body weight, and indicators for diabetes, lung disease, stroke, high blood pressure, cancer, physical inactivity, mobility limitation, and the Euro-D depression scale), the test scores in the first wave, some interview characteristics (the willingness to participate to the survey, as reported by the interviewer), information on social activities (indicators for the different activities) and on parents' last occupations (indicators for a white-collar father and for a blue- or a white-collar mother), plus a full set of country indicators.

Table 6 shows, for each of our three cognitive domains, the estimates from a simple OLS regression (first column) and those obtained from the second step of Heckman's procedure (second column), which simply adds a selection-correction term (the inverse Mills ratio). Following standard practice to ensure comparability across domains, test score differences have been standardized to give a mean of zero and a standard deviation of one. The standard error for the OLS are robust to the presence of heteroskedasticity.

The intercept in the model represents the mean score difference for the baseline individual (essentially the same as in Section 4.1). In the case of OLS, this intercept is negative for fluency and recall, but positive for numeracy. After introducing the selection-correction term, the intercept becomes more negative for fluency and recall but remains positive and statistically significant at the 1% level for numeracy. Interestingly, the negative coefficients on the indicators for Greece and Spain in recall and fluency suggest a stronger cognitive decline for Mediterranean countries.

A t-test of significance of the selection-correction term (labeled as `mills` in the table) provides a simple test of the MAR assumption. This assumption cannot be rejected at conventional significance levels for numeracy, but is rejected for fluency and recall (at the 5% level for recall and at the 1%

level for fluency). In all cases, introducing the selection-correction term changes the size and significance of several coefficients with respect to OLS.

Finally, we use the estimated coefficients to construct the curves in Figure 5. The figure plots the predicted mean score differences for our baseline individual by age and education level. The figure shows a clear positive effect of education for all cognitive domains. However, the magnitude of the effect is very different across domain. In particular, numeracy seems to be affected most, while recall seems to be affected least. This result is consistent with both the available literature and the cross-sectional evidence.

5 Conclusions

In this paper we investigate the relation between aging, cognitive functions and social, behavioral and environmental factors, in particular education, using the first two waves of SHARE, which has the unique feature of providing measures of cognitive functions for a representative sample of people aged 50+ in eleven European countries.

Our results indicate that cognitive abilities are better treated as a multidimensional concept. They also show that education plays a fundamental role in explaining heterogeneity in the level of cognitive abilities and in delaying their age-related decline associated with the natural process of neurodegeneration. However, not less important is the role of social and physical activity and of home environment in early life, proxied here by body height and parents' last occupation.

Sample attrition is an important feature of the data, leading individual with lower cognitive scores in the first wave to drop out of the panel. As a consequence, our results confirms the presence of selection bias in the OLS estimates, compared with the corresponding results of the Heckman two step procedure.

Finally, we find large and systematic differences in measured cognitive function across European countries, with people living in Mediterranean countries showing lower test scores in all cognitive domains and evidence of a stronger decline over a two-year period.

References

Adam S., Bay C., Bonsang E., Germain S., and Perelman S. (2006), "Occupational activities and cognitive reserve: a frontier approach applied to the Survey of Health, Ageing and Retirement in Europe (SHARE)", CREPP WP 2006/05, University of Liege.

Banks J., and Oldfield Z. (2007), "Understanding pensions: cognitive functions, numerical ability and retirement saving". *Fiscal Studies*, 28: 143–170.

Bernstein A., Penner L. A., Clarke-Stewart A., and Roy E. J. (2006), *Psychology* (6th ed.). Houghton Mifflin, Boston.

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

Brandt J., Welsh K. A., Breitner J.C., et al. (1993), "Hereditary influences on cognitive functioning in older men: a study of 4000 twin pairs". *Archives of Neurology*, 50: 599-603.

Case A., and Paxson C. H. (2008), "Height, health and economic development". *American Economic Review*, 98: 463-467.

Cervilla J. A., Prince M., Joels S., Lovestone S., and Mann A. (2000), "Long-term predictors of cognitive outcome in a cohort of older people with hypertension". *British Journal of Psychiatry*, 177: 66–71.

Cervilla J. A., Prince M., and Mann A. (2000), "Smoking, drinking, and incident cognitive impairment: a cohort community based study included in the Gospel Oak project". *Journal of Neurology, Neurosurgery, and Psychiatry*, 68: 622–626.

Christelis D., Jappelli T., and Padula M. (2006), "Cognitive abilities and portfolio choice". CEPR Discussion Papers 5735, London.

Contoyannis P., Jones A. M., and Rice A. N. (2004), "The dynamics of health in the British Household Survey". *Journal Of Applied Econometrics*, 19: 473503.

Cunha F., and Heckman J. (2007), "The Technology of skill formation". *American Economic Review*, 97: 31–47.

de Jager C. A., Budge M. M., and Clarke R. (2003), "Utility of TICS-M for the assessment of cognitive function in older adults". *International Journal of Geriatric Psychiatry*, 18: 318-324.

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

Ferri C. P., Prince M., Brayne C., Brodaty H., Fratiglioni L., Ganguli M., Hall K., Hasegawa K., Hendrie H., Huang Y., Jorm A., Mathers C., Menezes P. R., Rimmer E., and Scazufca M. (2005), "Global prevalence of dementia: a Delphi consensus study". *Lancet*, 366: 2112–2117.

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

Heckman J. (1979), "Sample selection bias as a specification error". *Econometrica*, 47, 153–161.

Horn J. L., and Cattell R. B. (1967), “Age differences in fluid and crystallized intelligence”. *Acta Psychologica*, 26, 107-129

Kaplan G. A., Turrell G., Lynch J. W., Everson S. A., Helkala E., and Salonen J. T. (2001), “Childhood socioeconomic position and cognitive function in adulthood”. *International Journal of Epidemiology*, 30: 256–263.

Knudsen E., Heckman J., Cameron J., and Shonkoff J. P. (2006), ” Economic, neurobiological and behavioral perspectives on building America’s future workforce”. *World Economics, NTC Economic & Financial Publishing*, Oxfordshire, United Kingdom.

Jones A. M., Koolman X., and Rice N. (2005), “Health-related non-response in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models”. Health, Econometrics and Data Group (HEDG) Working Papers 05/05, Department of Economics, University of York.

Jönsson L., and Berr C., (2005), “Cost of dementia in Europe”. *European Journal of Neurology*, 12: 50-53.

Le Carret N., Lafont S., Letenneur L., Dartigues F., Mayo W., and Fabrigoule C. (2003), “The effect of education on cognitive performances and its implication for the constitution of the cognitive reserve”. *Developmental Neuropsychology*, 23: 317–337.

Leibovici D., Ritchie K., Ledesert B., and Touchon D. (1996), “Does education level determine the course of cognitive decline?”. *Age and Ageing*, 25: 392–397.

Maraldi C., and Pahor M. (2006), “Physical exercise, cognitive decline and dementia in older adults “. *Giornale di Gerontologia*, 54: 186–194.

Meyer J. S., Rauch G. M., Crawford K., Rauch R. A., Konno S., Akiyama H., Terayama Y., and Haque A. (1999), “Risk factors accelerating cerebral degenerative changes, cognitive decline, and dementia”. *International Journal of Geriatric Psychiatry*, 14: 1050–1061.

Nicoletti C., Peracchi F. (2005), Survey response and survey characteristics: Microlevel evidence from the European Community Household Panel, *Journal of Royal Statistical Society A*, 168: 1–19.

Park D. (1999), “Cognitive aging, processing resources, and self-report”. In Schwarz N., Park D.C., Knauper B., and Sudman S. (Eds.), *Cognition, Aging, and Self-reports*, Psychology Press, Philadelphia.

Richards M., Shipley B., Furher R., and Wadsworth M. E. (2004), “Cognitive ability in childhood and cognitive decline in mid-life: longitudinal birth cohort study”. *British Medical Journal*, doi:10.1136/bmj.37972.513819.EE.

Rubin D. B. (1976), “Inference and missing data”. *Biometrika*, 63, 581–592.

Salthouse T. A. (1985), *A Theory of Cognitive Aging.* Amsterdam, North-Holland.

Salthouse T. A. (1996), “The processing-speed theory of adult age differences in cognition”, *Psychological Review*, 103: 403–428.

Schaie K. W. (1986), “Individual differences in rate of cognitive change in adulthood”. Paper presented at a Convocation: The Power of Years in honor of James E. Birren, University of Southern California, Los Angeles, CA.

Schaie K. W. (1989), “The hazards of cognitive aging”. *The Gerontologist*, 29: 484–493.

Stern Y. (2002), “What is cognitive reserve? Theory and research application of the reserve concept”. *Journal of the International Neuropsychological Society*, 8: 448–460.

Stern Y. (2003), “The concept of cognitive reserve: A catalyst for research”. *Journal of Clinical and Experimental Neuropsychology*, 25: 589–593.

Trouton A., Stewart R., and Prince M. (2006), “Does social activity influence the accuracy of subjective memory deficit? Findings from a British community survey”. *Journal of the American Geriatrics Society*, 54: 1108–1113.

Vella F. (1998), “Estimating models with sample selection bias: A survey”. *Journal of Human Resources*, 33: 127–169.

Zamarro G., Meijer E., and Fernandes M. (2008), “Mental Health and Cognitive Ability”. In Börsch-Supan A., Brugiavini A., Jürges H., Kapteyn A., Mackenbach J., Siegrist J., and Weber G. (eds.), *First Results from the Survey of Health, Ageing and Retirement in Europe (2004-2007)*. Mannheim Research Institute for the Economics of Aging (MEA).

Table 1: Samples sizes by country and sex.

	Cross-sectional		Longitudinal		Attrition rate	
Country	Male	Female	Male	Female	Male	Female
Austria	722	936	467	617	.3531	.3408
Belgium	1339	1468	930	1046	.3054	.2874
Denmark	688	751	507	541	.2630	.2796
France	831	965	462	576	.4440	.4031
Germany	1002	1075	487	539	.5139	.4986
Greece	810	952	650	786	.1975	.1743
Italy	868	1041	573	707	.3398	.3208
Netherlands	980	1116	511	636	.4785	.4301
Spain	688	933	387	527	.4375	.4351
Sweden	1078	1221	679	785	.3701	.3570
Switzerland	412	446	285	321	.3082	.2802
Total	9418	10904	5938	7081	.3695	.3506

Table 2: Mortality and attrition rates.

	Attrition		Mortality	
Age	Male	Female	Male	Female
50	0.383	0.362	0.010	0.005
51	0.346	0.349	0.011	0.006
52	0.365	0.362	0.012	0.006
53	0.377	0.337	0.013	0.007
54	0.389	0.408	0.014	0.007
55	0.337	0.343	0.015	0.008
56	0.427	0.336	0.016	0.008
57	0.336	0.334	0.017	0.009
58	0.353	0.300	0.019	0.010
59	0.385	0.357	0.021	0.011
60	0.397	0.362	0.022	0.011
61	0.355	0.328	0.024	0.012
62	0.323	0.344	0.027	0.013
63	0.385	0.325	0.029	0.015
64	0.344	0.294	0.032	0.016
65	0.316	0.321	0.035	0.017
66	0.354	0.294	0.038	0.019
67	0.331	0.334	0.043	0.021
68	0.366	0.380	0.047	0.024
69	0.339	0.360	0.052	0.027
70	0.345	0.364	0.057	0.030
71	0.397	0.373	0.064	0.034
72	0.377	0.386	0.071	0.038
73	0.399	0.354	0.077	0.042
74	0.385	0.399	0.085	0.048
75	0.359	0.347	0.095	0.055
76	0.420	0.346	0.105	0.062
77	0.414	0.340	0.116	0.070
78	0.378	0.435	0.129	0.080
79	0.518	0.425	0.144	0.090
80	0.410	0.401	0.160	0.104
Total	0.369	0.351	0.051	0.029

Table 3: Summary statistics. Final samples.

		Cross-sectional			Longitudinal		
Variable	Description	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
recall1	Recall score in Wave1	20322	4.93	1.79	13019	5.03	1.72
fluency1	Fluency score in Wave1	20322	19.12	7.35	13019	19.48	7.15
numeracy1	Numeracy score in Wav1	20322	8.17	4.42	13019	8.30	4.3
recall2	Recall score in Wave2				13019	5.12	1.74
fluency2	Fluency score in Wave2				13019	19.5	7.35
numeracy2	Numeracy score in Wave2				13019	8.51	4.38
age	Age	20322	63.0	8.4	13019	62.9	8.3
educ1	Dropt High School (HS)	20322	0.497		13019	0.493	
educ3	College degree	20322	0.195		13019	0.202	
sick	Permanently sick	20304	0.034		13017	0.03	
unemp	Unemployed	20304	0.034		13017	0.034	
emp	Employed	20304	0.3		13017	0.302	
homemaker	Home maker	20304	0.154		13017	0.152	
nospouse	Without spouse	20322	0.255		13019	0.261	
hsize	Household size	20322	2.20	0.97	13019	2.20	0.98
child_in	Children in the household	20322	0.37	0.75	13019	0.38	0.76
urban	Urban area	20322	0.476		13019	0.47	
will	Willing to answer	20322	0.897		13019	0.927	
pc_income	Per capita income	20322	21703	23483	13019	21316	22325
height	Body height	20158	168.3	9.0	12943	168.1	8.9
weight	Body weight	20131	75.1	14.2	12934	75.2	14.2
diabetes	Diabetes	20322	0.093		13019	0.088	
lung_disease	Lung disease	20322	0.049		13019	0.049	
stroke	Stroke	20322	0.033		13019	0.031	
high_pressure	High pressure	20322	0.313		13019	0.319	
cancer	Cancer	20322	0.052		13019	0.048	
mobility	Mobility limitations	20315	1.28	1.95	13016	1.23	1.87
phactiv	Physical inactivity	20314	0.079		13017	0.066	
eurod	Euro-D depression scale	20203	2.21	2.19	12967	2.20	2.14
voluntary	Voluntary	20283	0.131		13015	0.142	
cared	Cared sick	20283	0.066		13015	0.07	
helpfam	Help family	20283	0.25		13015	0.267	
training	Training course	20283	0.07		13015	0.076	
club	Club or sport member	20283	0.197		13015	0.204	
religious	Religiuos partecipation	20283	0.116		13015	0.127	
political	Political activity	20283	0.045		13015	0.051	
f_whitecollar	White collar father	19269	0.22		12390	0.223	
m_whitecollar	White collar mother	19285	0.064		12404	0.066	
m_bluecollar	Blue collar mother	19285	0.422		12404	0.441	
iv_age	Interviewer's age	19781	43.7	13.1	12713	43.1	13.4
iv_female	Female interviewer	19820	0.643		12737	0.646	
iv_edu_1	Interviewer without HS	19647	0.104		12617	0.097	
iv_edu_3	Interviewer with college degree	19647	0.457		12617	0.472	

Table 4: Probit estimates for the probability of partecipate to the second wave.

Variable	Pooled	Male	Female
female	0.036		
age1	0.002	0.007	-0.001
age2	-0.062 ***	-0.074 **	-0.051 *
age3	0.000	-0.001	0.002
educ1	0.035	0.030	0.047
educ3	0.005	-0.019	0.036
sick	-0.047	0.017	-0.099
unemp	0.010	0.083	-0.061
emp	-0.026	-0.026	-0.028
homemaker	-0.051	-0.499 *	-0.051
nospouse	-0.045	-0.088	-0.007
hsize	-0.124 ***	-0.136 **	-0.109 *
child_in	0.170 ***	0.201 ***	0.135 **
lpc_income	-0.023	-0.015	-0.028
urban	-0.077 **	-0.065 *	-0.087 **
will	0.427 ***	0.447 ***	0.409 ***
at	-0.172 **	-0.124	-0.210 **
de	-0.579 ***	-0.542 ***	-0.601 ***
sw	-0.246 ***	-0.192 *	-0.290 ***
nl	-0.431 ***	-0.444 ***	-0.418 ***
es	-0.284 ***	-0.276 ***	-0.280 ***
fr	-0.294 ***	-0.329 ***	-0.263 ***
dk	0.033	0.104	-0.029
gr	0.478 ***	0.494 ***	0.474 ***
ch	-0.058	-0.040	-0.065
be	-0.055	-0.050	-0.053
recall1	0.013 ***	0.011	0.015 **
fluency1	0.012 ***	0.009 ***	0.013 ***
Numer1	-0.063	0.020	-0.107
Numer2	-0.048	-0.010	-0.073
Numer4	0.042	0.059	0.033
Numer5	-0.045	0.022	-0.127 **
lnh	-1.177 ***	-1.333 **	-1.171 **
lnw	0.309 ***	0.437 ***	0.223 *
diabetes	-0.080 *	-0.074	-0.084
lung_disease	0.058	0.093	0.035
stroke	0.043	0.048	0.054
high_pressure	0.051 *	0.031	0.069 *
cancer	-0.152 ***	-0.272 ***	-0.067
mobility	-0.009	-0.029 **	0.002
phinactiv	-0.147 ***	-0.193 **	-0.116 *
eurod1	0.017 **	0.019 *	0.016 *
voluntary	0.112 ***	0.078	0.142 **
cared	0.003	0.091	-0.051
helpfam	0.074 **	0.085 *	0.060
training	0.029	0.033	0.023
club	0.059 *	0.065	0.050
religious	0.023	0.077	-0.014
political	0.076	0.132 *	-0.036
f_whitecollar	0.007	-0.012	0.027
m_whitecollar	0.029	0.080	-0.014
m_bluecollar	0.012	-0.008	0.030
iv_age1	0.019 **	0.024 **	0.013
iv_age2	-0.020 **	-0.026 **	-0.015
iv_female	-0.025	-0.017	-0.034
iv_edu1	-0.012	-0.014	-0.014
iv_edu3	0.096 ***	0.075 *	0.113 ***
cons	0.612 ***	0.712 **	0.524 *
N	17596	8243	9353
R^2_p	0.06	0.06	0.06
ll	-10761	-5075	-5660

*: p-value $< .05$; **: p-value $< .01$; ***: p-value $< .001$

Table 5: Test scores in the first wave by sex.

	Recall		Fluency		Numeracy	
	Male	Female	Male	Female	Male	Female
age1	-0.025 ***	-0.026 ***	-0.010 ***	-0.018 ***	-0.008 **	-0.015 ***
age2	-0.039 **	-0.067 ***	-0.044 ***	-0.041 ***	-0.057 ***	-0.018
age3	0.001	-0.001	-0.002	0.000	-0.000	0.001
educ1	-0.258 ***	-0.258 ***	-0.171 ***	-0.222 ***	-0.349 ***	-0.353 ***
educ3	0.172 ***	0.130 ***	0.218 ***	0.188 ***	0.213 ***	0.190 ***
sick	-0.073	-0.067	-0.121 *	-0.088	-0.007	-0.148 **
unemp	0.031	-0.016	-0.024	-0.052	-0.007	-0.001
emp	0.044	0.001	0.048	0.067 *	0.089 ***	0.003
homemaker	0.209	-0.039	-0.047	-0.017	-0.155	-0.105 ***
nospouse	-0.089 *	-0.099 ***	-0.028	-0.030	-0.063	-0.098 ***
hsize	-0.029	-0.000	-0.020	-0.007	-0.021	-0.069 *
child_in	0.017	-0.012	0.029	-0.005	0.052	0.074 *
lpc_income	0.045 ***	0.027 ***	0.050 ***	0.055 ***	0.065 ***	0.048 ***
urban	0.052 **	0.044 *	-0.034	-0.037 *	0.043 *	0.041 *
will	0.262 ***	0.305 ***	0.211 ***	0.269 ***	0.176 ***	0.200 ***
at	0.237 ***	0.382 ***	0.831 ***	0.910 ***	0.325 ***	0.558 ***
de	0.301 ***	0.472 ***	0.572 ***	0.587 ***	0.227 ***	0.339 ***
sw	0.205 ***	0.491 ***	0.905 ***	0.975 ***	0.256 ***	0.223 ***
nl	0.139 ***	0.442 ***	0.374 ***	0.526 ***	0.351 ***	0.308 ***
es	-0.190 ***	-0.188 ***	0.257 ***	0.273 ***	-0.174 ***	-0.214 ***
fr	-0.024	0.152 ***	0.601 ***	0.524 ***	0.026	-0.017
dk	0.152 ***	0.490 ***	0.617 ***	0.705 ***	0.009	0.004
gr	0.247 ***	0.211 ***	0.016	0.006	0.282 ***	0.176 ***
ch	0.297 ***	0.559 ***	0.420 ***	0.623 ***	0.259 ***	0.414 ***
be	0.078 *	0.289 ***	0.496 ***	0.544 ***	0.111 ***	0.055
lnh	1.383 ***	0.688 **	1.775 ***	1.090 ***	2.067 ***	1.873 ***
lnw	-0.183 **	-0.074	-0.107	-0.048	-0.043	-0.045
diabetes	-0.045	-0.049	-0.053	-0.062 *	-0.052	-0.090 ***
lung_disease	0.080	-0.025	0.098 *	-0.003	-0.007	0.008
stroke	-0.199 ***	-0.146 *	-0.189 ***	-0.134 **	-0.080	-0.107 *
high_pressure	0.043 *	0.027	0.000	-0.014	-0.025	-0.005
mobility	-0.012	-0.018 ***	-0.033 ***	-0.018 ***	-0.031 ***	-0.023 ***
phactiv	-0.242 ***	-0.193 ***	-0.238 ***	-0.208 ***	-0.023	-0.037
eurod	-0.037 ***	-0.032 ***	-0.019 ***	-0.009 *	-0.047 ***	-0.036 ***
cancer	0.099 *	0.086 *	0.024	0.067 *	0.057	0.069 *
voluntary	0.017	0.096 ***	0.073 *	0.111 ***	0.050	0.067 *
cared	0.015	0.003	0.104 *	0.064 *	0.001	0.010
helpfam	0.047 *	0.082 ***	0.080 ***	0.139 ***	0.026	0.031
training	0.125 ***	0.046	0.076	0.103 ***	0.054	0.065 *
club	0.078 ***	0.053 *	0.117 ***	0.107 ***	0.089 ***	0.069 ***
religious	0.031	0.053 *	-0.018	0.012	-0.035	0.017
political	0.048	0.132 **	0.117 ***	0.046	0.129 ***	0.067
f_whitecollar	0.106 ***	0.080 ***	0.104 ***	0.099 ***	0.117 ***	0.090 ***
m_whitecollar	0.060	0.039	0.136 ***	0.069	0.050	0.021
m_bluecollar	0.012	-0.001	-0.018	0.002	0.006	-0.006
cons	-0.384 ***	-0.281 ***	-0.622 ***	-0.644 ***	0.005	-0.175 ***
N	8522	9707	8522	9707	8522	9707
R^2	0.25	0.3	0.29	0.36	0.27	0.28
ll	-10481	-12080	-10565	-11489	-10299	-11989

*: p-value < .05; **: p-value < .01; ***: p-value < .001

Table 6: Differences in test scores between waves.

	Recall		Fluency		Numeracy	
	Heckman correction		Heckman correction		Heckman correction	
	No	Yes	No	Yes	No	Yes
female	0.182 ***	0.183 ***	0.010	0.006	-0.162 ***	-0.166 ***
age1	-0.018 ***	-0.017 ***	-0.012 ***	-0.011 ***	-0.013 ***	-0.013 ***
age2	-0.033 *	-0.050 **	-0.036 **	-0.058 ***	-0.000	0.010
educ1	-0.042	-0.035	-0.092 ***	-0.079 ***	-0.179 ***	-0.177 ***
educ3	0.064 **	0.067 *	0.091 ***	0.103 ***	0.084 ***	0.088 ***
sick1	-0.012	-0.003	-0.055	-0.062	-0.087	-0.094
unemp	0.003	0.003	-0.016	-0.029	-0.111 *	-0.112 *
emp	0.014	0.014	0.029	0.018	-0.019	-0.011
homemaker	-0.024	-0.031	0.022	0.003	-0.019	-0.018
nospouse	0.031	0.020	-0.023	-0.040	-0.021	-0.019
hsize	0.088 **	0.056	0.014	-0.032	0.025	0.042
child_in	-0.085 *	-0.040	-0.033	0.031	-0.003	-0.025
lpc_income	0.010	0.002	0.014	0.005	0.031 **	0.035 ***
urban	0.059 **	0.040	0.027	0.003	-0.024	-0.014
will	0.060	0.171 *	0.120 ***	0.288 ***	0.056	-0.006
at	0.122 **	0.080	0.246 ***	0.189 ***	0.198 ***	0.231 ***
de	0.036	-0.111	0.303 ***	0.111	0.173 ***	0.278 ***
sw	0.112 *	0.061	0.401 ***	0.331 ***	0.129 **	0.169 **
nl	0.115 **	-0.007	0.175 ***	0.020	0.239 ***	0.318 ***
es	-0.119 **	-0.188 ***	-0.068	-0.164 **	-0.118 **	-0.072
fr	-0.094	-0.165 **	0.245 ***	0.146 *	0.011	0.053
dk	0.188 ***	0.197 ***	0.250 ***	0.265 ***	0.070	0.061
gr	-0.048	0.045	-0.146 ***	-0.024	0.331 ***	0.265 ***
ch	0.106 *	0.061	0.260 ***	0.236 ***	0.333 ***	0.330 ***
be	0.036	0.026	0.278 ***	0.263 ***	0.049	0.055
recall1	-0.167 ***	-0.164 ***	0.031 ***	0.036 ***	0.018 ***	0.017 ***
fluency1	0.013 ***	0.016 ***	-0.089 ***	-0.085 ***	0.009 ***	0.007 ***
Numer1	-0.163 ***	-0.180 ***	-0.126 ***	-0.158 ***	0.925 ***	0.933 ***
Numer2	-0.043	-0.050	-0.050 *	-0.072 *	0.627 ***	0.633 ***
Numer4	0.075 ***	0.091 ***	0.081 ***	0.091 ***	-0.525 ***	-0.532 ***
Numer5	0.111 ***	0.102 ***	0.135 ***	0.114 ***	-1.130 ***	-1.125 ***
lnh	-0.026	-0.443	0.384	-0.197	0.597 *	0.783 **
lnw	-0.041	0.044	-0.109 *	0.003	-0.101	-0.164 *
diabetes	-0.065 *	-0.092 **	-0.076 **	-0.099 **	-0.018	0.001
lung_disease	0.031	0.049	0.024	0.039	-0.015	-0.026
stroke	-0.060	-0.047	-0.013	0.014	-0.076	-0.080
high_pressure	0.001	0.014	-0.032	-0.015	0.033	0.022
cancer	-0.008	-0.049	-0.018	-0.085	0.076	0.105 *
mobility	-0.009	-0.013 *	-0.002	-0.006	-0.018 ***	-0.017 **
phinactiv	-0.012	-0.053	-0.094 **	-0.144 ***	-0.016	0.012
eurod	-0.016 ***	-0.013 *	-0.004	0.002	-0.010 *	-0.012 *
voluntary	0.097 ***	0.125 ***	0.047	0.088 **	0.028	0.011
cared	-0.001	-0.016	0.094 **	0.090 *	-0.016	-0.019
helpfam	0.009	0.027	0.049 *	0.067 **	0.034	0.019
training	0.035	0.042	0.064	0.073 *	-0.000	0.003
club	0.061 **	0.076 **	0.069 **	0.095 ***	0.026	0.024
religious	-0.028	-0.023	0.025	0.038	-0.096 ***	-0.104 ***
political	-0.007	0.009	0.040	0.080	0.082 *	0.064
f_whitecollar	0.056 **	0.058 *	0.071 ***	0.071 **	0.033	0.028
m_whitecollar	0.023	0.029	0.117 **	0.128 ***	0.031	0.024
m_bluecollar	0.019	0.019	0.027	0.033	0.002	0.002
mills		0.479 *		0.656 **		-0.323
cons	-0.474 ***	-0.809 ***	-0.359 ***	-0.829 ***	0.280 ***	0.491 **
N	11832	11438	11832	11438	11832	11438
R^2	0.20	0.21	0.25	0.25	0.24	0.25
ll	-15364	-14871	-14946	-14426	-15114	-14634

*: p-value < .05; **: p-value < .01; ***: p-value < .001

Figure 1: Mean test scores by sex and age.

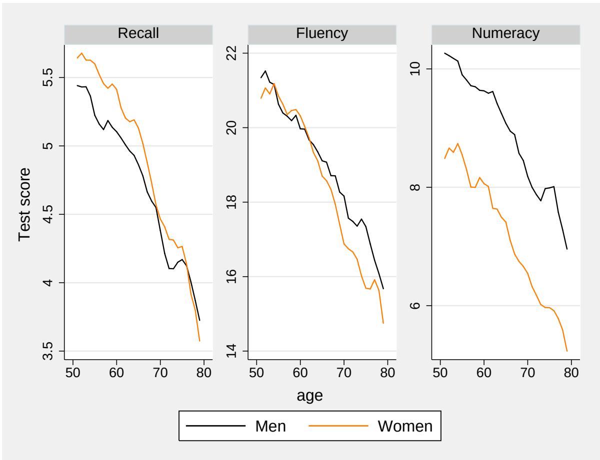


Figure 2: Mean test scores by age and education level.

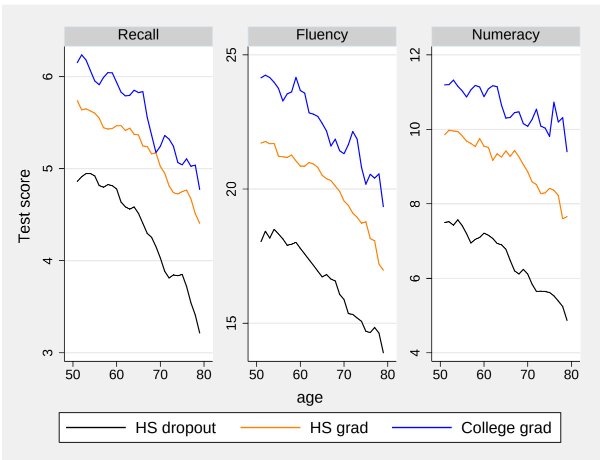


Figure 3: Mean test scores by country and age.

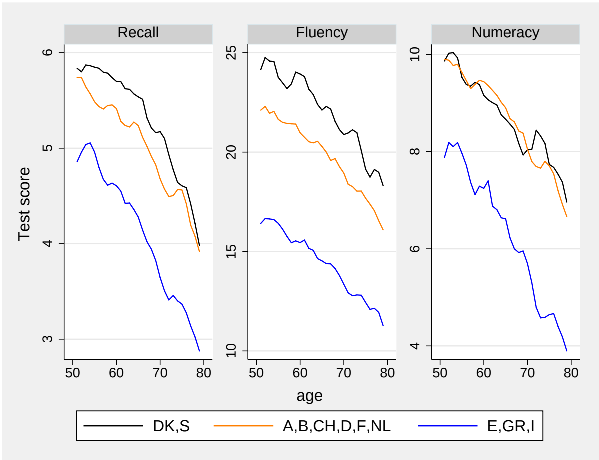


Figure 4: Mean differences in test scores (wave2-wave1) by sex and age.

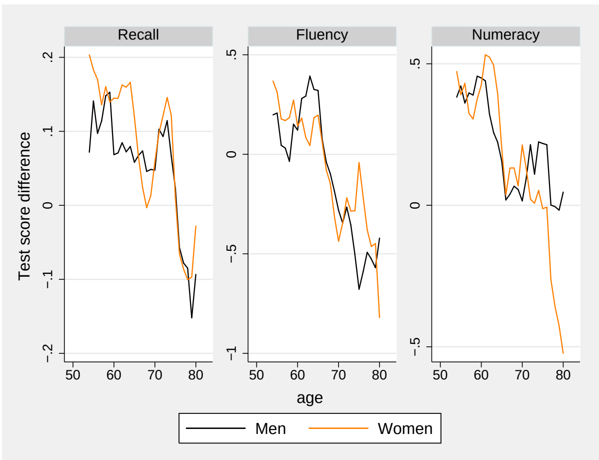
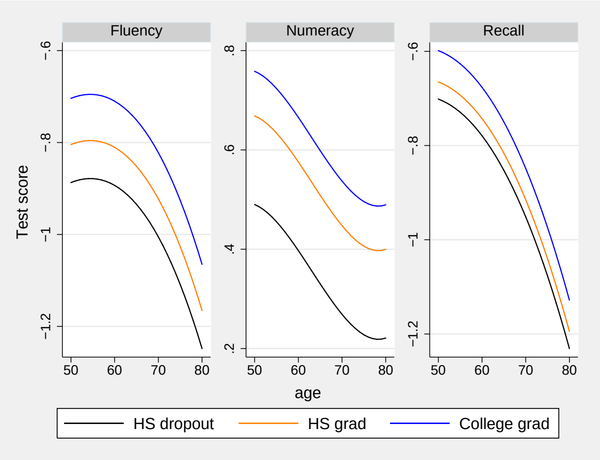


Figure 5: Predicted differences in test scores (wave2-wave1) by age and education level.



A Description of the key variables in SHARE

A.1 Cognitive functions

Recall: To measure memory, the interviewer reads a list of ten items. Immediately and then after a few minutes (after the fluency and numeracy questions), the respondent is asked which one he or she remembers. The list consists of the following items: butter, arm, letter, queen, ticket, grass, corner, stone, book, stick. The recall score is the number of items recalled by the respondent.

Numeracy: Possible answers are shown in a card while the interviewer is instructed not to read them out to the respondent:

1. *"If the chance of getting a disease is 10 per cent, how many people out of one thousand would be expected to get the disease?"* The possible answers are 100, 10, 90, 900 and another answer.

2. *"In a sale, a shop is selling all items at half price. Before the sale a sofa costs 300 Euro. How much will it cost in the sale?"* The possible answers are 150, 600 and another answer.

3. *" A second hand car dealer is selling a car for 6,000 Euro. This is two-thirds of what it costs new. How much did the car cost new?"* The possible answers are 9,000, 4,000, 8,000, 12,000, 18,000 and another answer.

4. *" Lets say you have 2,000 Euro in a saving account. The account earns ten per cent interest each year. How much would you have in the account at the end two years?"* The possible answers are 2,420, 2,020, 2,040, 2,100, 2,200, 2,400 and another answer.

The "raw" total score is constructed as follow: If a person answers (1) correctly she is then asked (3) and if she answers correctly again she is asked (4). Answering only to (1) correctly results in a score of 3; answering (3) correctly but not (4) results in a score of 4; while answering (4) correctly results in a score of 5. On the other hand if she answers (1) incorrectly she is directed to (2). If she answers (2) correctly she gets a score of 2 while if she answers (2) incorrectly she gets a score of 1. The procedure of inclusion of time length is as follow. Except for the respondents with score equal to 0, we divide the respondents according to the "raw" total score (from 1 to 4). In each score group, we divide the respondents according to the quartile distribution of the time spent to answer, generating other four subgroups for each of the score groups. In that way, we are able to construct a new total score that ranges form 0 to 16. **Fluency** The indicator of fluency is based on the following question: *"I would like you to name as many different animals as you can think of. You have one minute to do this."* The indicator corresponds to the number of valid animals named by the respondent. Any member of the animal kingdom, real or mythical is considered a valid answer. Repetitions and proper nouns are instead invalid.

A.2 Working status, social activities and income

Working status

The question refers to employment status: (1) self-employed, (2) dependent employee, (3) retired, (4) unemployed, (5) out of the labor force.

Social activities

The SHARE questions on which we base our indicator of social activity is as follows: *"Have you done any of these activities in the last month?"*.

Possible answers are: 1. voluntary or charity work; 2. care for a sick or disabled adult; 3. help for family, friends or neighbors; 4. attendance of an educational or training course; 5. participation in a sport, social or other kind of club; 6. taking part in a religious organization (church, synagogue, mosque etc.); 7. taking part in a political or community-related organization; 8. None of these.

Income

This variable is constructed on the basis of the SHARE variable `hgtincv_p`. It corresponds to the gross household income of the respondent, PPP-adjusted in Euro, and divided by household size to obtain a real per-capita amount .

A.3 Health

Self-assessment questions for each health domain are the following.

Chronic conditions

The detailed definition of the chronic conditions used in this paper are the following:

`high_pressure`: high blood pressure; `high_cholesterol`: high blood cholesterol; `stroke`: stroke or cerebral vascular disease; diabetes: diabetes; lung_disease: chronic lung disease; cancer: cancer or malignant tumour, including leukaemia or lymphoma, but excluding minor skin cancers.

Euro-D depression

The Euro-D scale of depression takes into account the following concepts: depression, pessimis, suicidality, guilt, sleeep problem, interst, irritability, appetite, fatigue, concentration enjoyment and tearfulness. The scale range from 0 to 12.

Mobility (mobility limitation)

This variable is based on the SHARE variables ph048_1 to ph048_11. It corresponds to the number of limitations with mobility, arm function and fine motor function reported by each individual.

Phinactiv (physical inactivity)

This variable is constructed on the basis of SHARE variables br015_ and br016_ regarding levels of vigorous and moderate physical activity, respectively. Physical inactivity is defined as never or almost never engaging in neither moderate nor vigorous physical activity.

A.4 Parent's occupation, household information and interviewers observation

Parent's occupation

SHARE asks respondents for their own, their former partners and their parents occupation.
SHARE uses the current (1988) International Standard Classification of Occupations (ISCO-88) by the International Labour Organization (ILO) to organize jobs into groups and international comparisons. The major groups in the ISCO-88 classification are: (1) managers; (2) professionals; (3) technicians and associate professionals; (4) clerical support workers; (5) service and sales workers; (6) skilled agricultural, forestry and fishery workers; (7) craft and related trades workers; (8) plant and machine operators, and assemblers; (9) elementary occupations; (0) armed forces occupations.

In our work we create one indicator for the father's last occupation (f_whitecollar) corresponding to groups (1)–(3). As for the mother's last occupation, because homemaker was reported by more than 50% of the respondents, we constructed two indicators: m_whitecollar corresponding to groups (1)–(3), and m_bluecollar corresponding to the others occupation excluding homemaker, used as reference category.

Child_in

This variable correspond to the SHARE variable `child_in`. It indicates the presence of children or children-in-law in the household.

Urban

The indicato `urban` is based on SHARE variable `ho037_` that corresponds to the question: "How would you describe the area where you live?". Possible answers are: (1) a big city; (2) the suburbs or outskirts of a big city; (3) a large town; (4) a small town; (5) a rural area or village. Our indicator takes value 1 if respondent answers (1), (2), or (3).

Willingness to answer

The indicator `will` is based on the SHARE variable `iv004_` that corresponds to the question: "How would you describe the willingness of the respondent to answer?". Possible answers are: (1) very good; (2) good; (3) fair; (4) bad; (5) good in the beginning, got worse during the interview; (6) bad in the beginning, got better during the interview. Our indicator takes value 1 if the variable `iv004_` is equal to (1), (2) or (6).

B Tables

B.1 Age and education coefficients by country (wave 1).

	Country	N	Age	Age^2	Educ1	Educ3
Fluency	At	1548	-0.005	-0.141 ***	-0.267 ***	0.239 ***
	Be	2558	-0.013 ***	0.003	-0.238 ***	0.260 ***
	Ch	778	-0.019 *	-0.02	-0.165 **	0.097
	De	1950	-0.007	-0.046	-0.179 ***	0.329 ***
	Dk	1337	-0.007	-0.022	-0.195 ***	0.220 ***
	Es	1422	-0.013 **	-0.054 **	-0.324 ***	0.065
	Gr	1697	-0.019 ***	-0.045 **	-0.159 ***	0.006
	It	1772	-0.019 ***	0.01	-0.265 ***	0.125
	Nl	1999	-0.018 ***	-0.011	-0.176 ***	0.184 ***
	Sw	2253	-0.015 **	-0.060 *	-0.136 ***	0.212 ***
Recall	At	1548	-0.016 **	-0.026	-0.206 ***	0.1
	Be	2558	-0.026 ***	-0.036	-0.237 ***	0.262 ***
	Ch	778	-0.018	-0.057	-0.253 ***	0.045
	De	1950	-0.022 ***	-0.005	-0.196 ***	0.177 ***
	Dk	1337	-0.021 **	-0.117 ***	-0.219 ***	0.217 ***
	Es	1422	-0.034 ***	-0.046	-0.380 ***	0.102
	Gr	1697	-0.037 ***	-0.025	-0.353 ***	-0.019
	It	1772	-0.029 ***	-0.028	-0.354 ***	0.393 ***
	Nl	1999	-0.027 ***	-0.068 ***	-0.166 ***	0.144 **
	Sw	2253	-0.012 *	-0.121 ***	-0.220 ***	0.098 **
Numeracy	At	1548	-0.013 *	-0.084 **	-0.410 ***	0.282 ***
	Be	2558	-0.012 *	-0.024	-0.375 ***	0.177 ***
	Ch	778	-0.007	-0.046	-0.193 ***	0.211 **
	De	1950	-0.006	-0.01	-0.388 ***	0.132 ***
	Dk	1337	-0.006	-0.012	-0.328 ***	0.301 ***
	Es	1422	-0.008	-0.072 **	-0.399 ***	-0.094
	Gr	1697	-0.032 ***	-0.047	-0.446 ***	0.276 ***
	It	1772	-0.019 ***	-0.04	-0.485 ***	0.133
	Nl	1999	-0.006	-0.039	-0.350 ***	0.216 ***
	Sw	2253	0	-0.029	-0.190 ***	0.225 ***

*: p-value $< .05$; **: p-value $< .01$; ***: p-value $< .001$

B.2 Differences in cognitive tests between waves by sex.

	Recall		Fluency		Numeracy	
	Man	Women	Man	Women	Man	Women
age1	-0.011 *	-0.020 ***	-0.011 *	-0.011 **	-0.012 **	-0.014 ***
age2	-0.071 **	-0.033	-0.055 *	-0.052 *	0.025	-0.001
educ1	-0.072	-0.012	-0.069 *	-0.082 **	-0.208 ***	-0.150 ***
educ3	0.072	0.041	0.072 *	0.146 ***	0.107 **	0.063
sick	-0.065	0.096	-0.216 **	0.095	-0.144	-0.026
unemp	0.023	0.047	-0.010	-0.044	-0.054	-0.164 *
emp	0.075	-0.038	-0.002	0.048	-0.006	-0.007
homemaker	-0.444	-0.014	0.170	0.051	0.476	-0.019
nospouse	-0.019	0.049	-0.031	-0.028	-0.006	-0.020
hsize	0.021	0.092 *	-0.009	-0.030	0.065	0.035
child_in	0.030	-0.106 *	0.019	0.017	-0.039	-0.030
lpc_income	0.009	0.001	0.005	0.008	0.038 **	0.033 *
urban	0.027	0.059 *	0.013	-0.001	-0.020	-0.003
will	0.302 **	0.045	0.283 **	0.261 **	-0.044	0.011
at	0.065	0.110	0.154 *	0.232 ***	0.156 *	0.283 ***
de	-0.196	-0.005	0.245 *	0.049	0.257 *	0.309 **
sw	0.071	0.093	0.371 ***	0.329 ***	0.238 **	0.118
nl	-0.142	0.109	0.129	-0.036	0.365 ***	0.271 **
es	-0.197 *	-0.166 *	-0.031	-0.259 ***	-0.072	-0.070
fr	-0.236 *	-0.105	0.296 **	0.068	0.079	0.037
dk	0.230 **	0.199 ***	0.282 ***	0.259 ***	0.104	0.008
gr	0.217 *	-0.124	0.012	-0.085	0.315 ***	0.212 *
ch	0.038	0.076	0.263 ***	0.232 **	0.300 ***	0.343 ***
be	0.029	0.030	0.281 ***	0.257 ***	0.016	0.084
recall1	-0.169 ***	-0.162 ***	0.037 ***	0.034 ***	0.015 **	0.017 ***
fluency1	0.018 ***	0.013 ***	-0.088 ***	-0.082 ***	0.006 *	0.008 *
Numer1	-0.091	-0.185 ***	-0.015	-0.218 ***	1.001 ***	0.897 ***
Numer2	-0.101	-0.005	-0.084	-0.066	0.682 ***	0.611 ***
Numer4	0.078 *	0.105 ***	0.096 **	0.084 **	-0.534 ***	-0.533 ***
Numer5	0.099 *	0.141 **	0.161 ***	0.052	-1.143 ***	-1.108 ***
lnh	-0.814	0.063	0.068	-0.227	1.126 *	0.547
lnw	0.153	-0.011	0.134	-0.075	-0.230	-0.138
diabetes	-0.101 *	-0.072	-0.071	-0.117 *	-0.017	0.020
lung_disease	0.105	0.011	0.078	-0.003	0.030	-0.096
stroke	-0.036	-0.053	0.024	0.009	0.006	-0.182 **
high_pressure	0.048	-0.020	-0.022	-0.007	0.011	0.033
cancer	-0.109	-0.045	-0.118	-0.059	0.232 **	0.040
mobility	-0.024	-0.012	-0.006	-0.004	-0.006	-0.021 **
phinactiv	-0.108	-0.019	-0.100	-0.156 **	-0.003	0.012
eurod	-0.009	-0.015 *	-0.010	0.009	-0.020 *	-0.009
voluntary	0.157 ***	0.076	0.057	0.115 **	-0.024	0.038
cared	-0.014	0.000	0.163 **	0.034	0.009	-0.035
helpfam	0.018	0.037	0.070 *	0.054	0.027	0.010
training	0.017	0.071	0.077	0.056	0.036	-0.023
club	0.121 **	0.035	0.152 ***	0.028	0.022	0.038
religious	0.027	-0.039	-0.010	0.071 *	-0.057	-0.129 ***
political	0.016	0.034	0.058	0.106	0.031	0.112
f_whitecollar	0.026	0.077 **	0.044	0.092 **	0.060	0.002
m_whitecollar	0.019	0.060	0.023	0.211 ***	-0.106	0.127 **
m_bluecollar	-0.068	0.080 **	-0.023	0.079 **	-0.034	0.029
mills	0.920 *	0.007	0.468	0.671 *	-0.383	-0.341
cons	-1.138 ***	-0.304	-0.756 **	-0.845 ***	0.554 *	0.305
N	5266	6172	5266	6172	5266	6172
R^2	.22	.2	.25	.24	.27	.25
ll	-6812	-8026	-6716	-7882	-6764	-7606

*: p-value $< .05$; **: p-value $< .01$; ***: p-value $< .001$