CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Does growing up surrounded by books improve our cognitive skills as seniors?

Bachelor's thesis

Author: Julie Kovácsová

Study program: Bakalářský program Ekonomie a finance

Supervisor: PhDr. Mgr. Jana Votápková, Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, January 3, 2023

Julie Kovácsová

Abstract

This thesis examines whether growing up surrounded by books has an impact on cognitive skills later in life. Data from SHARE - Survey of Health, Ageing and Retirement in Europe are used. The variable 'books when ten' is used to determine its effect on variable 'numeracy'. We used also other variables regarding the respondent's age, gender, mental health or family situation. The results of this study show that those who were surrounded by a higher number of books as children perform better at a specific task designed to test cognitive ability when presented with it at an age of over 50 years. The results were obtained using logit model. The finding that access to books can slow down cognitive decline later in life can be used to argue for greater focus on reading and availability of books in schools as well as greater access to books through public libraries, or act as a hint for parents aiming to give their children the greatest possible introduction to life.

Keywords books, education, congnitive skills, SHARE

Title Does growing up surrounded by books improve our cognitive skills as seniors?

Abstrakt

Tato práce zkoumá, zda má vyrůstání obklopené knihami vliv na kognitivní schopnosti v pozdějším věku. Jsou použita data z průzkumu SHARE - Survey of Health, Ageing and Retirement in Europe. Proměnná "knihy v deseti letech" je použita ke zjištění jejího vlivu na proměnnou "početní dovednosti". Použili jsme také další proměnné týkající se věku, pohlaví, duševního zdraví nebo rodinné situace respondenta. Výsledky této studie ukazují, že ti, kteří byli v dětství obklopeni větším počtem knih, dosahují lepších výsledků ve specifickém úkolu určeném k testování kognitivních schopností, když je jim předložen ve věku nad 50 let. Výsledky byly získány pomocí logistické regrese. Zjištění, že přístup ke knihám může zpomalit pokles kognitivních schopností v pozdějším věku, lze využít jako argument pro větší důraz na četbu a dostupnost knih ve školách i pro větší přístup ke knihám prostřednictvím veřejných knihoven nebo může sloužit jako nápověda pro rodiče, jejichž cílem je poskytnout svým dětem co nejlepší vstup do života.

Klíčová slova knihy, vzdělání, kognitivní schopnosti, SHARE

Název práce Ovlivňuje přístup k literatuře v dětství kongnitivní schopnosti člověka v seniorském věku?

Acknowledgments

I am very grateful to my supervisor PhDr. Mgr. Jana Votápková, Ph.D.for her patience, time and willingness to help.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Kovácsová, Julie: *Does growing up surrounded by books improve our cognitive skills as seniors?*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 50. Advisor: PhDr. Mgr. Jana Votápková, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

SHARE Survey of Health, Ageing and Retirement in Europe

Chapter 1

Introduction

As many people experience cognitive decline associated with either advancing age or due to other factors, it is important to look into how cognitive decline can be avoided or at least mitigated. Many studies and research papers have been investigating what influences cognitive impairment.

The objective of this thesis is to study the effect of ease of access to books as a child on cognitive ability in late adulthood. Early childhood literature exposure will be in our main interest. We will look at how having access to books might impact a person's cognitive abilities in adulthood, specifically numeracy, focusing on the differences between those who grew up in book-rich homes and those who did not.

The focus is specifically centered around access to books, not reading itself. More specifically, one of the variables in the research was the number of books a person had in their home at age ten, not how many books a person read (this is a possible extension of our work, however the dataset we are working with does not provide us with such information). Access of books is undoubtedly linked with reading itself, however, the aim of this thesis is to determine whether the simple fact of being able to access books easier than others offers in itself an advantage in being linked with a higher cognitive performance in adults over the age of 50.

The paper considers the implications of these findings for the promotion of early childhood literacy and improved cognitive functioning in older adults. Our findings might influence literacy initiatives or to form educational policies regarding the importance of book exposure.

The studies and research papers show that cognitive flexibility develops rapidly during childhood and continues to improve during adolescence and young adulthood and is even related to health status during adulthood (Buttelmann & Karbach 2017), but declines with advancing age (Salthouse 2009). Next to physical health, being independent and having a sharp mind is considered important within the elderly generation. Better cognitive skills development due to book exposure as a child definitely has an impact on life quality as cognitive decline is considered a major health and social issue (Deary *et al.* 2009a). It is therefore important to think about future cognitive skills and the possible ways to improve them as early as possible.

The thesis is structured into six chapters. In the first chapter, we are providing brief overview of the context of the study. This is followed by the second chapter in which we go through the literature and research that is relevant to out study. The third chapter describes the process of obtaining our data and describing the whole data set as well as individual variables used in our research. Chapter four describes methodology used in our analysis, specifically logistic regression model and chapter five goes through the results of our study and states the limitations. In the final chapter, we provide final overview of our research and our conclusions.

Chapter 2

Literature review

The aim of this chapter is to describe, analyze and give an overview of the existing studies and research papers related to the topic of this thesis and identify gaps in the current state of scientific findings as well as to summarize the most important determinants of cognitive decline and then suggest a solution to possibly mitigate the impact of these determinants in our thesis and in real life.

2.1 Cognition throughout the lifetime

Existing research revealed countless reasons for cognitive decline, many of which can not be influenced by people themselves. We are talking about a link between genetics and cognitive ageing. Study Deary *et al.* (2009a) has estimated that the heritability of general cognitive ability is around 50% (by using data from twins and families with adopted children) (Deary *et al.* 2009b), increasing from childhood to adulthood and into old age, but decreasing in very old age. Studies have looked for associations between variations in specific genes and age-associated cognitive decline, finding a link between these genes and cognitive impairment. There are genetic influences on age-associated cognitive decline (Deary *et al.* 2004).

We, unfortunately, do not have an access to data to study cognitive decline linked to genetics and there is no way to eliminate or determine the influence of genetics on our data set and results. Furthermore, not even these studies focused on genetics, did come to a final conclusion which genes specifically influence what kind of cognitive decline - also because different genes might be

important for different populations or age groups (Deary *et al.* 2004). In this thesis, determinants that are more easily measurable or influenced and more easily implemented to have cognition enhancing or deteriorating influence will be studied.

2.1.1 Influence of age

During the preschool years, cognitive flexibility develops quickly and carries on to enhance as an individual grows into adolescence and young adulthood. This ability has been associated with many positive outcomes in life, like academic success and overall health (Buttelmann & Karbach 2017). However, cognitive abilities often decline with age. Beginning in healthy, educated adults in their 20s and 30s, some aspects of age-related cognitive decline can be seen, and this decline increases with age, from 18 to 60 years old (Salthouse 2009).

According to Deary *et al.* (2009a), verbal ability, numerical abilities and general knowledge are not as affected by age, while memory, executive functions, processing speed, and reasoning decline from middle age on. In this thesis, we will be examining numeracy as a dependent variable to possibly avoid the influence of age related memory impairment on the research. The evidence of this theory and further thought is discussed in the following chapters. This approach has proven to be valid as we have found only little influence of age on numeracy test that we used as an indicator of cognitive skills.

Existing studies have identified many other determinants that can have either a positive or negative impact on someone's cognition, that are rather easily influenced by the person themselves and their behaviour or lifestyle. The findings of study by Lipnicki *et al.* (2019) imply that education, smoking, physical activity, diabetes, and stroke may all be alterable aspects correlated with cognitive decline. Regulating them could reduce global rates of cognitive decline. Many studies identified steps that can be taken to slow down the process of cognitive decline.

2.1.2 Exposure to books

A study by Sikora *et al.* (2019) shows that adolescent exposure to books had a positive effect on adult literacy, numeracy, and technology skills. The effect was stronger in more developed countries, and it increased with the number

of books available during adolescence and the amount of time spent reading. The study also found that the effects of adolescent reading on adult skills were greater in countries with higher levels of educational attainment.

This research paper examines the influence of adolescent reading on adult literacy, numeracy, and technology skills in 31 societies and is based on data from the Programme for the International Assessment of Adult Competencies, a survey of adult skills conducted by the Organisation for Economic Co-operation and Development. The findings are based on a sample of over 25,000 adults aged 16 to 65 who were surveyed in 31 countries. The survey included questions about the number of books in the respondent's home during adolescence, from ages 10 to 15, and the amount of time spent reading books during adolescence. Questions about adult literacy or technology skills were included. Unlike our thesis, this paper takes into consideration also the amount of time spent reading during adolescence and is conducted on a different age spectrum of both children and adults as well as in different regions.

The study from Manu *et al.* (2019) suggests that providing children with access to books is an important factor in helping to develop their literacy and numeracy skills. This study however assessed children aged between 36 to 59 months and tested skills such as naming letters or numbers, without reflecting their development later in life.

As a cognitively stimulating activity, reading books also results in a reduced risk of dementia developing. (Anme *et al.* 2013). There is evidence that taking part in mentally stimulating leisure activities may help reduce the danger of cognitive decline in later life. It may be beneficial for governments and health services to access books to all people and prioritize getting them involved in stimulating activities, such as reading throughout their lifespan as a primary means of prevention (Yates *et al.* 2016). It was also proven that more frequent engagement in mentally stimulating activity in old age also leads to better cognitive functioning (Wilson *et al.* 2012), therefore encouraging people of all age groups to own books and to read them is beneficial.

2.1.3 Importance of education

A study came up with an interesting findings about the relationship between education and books availability, adolescents with lower secondary education

who grew up with many books become just as literate, numerate, and technologically proficient in adulthood as university graduates who grew up with fewer books (Sikora *et al.* 2019). Education in relation to book exposure is therefore considered an important indicator of cognitive skills and we will examine it in our research.

2.1.4 Influence of having children and grandchildren

Existing research suggests that childbearing may have a negative effect on cognitive health in later life (Bonsang & Skirbekk 2022). This study examined the association between childbearing and cognitive health in later life using an instrumental variable approach. The study found that having three or more children was associated with lower cognitive scores, as well as with a faster rate of cognitive decline. Furthermore a study examining an effect on grandparents came to a conclusion that providing child care and being intensively engaged as a grandparent resulted in lower cognitive scores compared to grandparents who are less engaged in childcare (Arpino & Bordone 2014). We did neither confirmed nor disproved this as none of these variables showed significant impact on the dependent variable in our research.

2.2 Health and other factors

Cognitive flexibility is even related to health status during adulthood, according to Buttelmann & Karbach (2017). A study by Ratey & Loehr (2011) states that physical activity can improve cognitive performance in adults and how it can be used to improve health and productivity. This study also looks at the mechanisms of how physical activity can improve cognition, and how physical activity during early and mid-life can help preserve cognitive abilities later on. Findings have also indicated that by improving the diet of elderly individuals, it could potentially delay or slow down cognitive decline related to increasing age (Deary *et al.* 2009a).

Diet as a factor determining the state of people's cognition is further described in the research paper by Reichelt *et al.* (2017), who state that consuming a bad diet may increase the chances of developing cognitive decline, while nutrients in one's diet can help protect against such decline. This study was focused on a vitamin intake as the strongest factor improving cognition,

furthermore an adequate intake of monounsaturated fatty acids and cholesterol was significantly associated with decreased risk of cognitive impairment.

Chapter 3

Data description

3.1 Data Description

The purpose of this chapter is to present the data set we are working with and identify and describe all of the important variables used in our model.

3.1.1 Origin of the data

Our data comes from a Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a project intended for studying effects of health, social and economic living conditions and environment policies over the course of lives of European citizens. The SHARE data set originally consists of 530 000 in depth interviews with 140 000 residents of 28 European countries and Israel, all of whom are aged 50 or older. The SHARE project is considered the largest pan-European social science study. The data can be accessed through the SHARE project website and is available free of charge for scientific or teaching purposes. SHARE also provides us with a release guide document containing a description of all variables included in the data set and their possible values. (Börsch-Supan *et al.* 2013)

The original easySHARE data set consists of panel data and provides us with 368 806 observations containing 109 variables. Given that the independent variable that is essential for our research - $books\_age10$ is included only in wave three and five and the essential dependent variables describing cognitive functions (these variables are: recall of words, orientation to date, numeracy score - description of these variables is provided in subsection 3.1.2) are not included in wave three, therefore only the data included in wave five have been

included in this data analysis. After excluding every wave other than wave five, we were left with 64 483 observations of 110 variables. The number of observations was further decreasing throughout the process of adjusting the data set. The description of this process is included in the following chapters. For the purpose of this paper, additional restrictions were used in determining

the final data set. First, not all of the 110 variables were used. Second, the original data set includes values between years 2004 and 2019, however, due to the elimination of certain waves this paper only deals with data from the year 2013. As such, the data analysed in this thesis is cross sectional. The process of elimination will be described in greater detail in the following subsections.

3.1.2 Dependent and independent variables

This section intends to describe all variables used in our model in detail. The variables are as follows (Börsch-Supan *et al.* (2013)):

- **Numeracy score 1 (percentage)** - this dependent variable ranges from 0 (bad) to 5 points (good) and describes mathematical performance (in the form of percentage calculation) of the study respondents. We transformed this variable into a dummy as further described in the methodology section in chapter 4.

  The model is built on this dependent variable because it best reflects the cognitive abilities of the respondents in terms of the construction of this model. However, other dependent variables could have been used as well. For example variable *recall of words* or *orientation to date*.

  - **Recall of words, first trial** - variable *recall_1* ranges from 1 to 10 and describes number of words that had been recalled in the first trial of the word recall task.
  - **Orientation to date** - variable *orienti* describes the ability to orientate oneself in key dates, months, years or week days and ranges from 0 to 4 where 4 is the best possible result.

    Both of these variables are greatly influenced by the state of the respondents' memory and their ability to perform recall tasks and thus neither was included in the model. Further reasoning as to why

the variable *Numeracy score* is considered to best depict the levels of cognitive skills is included in the sections below.

- **Age at interview** - *Age* of the respondent in the time of the interview, this variable is measured in years.

  This variable was used because age greatly affects cognitive skill and performance of the study participants. Cognitive skill tends to decline with age and studies show that older adults have significantly lower scores on tests of cognitive functioning than younger adults, especially those examining their memory (Diabetes.co.uk (2022)).

  However, this decline can vary person to person. It is therefore important to take age into account in the thesis and compare it to the variable *Numeracy score* as it is a variable that is least connected to memory alone as age-related changes in the brain, such as decreased plasticity, can lead to declines in memory (Craik 2012).

- **Gender** - *female* is a categorical dummy variable (a dummy variable is a binary variable where the only possible outcome is yes/true/1 and no/false/0), in our case value 0 means the respondent is a male, 1 means the respondent is a female.

  This variable was used to identify disparities in outcomes of cognitive skills and impacts of other variables between genders.

- **Number of books when ten** - variable *books_age*10 captures data on the amount of books present in the respondents' household when they were ten years old. This number does not include magazines, newspapers, or school books. This variable is measured on a scale ranging from 1 to 5, where

  1: none or very few ($0 - 10$ books)

  2: enough to fill one shelf ($11 - 25$ books)

  3: enough to fill one bookcase ($26 - 100$ books)

  4: enough to fill two bookcases ($101 - 200$ books)

  5: enough to fill two or more bookcases (more than 200 books)

  This variable is the basis of our research - to determine whether being

surrounded by books as a child helps with cognitive skills later in life. Based on existing research the assumption is that reading does affect cognitive skills of an individual in a positive way.

- **Depression scale EURO-D** - variable *eurod* describes depression on a scale from 0 to 12. 0 meaning that the respondent is not depressed and 12 meaning the respondent is very depressed. The question consists of depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. All these factors combined contributed to the final variable.

  Depression has been found to cause deficits in executive functioning, including impairments to memory, concentration, decision-making, and problem-solving abilities. Such executive functioning is necessary for proper planning, organization or task completion. (Surguladze *et al.* 2005). Therefore we included this variable to assess its importance and influence and the expected outcome is that depression will influence cognitive skills in a negative way.

- **Drinking behavior** - *br*010_*mod* is a categorical variable describing the frequency of which the study participant drank any alcoholic beverage in the past three months (specifically beer, cider, wine, spirits or cocktails).

  This variable is included because studies have shown that people with alcoholism have lower scores on cognitive tests than people who do not drink. Alcoholism has been linked to a decrease in cognitive skills, such as attention, memory, and executive functioning (Evert & Oscar-Berman 1995), (Woods *et al.* 2016). Among the many effects of alcohol consumption is the decrease the amount of gray matter in the brain, which can cause cognitive deficits (Dalal & Das 2018), therefore it is important to include it into the research and examine its influence further. The expected outcome is a negative influence on cognitive skills.

  The scale on which the frequency of alcohol consumption was measured:

1: not at all

2: less than once a month

3: once or twice a month

4: once or twice a week

5: three or four days a week

6: five or six days a week

7: almost every day

- **Years of education** - *eduyears_mod* gives us a number of years of education of a given study participant. This variable is measured in years.

  Education and mental capacity are undoubtedly two closely linked variables. The question might be if a person gains better educational results with better mental capacity or the other way around. The assumption is that the better education a participant has, the greater probability of acquiring better cognitive test results.

- **Number of children** variable *ch*001_ indicates number of alive children (including foster children, adopted children and also step children).

  Existing research suggests that there was a significant negative association between the number of children a woman had and her cognitive health in later life. However, the effect was small and the association was not significant when a woman's education level was taken into account (Bonsang & Skirbekk 2022). The findings suggest that while there may be a small negative effect of childbearing on cognitive health in later life, other factors such as education level play a more important role (Schneeweis *et al.* 2014).

- **Number of grandchildren** variable *ch021_mod* giving us the total number of grandchildren of a respondent.

  This variable is potentially interesting to examine and to determine if the number of grandchildren also might have a negative effect on cognitive skills similarly to the number of children and if grand-parenting as opposed to parenting can be considered a brain stimulating activity.

3.1.3 Other variables

As stated above, the data set consists of more than a hundred variables, many of which we were not able to use in our data analysis or describe in the thesis. The easySHARE release guide document divides variables into seven categories: demographics, household composition, social support and network, childhood conditions, health and health behaviour, functional limitation indices and work and money. The aim was to include at least one variable from each category in the model and data analysis.

3.1.4 Specific values and missing codes

The data set includes numbers of specific values. These values might have a different origin but the goal is to eliminate them all in order to avoid model bias. The process of getting rid of these strange values is discussed in greater detail in the Methodology section in Chapter 4.

Specifically these values were detected while examining the data set:

- value -3 meaning "implausible value or value is suspected to be wrong"
- value -12 meaning "don't know or refused to answer"
- value -13 meaning "not asked in this wave"
- value -14 meaning "not asked in this country"
- value -15 meaning "no information",

These values are quite common in the data set and after their removal the number of observations decreased.

Chapter 4

Methodology

I this section the data set and methodology how the data were prepared and then analysed will be introduced and described.

4.1 Data analysis

As stated above, the data comes from SHARE and all models and the data analysis were made in R.

4.1.1 Cleaning the data

Before conducting any analysis the data need to be cleared. All of the variables that are used are described in Chapter 3 . The description of specific negative values that signify a state of the contestant of the survey instead of the results of the measured variable is included in the same chapter. These values have to be omitted from the analysis because their inclusion would have the effect of taking into account missing values (that have been labeled as negative values) and the results would be biased and would not present a satisfactory reflection of reality.

The first step of preparation of the data for analysis was looking at the data and seeing that not all variables were included in all of the waves and therefore only wave 5 satisfies all requirements for our research question and our model. This step alone reduced number of observations from 365 806 to 66 188 but in the end we ended up with 11 359 observations of 11 variables, which is still quite a decent number of observations that should ensure reliability and

precision of the model. Detailed steps that led to this decrease of observations is described in this and the following chapter.

Next steps included removal of the missing and error values. Some kind of condition was made for most of the variables, the following example is describing the variable *books_age*10. The condition of non negativity has been implemented in this case and as we can see from the summary tables below (Table 4.1 and Table 4.2), this change removed values smaller than 1 as these values do not belong in the data set and were changing the outcomes, now the variable only takes on values that make sense and are described in the release guide document (the lowest meaningful value is 1 meaning that none or very few books were present in the household), this step also of course slightly reduced the number of observations.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-15.000	1.000	2.000	2.190	3.000	5.000

Table 4.1: Number of books at age ten, before

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	2.000	2.367	3.000	5.000

Table 4.2: Number of books at age ten, after

Similar conditions were implemented for most of the other variables in order to eliminate missing codes and meaningless values. This issue will be further discussed in Chapter 5 including the process and reasons behind setting a boundary for determining meaningful values as well as the process of outliers removal.

4.2 Model

For the purposes of analyzing the data set, we have decided to use the logit model.

4.2.1 General description of logit model

Logit or logistic regression is a discrete choice model used to analyze binary dependent variables (those taking on values 1 or 0). To correctly estimate a model for a binary response variable, traditional linear regression is no longer a valid option as it is not bounded to the range of 0 to 1. Logistic regression approaches the problem from a probability point of view by estimating the probability of $Y = 1$ given our set of explanatory variables. Given that only binary response variables are considered, the model can be constructed in the following way. In linear regression, the intercept is in the form of expected value of the dependent variable given the data. In case of logistic regression the interpretation is in form of probability:

$$P\left(Y = 1 \mid X_1, X_2, \ldots, X_k\right) = F\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k\right)$$

where Y is the dependent variable, $X_1, ..., X_k$ are the independent variables and F denote the logistic function that ensures that model the probability. The logistic function is a non-linear function that scales the values to strictly 0 to 1 range. Hence, the resulting estimate is the value of the logistic cumulative distribution function. The logistic function is defined in a following way:

$$F(x) = \frac{exp(x)}{1 + exp(x)}$$

Given that distribution of variables used is known or at least the assumption is made. Then it is possible to employ maximum likelihood estimation to determine the values of the coefficients.

The estimation of parameters is done using the Maximum Likelihood Estimation. The goodness of fit can be measured by pseudo-R^2, which is defined as likelihood ratio index (the higher the value, the better fitting the model is). The idea is to compare the maximized value of lnL of the model with the lnL_0 that denote the value of log-likelihood of a model with only an intercept.

$$LRI = 1 - \frac{\ln L}{\ln L_0}$$

The higher the likelihood ratio, the stronger the relationship between the independent variable and dependent variable is, likelihood ratio being larger than one implies significance, ratio smaller than one means otherwise.

4.2.2 Model description

For the purposes of our research and our model we established new variable *numeracy_dummy* with a threshold that every score of 4 and higher (variable $Numeracy\_dummy \geq 4$) is considered true (as if the study participant passed the test - value 1) and every value smaller than 4 is considered fail ($Numeracy\_dummy = 0$). This decision is based on descriptive statistics and the fact that from all of the observations, slightly more than a half has a worse result than 4. We have decided to use logit model. Based on the fact that we are dealing with a discrete choice problem (independent variable $Numeracy\_dummy$ now obtains values 1 and 0). Based on descriptive statistics we have decided that logistic function is the best fit for distribution of our independent variable. Assuming that the initial distribution of the dependent data (*numeracy_1*) can be extreme, that means range from 0 to 5, and hence this is better reflected by logistic distribution.

4.2.3 Equation and hypothesis

The formula describing the investigated relationship is as follows:

$$\begin{aligned}\textbf{numeracy\_dummy} &= \beta_0 + \beta_1 * age + \beta_2 * female + \beta_3 * books\_age10 \\ &+ \beta_4 * eurod + \beta_5 * br010\_mod + \beta_6 * eduyears\_mod + \beta_7 * ch001\_ \\ &+ \beta_8 * ch021\_mod\end{aligned}$$

The hypothesis is that being surrounded by books throughout childhood has a significant positive effect on cognitive abilities later in life. The hypothesis will be tested using logit model and verified.

Chapter 5

Results and discussion

5.1 Results

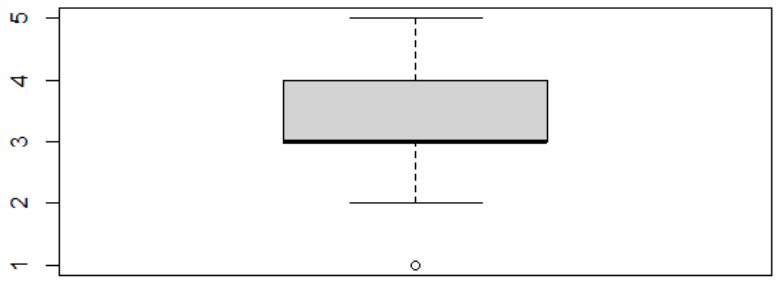
The discussion of results and their interpretation will be presented in the following chapter. Tables, figures, graphs and descriptive analysis results are included to provide a better understanding of the data, results and the process which led to obtain them.

5.1.1 Descriptive statistics of the data set

In this subsection, the results of descriptive analysis will be presented. Among others, we will take a closer look at how the distribution of the variables looks like or how the abnormalities of the distribution influence our data. This section is highly important to fully understand the data in order to build the model properly. We will also identify outliers and decide whether to include or exclude them from the model. We will now go through respective variables and discuss the descriptive analysis. Please note that the detailed description of the variables is already provided in Chapter 3.

- dependent variable **Numeracy**, this variable represents the success or failure in a cognitive test investigating numeracy and percentage calculation. This variable was originally ranging from 0 to 5 (see Figure 5.1). While inspecting the distribution of this variable, a few outliers were identified, however it is not necessary to omit these observations as they are not significantly different then the rest of the data and will not lead to false conclusions. Removal of missing values (those smaller than one) is sufficient.

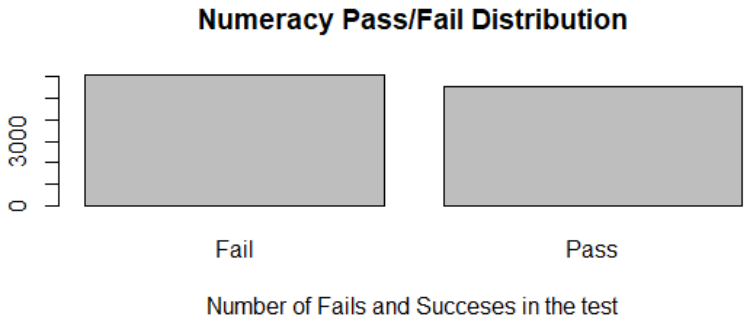
Figure 5.1: Numeracy - Boxplot



For the purposes of the thesis, this variable was transformed into a dummy with a threshold level of 4 for passing the test. Every value lower than 4 is considered a failure. This transformation was made in order to meet the assumptions of the logit model.

This resulted into roughly 52 percent observations being false (meaning the respondent failed a test) and approximately 48 percent observations being true (meaning the respondent passed the test).

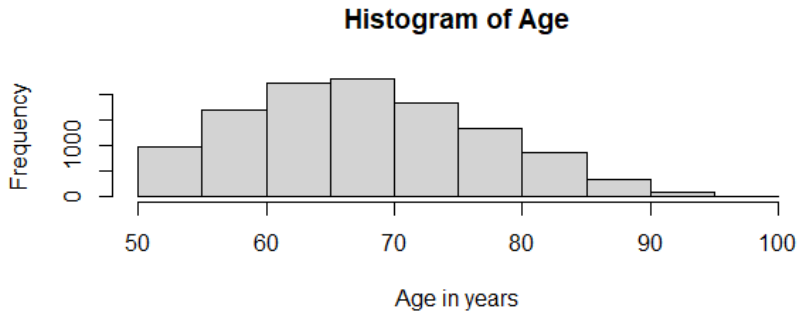
Figure 5.2: Numeracy - Success rate



- independent variable **Age** - this variable originally attained a few values lower than 50 years but an arbitrary threshold was implemented to determine whether a respondent is old enough to be taken into account. The existing literature did not reach a clear conclusion as to when the cognitive decline starts and some studies suggest that it might start as early as in respondent's thirties (Salthouse 2009), nevertheless the majority of research papers on this topic suggest that it is more likely to observe signs of cognitive decline later in life. We set a boundary to 50 years and it seems reasonable.

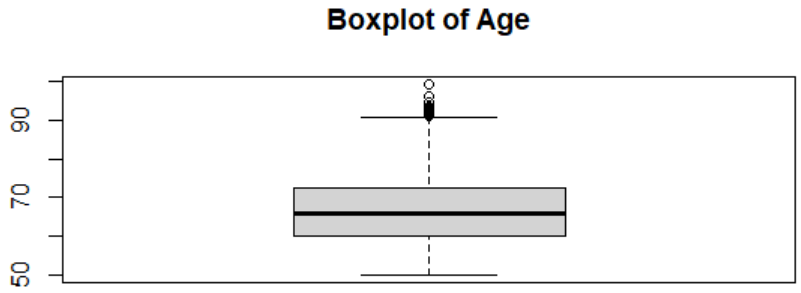
The histogram below shows the distribution of age among the study participants. It is obvious that most of the participants are aged between sixty and seventy years old and then the number of people representing other age categories gradually decreases as the age increases.

Figure 5.3: Age - histogram



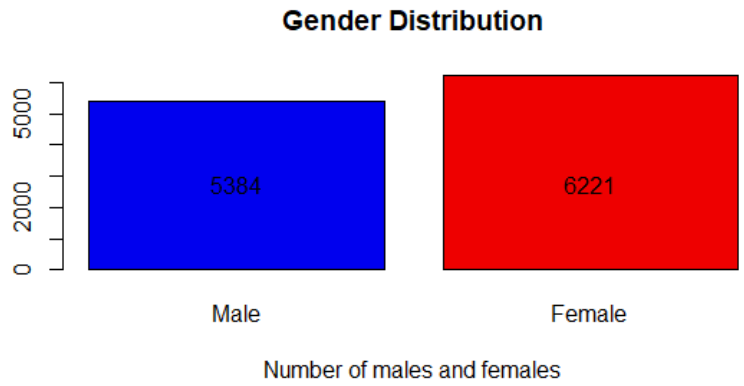
There are only 97 people representing the last category of people aged ninety to a hundred years old. Representatives of this age category are considered to be outliers but we have decided not to exclude them as they will provide important information about the population.

Figure 5.4: Age - boxplot



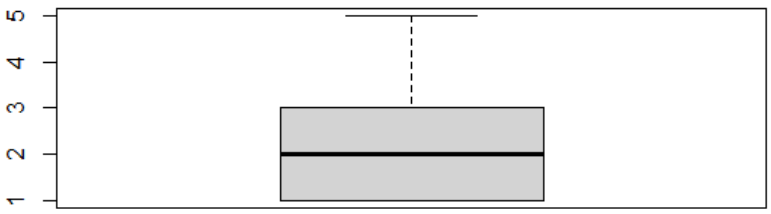
- independent variable **female** - as this is a dummy variable, no outliers are expected to occur as missing values were omitted. Number of males and females is depicted in the following figure. The larger number of women in our data set might be caused by the fact that women live longer than men, therefore more women got the chance to be part of the study.

Figure 5.5: Gender - distribution



- independent variable **Number of books when 10** - this is the variable of our main interest. No outliers were detected (also probably due to the low scale from 1 to 5) and only values lower than zero had to be removed, as they denoted missing values. Median is in this case equal to 2 and mean is equal to 2.288.

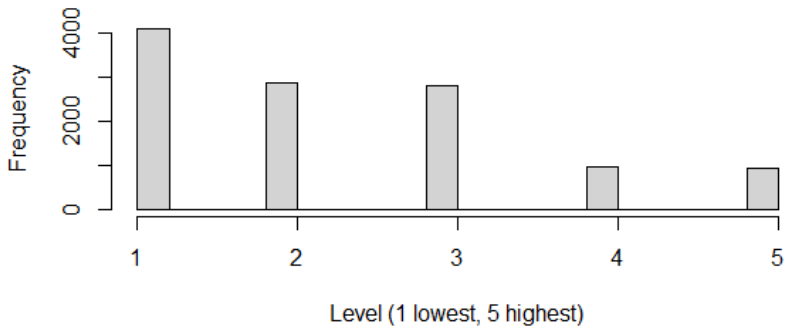
Figure 5.6: Books at age ten - boxplot



This variable is non-linear, meaning that it does not change in a linear fashion, it does not follow a linear trend (see Chapter 3 for a detailed description and a scale on which this variable is measured).

The largest portion of respondents stated that their household kept less than 10 books during their childhood, as seen from Figure 5.8. Table 5.1 shows the exact frequency of which participants reported the number of books in their household.

Figure 5.7: Books at age ten - histogram

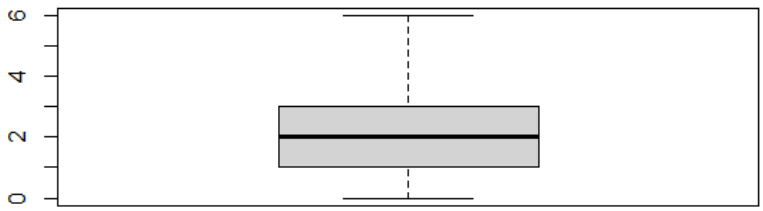


Number of books	0-10	11-25	26-100	101-200	200+
Frequency	4 082	2 859	2 811	941	912

Table 5.1: Books when ten, frequency of responses

- independent variable **Depression** - this variable is also non-linear (meaning that person with depression score of 10 is not twice as depressed as a person reporting a level of depression of 5). But of course the higher the number the more depressed the person is. In this case we have decided to exclude the outliers, the reasoning is that respondents may tend to identify themselves as maybe more depressed than they really are and hence this method of self-evaluation might skew the data, the upper bound was identified as value 6, all values above 6 were excluded and the new distribution is depicted in the following figure.

Figure 5.8: Depression - boxplot

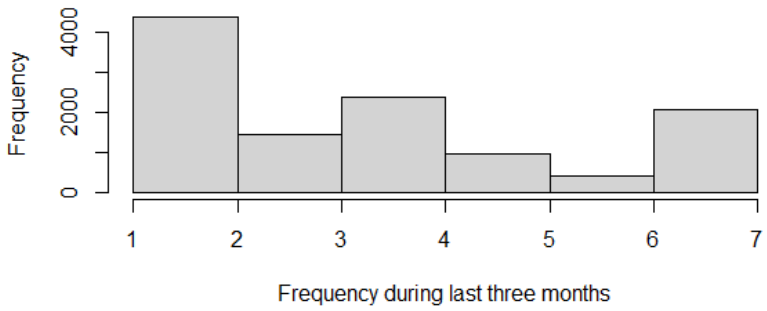


There are interesting differences in depression between genders. It is obvious from our data set that women tend to identify themselves as more

depressed than men. This might mean that males are really less depressed than females, but it could also mean that they just tend to withhold their feelings and mental health condition. This does not necessarily mean that males have better overall mental health in general (Surguladze *et al.* 2005). As this variable is based on self-evaluation, the results provided might not be absolutely reliable. Positive correlation relationship between depression and age has been identified, however it is not significant.

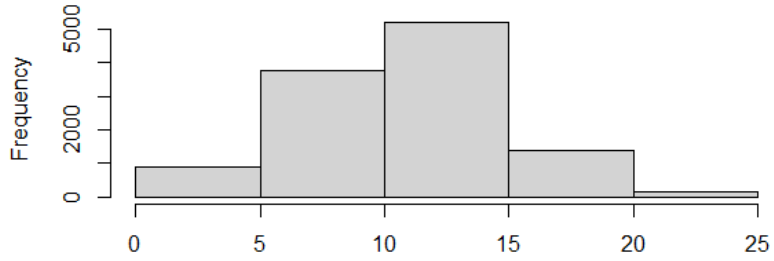
- independent variable **Drinking behavior**, we did not identify any outliers regarding this variable. Derived from the data, 2057 people responded that they drink alcohol almost every day, this number is surprisingly high, on the other hand, 3291 respondents stated that they did not consume any alcohol at all in the past three months.

Figure 5.9: Alcohol - histogram



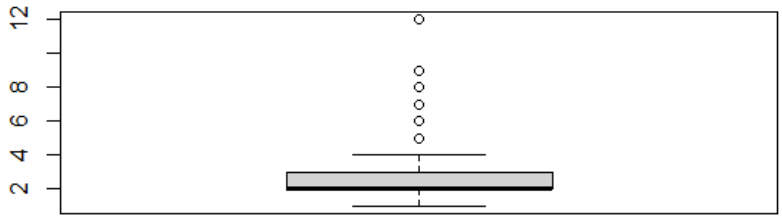
- independent variable **Years of education** shows that most people spent between 10 and 15 years in school in total. There are some values indicating that a respondent spent more than 20 years in school.

Figure 5.10: Education - histogram



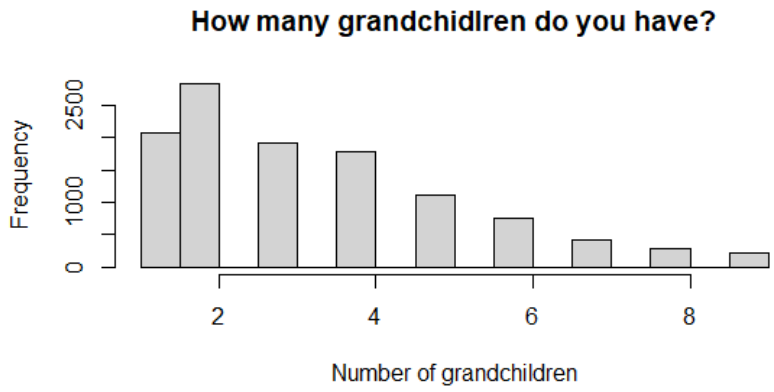
- independent variable **Number of children** - there are some outliers (see Figure 5.11 below), but they are essentially reasonable and should not be caused by an error in the measurement, therefore there is no need to remove them.

Figure 5.11: Children - boxplot



- independent variable **Number of grandchildren**, this variable contained a significant amount of outliers, we therefore decided to calculate upper bound and exclude all values exceeding this bound. We ended up with the following distribution of number of grandchildren.

Figure 5.12: Grandchildren - histogram



5.1.2 Correlation matrix

We used a correlation matrix to identify any correlation coefficients and relationships between all of our potentially independent variables. The Correlation matrix tables A.2 and A.3 did not reveal any significant correlation between our variables. Any additional adjustments of the data set were not necessary due to their correlation.

Interesting relationships are between variables describing number of books and years of education, number of children and grandchildren and between

number of grandchildren and age. None of these correlations are considered significant and all of these correlations were expected to occur.

5.1.3 Interpretation of logit model

Due to the non-linearity of logit models, the results can not be interpreted directly and marginal effects have to be computed.

We have decided to use only logit since we expect that the extremes are possible and rather probable. We are also more interested in depicting the extremes rather than the body of the probability distribution (for modelling of the body of the distribution, probit model would be more suitable)

Since the logit transforms the distribution using probability, we have to transform the estimates back in order to obtain marginal effect values of respective independent variables. We therefore have to create a function to derive marginal effect.

When interpreting the coefficients of the logistic regression, it needs to be taken into account the non-linearity imposed by the logistic function and thus the coefficient estimates cannot be interpreted as in traditional linear regression. Instead, to quantitatively interpret the coefficients, marginal effects are used. It is possible to compute it at sample means or as a mean of individual marginal effects.

Values then can be interpreted as marginal effects, which means that if x changes, so does the dependent variable according to following 'slopes'.

5.1.4 Discussion of results

Table 5.2 below shows the main results. The column on the left side shows the results of logit model with estimates values, values of standard error (measure of a variability) and significance indicators. The right column then shows values of marginal effect at mean of all variables, that are represented in rows. We will now discuss results of all independent variables included in our model.

- **books_age10**, the logit model evaluated this variable as highly statistically significant at $p < 0.01$ level (p-value indicates how high is the probability that the results are due to a chance, p-value smaller than

0.01 indicates a strong evidence) and therefore has a great impact on our dependent variable. The marginal effects calculated show a positive effect meaning that the average person's (that was surrounded by books as a child) chances of passing the numeracy test increases by 11.687% (regardless of how a person whose household did not have many books did on the test).

These results confirm our hypothesis that growing up surrounded by books has a positive impact on cognitive skills later in life. This finding can have a great impact on behaviour of parents of young children or the schooling system. Knowing that we can influence our children's cognitive skills development and therefore their overall well-being, can change our behaviour and we can lean towards promoting e.g. greater access to public libraries.

- **age**, this variable was evaluated as statistically significant on $p < 0.01$ level and the marginal effects at mean show a significant negative effect of age on cognitive skills. This result is not at all surprising as many studies and research papers show that cognitive skills tend to decline with age.

- **female**, this gender describing variable is also considered statistically significant on $p < 0.01$ level and the marginal effect indicate that there is a significant negative relationship.

- **eurod**, this depression describing variable is also considered statistically significant on $p < 0.01$ level and marginal effects indicate that depression has a significant negative effect, meaning that for the average depressed person, depression reduces the chance of passing the numeracy test by 3.957% percent, compared to a person whose mental health was fine. If the respondent's mental health deteriorates and all other parameters remain same, their chances to pass the numeracy test will decrease by 4 percent.

  This result was also anticipated, we did not expect that depression would improve cognitive skills of our respondents.

- **br010_mod** this alcohol drinking behaviour describing variable is also considered statistically significant on $p < 0.01$ level and marginal effects indicate that drinking alcohol slightly increases the probability of passing a cognitive skills test.

This result is quite surprising but studies that confirm cognitive decline related to alcoholism talk about excessive amounts of alcohol, in our case, mean value of this variable is only 3.527 and median value is 3, implying that most people only drink once or twice a month. These small amounts of alcohol consumption are proved to be beneficial and protect an elderly against dementia or Alzheimer's disease (Peters *et al.* (2008)). Small amounts of alcohol are not harmful - on the contrary, we confirmed their benefit.

- **eduyears_mod** this variable indicating years of education of a study participant is also considered statistically significant on $p < 0.01$ level and marginal effects indicate that there is a positive effect of being educated on cognitive skills.

  The main reason might be that these people are well educated and thus thanks to education they have a better trained thought process and brain and are simply better at solving any tests.

  This variable might be influenced by the fact that this person is from academical background (we do not have variable discussing for example parent's education), therefore there might be an important instrumental variable. Furthermore, we can assume that well educated children have well educated parents.

- **ch001_** and **ch021_mod** (number of children and number of grandchildren) - none of these variables is statistically significant which is quite surprising as we know about a study using SHARE data and examining an influence of having children on cognitive skills (the results stated that having children has a negative effect on cognition). This result might be due to a different approach in defining cognitive skills in this thesis, our dependent variable might me too specific and therefore we obtained different results. We also used other methods to determine the influence as well as differently modified dataset.

- value of a constant is not important for the interpretation of the results in our case.

Our hypothesis has therefore been confirmed, growing up surrounded by books really does have a significantly positive effect on cognitive skills later

in life. In addition, we have uncovered other significantly beneficial or non-beneficial variables some of which we can further use for policy making processes or simply to influence individual's behaviour and decision making in everyday life.

Table 5.2: Logit model results and marginal effect at mean

	Logit model results and estimates:	
	numeracy_dummy	Marginal effect at mean
age	−0.017***	-0.008316
	(0.002)	
female	−0.436***	-0.207153
	(0.043)	
books_age10	0.246***	0.116872
	(0.018)	
eurod	−0.083***	-0.039570
	(0.012)	
br010_mod	0.071***	0.033940
	(0.010)	
eduyears_mod	0.128***	0.061051
	(0.006)	
ch001_	0.041*	0.019398
	(0.023)	
ch021_mod	−0.013	-0.006257
	(0.012)	
Constant	−0.825***	x
	(0.199)	
Observations	11,359	
Log Likelihood	−7,010.184	
Akaike Inf. Crit.	14,038.370	
Note:	*p<0.1; **p<0.05; ***p<0.01	

5.1.5 Shortcomings and limitations of the research

This section lists possible shortcomings and improvements of this thesis.

- We are working with cross sectional data due to the reason that many waves were deleted and we therefore do not have a comparison of specific individuals and their development throughout the years (this was because only one wave included variables needed to answer the research question properly). Panel data would be probably much more interesting to study as we could watch cognitive decline of specific people in time.
- All our data comes from year 2013, repeating research on data that are more up to date when they are released (and if) could show some interesting development as well (for example life expectancy has risen in the last ten years according to WHO, so we could be seeing higher number of older people in more recent datasets).
- The data set comes from mostly European countries, studying accessibility of books in other continents and even non developed countries might be interesting.
- Dependent variable *Numeracy* might or might not be the best indicator of cognitive skills or cognitive decline. It should be interesting to include other variables describing cognitive skills into the research next time.
- Some of our variables come from a retrospective self-assessment which might not be the most efficient way to obtain a reliable data set.

Chapter 6

Conclusion

We tested a hypothesis that access to books has a positive effect on cognitive performance in late adulthood, that is in our case represented by numeracy test. The results show a positive effect of access to books in childhood on cognitive performance in late adulthood, specifically that being surrounded by books means 11.687% increase of chances of passing the test with significance level of $p < 0.01$.

This thesis used data from SHARE (Survey of Health, Ageing and Retirement in Europe), we selected eleven variables and worked with cross-sectional data consisting of eleven thousand observations from year 2013. To obtain the results, we used logistic regression and marginal effect at mean coefficients in order to determine the degree of influence of our independent variables. We used one dependent variable describing the numeracy test and several independent variables regarding participant's age, gender, mental health status, drinking behaviour or family situation.

The main contribution of this work is understanding what impact growing up surrounded by books alone has and how we can use it to create better environment for children to grow up in. The finding that access to books can slow down cognitive decline can improve school curricula, public library's policies or parent's behaviour and possibly encourage young individuals to spend more time with books as they will experience the benefits many years after their childhood ends.

Further research should focus more on the effects access to books has, for example its effect on the maximum achieved level of education or relationship to books later in life. Future research should also extend to examine the effect

of access to books and reading itself not only in childhood, but throughout the whole life and if exposure to books helps to create a habit of reading and how strong the relationship is between being exposed to books with actually reading them. This is to determine the full extent of the effect of being surrounded by books as the results at hand might suggest that general access to books is beneficial.

The effect of being surrounded by books should also be tested in respect to non-numeracy tasks, for example memory tasks or reading comprehension. Further research with broad panel data would also be beneficial. This thesis uses cross sectional data, a more complete and broader data set would undoubtedly give a better insight into the research question and help to understand how cognitive skills evolve in time. Results from a similar research from non-developed countries with lower measure of access to books and education (as availability of books and education in Europe is high) might be interesting as well and the recommendations may differ (such as development of schooling or building of public libraries to provide access to books).

Bibliography

Anme, T., R. Shinohara, Y. Sugisawa, T. Watanabe, K. Tokutake, Y. Mochizuki, K. Morita, E. Tomisaki, E. Tanaka, B. Wu *et al.* (2013): "Social interaction and dementia prevention: six-year follow-up study." *J Public Health Front* **2(2)**: pp. 109–113.

Arpino, B. & V. Bordone (2014): "Does grandparenting pay off? the effect of child care on grandparents' cognitive functioning." *Journal of Marriage and Family* **76(2)**: pp. 337–351.

Bonsang, E. & V. Skirbekk (2022): "Does childbearing affect cognitive health in later life? evidence from an instrumental variable approach." *Demography* **59(3)**: pp. 975–994.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, & S. Zuber (2013): "Data resource profile: the survey of health, ageing and retirement in europe (share)." *International journal of epidemiology* **42(4)**: pp. 992–1001.

Buttelmann, F. & J. Karbach (2017): "Development and plasticity of cognitive flexibility in early and middle childhood." *Frontiers in psychology* **8**: p. 1040.

Craik, F. (2012): *Aging and cognitive processes*, volume 8. Springer Science & Business Media.

Dalal, P. & A. Das (2018): "Cognitive problems in alcohol use: a review." *MOJ Addict Med Ther* **5(6)**: pp. 262–264.

Deary, I. J., J. Corley, A. J. Gow, S. E. Harris, L. M. Houlihan, R. E. Marioni, L. Penke, S. B. Rafnsson, & J. M. Starr (2009a): "Age-associated cognitive decline." *British medical bulletin* **92(1)**: pp. 135–152.

Deary, I. J., W. Johnson, & L. M. Houlihan (2009b): "Genetic foundations of human intelligence." *Human genetics* **126(1)**: pp. 215–232.

Deary, I. J., A. F. Wright, S. E. Harris, L. J. Whalley, & J. M. Starr (2004): "Searching for genetic influences on normal cognitive ageing." *Trends in cognitive sciences* **8(4)**: pp. 178–184.

Diabetes.co.uk (2022): "Cognitive ability in older adults is dependant on family size, study finds." .

Evert, D. L. & M. Oscar-Berman (1995): "Alcohol-related cognitive impairments: An overview of how alcoholism may affect the workings of the brain." *Alcohol Health and Research World* **19(2)**: p. 89.

Hlavac, M. (2015): "stargazer: beautiful latex, html and ascii tables from r statistical output." *Available at the Comprehensive R Archive Network (CRAN) https://CRAN.R-project.org/package=stargazer* .

Lipnicki, D. M., S. R. Makkar, J. D. Crawford, A. Thalamuthu, N. A. Kochan, M. F. Lima-Costa, E. Castro-Costa, C. P. Ferri, C. Brayne, B. Stephan *et al.* (2019): "Determinants of cognitive performance and decline in 20 diverse ethno-regional groups: A cosmic collaboration cohort study." *PLoS medicine* **16(7)**: p. e1002853.

Manu, A., F. Ewerling, A. J. Barros, & C. G. Victora (2019): "Association between availability of childrenâ€™s book and the literacy-numeracy skills of children aged 36 to 59 months: secondary analysis of the unicef multiple-indicator cluster surveys covering 35 countries." *Journal of global health* **9(1)**.

Peters, R., J. Peters, J. Warner, N. Beckett, & C. Bulpitt (2008): "Alcohol, dementia and cognitive decline in the elderly: a systematic review." *Age and ageing* **37(5)**: pp. 505–512.

Ratey, J. J. & J. E. Loehr (2011): "The positive impact of physical activity on cognition during adulthood: a review of underlying mechanisms, evidence and recommendations." .

Reichelt, A. C., R. F. Westbrook, & M. J. Morris (2017): "impact of diet on learning, memory and cognition."

Salthouse, T. A. (2009): "When does age-related cognitive decline begin?" *Neurobiology of aging* **30(4)**: pp. 507–514.

Schneeweis, N., V. Skirbekk, & R. Winter-Ebmer (2014): "Does education improve cognitive performance four decades after school completion?" *Demography* **51(2)**: pp. 619–643.

Sikora, J., M. D. Evans, & J. Kelley (2019): "Scholarly culture: How books in adolescence enhance adult literacy, numeracy and technology skills in 31 societies." *Social science research* **77**: pp. 1–15.

Surguladze, S., M. J. Brammer, P. Keedwell, V. Giampietro, A. W. Young, M. J. Travis, S. C. Williams, & M. L. Phillips (2005): "A differential pattern of neural response toward sad versus happy facial expressions in major depressive disorder." *Biological psychiatry* **57(3)**: pp. 201–209.

Wilson, R. S., E. Segawa, P. A. Boyle, & D. A. Bennett (2012): "Influence of late-life cognitive activity on cognitive health." *Neurology* **78(15)**: pp. 1123–1129.

Woods, A. J., E. C. Porges, V. E. Bryant, T. Seider, A. Gongvatana, C. W. Kahler, S. de la Monte, P. M. Monti, & R. A. Cohen (2016): "Current heavy alcohol consumption is associated with greater cognitive impairment in older adults." *Alcoholism: clinical and experimental research* **40(11)**: pp. 2435–2444.

Yates, L. A., S. Ziser, A. Spector, & M. Orrell (2016): "Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis." *International psychogeriatrics* **28(11)**: pp. 1791–1806.

Appendix A

Appendix A

Statistic	N	Mean	St. Dev.	Min	Max
numeracy_1	11,359	3.417	1.052	1	5
age	11,359	67.644	9.232	50.000	99.200
female	11,359	0.537	0.499	0	1
books_age10	11,359	2.292	1.239	1	5
eurod	11,359	1.957	1.728	0	6
br010_mod	11,359	3.527	2.149	1	7
eduyears_mod	11,359	11.186	3.923	1	25
ch001_	11,359	2.399	1.043	1	12
ch021_mod	11,359	3.321	1.981	1	9
wave	11,359	5.000	0.000	5	5
numeracy_dummy	11,359	0.478	0.500	0	1

Table A.1: Summary of all variables

All stargazer tables: Hlavac (2015)

	age	female	books_age10	eurod	br010_mod	eduyears_mod	ch001_	ch021_mod
age	1	-0.086	-0.148	0.036	-0.019	-0.177	-0.073	0.256
female	-0.086	1	0.028	0.166	-0.274	-0.078	-0.009	0.019
books_age10	-0.148	0.028	1	-0.098	0.107	0.418	0.053	-0.008
eurod	0.036	0.166	-0.098	1	-0.112	-0.111	-0.0001	0.009
br010_mod	-0.019	-0.274	0.107	-0.112	1	0.147	-0.014	0.007
eduyears_mod	-0.177	-0.078	0.418	-0.111	0.147	1	0.007	-0.064
ch001_	-0.073	-0.009	0.053	-0.0001	-0.014	0.007	1	0.458
ch021_mod	0.256	0.019	-0.008	0.009	0.007	-0.064	0.458	1
numeracy_dummy	-0.128	-0.131	0.248	-0.124	0.146	0.310	0.026	-0.039

Table A.2: Correlation matrix, part one

	numeracy_dummy
age	-0.128
female	-0.131
books_age10	0.248
eurod	-0.124
br010_mod	0.146
eduyears_mod	0.310
ch001_	0.026
ch021_mod	-0.039
numeracy_dummy	1

Table A.3: Correlation matrix, part two

Appendix B

Appendix B

Data and source codes are available upon request.

SHARE ACKNOWLEDGEMENTS

This paper uses data from the generated easySHARE data set (DOI: 10.6103/SHARE.easy.710), see Gruber et al. (2014) for methodological details.

The easySHARE release 7.1.0 is based on SHARE Waves 1, 2, 3, 4, 5, 6 and 7 (DOIs: 10.6103 /SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, 10.6103/SHARE.w7.710).

This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6, 7 and 8 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, 10.6103/SHARE.w7.710, 10.6103/SHARE.w8.710), see Börsch-Supan et al. (2013) for methodological details. (Börsch-Supan *et al.* (2013))

The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA NA °211909, SHARE-LEAP: GA NĂ, °227822, SHARE M4: GA NĂ, °261982, DASISH: GA NĂ, °283646) and Horizon 2020 (SHARE-DEV3: GA NĂ, °676536, SHARECOHESION: GA NĂ, °870628, SERISS: GA NĂ, °654221, SSHOC: GA NĂ, 823782) and by DG Employment, Social Affairs Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).