Lifebrain

D.4.1
Epidemiological analyses of cognitive functions across European cohorts

Project title:	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
Due date of deliverable:	30th June 2018
Submission date of deliverable:	29th June, 2018
Leader for this deliverable:	Frisch Centre

This project has received funding from the *European Union's Horizon2020 research and innovation programme* under grant agreement *No 732592.*

Contributors to deliverable:	Name	Organisation	Role / Title
Deliverable Leader	Ole Rogeberg	Frisch Centre	PI
Reviewer(s)	William Baaré	REGIONH	WP2 leader
	Christian A Drevon	Vitas	WP5 leader
Final review and approval	Barbara B. Friedman	UiO	Administrative coordinator

Document History				
Release	Date	Reason for Change	Status (Draft/In-review/Submitted	Distribution
1.0.	21 June, 2018	First draft to the Executive Board	Draft	Slack, email
2.0.	28th June, 2018	Revised version based on the EB comments	In-revision	
3.0.	29th June, 2018	Final version	Submitted	Participant portal

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

Executive Summary

The Lifebrain consortium pools a set of participant cohorts recruited in longitudinal studies from different European countries, and provides "deep" data on relatively small samples that may not be representative of the general population in the different countries. The Survey of Health, Ageing and Retirement in Europe (SHARE) has less detailed and extensive measures of cognitive functions, but a larger sample recruited to be representative of the population aged 50+ years in the respective countries. To provide a context for assessing the patterns in the Lifebrain cohorts, this report presents a series of descriptive analyses of the patterns in the SHARE data.

The SHARE data contain observations of more than 120 000 individuals from 27 countries and Israel, many of whom participated in multiple waves of this ongoing study. Using the easySHARE data including a subset of variables measured consistently over time; we assess how performance on tests of orientation, recall (immediate and delayed), and numeracy vary by age, and how the age trajectory differs across identifiable groups defined by sex, country, household income decile and educational level.

For orientation, participants are asked to give the current year, month, date and day of week. The ability to answer all items correctly decreases with age, with some 50% unable to answer all questions correctly at age 90 years. The measure shows substantial variation across countries, with about 15% of Swiss respondents aged 80-90 years failing at least one item whereas the same is true for more than 40% of the Spanish respondents in the same age group. Substantial differences were also seen across educational groups, whereas the sexes performed similarly.

For recall, an improved performance at low but not at high ages was seen for both females relative to men and for individuals in high income households relative to lower income households. Differences between educational groups persisted into high age groups. These relative differences between groups were typically similar across countries, although the average age-related patterns themselves showed substantial variation across countries.

For numeracy, and opposite to recall, males generally scored higher than women and this difference appeared to persist at high ages. The differences between participants in different household income deciles were similar to what was seen for recall, with improved performance in participants from high income households at young ages and smaller differences at high ages. The pattern by educational level was also similar to that of recall, with persistently raised performance at all ages among those with higher educational attainment. For numeracy, too, there was evidence of country level differences in both average levels and trajectories. However, analyses of how the group differences themselves differed by country were hard to interpret due to the more limited variation

in the numeracy indicator, which resulted in noisy estimates and non-credible age-related patterns in several cases.

Finally, assessing the associations between cognitive scores and indicators of daily functions, all measures of cognitive ability have statistically significant associations with daily living and functions, with orientation measures having a particularly large association with the ability to perform daily living tasks (such as “eating, cutting up food”, “taking medications” and “managing money”).

List of acronym/abbreviations

ERIC	European Research Infrastructure Consortia
NIPH	Norwegian Institute of Public Health
SHARE	Survey on Health And Retirement Europe
UiO	University of Oslo

1. Introduction

The Lifebrain consortium aims to pool a set of participant cohorts recruited into longitudinal studies of cognition and imaging (magnetic resonance imaging = MRI) of the brain across Europe. These data are "deep" but potentially "narrow"; there are a lot of data on each participant, including cognitive measures and MRI scans, but the samples are small relative to the populations studied and may not be representative of the respective populations.
We analyse data from the Survey of Health, Ageing and Retirement in Europe, data that are "shallow" but "broad," in the sense of having fewer measures of cognitive function but a larger sample recruited to be representative of the population above 50 years of age in participating countries.

The purpose of the SHARE analyses is to examine

a) how measures of cognitive functions are associated with age, sex, and socioeconomic status;
b) how these associations differ across countries; and
c) to what extent they are associated with differences in health and everyday functions.

This new information will be valuable for an integrated view of the Lifebrain cohorts, e.g. by indicating possible differences between nations or groups that can be assessed in greater detail using the deeper data available in these smaller cohorts. The analyses will also be helpful in assessing whether national differences apparent in the Lifebrain data are consistent with the patterns in the population-representative SHARE data. If they are not, this may suggest that the differences in the Lifebrain data might reflect differences in national sampling schemes and population representativeness.

1.1. Description of deliverable

D4.1 - Epidemiological analysis across European cohorts
Lead: Frisch; Participants: UIO, NIPH (M12-M18)
Task: Use multivariate regression analysis and multilevel/hierarchical models on existing European Research Infrastructure Consortia (ERICs; data from the Survey of Health, Ageing and Retirement in Europe, SHARE) to analyse

1. how cognitive performance differs by socioeconomic status (SES), sex and age
2. how these patterns differ across regions/countries
3. the extent to which the cognitive outcomes are associated with differences in health and everyday functions

1.2. Objectives

Present the results of descriptive analyses of associational patterns in the SHARE data relevant to the Lifebrain consortium.

2. Overview of data and approach

2.1. Data source

We use SHARE-data, a cross-national panel database with individual level data on health, socio-economic status and social and family networks including more than 120,000 individuals aged 50 or older from 27 European countries and Israel (Figure 1 and Figure 2).

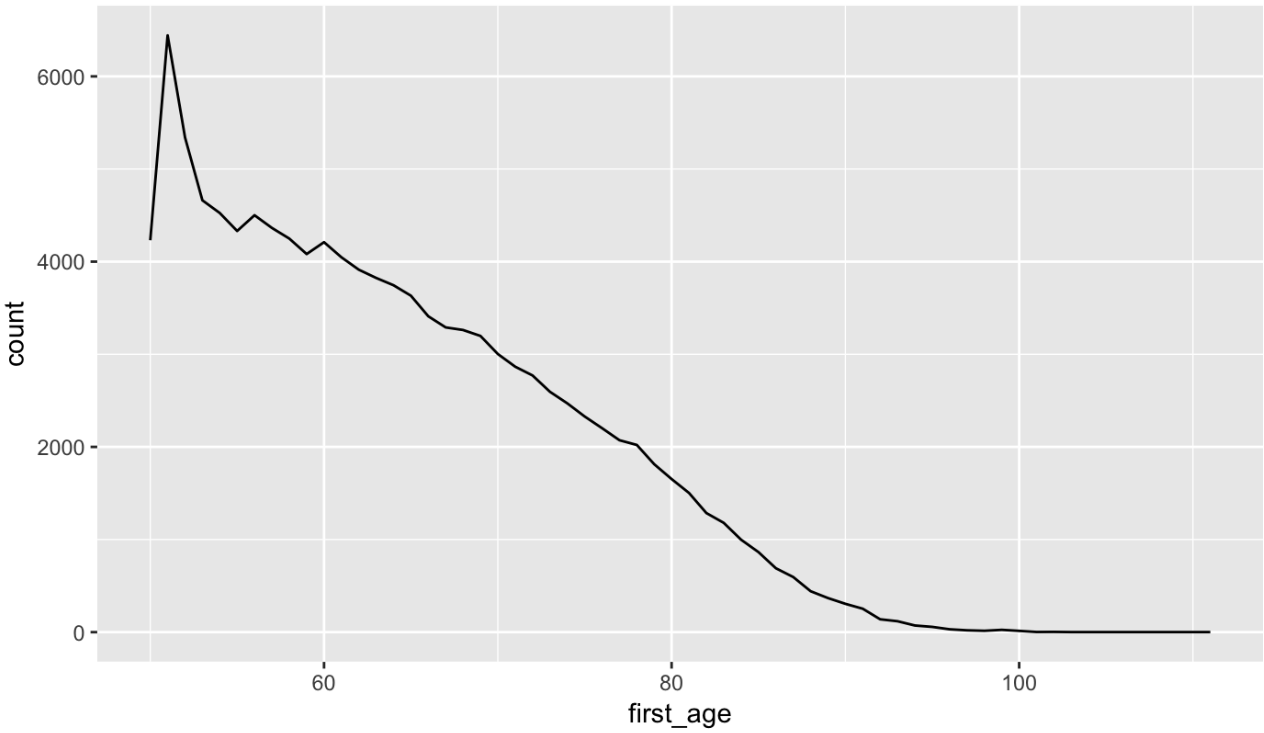

Figure 1. Age distribution in the SHARE sample. The figure shows the number of participants aged 50+ in the sample by integer age at recruitment

The sample recruitment strives to achieve a population-representative sample from each country, and the measures of cognitive and physical performance combined with a broad set of covariates capturing background factors allowing assessment of across and within-individual change trajectories and how these correlate across groups defined by cohorts, countries, socio-economic levels, etc.

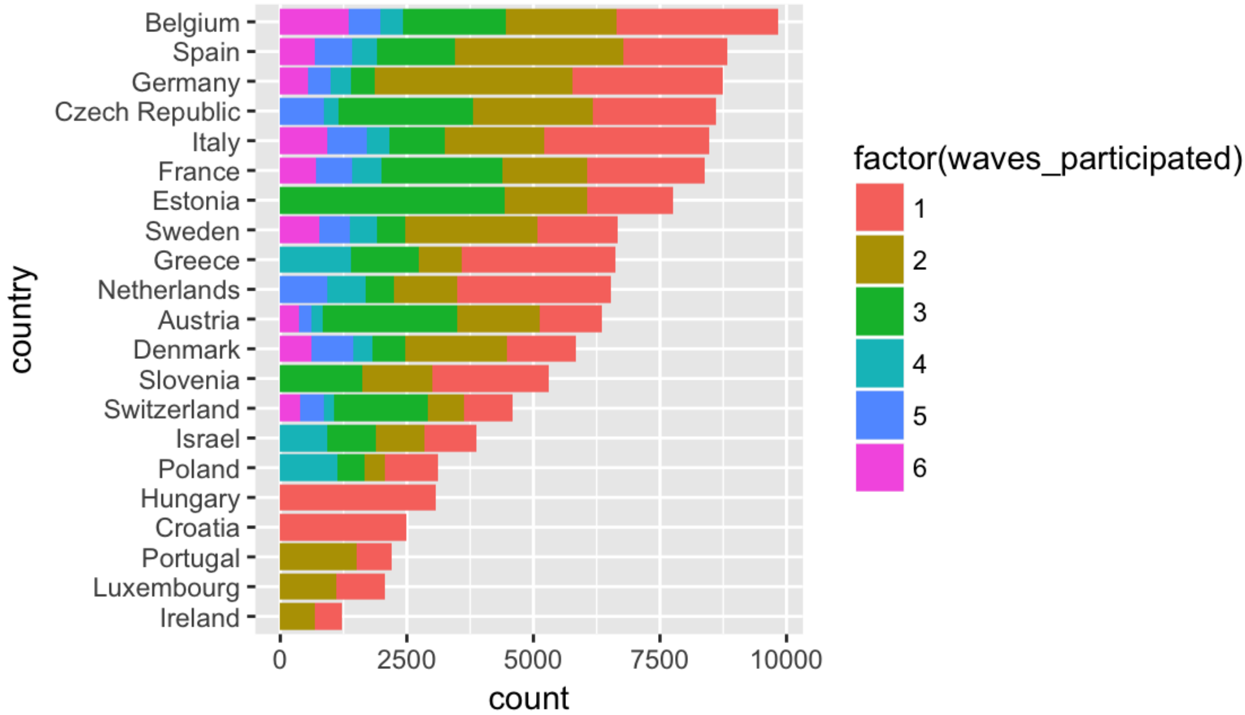

Figure 2. Number of participants by country and number of waves participated. The figure shows the full sample size from each country, broken down by the number of waves for each participant.

Whereas the survey instruments that provide indicators of cognitive and physical performance are cruder and less informative than the in-depth testing employed in the Lifebrain cohorts, the large population-representative samples may nonetheless compensate sufficiently for the SHARE-data to identify associations that can be assessed in more detail within the Lifebrain data.
We use easySHARE. This dataset contains a subset of the variables in a single longitudinal dataset, avoiding "the need for complex merging of waves and modules[1]". This approach was chosen due to the complexity of combining the wave-data directly, as this would require extensive work to assess comparability of items over time, etc. Illustrating this, the Stata code used to generate the panel data set runs to more than 4700 lines of code.

2.2. Outcome variables

The easySHARE dataset contains four variables reflecting cognitive function:

- orienti - orientation to date, month, year and day of week
- recall 1 - 10 words list learning first trial
- recall 2 - 10 words list learning delayed recall
- numeracy - numeracy scores (a second indicator ("numeracy2") is available in waves 4-6)

[1] easyShare release guide 6.0.0

Orientation (**orienti**) is a variable capturing "orientation to date, month, year and day of week", and measures how many of these four time-indicators people could correctly identify.
Both the recall tests provide scores ranging from 0 to 10, denoting the number of items recalled from a set of 10 immediately after hearing a list (**recall1**) and with a delay (**recall2**).
The numeracy tests include a set of mathematical problems the respondents were asked to solve in their head. **Numeracy 1** contains 4 questions involving percentages and shares, **Numeracy 2** involves subtraction (repeatedly subtracting 7 from 100). The numeracy 1 items vary in complexity, as shown by the two following items:

- *If the chance of getting a disease is 10 percent, how many people out of 1000 (one thousand) would be expected to get the disease?*
- *A second hand car dealer is selling a car for 6,000 [some currency]. This is two-thirds of what it costs new. How much did the car cost new?*

For health and daily life functions, the easySHARE data contain several summary measures of health and functioning relevant for older populations. We employ the following:

- Daily living (adla - activities of daily living). Items like "Dressing, including putting on shoes and socks", "Bathing or showering", "Eating, such as cutting up your food"
- Living – instrumental: (iadzla - instrumental activities of daily living index 2). Items like "making telephone calls", "Taking medications", "Managing money"
- Finemotor (finemotor). Items like “picking up small coin from table”, “Cutting up food”, “dressing including shoes and socks”
- Grossmotor (grossmotor). Items such as “Walking 100 meters”, “Walking across room”, “getting in or out of bed”, “bathing/showering”, “Climbing one flight of stairs”
- LGmuscle (lgmuscle - large muscle index). Items such as "sitting for about two hours", "Getting up from a chair after sitting for long periods", "Pulling or pushing large items like a living room chair"
- Mobility (mobilityind). Items such as "walking 100 metres", "walking across a room" and using stairs

Due to the low number of respondents aged below 50 years of age (individuals recruited due to being spouses of other participants) and above 90 (few individuals, and often affected by strong floor effects in the cognitive measures) were excluded from the regression sample.

2.3. Background variables

Given the Lifebrain focus on ageing and lifespan, patterns relating to age are a primary interest. These patterns may themselves differ by other background variables. The ones employed in this report are:

- Sex
- Household income decile (current household income relative to own country participants). This measure was missing from wave 3. A simple interpolation and rounding was used to assign scores for wave 3 for participants observed in multiple waves.
- Education – based on the ISCED-97 coding of education
- Country

2.4. Analysis strategy

Our main focus is to assess how the association between cognitive function and age differs across different groups. The analyses are descriptive and causal inference based on the identified associations is not warranted. To see this, it may be helpful to describe briefly a "causal" account of different factors that will be reflected in the data.

2.5. Limitations precluding causal inference

The data include of a large number of individuals of different ages, from different countries, with different covariates (household income, education, sex), scored on different measures of cognitive function (orientation, recall, numeracy). The different measures of cognitive function are assumed to be imperfect measures of some latent ability/trait (e.g. memory, cognitive ability). We assume that these latent traits have a distribution across individuals, and that some of this variation is systematically related to age and observed covariates.

Change over time

We expect the mental capacities reflected in the cognitive scores to change over time within individuals due to ageing. This change may differ across participants in ways systematically related to background and current environment/behaviour. Such effects may be bidirectional: individuals with poor cognitive performance or rapid decline in cognitive functioning may be selected out of employment, for instance, whereas employment may support cognitive abilities by requiring workers to perform cognitively stimulating tasks, enriching their social environment through colleagues, or involving physical activity with beneficial effects on health and daily functioning. Because the characteristics of jobs have changed over time due to automation, technological progress, outsourcing of routine work to non-European countries, etc., such mechanisms may also have shifted over time and differ across cohorts and nations.

Cohort differences

An additional issue concerns cohort differences. As evidenced most famously by the Flynn effect, cognitive scores on the same test conducted with participants at a fixed age may show trends over time across birth cohorts. Because a substantial part of the age-variation in the SHARE data is across individuals, such cohort differences will potentially confound age-related patterns. As an extreme illustration, if IQ scores were constant after conscription testing, an across-participant assessment of age-related decline would display the Flynn-effect in reverse, with lower scores of old participants reflecting pure cohort differences rather than actual within-individual decline.

To assess the presence of actual within-individual changes, we may use individual fixed effect models adjusting for constant individual level differences. This is not sufficient to avoid the cohort issue, however, as the within-individual variation in age is substantially less than the age-span across which we want to assess functioning: While a participant may be observed repeatedly across a 10 year period, we are interested in seeing how the outcome measures change across a 40 year period. The within-individual estimates of change across the full age span is consequently based on young cohorts in the lower part of the age range and old cohorts in the higher part of the range. To see how this matters, consider a model where individual ability decreases at an increasing rate after some “peak year,” and where this decline is simply shifted to a higher age for later cohorts due to improved health and functioning. In this case, the age-related curve inferred from within-family variation would fail to represent the true shape of the decline and would compress and distort the age-related decline as younger cohorts showed less decline than the older cohorts at the same age.

Retest effects

Repeated measures of the same individual introduce further issues in the cognitive scores, as participants show “retest” effects, whereby they systematically do better in later tests. Whereas this can be “controlled for” in multivariate analyses, this will necessarily rely on further assumptions, such as “retest effects are on average the same across identifiable subgroups in the data”. If this is incorrect, the adjustment will influence the shape of identified age trajectories. If, for instance, retest effects are systematically smaller for high performers, perhaps because these people tend to work in occupations where the relevant cognitive skills are frequently exercised, this would lead to excessive declines estimated for these groups. Alternatively, if retest effects are smaller at advanced ages, because the task familiarity and learning from one testing session decays more rapidly when cognitive skills are reduced, we would systematically overestimate the decline at higher ages.

Group composition differences

Another issue is introduced when we assess age-related decline across groups changing composition over time. This is evidently an issue for education, where young cohorts systematically have higher educational levels (Figure 3). If there is sorting on cognitive ability into education, and if this sorting has weakened or shifted over time as educational systems have expanded, then age-related trajectories associated with specific educational levels will be confounded.

If selection on cognitive ability has weakened due to expansion of the educational system, the age-related decline observed across individuals with high education would be underestimated, with the high performance of old cohorts reflecting a more selected high-performance subpopulation rather than an actual difference in decline.

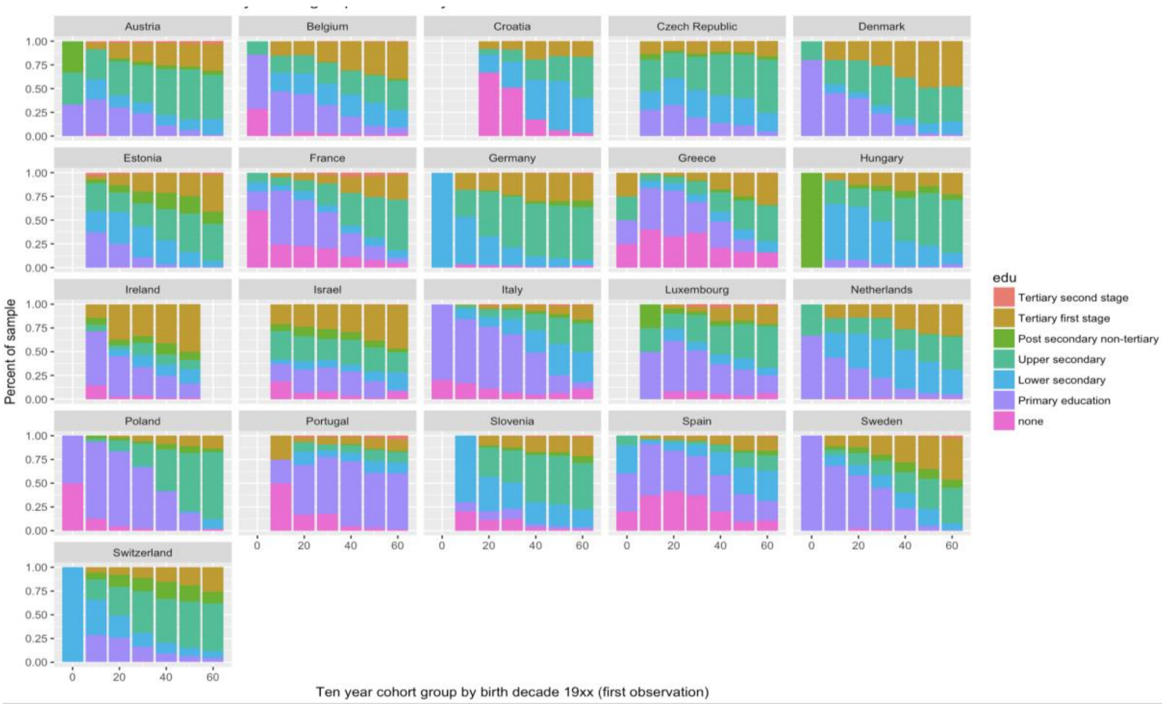

Figure 3. Educational level by birth cohort decade and country. We group respondents by country and birth cohort (10 year groups; Those born 1900-1910, 1910-1920, …) and show the distribution of educational level within each cohort.

A related issue should be noted with household income decile. Because this income shifts over time and is systematically reduced with age, individuals in highest income deciles at young ages are not directly comparable to those with the highest income decile at higher ages (Figure 4). Individuals in a high-income decile at high ages will differ systematically from the individuals in a high-income decile at lower ages, as those with declining cognitive function would be expected to shift downwards in the income distribution as their productivity decline.

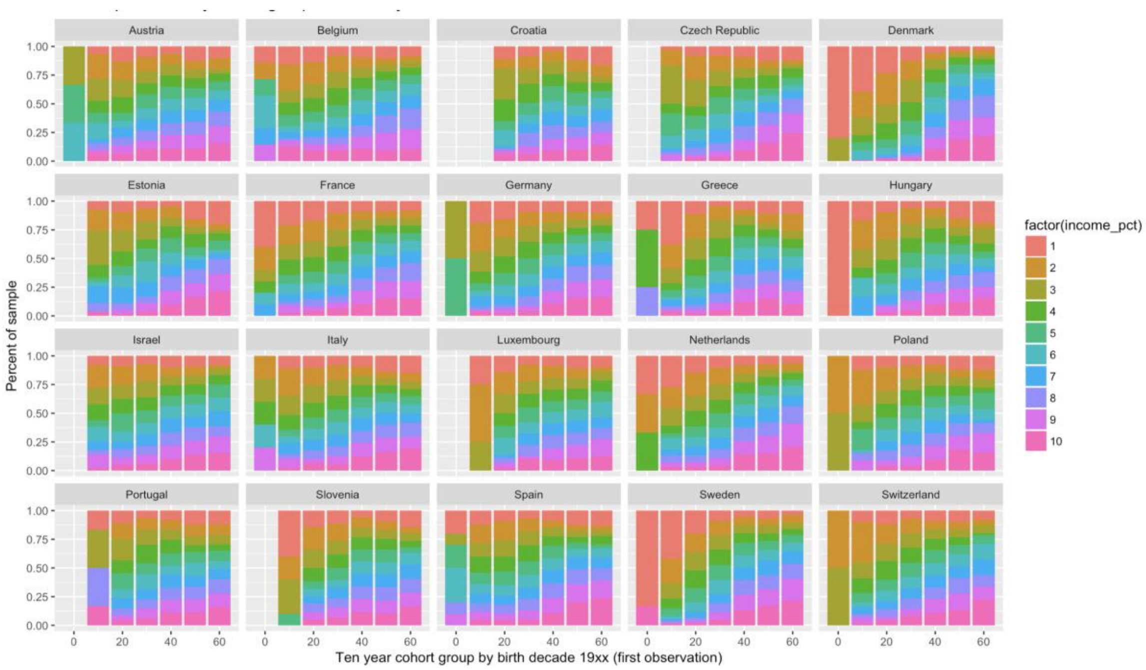

Figure 4. Current household income decile at first observation by country and birth cohort decade. We group respondents by country and birth cohort (10 year groups; Those born 1900-1910, 1910-1920, …) and show the distribution of household income decile at first measurement within each cohort.

Representativeness

Whereas the SHARE project is focused on achieving a population-representative sample from each country, we cannot exclude the possibility that the sampling yields data more or less representative of the underlying population at different ages or for different subgroups. This may be the result of systematic selection bias (e.g. sample inclusions criteria leaving out groups with pathologies, or systematic selection on unobserved factors when participation is voluntary). With repeated observations over time (panel data), there may be systematic selection out of participation or drop out (attrition bias).

2.6. Statistical models and presentation of results

Whereas the above comments note some serious limitations regarding how the results may be interpreted, they should not be misconstrued as robbing the present analyses of any relevance. By identifying patterns of association, we provide a context where more fine-grained analyses of the Lifebrain cohort can be evaluated.

We use multivariate least-squares regressions on "all data," combining both across and within-individual variation. These models allow us flexibly to specify patterns of age-related decline and their variation with observed variates.

The number of coefficients in such models, however, expands greatly as we allow interactions at higher levels to assess, e.g., how the age-score correlations change when interacted with educational level, and how these interactions vary by country (three level interaction terms). To display the results, we present plots of how the predicted average score (with confidence intervals) across age differs by different groups.

Central issues in these models are how we model the age trajectory and retest effects. To avoid artefacts in the results driven by overly narrow functional forms on the trajectories (e.g. second or third-order polynomials in age), we use a spline formulation that typically results in more stable estimates in line with the patterns in the underlying data. These splines are in turn interacted with variates of interest in separate models.

In addition, we assess systematic group differences in age-related trajectories using within-individual variation in scores. In these data, the retest effects are particularly difficult to control for, because the “first score” of participants aged 60 years cannot be used to infer the retest-effect of participants who have their second or third score measured at this age. Because poor controls for retest effects will bias the slope identified from within-individual variation, these analyses are not intended to provide accurate estimates of the age-related decline trajectory, but rather to assess whether there are differences across groups, and how these relate to the patterns in the other analyses.

Whereas the most common method for analysing within-unit variation remains regression models with fixed effects (equivalent to estimating separate constant terms for each unit), these models proved overly sensitive to the noisy data for cognitive outcomes. As an alternative strategy, we specified a Bayesian model with hierarchical Gaussian processes across groups (sex, country, education, income deciles).

We use a spline as the basis for estimating an “average age-trajectory” for all participants, whereas the development within each group is allowed to differ from this average trajectory. The deviation from the average age-trajectory is modelled as a Gaussian Process with an exponentiated quadratic kernel, which serves to smooth parameter values. E.g. if a group has a higher decline than average at age 60, 61, 63 and 64 years but a low number of observations indicates an *increase* at age 62 years, then this parameter will be “pulled into place” by the parameter estimates for the surrounding ages. The data are used to estimate the extent to which parameters are similar and how far the covariance extends.

The data for this model are individual change observations. For instance, an individual observed three times would be represented with two observations; one for the change from the first to the second observation, and one for the change from the second to the third observation. The group assigned to each change observation is based on the covariates measured at the first of the two time-points involved. For sex, country and educational level this will not matter, as these are typically fixed over time within an individual. For income decile, however, this means that we assess whether people in different income deciles will be predicted to change differently in the future. If we instead had used the second observed income decile for each change observation, we would be assessing whether people in different deciles would be predicted to *have changed* differently in the recent past.

3. Results pooled across countries

3.1. Orientation

An imperfect score on the orientation test may signal the beginning of serious cognitive impairment (e.g. beginning dementia). We see limited variation in the scores at “low” ages with almost 90% getting a perfect score in their 50s. For this measure, we restrict ourselves to displaying how scores differ in the 80-90 years age category across different groups. The ability to answer all four questions correctly declines with increasing rapidity across age groups (Figure 5), although the share with imperfect scores in the 80-90 age group shows substantial variation across countries (Figure 6), education (Figure 7), and to a lesser extent by sex (Figure 8) and income (Figure 9).

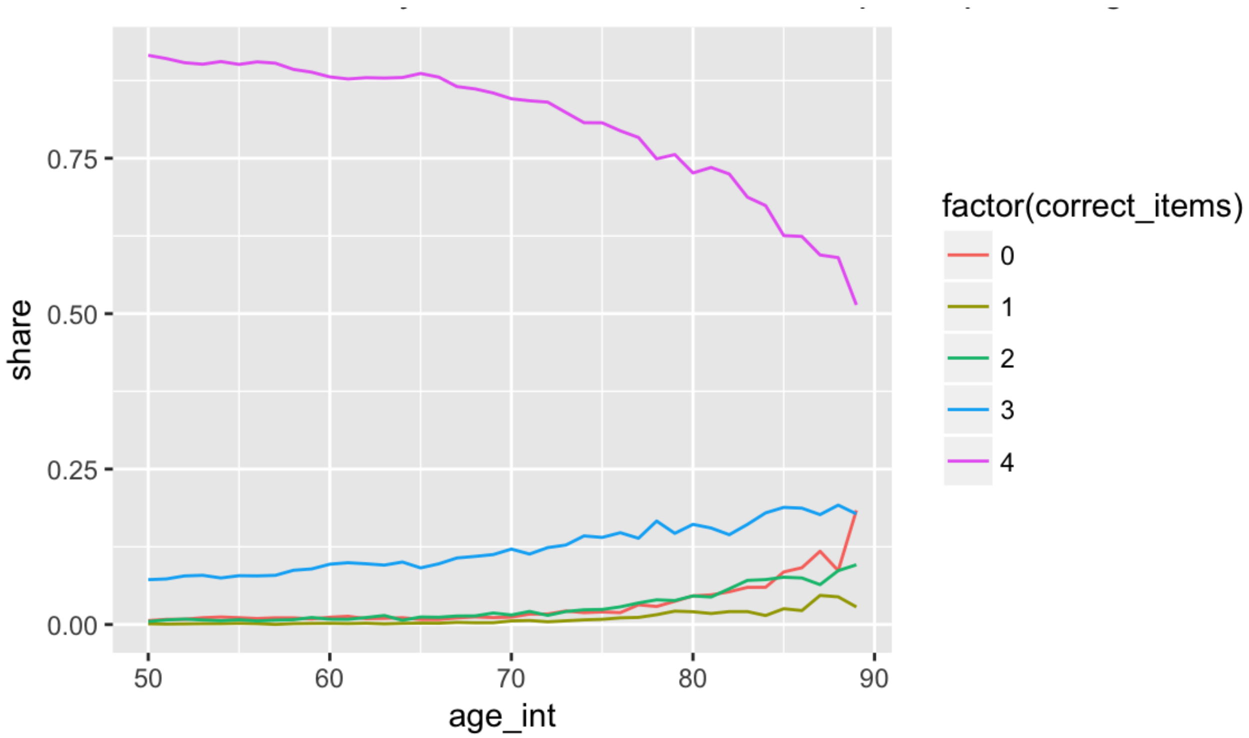

Figure 5. Share with different orientation scores by age at first observation. We display the share with 0, 1, 2, 3 or 4 correct items on the orientation test by integer age.

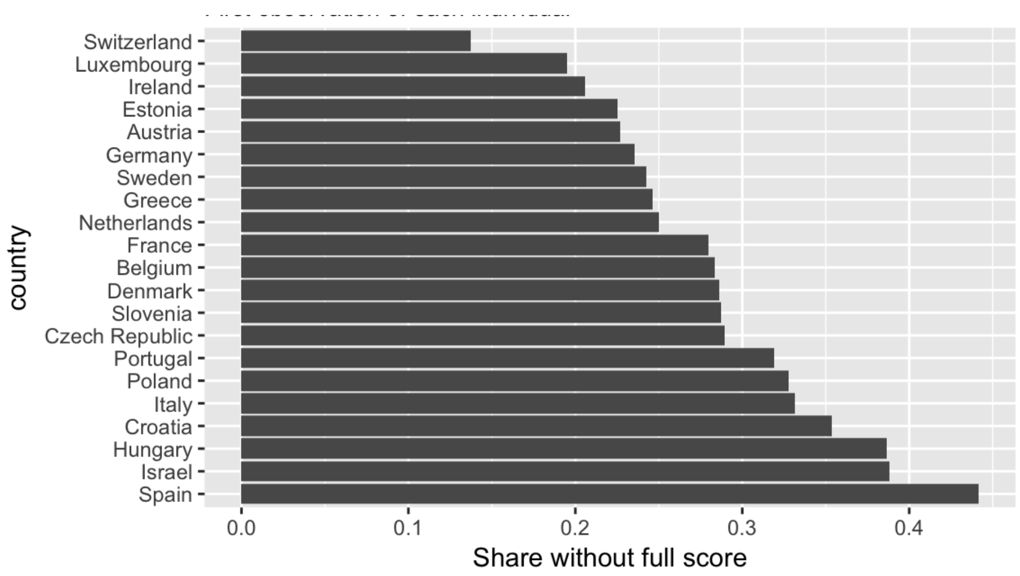

Figure 6. Share of participants aged 80-90 years with non-perfect orientation score at first observation by country

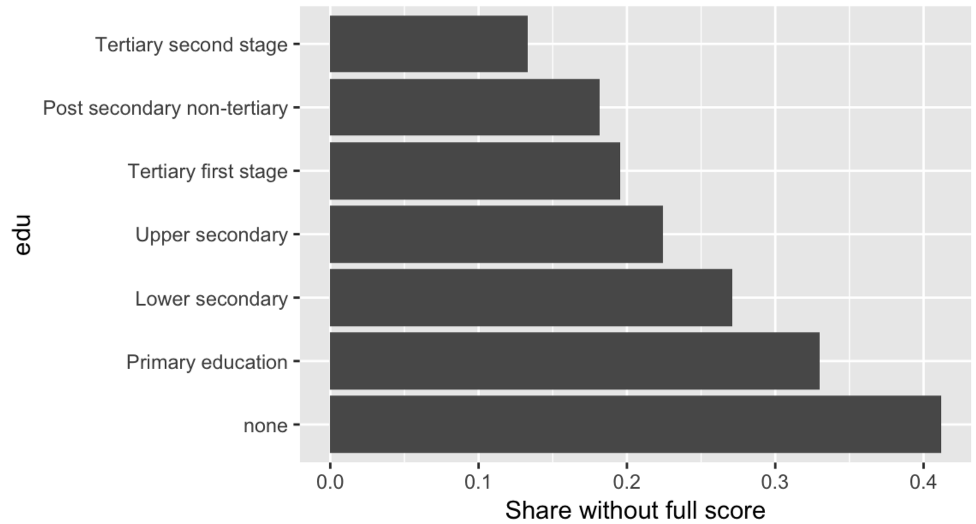

Figure 7. Share of participants aged 80-90 years with non-perfect orientation score at first observation by educational level. Categories are ordered by the share without full score.

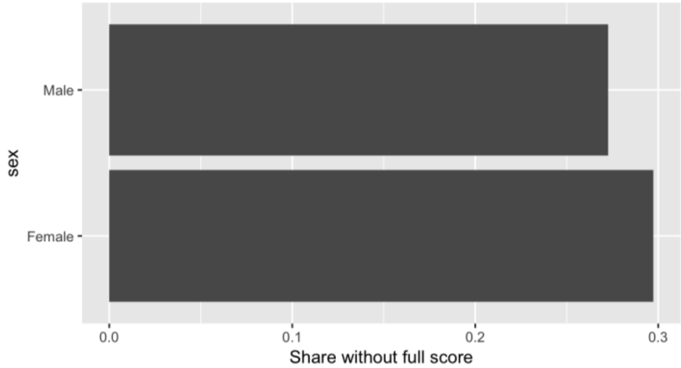

Figure 8. Share of participants aged 80-90 years with non-perfect orientation score at first observation by sex

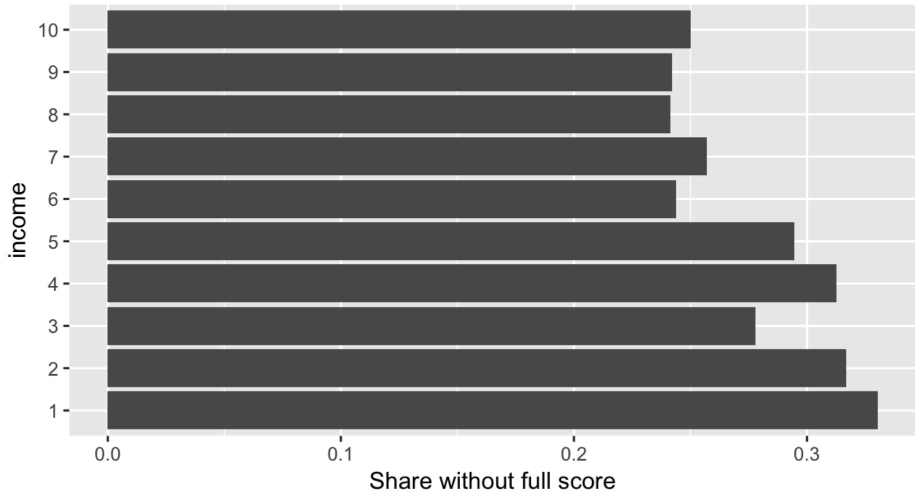

Figure 9. Share of participants aged 80-90 years with non-perfect orientation score at first observation by income decile

3.2. Recall

Assessing the score distribution of the two recall measures for the first observation of each individual, the scores on both are symmetrically distributed for the young age groups, with a tendency towards asymmetry at high ages. Based on a visual inspection of the distribution at advanced ages (Figure 10), the left-hand side of the distribution (low scorers) increases in length and thickens. Speculatively, this may be thought of as a normal distribution that – at advanced ages – is distorted by an increasing share of the population shifting into a second group characterized by more rapid decline in cognitive functioning. Consistent with this, the distribution of recall 1 score at different ages is largely symmetric at all ages for those with perfect scores on the orientation test (Figure 11), whereas the tendency towards a left-skewed asymmetric distribution is more pronounced the fewer orientation items a respondent successfully answered. In sum, this gives us a sample with a pronounced floor effect – especially on the delayed recall test – with more than 50% of respondents scoring 0 on the delayed test after age 90 years (Figure 12).

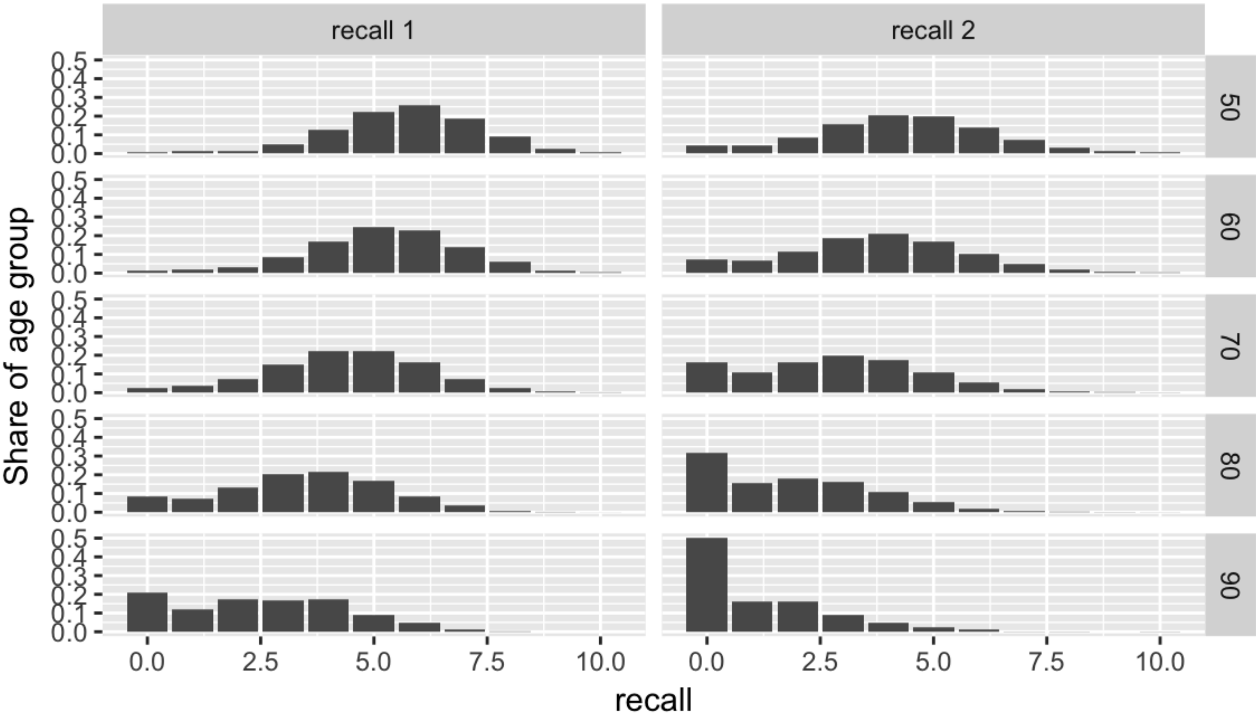

Figure 10. Distribution of recall scores by ten year age groups. The figure displays the distribution of the SHARE sample recalling 0, 1, …, 10 items at first participation.

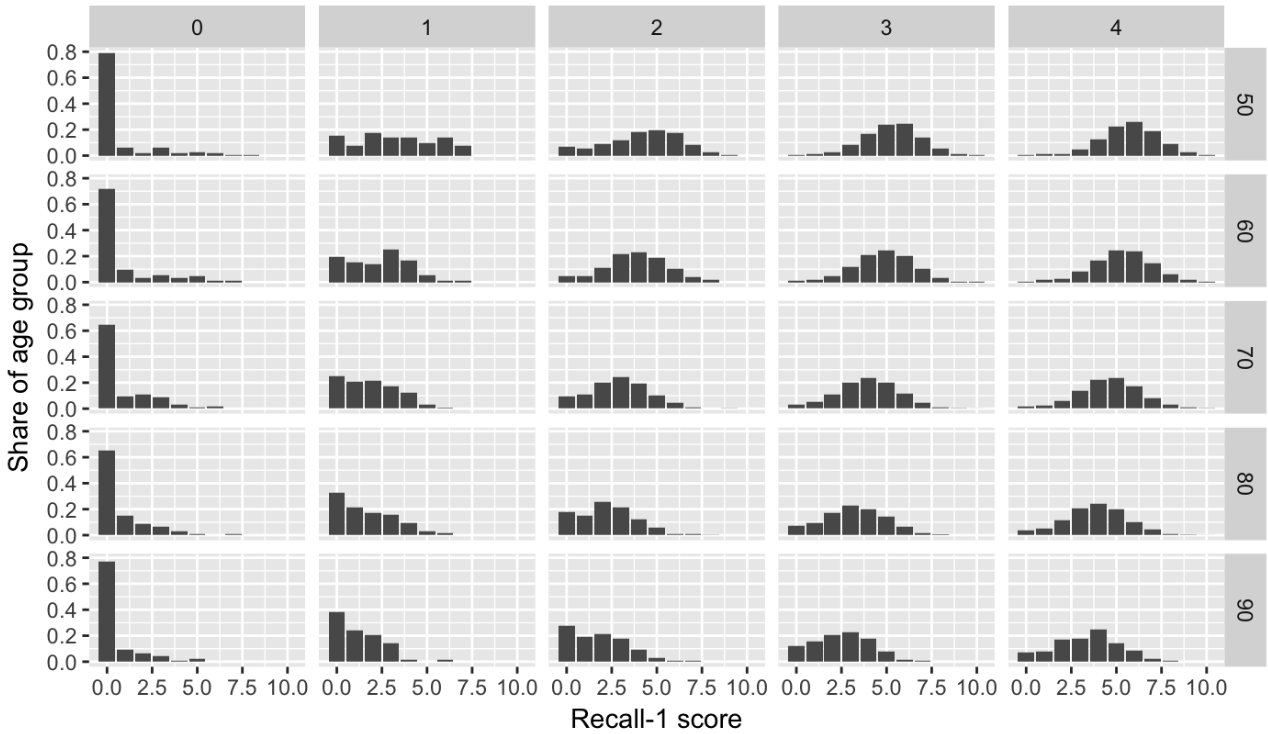

Figure 11. Score distribution for recall-1 by orientation score and age group. Each column shows the distribution of recall-1 scores for the subset of participants with the orientation score shown in the column heading.

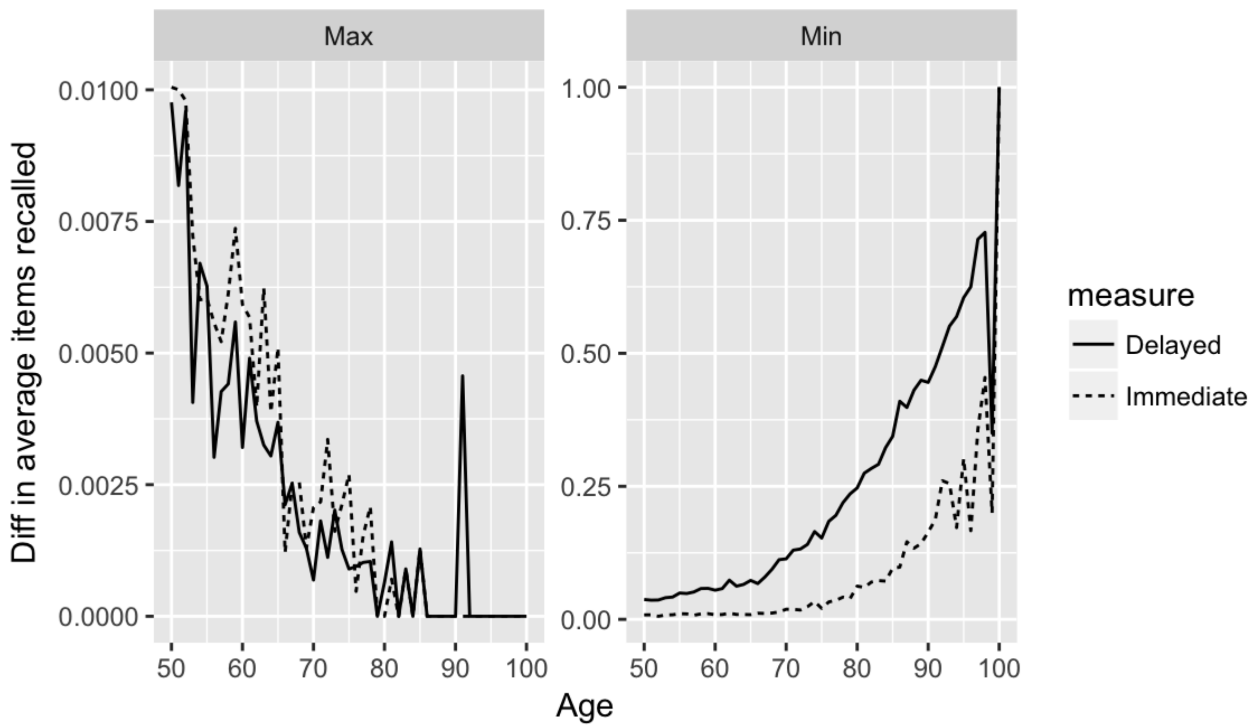

Figure 12. Share with maximum and minimum score on recall measures by age

Rather than perform analyses for each of the two recall scores separately, we analysed the sum of the two variables assuming that they are both measures of a similar underlying memory trait (the two scores have a correlation coefficient of 0.68 and decline in parallel – see Figure 13).

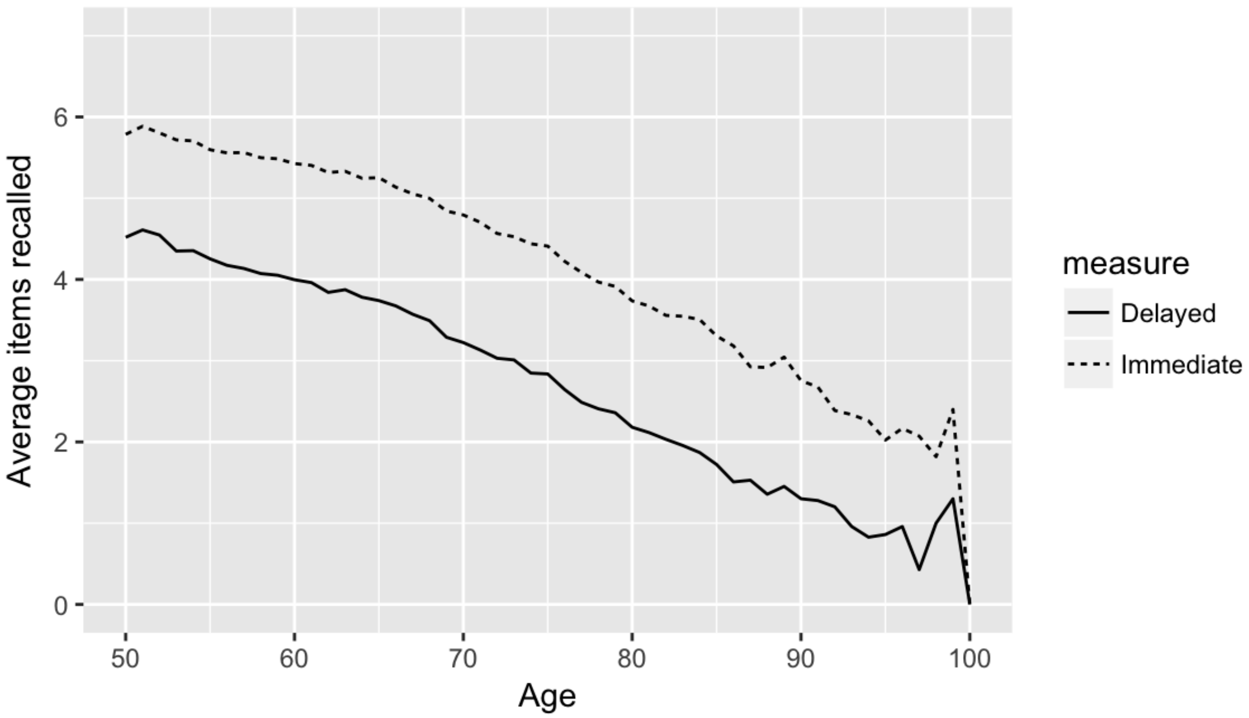

Figure 13. Average recall scores at first observation by age

There are clear retest effects that indicate improved scores for each repeated testing of the same individual. To show this, we split the data into a "pure across" sample using only the first observation of each individual and a "pure within" sample (only repeatedly sampled individuals). Estimating the decline curve in the within sample using an individual fixed effect model, we compare the observed declines over age between different models:

- Across variation only – no control for retest required, as we use only first observation of each individual
- Within variation only
  - No control – A model ignoring retest effects
  - Ever-tested – A model with a single dummy for "ever tested"
  - Fully flexible dummy specification – A set of dummy variables, one for each possible value of "prior tests"
  - Ever-tested with first two observations only – Uses only data from the first two observations of each individual, controlling for "ever tested" using a single dummy

The results show that a failure to control for previous tests generated an apparent score increase with age from the within-variation (Figure 14), and that this apparent increase was also present when we only controlled for being "ever tested." The fully flexible dummy specification results in a curve similar to the one found using only the first two observations from each individual but requires 5 parameters. Further testing (not shown) revealed that this model yields largely identical results to one requiring only two parameters: A dummy indicating "ever tested" and a coefficient for a linear effect of "number of previous tests." This specification captures the underlying pattern shown by the coefficients of the flexible dummy set, which indicates that the first retest effect is particularly large, whereas each later additional testing increases the score by a similar amount. This formulation was used for the remainder of the analyses.

Note that even the "fully flexible dummy specification" revealed a different decline curve from the within-individual change than what we found using the pure across-individual data. This may reflect cohort differences or variation in the retest-effect across ages; the retest coefficients average across age, and if retest effects are larger for young respondents this will underestimate the decline at young ages and overestimate the decline at high ages.

Figure 14. Age-related decline in recall indicated by across and within individual information

The analyses in Figure 14 identify the differences across age using a dummy set with one variable for each integer age. This ensures that the curves display the variation from the data without being forced into a specific functional form (e.g., a linear or polynomial change with age).

The drawback of such an approach is that it requires large samples, as seen by the non-credible trajectories at the advanced ages where sample sizes are small. As we want to interact the age-trajectory with an increasing number of variables, more structure needs to be imposed on the shape of the age trajectory. To avoid artefacts from a polynomial specification, which can lead to non-credible extrapolations at high ages, we use a quadratic spline function.

Pooling all the data, we control for retest effects and estimate group differences by adding a) a dummy set indicating which group an individual belongs to (a level effect) and b) interacting the age splines with the dummy set (allowing for different age trajectories). We find:

- Overall predicted difference in average recall by age is estimated with negligible confidence bands (Figure 15)
- Higher recall scores for females, with convergence at high ages (Figure 16). No statistically significant difference was found with the Bayesian model to compare males and females using within-individual changes (Figure 17).
- Higher recall scores for individuals with high household incomes and convergence at high ages (Figure 18). No statistically significant difference was found when with the Bayesian model to compare individuals with different income deciles (using the income decile at wave T when assessing changes from T to T + 1 for any individual observed in both).
- Higher recall scores for those with high education, with limited indication of convergence at high ages (Figure 19). The Bayesian model revealed no statistically significant differences between the two based on within-individual variation.

The lack of statistically significant differences across groups when using within-individual change data may have several explanations. The simplest explanation is a lack of statistical power for detecting group differences in the SHARE data when using a model that avoids strong functional assumptions by modelling group deviations from a "common trajectory" using Gaussian Processes. As shown by the results contrasting males and females (Figure 17), the curve for females is below that of males as we would expect from the least squares model, but the credibility intervals of the estimated trajectories are sufficiently large to create large overlap. Alternatively, it may reflect that there are other explanations behind the convergence seen in the least squares results, such as cohort-differences, sample representativeness, etc.

As noted, changing selection into education means that these trajectories should not be taken to establish a long-term protective effect of education, because the high ages come from cohorts where the selection to education by cognitive ability probably was stronger.

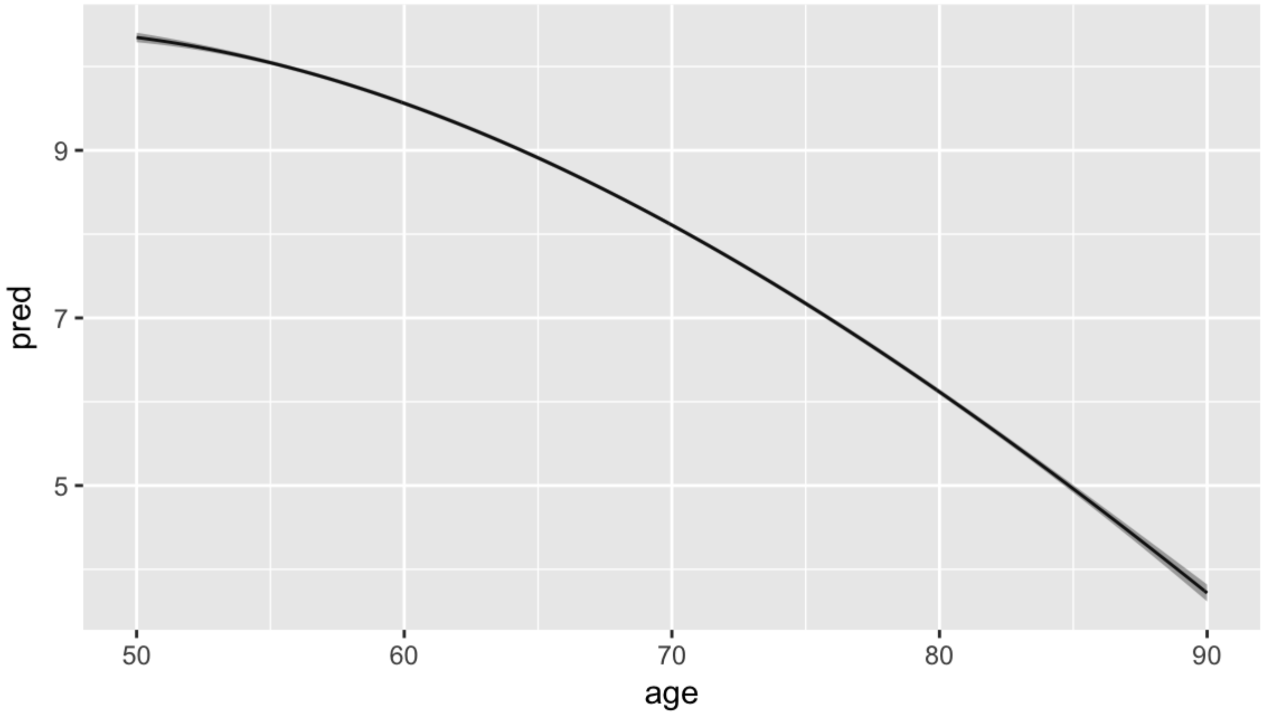

Figure 15 - Predicted trend in average recall across full sample

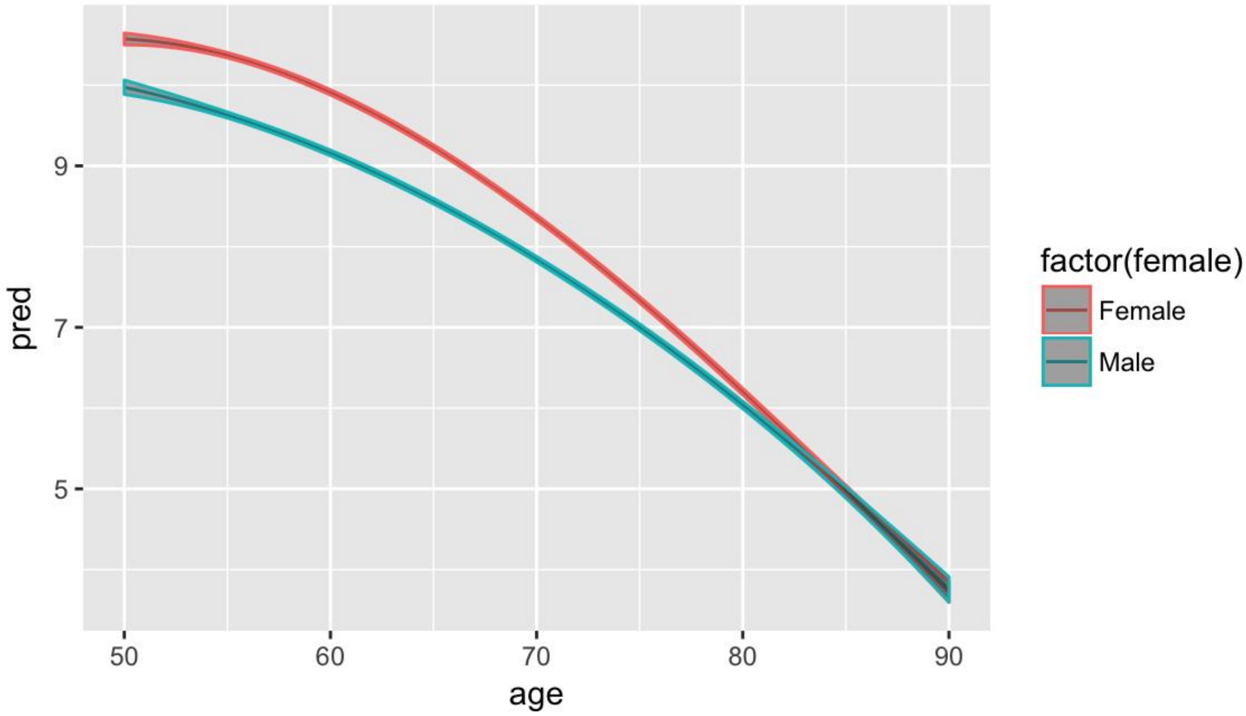

Figure 16. Predicted trend in average recall by sex

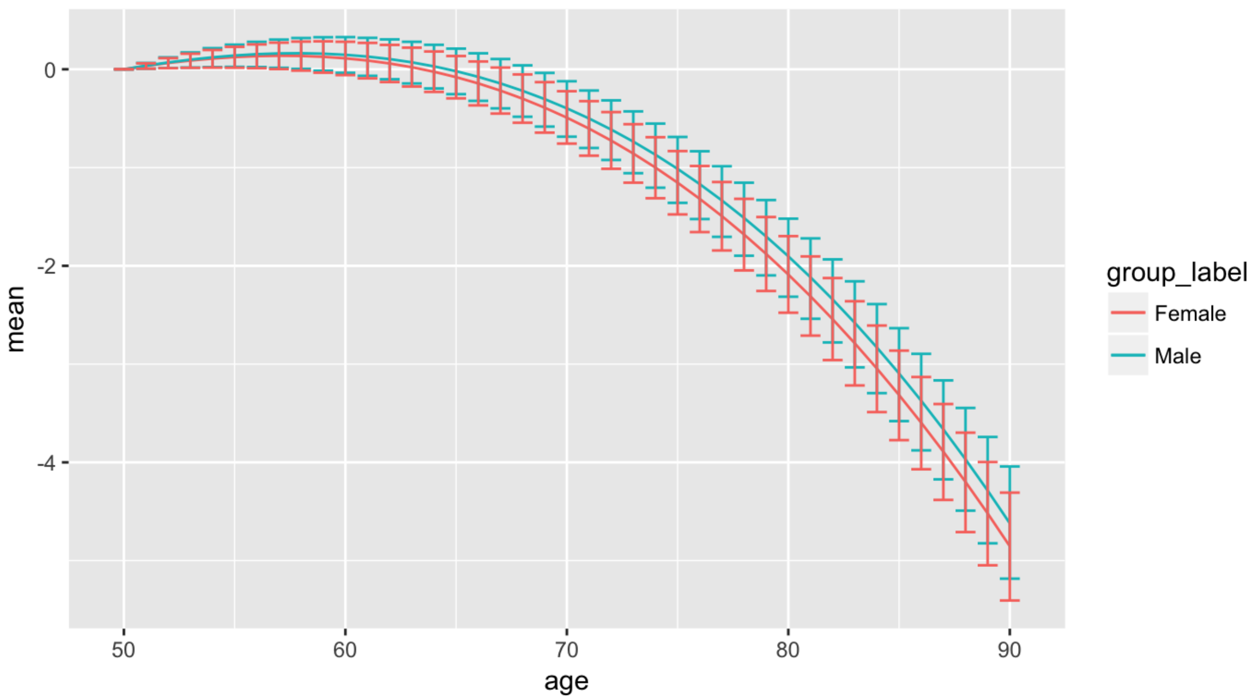

Figure 17. Variation in change trajectories by sex assessed using within-individual variation (Bayesian model)

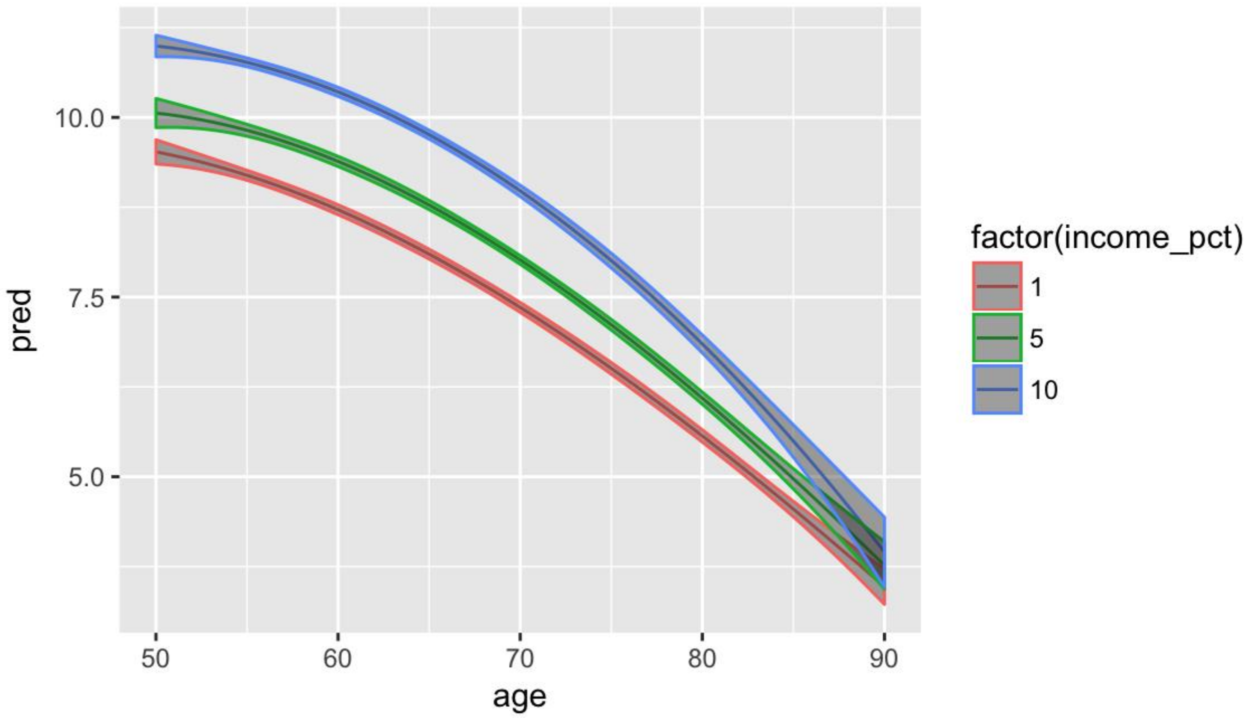

Figure 18. Predicted trend in average recall by three income deciles

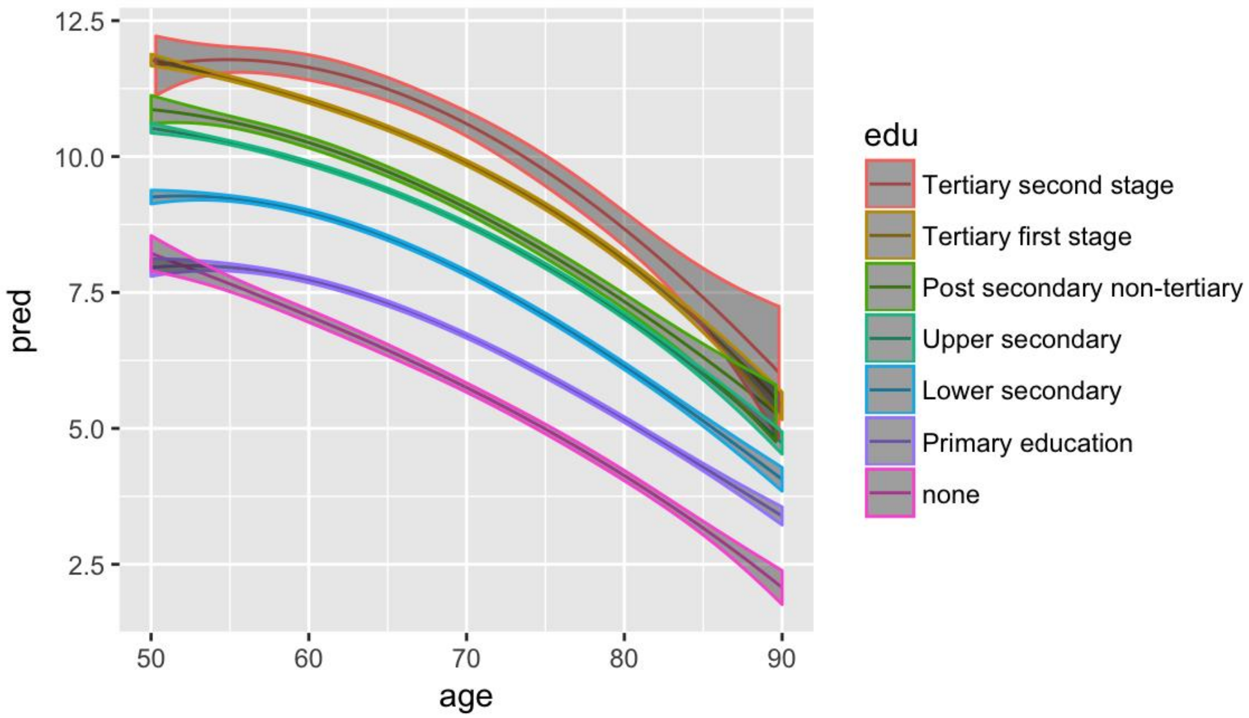

Figure 19. Predicted trend in average recall by educational level

3.3. Numeracy

For the two numeracy outcome measures, only the first measure is available throughout the whole data period. The second measure involved the task of repeatedly subtracting 7 from 100, was introduced in wave 3, and also has a limited amount of variation in scores at young ages with substantial ceiling effects in that more than 50% of the sample below age 80 years gets the highest possible score on the test (Figure 20).

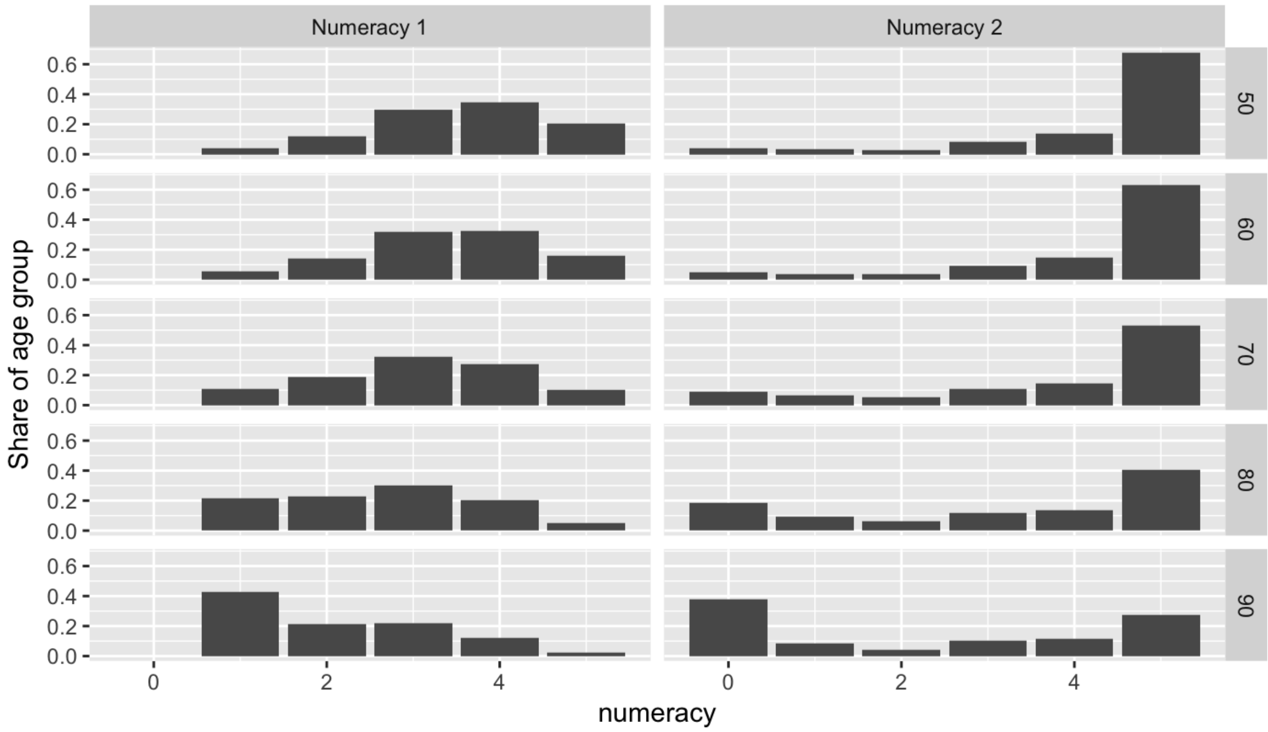

Figure 20. Distribution of numeracy scores by ten year age groups

Using only the scores on numeracy 1, a similar comparison of across- and within-individual variation to that conducted for recall, revealed no indication that retest effects were increasing with the number of previous tests, but a strong indication that being "ever-tested" influenced scores substantially (Figure 21).

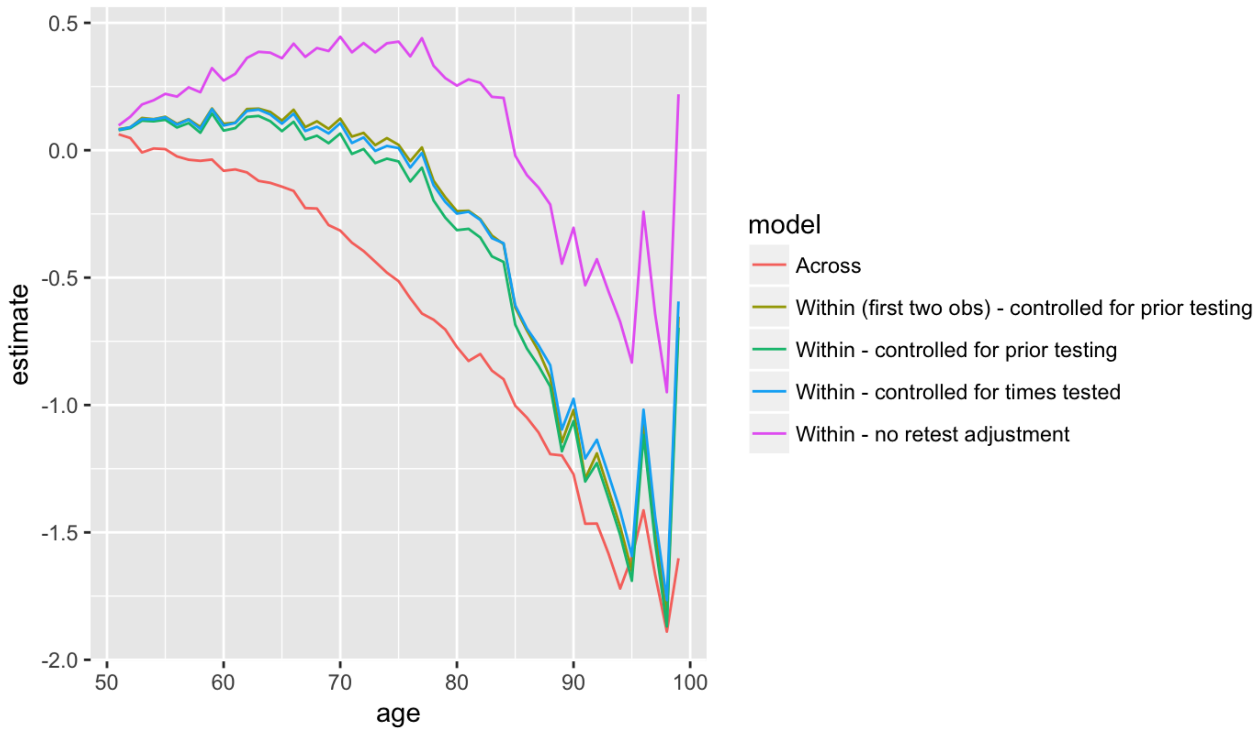

Figure 21. Age-related decline in numeracy 1 indicated by across and within individual information

For the remaining analyses, we pooled all data, controlled for "ever tested", and used dummy variables for groups to capture level differences and trajectory differences (age splines interacted with the group dummy variables). We observed:

- Predicted difference in average numeracy by age is estimated with negligible confidence bands (Figure 22)
- Higher numeracy scores for males, with limited convergence at high ages (Figure 23)
- Higher numeracy scores for individuals with high household incomes but convergence at high ages (Figure 24)
- Higher numeracy scores for those with high education, with limited indication of convergence at high ages (Figure 25). Note in particular that the predicted mean numeracy-1 score differs substantially, with most groups showing a lower decline than in the overall decline with age. In addition to the aforementioned changing selection across educational categories, this may suggest that the stronger age-related declines in numeracy-1 scores in earlier figures partly reflects lower education in the older cohorts.

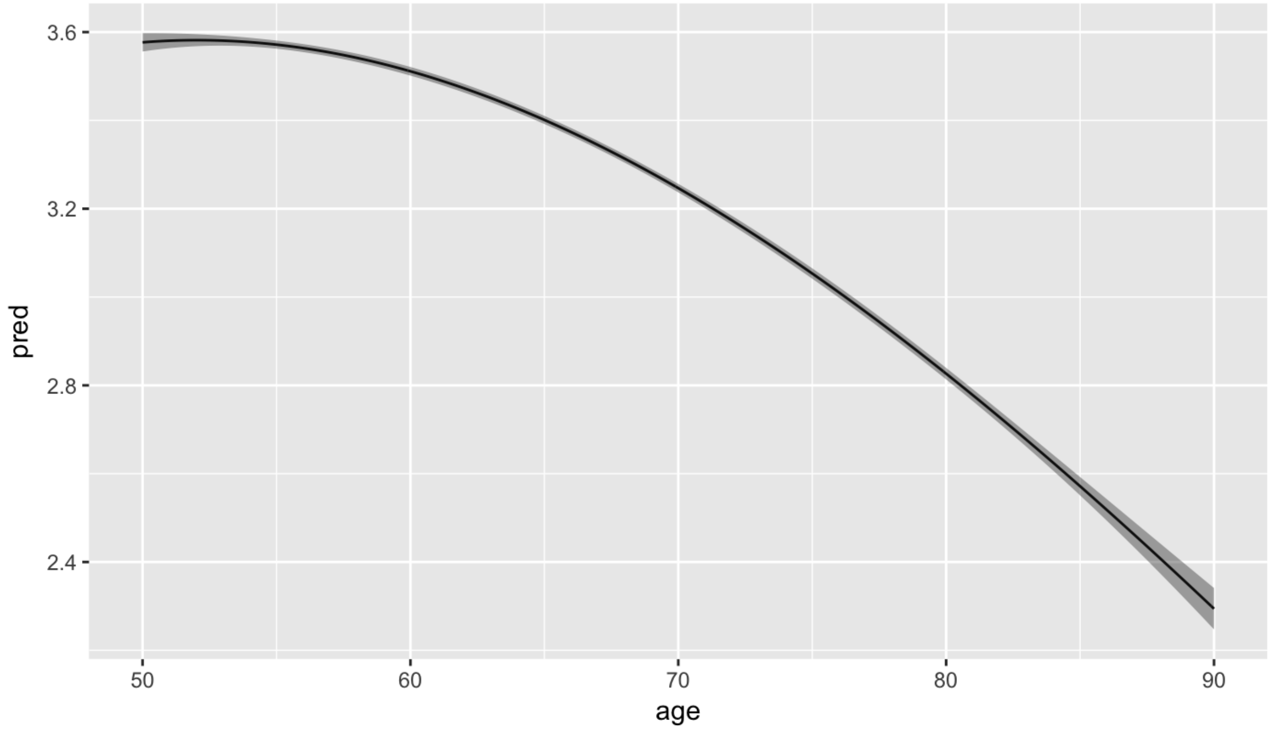

Figure 22. Predicted average numeracy by age across full sample

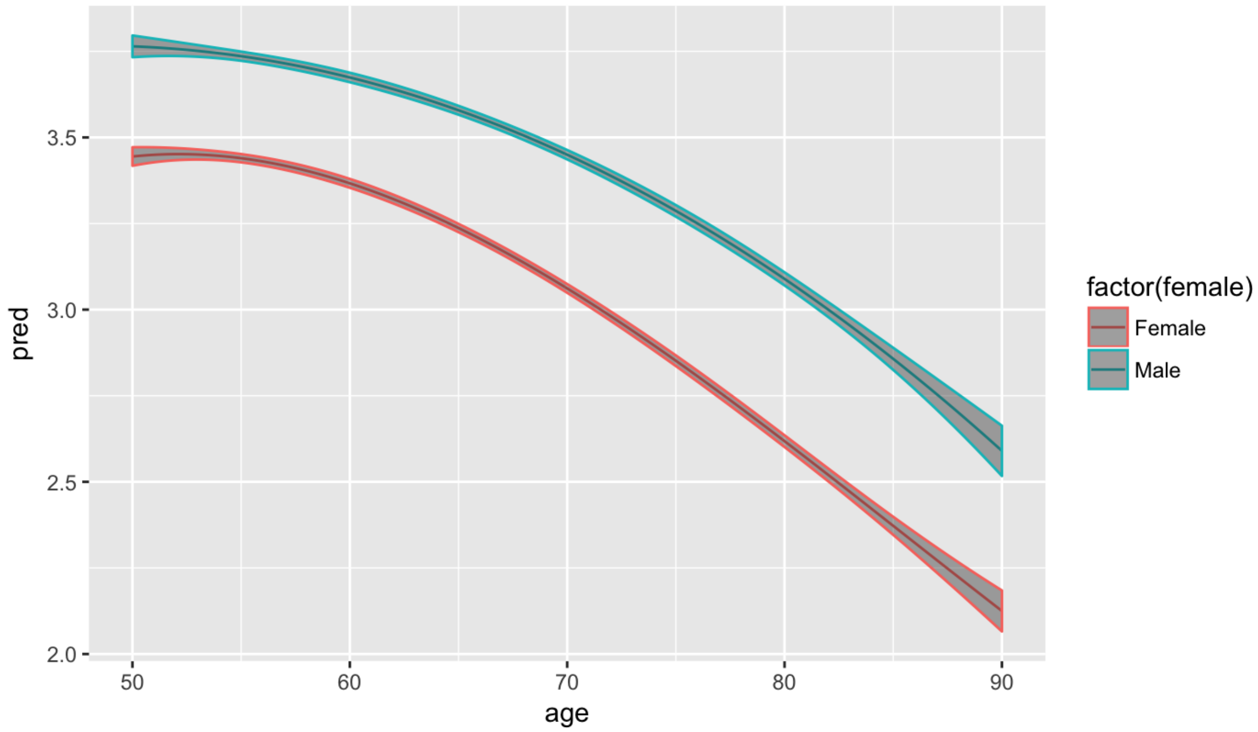

Figure 23. Predicted average numeracy by age and sex

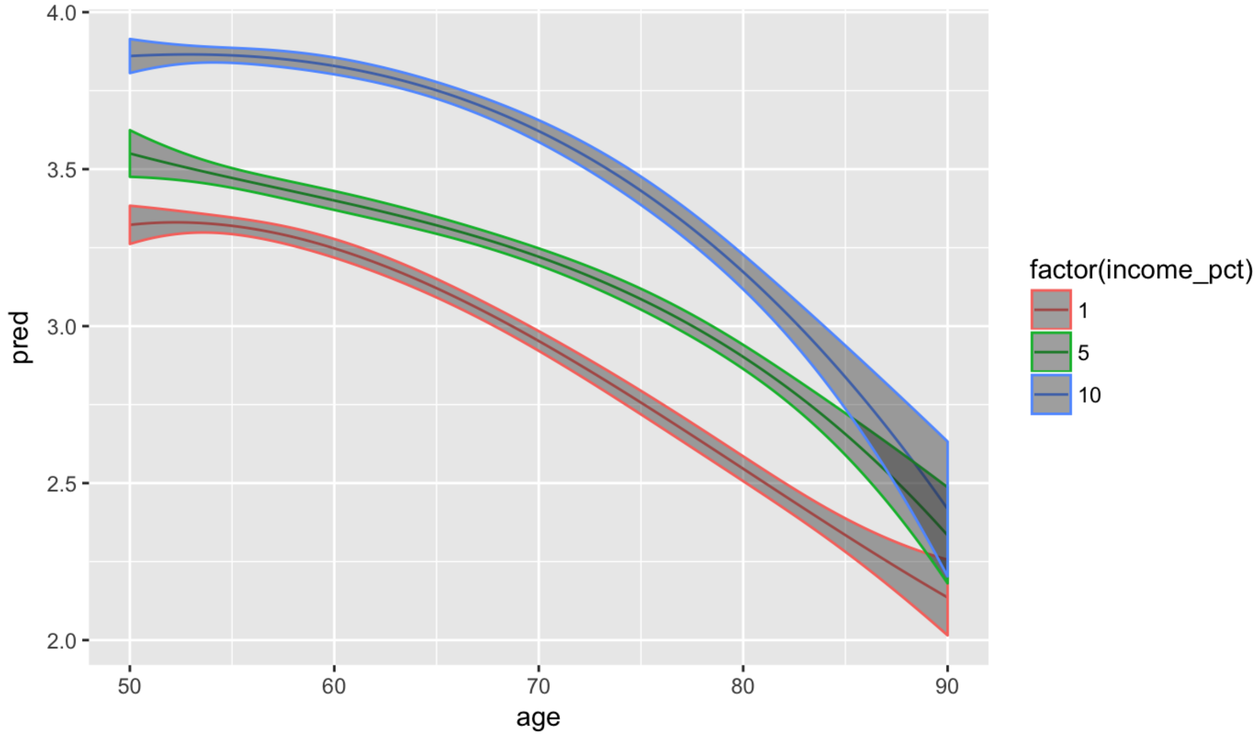

Figure 24. Predicted average numeracy by age for three income deciles

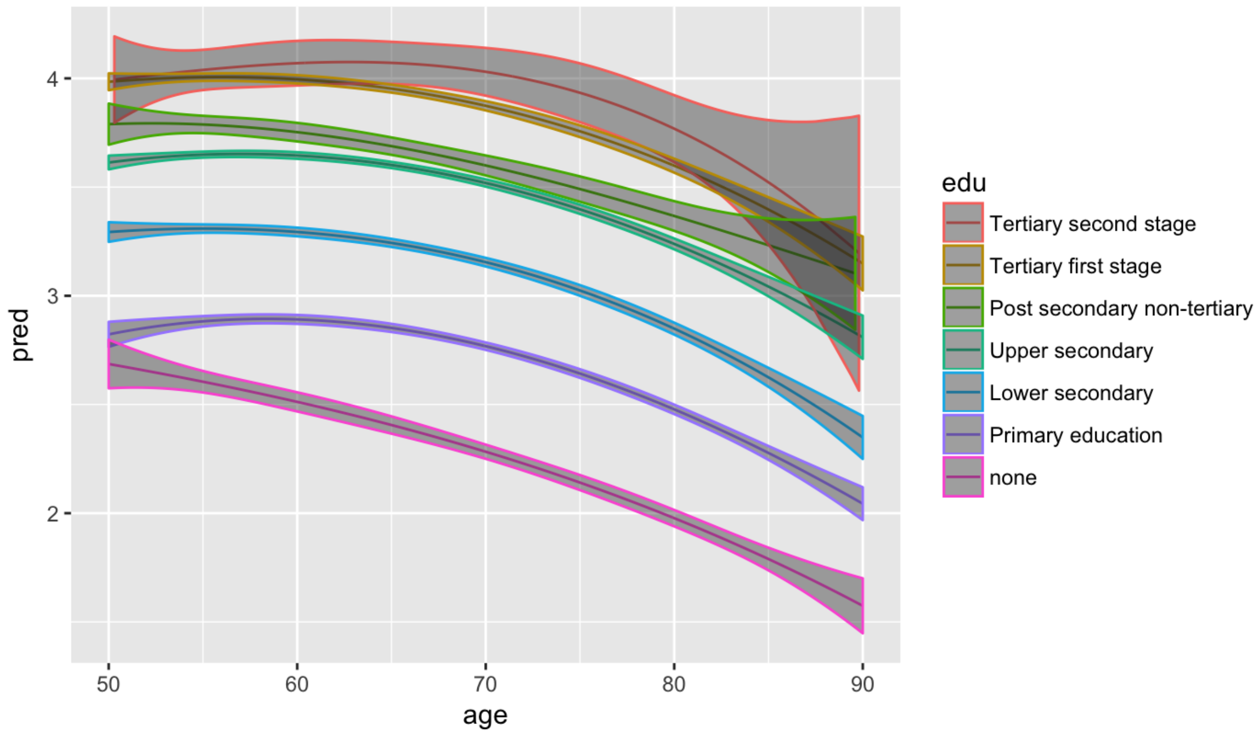

Figure 25. Predicted average numeracy by age and educational attainment

4. Country differences

4.1. Recall

To examine how the above patterns might differ by country, we allowed the level and development of the curves to differ by country by interacting the splines with the grouping variable. Whereas all countries showed the expected pattern of decline by age (Figure 26), the average predicted value differed between countries to a statistically significant extent, as can be seen by the nonoverlapping confidence bands when different nations are plotted in the same figure (Figure 27).

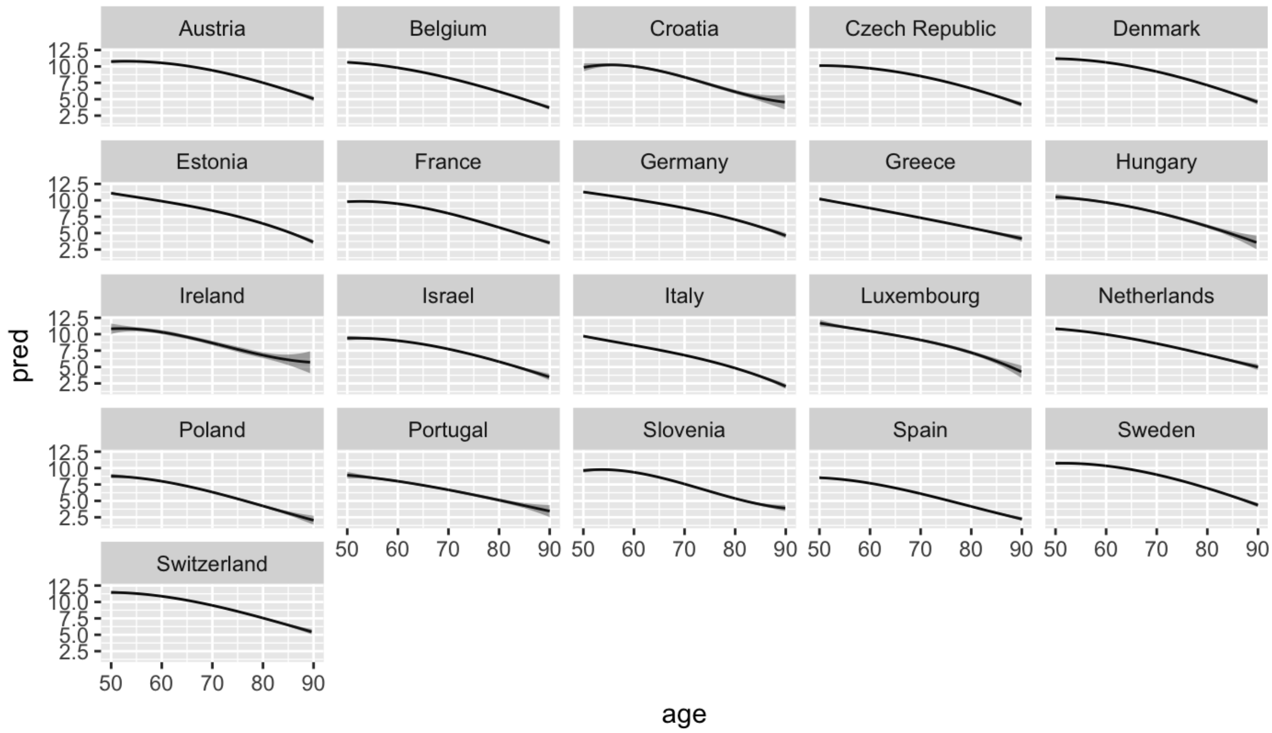

Figure 26. Predicted average recall by age and country

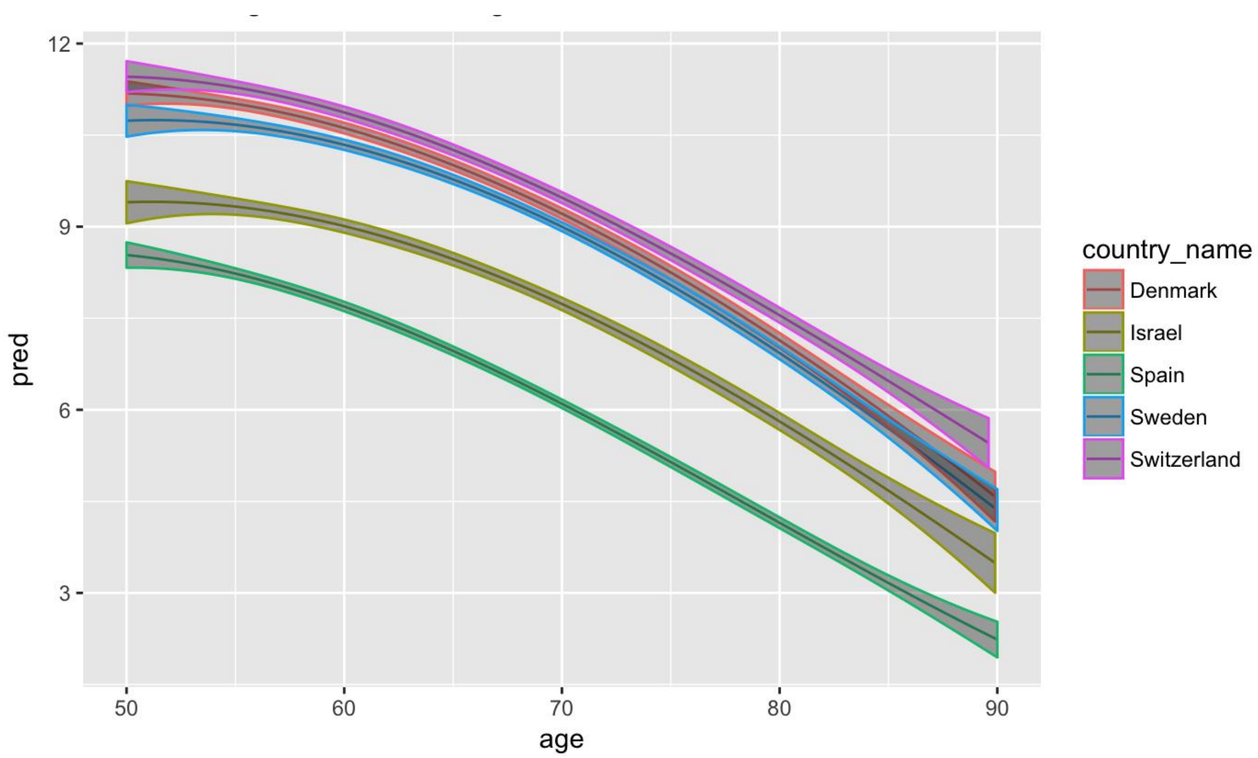

Figure 27. Predicted average recall by age for selected countries

Using the Bayesian model to estimate country-level differences in age-trajectories from within-individual variation, this model too suggests country-level differences in trajectories (Figure 28). The differences found here, however, do not show any correlation with the predicted difference (from age 50 to 90 years) when using the least squares regressions (Figure 29). The lower decline across the age-span identified in the Bayesian model should have limited emphasis given the difficulty of accurately adjusting for retesting when using individual variation alone.

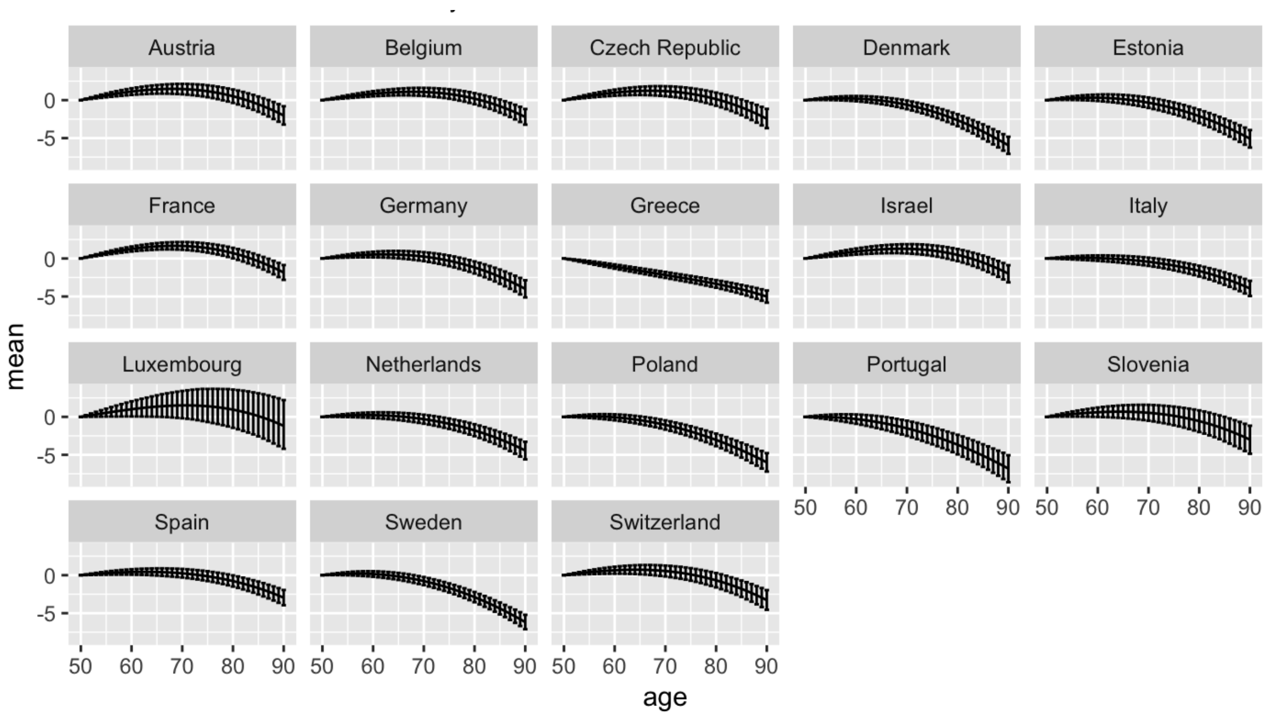

Figure 28. Variation in country level change trajectories assessed using within-individual variation (Bayesian model)

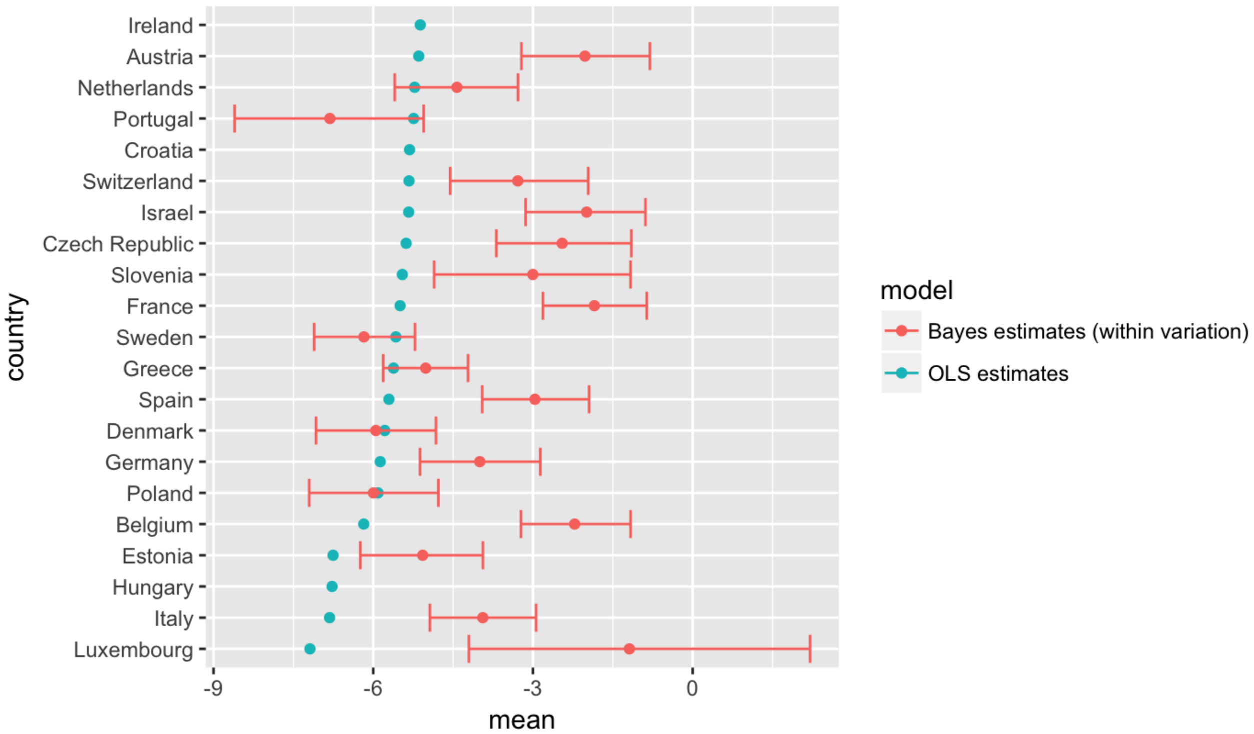

Figure 29. Comparison of predicted decline from 50 to 90 years of age (All variation to individual). The figure shows the decline predicted from the ordinary least squares regression as point estimates, whereas the decline across the same age span estimated in the bayesian model is shown with 95% credibility bounds using the samples drawn from the posterior distribution of the model parameters.

Separating the trajectories by sex, we see most countries reproducing the result that women display higher recall scores than men. Whereas a number of the country-level patterns are similar to the pooled results (reasonably stable difference between the sexes over time), some countries show larger differences at some age spans or no differences at certain ages (Figure 30). Given the uncertainties inherent in the data, however, it is probably best to view these country level differences with caution to avoid the risk of "explaining noise" in the estimates.

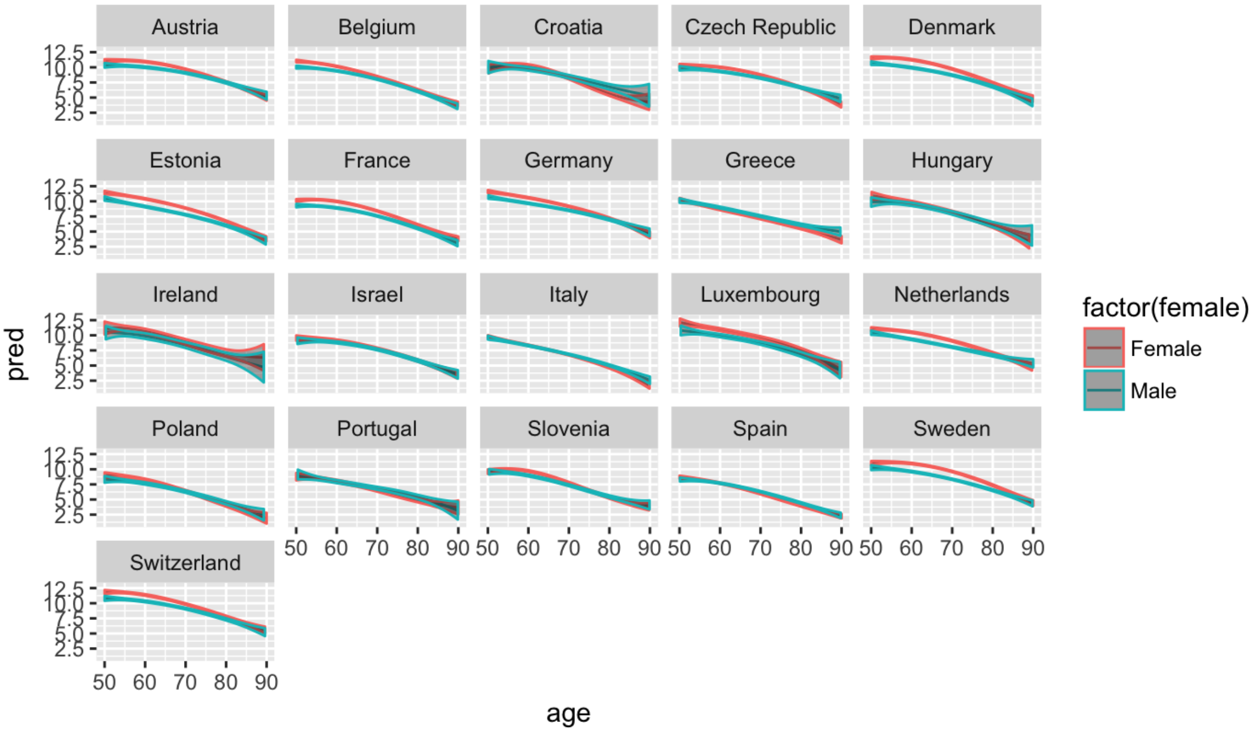

Figure 30. predicted average recall by age, sex and country

For income percentiles, we again observed the patterns from the pooled results appearing in most of the countries (Figure 31). For clarity, we here plot only the extreme groups. As seen in the confidence bands around the upper-income decile, the results for different income groups are imprecise at high ages when few of the respondents in the relevant age category are observed. Curves that are "shortened" relative to the full age-span reflect countries lacking observations from the displayed group in the relevant intervals.

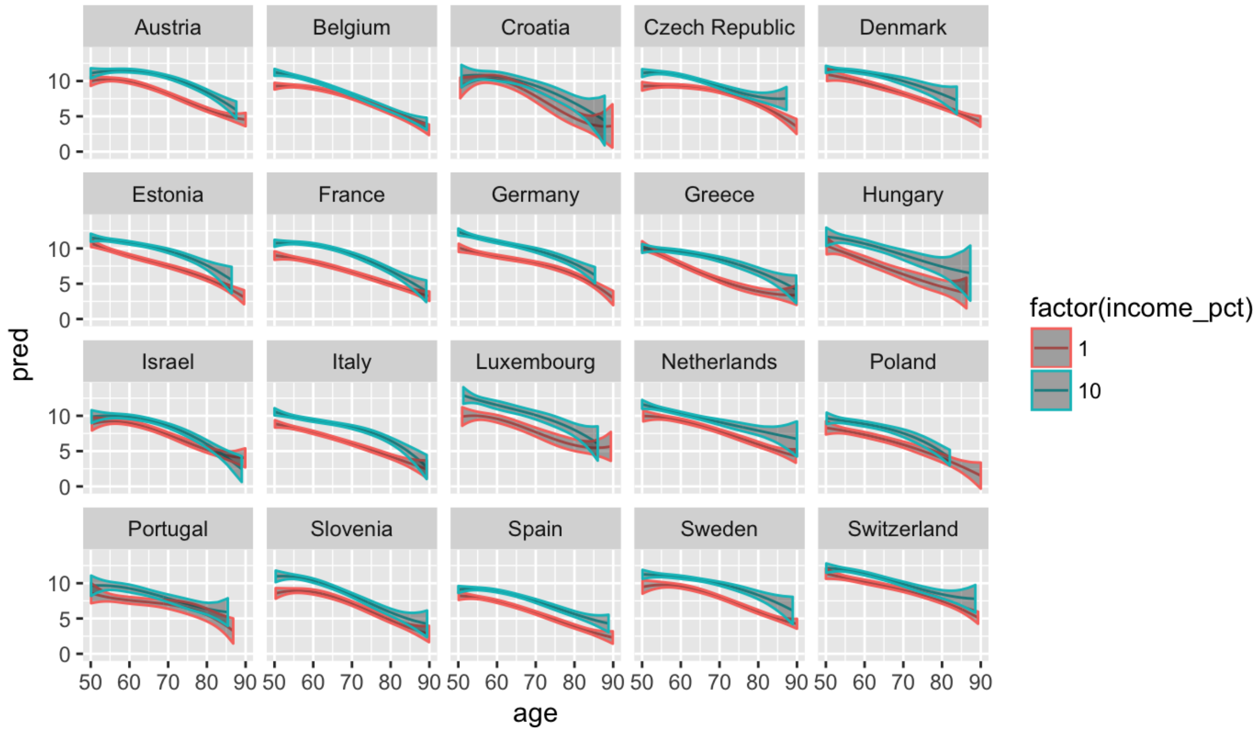

Figure 31. Predicted average recall by age and country for extreme income deciles

The trends across educational groups showed the difficulties due to changing educational attainment (and differing educational systems) across nations. To evaluate this, we first plotted the education-specific curves (without confidence intervals), which display a variety of odd and non-credible patterns (Figure 32). Choosing two reasonably common educational categories to display with confidence intervals, we see that this problem to some extent persisted (Figure 33).

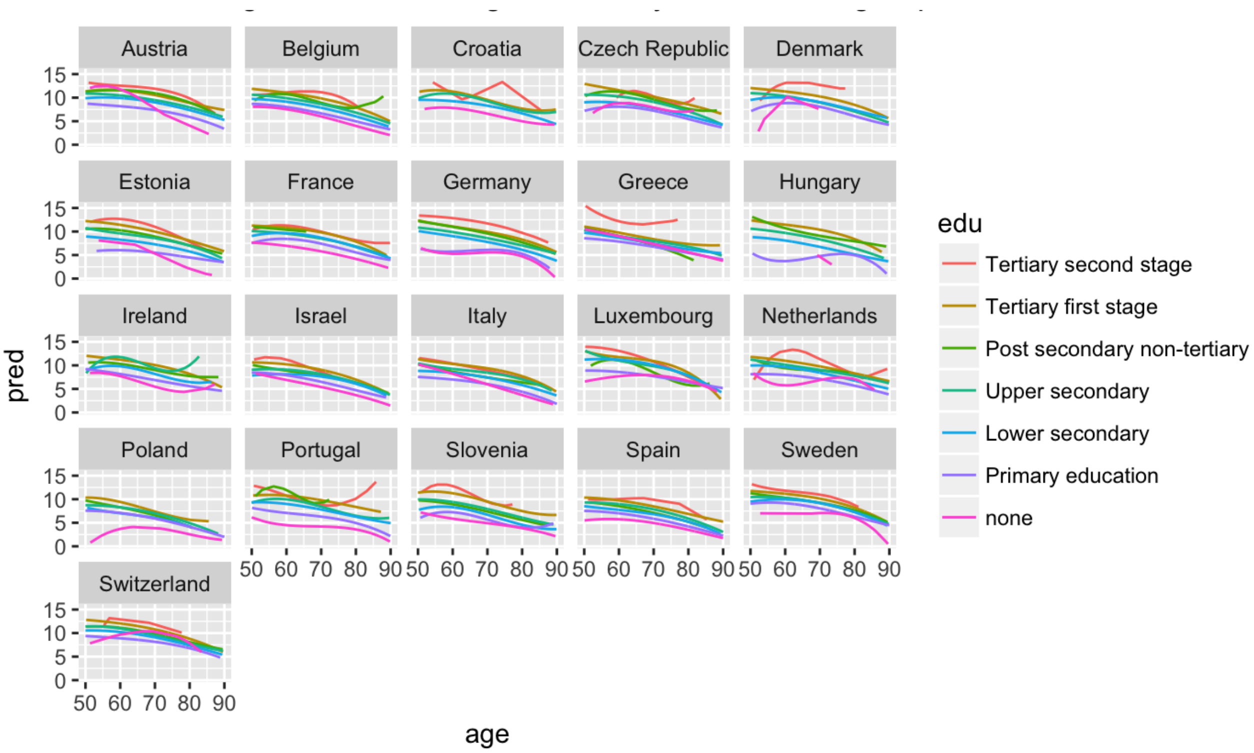

Figure 32. Predicted average recall by age, educational group and country

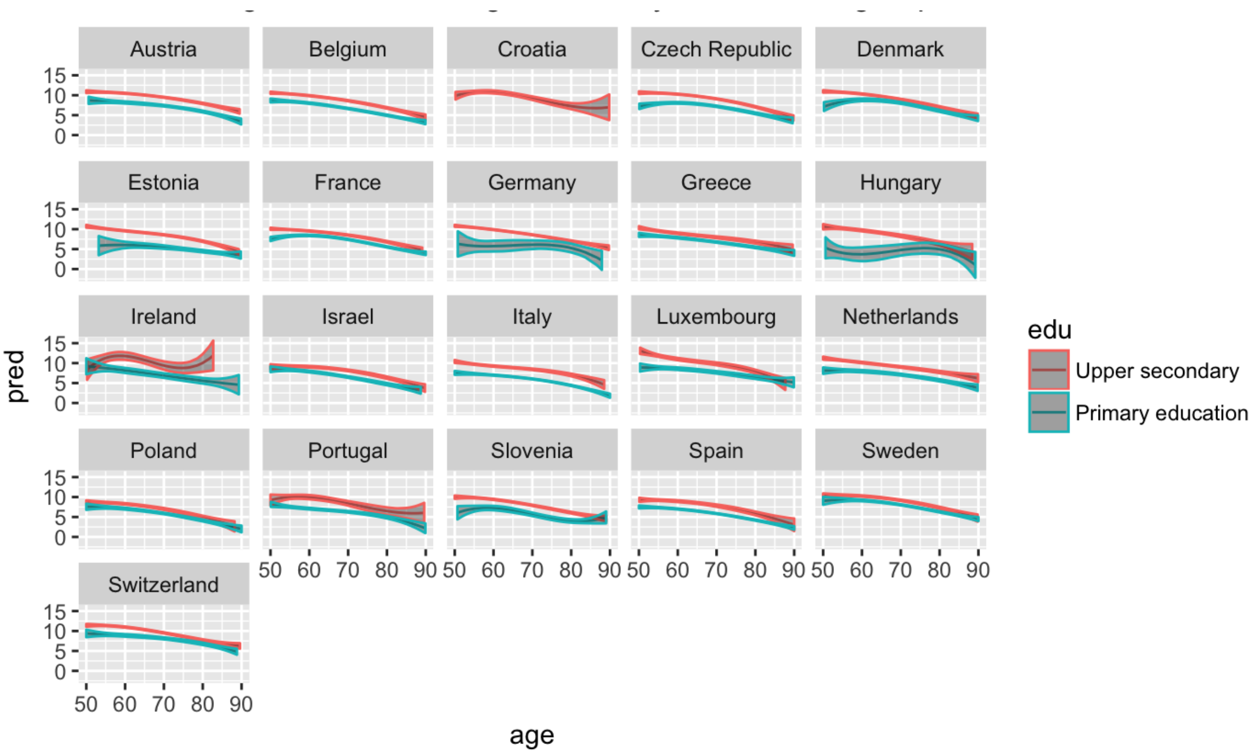

Figure 33. Predicted average recall by age and country for selected educational groups

4.2. Numeracy

For numeracy too, there were substantial differences in the relationship between age and predicted scores, although all countries showed evidence of declining scores by age (Figure 34 and Figure 35).

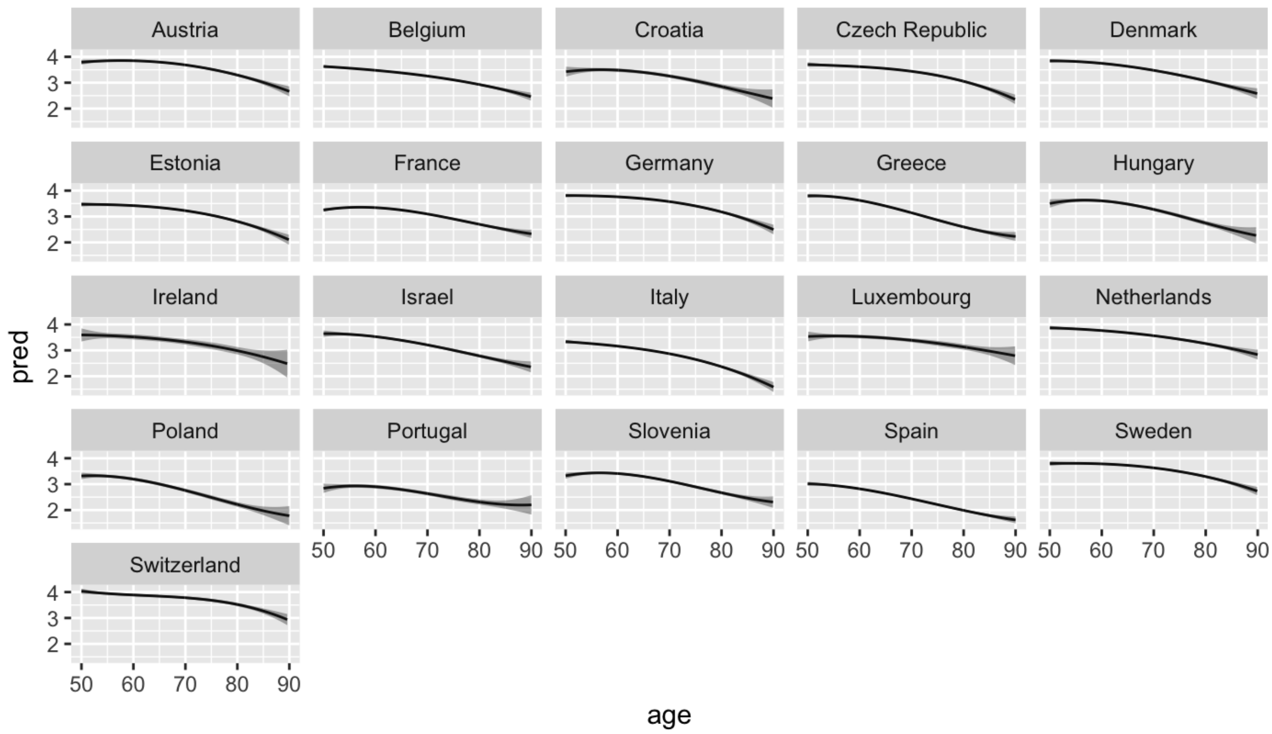

Figure 34. Predicted average numeracy by age and country

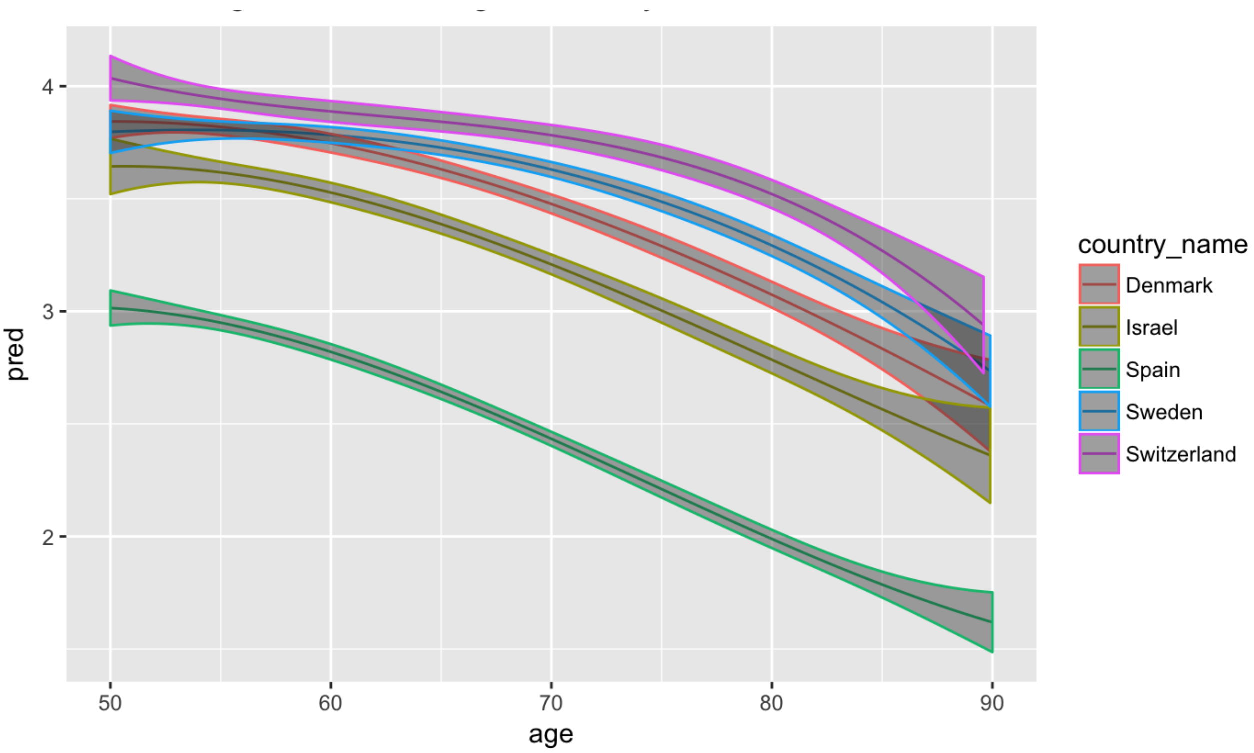

Figure 35. Predicted average numeracy by age for selected countries

Separating the trajectories by sex, we see most countries reproducing the result that men score higher on numeracy than women (Figure 36). Whereas many country-level patterns are similar to the pooled results (reasonably stable difference between the sexes over time), some country level data indicate larger differences at some age spans or no differences at certain ages. Given the inherent uncertainties in the data, however, it is probably best to view these country level differences with caution to avoid the risk of "explaining noise" in the estimates.

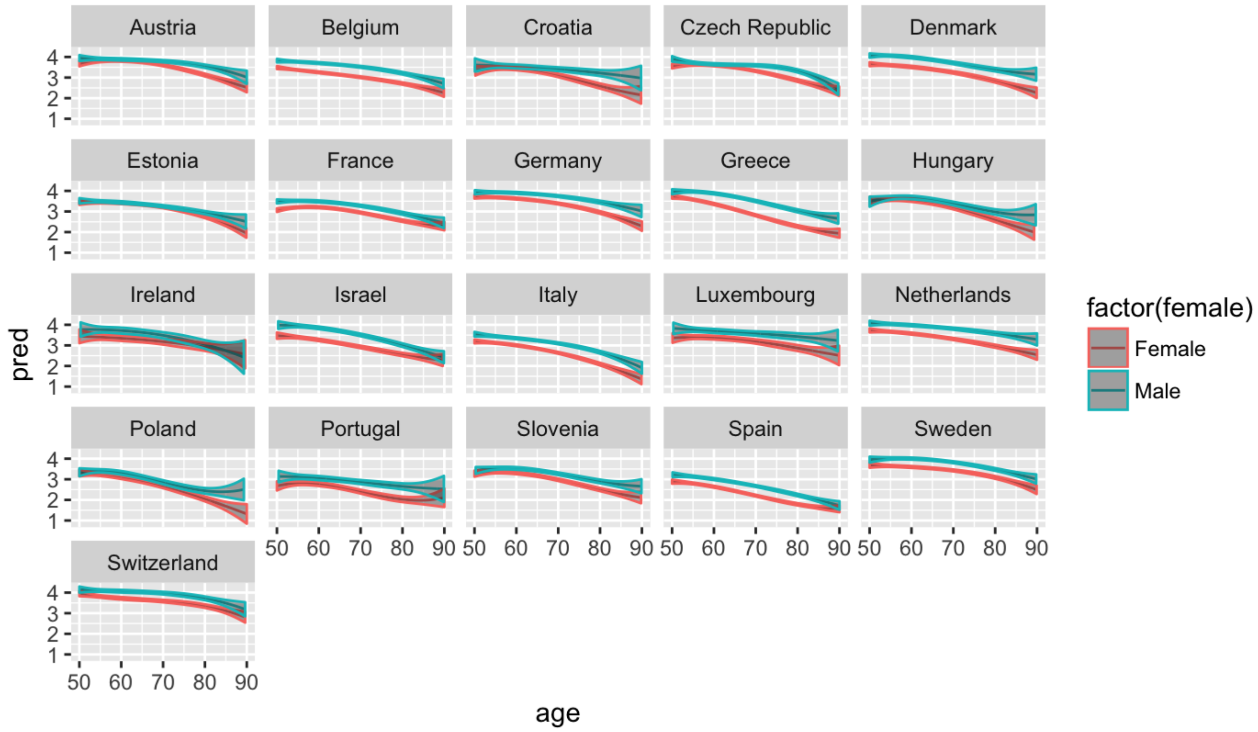

Figure 36. Predicted average numeracy by age, sex and country

Results by income group and educational category suggest that the more limited variation from the numeracy score seems to result in more unstable estimates in models with the current level of flexibility (Figure 37, Figure 38, and Figure 39).

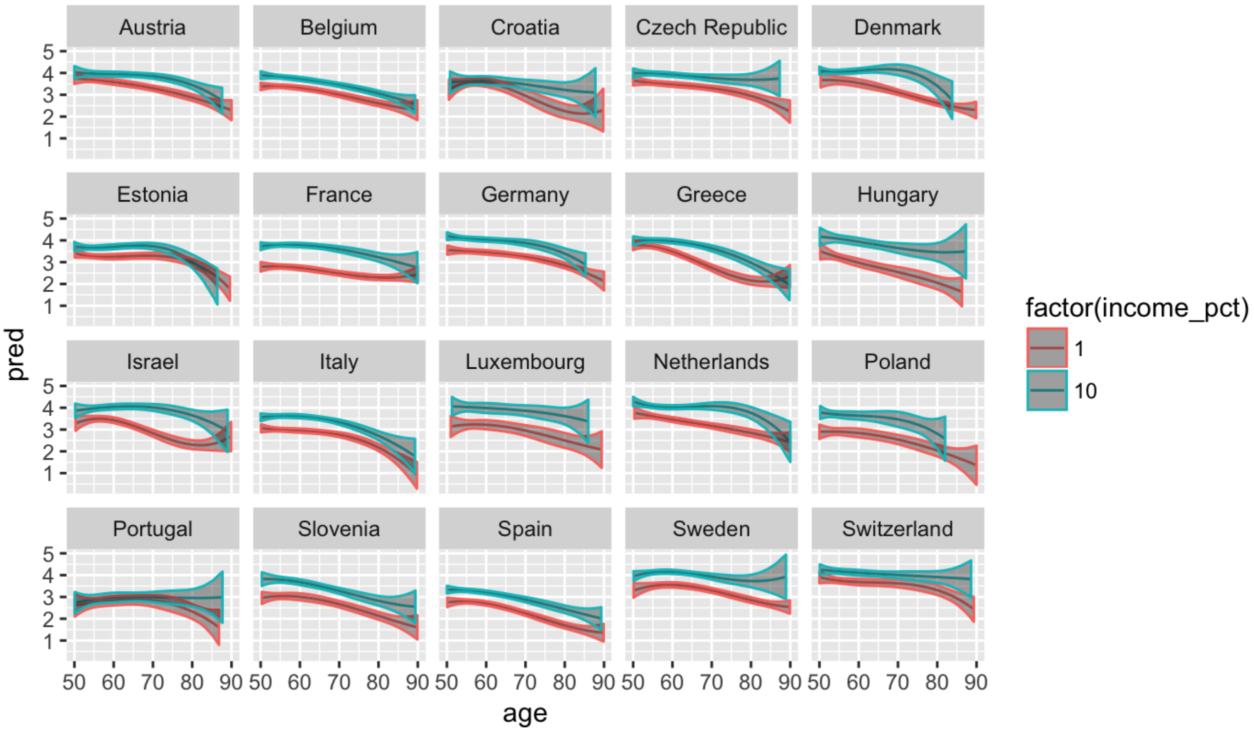

Figure 37. Predicted average numeracy by age and country for extreme income deciles

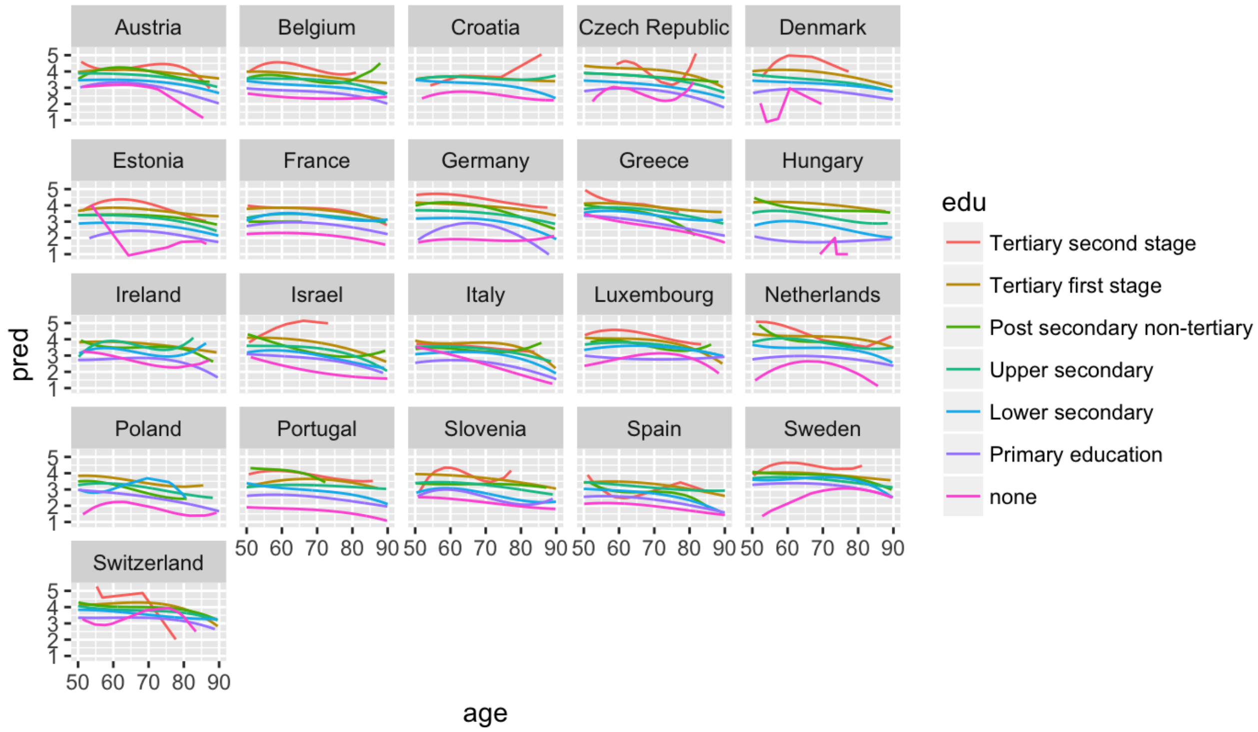

Figure 38. Predicted average numeracy by age, education and country

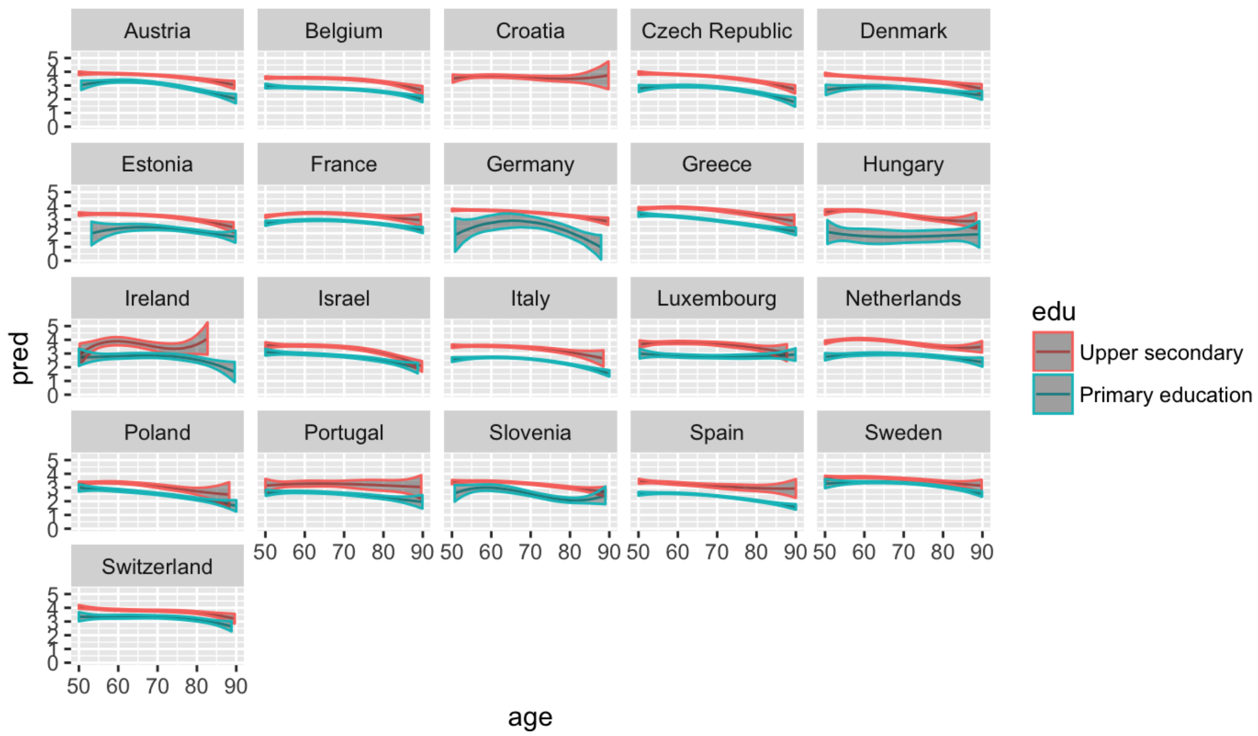

Figure 39. Predicted average numeracy by age and country for selected educational groups

5. Associations between cognitive scores and health/functioning

The different measures of health/functioning in the easySHARE data used are:

- Daily living (adla - activities of daily living – 4 items). Items like "Dressing, including putting on shoes and socks", "Bathing or showering", "Eating, such as cutting up your food"
- Living – instrumental (iadzla - instrumental activities of daily living index 2 – 5 items). Items like "making telephone calls", "Taking medications", "Managing money"
- Finemotor (finemotor – 3 items). Items like “picking up small coin from table”, “Cutting up food”, “dressing including shoes and socks”
- Grossmotor (grossmotor – 4 items). Items such as “Walking 100 meters”, “Walking across room”, “getting in or out of bed”, “bathing/showering”, “Climbing one flight of stairs”
- LGmuscle (lgmuscle - large muscle index – 4 items). Items such as "sitting for about two hours", "Getting up from a chair after sitting for long periods", "Pulling or pushing large items like a living room chair"
- Mobility (mobilityind – 4 items). Items such as "walking 100 metres", "walking across a room" and using stairs

Note that for these outcomes, a higher score indicates a larger number of problems, i.e. poorer functioning. This means that the average scores are expected to increase with age, reflecting an increased average number of items on which participants have reduced functioning.

To assess the associations between cognitive scores and these different measures of health/functioning, we specified a regression model with a simpler form for the age-trajectory (a second order polynomial) estimated on a dataset containing only the first observation of each individual (to avoid spurious patterns driven by retest-effects in the explanatory variables). Using both orientation, the two recall measures and numeracy-1 as linear predictors simultaneously orientation, immediate recall measures and numeracy-1 were statistically significantly associated with all outcome measures, while delayed recall was statistically significant associated with all except “daily living” (adla). To visualize the magnitude of the associations, we plotted the predicted age curves for a hypothetical individual with the highest score on a specific cognitive measure throughout the period relative to one with the lowest score on the same measure throughout the period (Figure 40). Note that, in reality, most people would have a mix of such curves as their cognitive scores would tend to fall with age. The figure shows that low scores are associated with more problems in daily living as well as reduced motor skills and mobility. Scores on orientation, immediate recall and numeracy, however, predict significantly larger differences in health/functioning than delayed recall. In general, the cognitive measures predict a larger impact on mobility and gross motor/large muscle indicators than they do on finemotor and daily living measures. The exception to this is orientation, which predicts large differences across most measures but particularly for daily living and living instrumental.

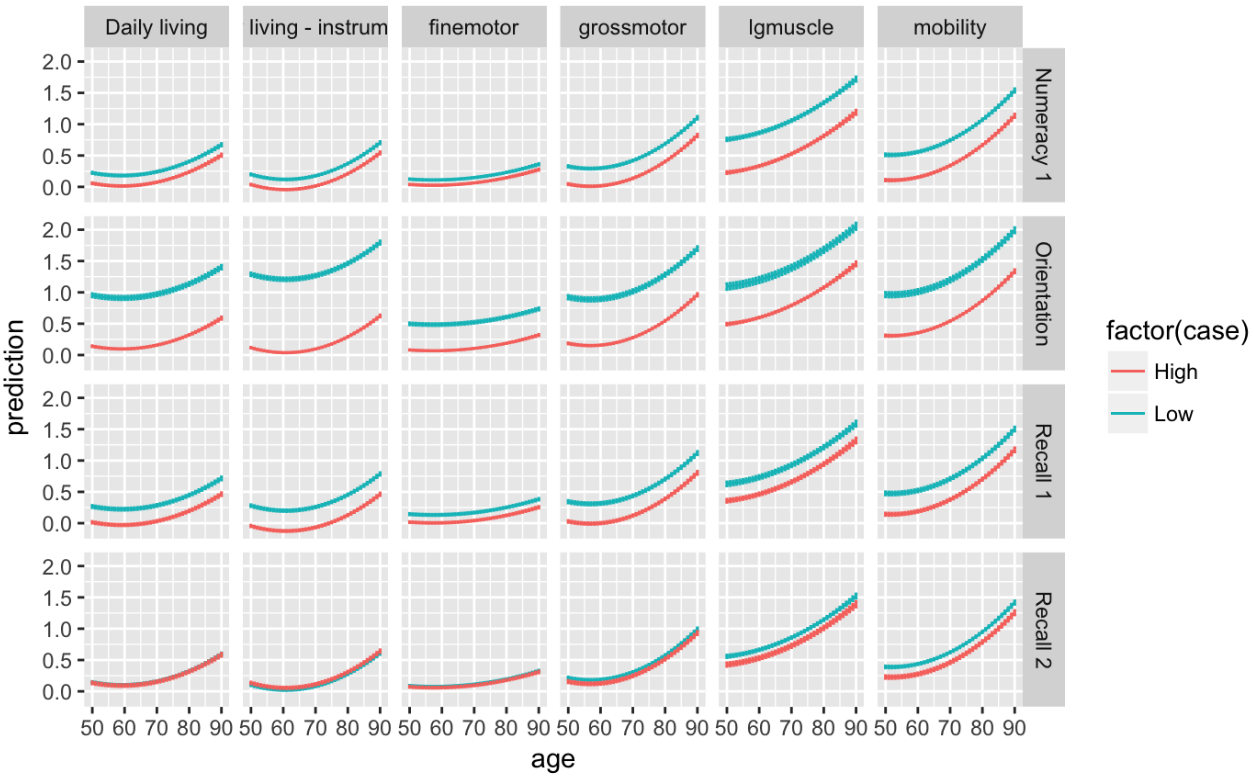

Figure 40. Predicted health/functioning issues by age – contrasting hypothetical ”persistent high scorer” with ”persistent low scorer” for each cognitive measure.

6. Conclusion

Cognitive outcome measures in the Survey of Health and Ageing in Europe can be used to describe differences in cognitive abilities across groups defined by age, sex, educational level, household income, and country. The identified patterns should be considered predictive and not causal: they show the expected average ability for individuals with some specified age, sex, income, etc. in the SHARE population. The patterns identified do, however, suggest that there may be substantial differences in average cognitive abilities between countries, educational and income groups, and that the outcome measures show different associations with age across these groups. We identify patterns that potentially may be useful for Lifebrain cohorts from different countries or European areas covered in both datasets.

Attempts to identify ways that group differences differ by country (using models with three-level interaction effects for age, group and nation) were less clear.

All cognitive outcome measures showed statistically significant and typically substantial associations with outcomes of daily functioning and health. As with the other analyses, the potential for causal conclusions is limited: reduced cognitive ability may curtail activities that promote physical health, reduced physical health may curtail activities that promote cognitive ability, or a “general functioning” factor may influence both physical and mental development over time.