Universität Paderborn
Fakultät für Wirtschaftswissenschaften

# Essays in Empirical Education and Health Economics

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
– Doctor rerum politicarum –

vorgelegt von
Beatrice Baaba Tawiah, Master of Science
geboren am 07.09.1989 in Accra, Ghana

Paderborn, 2023

# Acknowledgments

It takes a whole village to go through this journey of getting a doctorate and I am grateful for each and everyone who supported me directly and indirectly. I am grateful to my supervisor and co-author Hendrik Schmitz for his guidance and support over the years. I am also grateful to my co-supervisor Wendelin Schnedler for his helpful comments and suggestions, and the time he has dedicated to this dissertation. I thank Britta for her support, suggestions, guidance and friendship right from the first day. You really made me feel at home. I want to thank Valentin for his support. It was a pleasure co-authoring and co-teaching with you. I would like to thank Christiane who has always kept me informed and guided me through the legalities of the university. I am very grateful to my colleagues, Diana and Verena, for their support and encouragement especially during the final phase of this dissertation. I really appreciate my colleagues and professors from Q.4 for their support and the time spent in and out of the university. It is always fun to spend time together with you. For my family and friends, I do not know where to begin. You have been a great pillar all these years with your prayers, words of encouragement and support. I am extremely grateful for each one of you. I am eternally grateful to my dad, Mr John Kieran Tawiah, for making it possible for me to get to this point in my life. Your pieces of advice and pressure are always worth it. Finally, but most importantly, I am truly grateful for my husband, Nana. Thank you for the love, for believing in me, encouraging me and making time out of your extremely busy schedule just to read through my papers. You are indeed a rare gem.

I dedicate this dissertation to the glory of God and to my wonderful mother, Mrs Sophia Robertson Tawiah of blessed memory, who always encouraged me to believe in myself.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Context and Scope

This dissertation is made up of four individual contributions to the economics of education and health economics literature. Each of them can independently stand well on its own. They are, however, connected by common features. The greatest commonality is of methodological nature. Each research question is addressed empirically, using large micro data and modern microeconometric methods in all four contributions to this dissertation. Thematically, they are connected. Three of the contributions examine the influence of education on cognitive abilities, character traits and health. Thus, they contribute to the understanding of advantages/disadvantages of being educated. Two contributions examine factors, namely education and children, that influence health not only aggregately, but over the life-cycle. Hence, they contribute to a broader understanding of the influences on health which has not received much attention in the economic literature. One of the contributions which looks at the influence of education on health, therefore, is an intersection of the two main contributions.

The first of these studies, presented in Chapter 2 ("Does Education Improve Cognitive Performance Four Decades After School Completion?"), is a replication study of Nicole Schneeweis, Vegard Skirbekk and Rudolf Winter-Ebmer (*Demography*, 2014). Cognitive abilities are vital for decision making (Banks and Oldfield, 2007) and, hence, necessary for labour market, pension and retirement policies (Banks and Oldfield, 2007; Schneeweis et al., 2014). Education is considered to be an important determinant of cognitive abilities (Banks and Oldfield, 2007; Glymour et al., 2008; Brinch and Galloway, 2012; Kamhöfer et al., 2019), but there is sparse literature in this regard. This study replicates the findings of Schneeweis et al. (2014). The ability to reproduce original research findings increases confidence in the results and also enables scholars to understand the results (Nosek et al., 2015). Replication studies can be done either in a narrow sense, where consistency and accuracy of the data as well as the validity of the computations are checked, or in a wide sense, where the sustainability of the findings is tested by using data from other periods, countries, regions or other entities (Pesaran, 2003). In order to better understand the impact of education on cognitive abilities, this study does not only look at a narrow replication but also a wide replication by including more periods and countries as well as different variable definitions for years of education.

The second study of this dissertation, presented in Chapter 3 ("Does Education have an Impact on Patience and Risk Willingness?"), also looks at the impact of education. Here, we analyse the impact of education on patience (also known as time preference) and risk willingness. Patience and risk are important in economic analysis such as intertemporal contexts (Sun and Li, 2010; Kelleher, 2017). Education, indirectly, may be used to shape an individual's character traits through its organisational structure: the rules and regulations in an academic setting, what is taught and the attributes of teachers, who indirectly are role models. Patience encourages people to invest in modern human capital, but once in school, schooling should make people even more patient (Reyes-Garcia et al., 2007). The correlation between education and risk taking indicates that the ability to use information effectively is a characteristic of risk taking (Shaw, 1996). Although, there is available literature on the relationship between education and these two character traits, there is scarce literature on the causal effect of education on these character traits. This is the contribution of Chapter 3

Chapter 4 ("Life-cycle Health Effects of Compulsory Schooling") centres on both education and health, looking at an impact of education while simultaneously examining a factor that influences health. In the past few decades, the estimation of the effects of education on socio-economic outcomes has been an important part of applied microeconometrics. The effect of education on health has been well researched. Majority of the studies estimate aggregate effects over age groups with mixed results, mostly so statistically significant effects. Aggregate effects may miss relevant patterns since it is likely the effect of education on health is not constant over the life-cycle with less effect at younger ages (Kaestner et al., 2020). Hence, an estimated small and insignificant effect averaged over younger and older individuals does not necessarily imply that health is not causally affected by education. It is likely that the effect sets in late in life which is blurred, however, by a zero effect for younger individuals. Most of the studies on the effect over the life-cycle are descriptive. Chapter 4 contributes to the very scarce literature on the causal effects of education on health over the life-cycle.

The final contribution to this dissertation in Chapter 5 ("The Effect of Children on Health"), examines another factor that possibly influences health. In this chapter, I analyse the effect children have on parental health. Time, energy and, especially, financial resources are necessary requirements for taking care of children. Raising young children can be physically and mentally demanding for parents (Kruk and Reinhold, 2014). Parents are likely to experience some long-term health effects as a result of having children. According to (Beral, 1985), women with children are associated with a higher likelihood of dying from diabetes and circulatory diseases such as hypertension but decreased risks of breast, ovarian and endometrial cancers than women without children. The recent arrival of children may be stressful and possibly harmful to a father's health but as the family size stabilises, the presence of children can be associated with better paternal health (Bartlett, 2004). Literature on the effect, usually aggregate effects, of children on health is scarce. Chapter 5 does not only contribute to the scarce literature, but also introduces life-cycle health effects of children.

## 1.2 Data and Methods

All four studies address the respective research question with the help of data and modern microeconometric methods. Data from different sources are used. They include information collected in surveys, where the units of observation are observed at multiple (panel data) points in time. All datasets used provide individual (instead of aggregate) measurements that can be analysed using microeconometric methods. The main statistical method used is instrumental variable regression. Availability of relevant information in the data along with this method helps in solving the fundamental problem of causal inference. The variables of interest are mostly endogenous as a result of self selection. For instance, the number of years an individual decides to stay in school is usually the individual's decision, which may depend on factors such as interests and intelligence. Self selection introduces the problem of omitted variable bias, which results in inconsistent results. In order to solve this problem, instrumental variables, which satisfy necessary assumptions, are introduced to exogenously determine the endogenous variable. Introduction of policies and random occurrences, which are considered as natural experiments, are mostly used as instruments. In this section, I will briefly describe the data sources, instruments and methods used in this dissertation to resolve the fundamental problem of causal inference.

Chapter 2 ("Does Education Improve Cognitive Performance Four Decades After School Completion?") is based on data from the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a micro panel database on health, socio-economic status, social and family networks covering individuals aged 50 and older in most of the European Union and Israel (Börsch-Supan et al., 2013). SHARE provides information on the educational background and cognitive abilities. Individuals are followed overtime, since it is a panel data, which allows the monitoring of changes in cognitive abilities overtime. This is beneficial for the estimation of effects on cognitive decline. Given the information on educational background, we are able to obtain the years of education of respondents. To exogenously determine years of education, we utilise the compulsory schooling reform introduced in 9 different European countries at different times. Since the educational reforms took place before 1970, individuals needed for the analysis should have been born as early as in the 1920s. The SHARE data provides information on such individuals, which is one of its advantages. The compulsory schooling reforms are used to construct the instrumental variable which is used in a two-stage least squares model to estimate the effect of years of education on cognitive abilities.

Chapter 3 ("Does Education have an Impact on Patience and Risk Willingness?") is based on The German Socio-Economic Panel (SOEP) data. The SOEP, established in 1984, is a wide-ranging representative longitudinal study of private households which contains yearly information on around 30,000 respondents in nearly 15,000 households (Goebel et al., 2019). The SOEP has information on character traits and educational background. Questions regarding individual character traits such patience and risk willingness, which we use in this analysis, are asked. Although these are self-rated by the respondents, the measure for patience and risk willingness in the SOEP have been found to be reliable proxies for these character traits (Dohmen et al., 2011; Vischer et al., 2013). The SOEP provides information on the type of school and years of education as well as the state of schooling. This information helps in defining the instrumental variable, which is based on the German compulsory schooling reform which was implemented from 1946 to 1969 in different years by the various states in West Germany after World War II. The instrument is then used in a two-stage least squares model to estimate the effect of education on patience and risk willingness.

The SOEP data is the main data used in Chapter 4 ("Life-cycle Health Effects of Compulsory Schooling"), which looks at the effect of eduction health over the life-cycle. We also use the same German compulsory schooling reform as mentioned in Chapter 3. SOEP has collected health information of respondents which have been collected since its inception. Here, we exploit the panel nature of the SOEP data in our identification and estimation. The main advantage of using SOEP is that it covers a 36 year-period, which allows us to follow the individuals born around the reform periods over many decades, and estimate both short-run and long-run effects of education on health within the same framework and data set. We complement the SOEP data with the SHARE (same data described in Chapter 2) and the German National Educational Panel Study (NEPS): Starting Cohort Adults data. NEPS is a longitudinal dataset that provides information on the acquisition of education in Germany, and educational processes and trajectories across the entire life span (Blossfeld et al., 2011). We do so to increase the number of observations, especially in older ages. We group the observations into 5-year age groups. These age groups are interacted with years of education. Using interactions between age groups and the instrumental variable as instruments allows us to estimate the effects over the life-cycle using a two-stage least squares model.

Similar to Chapters 3 and 4, the SOEP data is also used in 5 ("The Effect of Children on Health"). This chapter deals with the effect of children on parental health. I look at both the aggregate effects and life-cycle effects. I make use of the health module in the SOEP taking information on BMI, mental health and physical health. There is also information on being diagnosed with certain illnesses in the SOEP, which I make use of. Aside the advantages of the SOEP mentioned in the two previous paragraphs, it also provides detailed information on the children of the respondents. Information such as the number of children of each respondent, the year of birth, month of birth and the gender of each child are available. This helps in defining the instrumental variables used to exogenously determine the number of children, which are twin birth and same-sex children. Having twins at second birth results in three children instead of two. Also, having the first two children being the same sex is likely to increase the probability of having a third child. Both instruments, therefore, help in exogenously determining those who have three or more children instead of two. Using these instruments, I estimate the aggregate and life-cycle effects of children on parental health using a two-stage least squares model and a reduced-form model respectively.

## 1.3 Summary of the Studies

This final section gives a summary of each of the four chapters in this dissertation. I briefly describe the research question, methods and data used in each chapter, and discuss the findings and its contributions.

### Chapter 2: Does Education Improve Cognitive Performance Four Decades After School Completion?

In the first chapter, we replicate the analysis of Schneeweis et al. (2014) to examine the effect of education on cognitive abilities. The analysis is based on the same European data they used. We try to draw almost the same sample they used and also extend the sample by including more survey waves and countries. Just as Schneeweis et al. (2014), we exploit compulsory schooling reforms implemented in different European countries to endogenously determine the years of schooling. We make use of two-stage least squares regression. Schneeweis et al. (2014) find a positive effect of education on memory scores and some evidence of a protective effect of education on the decline in verbal fluency. Our results support their findings when we use the same waves as they do, but also when we extend the sample by including more countries and interview waves, and use different variables for years of education.

## Chapter 3 ("Does Education have an Impact on Patience and Risk Willingness?")

In this chapter, we analyse the causal effect of education on patience (also known as time preference) and risk willingness. Using the German compulsory schooling reform, which took effect in West Germany after World War II increasing compulsory schooling from eight years to nine years, we exogenously determine the years of schooling. We use two-stage least squares regression to obtain causal effects. In line with the literature, the results show a positive effect of education on risk willingness mainly for those who were the immediate partakers of the reform. Contrary to the literature, a negative effect of education on patience is found. This effect is larger as more years around the pivotal years are considered. Our results do not only contribute to the few causal analyses in this area, but also indicate that the type of educational system may affect the impact education has on patience.

## Chapter 4: Life-cycle Health Effects of Compulsory Schooling

In the third chapter, we study the effect of education on health (hospital stays, number of diagnosed conditions, self-rated poor health, and obesity) over the life-cycle in Germany. We use the same compulsory schooling reforms as in Chapter 3 as a source of exogenous variation. We pool data from two different German panel studies and a European survey, and implement the two-stage least squares method. Our results suggest a positive correlation between health and education which increases over the life-cycle. We do not, however, find any positive local average treatment effects of an additional year of schooling on health or health care utilization for individuals up to age 79. An exception is obesity, where positive effects of schooling start to be visible around age 60 and become very large in age group 75-79. The results in age group 75-79 need to be interpreted with caution, however, due to small sample size and possible problems of attrition. We also find that the additional year of schooling does not necessarily influence the choice of healthier jobs. This gives us some understanding to why we hardly find any long-term effects of education on health. Our results contribute to the scarce literature on the life-cycle effects of education.

## Chapter 5: The Effect of Children on Health

In this final chapter, I analyse the causal effect of having children on parents' health. I do not only focus on general measures of health but also the likelihood to be diagnosed with certain illnesses. A representative longitudinal study of private households from German is used for this analysis. To exogenously determine the number of children, I use twin birth at second birth and same sex of first two children as instruments. Two-stage least squares and reduced-form regressions are used for this study. I find negative effects on the BMI of women, and negative effects on the mental and physical health of men. Looking at the effects over the life-cycle, I find that the BMI of women increases until age 64. Mental health starts to decline from age 75 and the effect on physical health fluctuates over the life-cycle. The results contribute to the few literature showing that children do negatively affect parental health and introduce new results that indicate that the effects are not necessarily constant over the life-cycle. The effect children have on parental health should be considered when health policies, fertility-related policies and labour supply policies are being drafted.

# Chapter 2

# Does Education Improve Cognitive Performance Four Decades After School Completion?

**A Replication Study of Nicole Schneeweis, Vegard Skirbekk and Rudolf Winter-Ebmer (*Demography*, 2014)***

## 2.1 Introduction

The sustainability of national social security and health systems is likely to be challenged as a result of the ageing population in Europe (Schneeweis et al., 2014). The declining importance of state provided social security and healthcare systems around the world is an indicator for the importance of individual or household decision-making skills of older individuals (Banks and Mazzonna, 2012). Cognitive abilities are essential for decision making (Banks and Oldfield, 2007) and, therefore, important for labour market, pension and retirement policies (Banks and Oldfield, 2007; Schneeweis et al., 2014). Recent literature suggests that education may be an important determinant of cognitive abilities (Banks and Oldfield, 2007; Glymour et al., 2008; Brinch and Galloway, 2012; Schneeweis et al., 2014; Crespo et al., 2014; Kamhöfer and Schmitz, 2016; Kamhöfer et al., 2019), but the evidence is still sparse. Being able to replicate studies that find an effect of education on cognitive abilities is of relevance. We, therefore, aim to replicate the findings of Schneeweis et al. (2014) and investigate the validity of the findings with more data.

Schneeweis et al. (2014) analyse the long-run effects of education on cognitive performance. To estimate the effects they implement a Two-Stage Least Square (2SLS) approach. For identification, this approach makes use of compulsory schooling reforms

*This chapter is joint work with Valentin Schiele and is published as a working paper:

implemented in six European countries. Schneeweis et al. (2014) find a positive effect of education on memory performance and also find that education reduces the decline in verbal fluency. They find stronger effects for men and individuals who had many books at home when growing up.

We first replicate the results based on how Schneeweis et al. (2014) did their analysis. We use the same waves and determine the years of education in a similar manner. We then extend the sample by including more compulsory schooling reforms and more interview waves. Due to the availability of more information that comes with including more reforms and waves, we check how robust the first-stage results are when we use different variables for years of education. We also replicate the second stage results using more reforms, all available interview waves and our variable for years of education. Our results are not far-fetched from that of Schneeweis et al. (2014).

## 2.2 Data and sample selection

Schneeweis et al. (2014) use data from the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a micro panel database covering most of the European Union and Israel (Börsch-Supan et al., 2013). Schneeweis et al. (2014) consider only individuals from Austria, Czech Republic, Denmark, France, Germany and Italy, who participated in one or more waves. They use the first wave (2004/2005), second wave (2006/2007) and fourth wave (2011/2012) for their baseline analysis. In further analyses, they also use the third wave (2008/2009) known as the SHARELIFE, which has information on individuals' life histories. They only consider individuals aged 45 or older, who were born in the country of residence or migrated before age 5. This ensures that they attended school in the country of residence in the early stages, when they could possibly be affected by the compulsory schooling reforms. For the baseline sample, they select individuals born between 1939 and 1956. They also consider three sub-samples: individuals born up to *(i.)* 10 years (sample 10) *(ii.)* 7 years (sample 7), and *(iii.)* 5 years (sample 5) before and after the pivotal birth cohort of the respective reform.

The tests Schneeweis et al. (2014) used to measure cognitive functioning are *Immediate and Delayed Memory*, *Fluency*, *Numeracy* and *Orientation-to-date*. As explained by Schneeweis et al. (2014): *Immediate Memory* measures the number of words a respondent recalls out of ten words directly after they are read (range: 0 to 10); *Delayed Memory* measures the number of words a respondent recalls out of the same ten words 5 to 10 minutes later after other interview questions have been asked (range: 0 to 10); *Fluency* (verbal fluency score) measures the number of animal names a respondent is able to state in a minute (range: 0 to 100); *Numeracy* measures the ability of a respondent to answer basic and more advanced mathematical questions from daily life (range: 1 to

5); and *Orientation-to-date* measures a person's ability to remember the correct date comprising the day of the month, month, year, and day of the week (range: 0 to 4). They conduct a level analysis which uses the current test score and a slope analysis which uses the difference in current and previous test scores. For the level analysis, Schneeweis et al. (2014) generate binary variables for numeracy and orientation since "numeracy and orientation have larger densities at the upper tail of the distributions, with 55% achieving either the highest or the second-highest value of numeracy and 89% showing a perfect orientation-to-date", but treat the other test scores as continuous variables. They define *Good Numeracy* to be 1 for individuals who achieve numeracy scores of 4 and 5, and *Good Orientation* to be 1 for individuals scoring 4 on the orientation variable. They also define the change in test scores as "*cognitive decline*, which we calculate by subtracting the cognitive outcome from the cognitive outcome in a previous wave. Thus, a positive value implies a decline in cognitive performance, and a negative value represents a performance improvement." Only individuals who participated in more than one of the cognitive assessments are considered in the slope analysis, therefore, there are fewer observations.

For the first part of the replication, we use the same waves as Schneeweis et al. (2014) but a current version from SHARE (release 8.0.0). There are some differences between this release and the one used by Schneeweis et al. (2014) (release 2.3.0 for waves 1 and 2, and release 1.1.1 for wave 4). The major difference in the data has to do with the variable for years of education. There were two variables for years of education in wave 2, the raw years of education as provided by the respondents and the corrected years of education which are raw years of education corrected by SHARE. Subsequent waves and the current releases of wave 2 only have the corrected years of education. Comparing the two variables, we do not find much difference between them for most countries except for Denmark, where there are vast differences. Schneeweis et al. (2014) use the raw years of education for Denmark in wave 2, however, we use the corrected years education for consistency and as advised by the SHARE team. This creates some differences in the results which will be shown and discussed in the following sections. We also try as much as possible to adjust the years of education similar to that of Schneeweis et al. (2014). They use information on years of education from waves 2 or 4 for those who participated in those waves, since educational degrees but not years of education were asked in wave 1. The years of education of those who only participated in wave 1 are calculated using country-specific conversion tables provided by SHARE. Additional corrections are made to the years of education based on educational qualifications for missing, zero or implausibly low values.

Table 2.1 shows the descriptive statistics of the sample for the level analysis from Schneeweis et al. (2014) and for our sample. The proportion of females are exactly the same for both samples and the difference in average ages is marginal. With individual years of education, there is a year difference in the average for Austria and 1.5 years

Table 2.1: Replication of Table 2 from Schneeweis et al. (2014) - Descriptive statistics of baseline sample: Level analysis

| | Female | | Age | | Years of Education: Individuals | | Years of Education: Compulsory | | Immediate Memory | | Delayed Memory | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| Austria | 0.57 | 0.57 | 62.24 | 61.97 | 10.26 | 11.24 | 8.28 | 8.27 | 5.67 | 5.64 | 4.37 | 4.36 |
| Czech Republic | 0.57 | 0.57 | 61.76 | 61.28 | 12.16 | 12.12 | 8.58 | 8.58 | 5.48 | 5.49 | 3.78 | 3.79 |
| Denmark | 0.52 | 0.52 | 59.37 | 58.92 | 12.09 | 13.62 | 5.77 | 5.77 | 5.96 | 5.96 | 4.85 | 4.84 |
| France | 0.55 | 0.55 | 60.13 | 59.64 | 11.99 | 12.16 | 8.46 | 8.46 | 5.39 | 5.41 | 4.07 | 4.08 |
| Germany | 0.54 | 0.54 | 59.76 | 59.27 | 13.33 | 13.31 | 8.26 | 8.27 | 5.96 | 5.95 | 4.45 | 4.43 |
| Italy | 0.56 | 0.56 | 60.88 | 60.44 | 8.71 | 9.04 | 6.11 | 6.11 | 4.94 | 4.94 | 3.46 | 3.45 |
| Total | 0.55 | 0.55 | 60.81 | 60.33 | 11.26 | 11.73 | 7.63 | 7.59 | 5.51 | 5.51 | 4.09 | 4.09 |

| | Fluency | | Good Numeracy | | Good Orientation | | Observations | | Individuals | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| Austria | 23.56 | 23.57 | 0.72 | 0.74 | 0.91 | 0.92 | 4,724 | 4,213 | 3,624 | 3,183 |
| Czech Republic | 22.11 | 22.21 | 0.60 | 0.62 | 0.88 | 0.89 | 5,448 | 4,984 | 4,571 | 4,126 |
| Denmark | 23.92 | 23.93 | 0.58 | 0.60 | 0.90 | 0.90 | 3,755 | 3,768 | 1,901 | 1,904 |
| France | 21.01 | 21.16 | 0.51 | 0.54 | 0.87 | 0.89 | 5,683 | 5,537 | 3,644 | 3,498 |
| Germany | 23.27 | 23.23 | 0.74 | 0.74 | 0.92 | 0.93 | 2,860 | 2,920 | 1,590 | 1,597 |
| Italy | 15.93 | 15.88 | 0.33 | 0.33 | 0.90 | 0.91 | 5,229 | 5,207 | 2,928 | 2,895 |
| Total | 21.33 | 21.32 | 0.57 | 0.58 | 0.90 | 0.90 | 27,699 | 26,629 | 18,258 | 17,203 |

*Note: Orig.* presents the original values from Schneeweis et al. (2014) and *Repl.* presents the values from our replication sample.

difference in the average for Denmark. The difference for Denmark is not surprising since the variable for wave 2 used for Denmark by Schneeweis et al. (2014) (raw years of education) is different from what we use (corrected years of education). The difference found in Austria could be a result of the reduction in the number of individuals and hence, the number of observations in our sample. The corrected years of education could have also undergone further corrections in the current versions. The average years of education for the rest of the countries are similar. The averages of the test scores are quite similar. Our sample size is slightly smaller than that of Schneeweis et al. (2014). For the slope analysis, the descriptive statistics are shown in Table 2.2. The average change in tests scores is similar for most tests except *Numeracy*. The direction of the changes are very similar. As mentioned above, the number of observations for the slope analysis is smaller than that of the level analysis. Czech Republic was included in the SHARE from wave 2. Schneeweis et al. (2014) do not include wave 4 data for the *Numeracy* and *Orientation* tests in their sample hence, the missing values for these tests for Czech Republic in the original table, but we do.[1] In our sample, we find that the

[1] Schneeweis et al. (2014) write that data on the test scores for numeracy and orientation are not available in wave 4. However, the release we use (but also earlier releases currently available at the SHARE website: `http://www.share-project.org/data-access.html`) includes these test scores in wave 4.

average change in the test scores for Numeracy and Orientation for Czech Republic is zero.

Table 2.2: Replication of Table 3 from Schneeweis et al. (2014) - Descriptive statistics of baseline sample: Slope analysis

| | Immediate Memory | | Delayed Memory | | Fluency | | Numeracy | |
|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| Austria | −0.01 | −0.00 | −0.24 | −0.21 | 0.36 | 0.32 | −0.06 | −0.06 |
| Czech Republic | −0.08 | −0.05 | −0.07 | −0.04 | −2.23 | −2.00 | – | −0.01 |
| Denmark | 0.02 | 0.01 | −0.15 | −0.15 | −0.25 | −0.25 | −0.14 | −0.05 |
| France | −0.29 | −0.30 | −0.49 | −0.49 | 0.99 | 1.00 | −0.08 | −0.04 |
| Germany | 0.05 | 0.06 | −0.21 | −0.21 | 0.22 | 0.16 | −0.03 | −0.04 |
| Italy | −0.22 | −0.23 | −0.22 | −0.22 | −0.12 | −0.12 | −0.05 | −0.02 |
| Total | −0.11 | −0.11 | −0.25 | −0.24 | 0.00 | 0.01 | −0.07 | −0.04 |

| | Orientation | | Duration | | Observations | | Individuals | |
|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| Austria | −0.06 | −0.03 | 39.50 | 39.94 | 1,100 | 1,030 | 741 | 668 |
| Czech Republic | – | −0.00 | 49.50 | 49.45 | 877 | 858 | 877 | 858 |
| Denmark | 0.01 | 0.00 | 42.35 | 42.29 | 1,854 | 1,864 | 1,326 | 1,331 |
| France | 0.03 | 0.01 | 40.18 | 40.07 | 2,031 | 2,031 | 1,379 | 1,376 |
| Germany | 0.00 | −0.01 | 43.01 | 42.94 | 1,268 | 1,321 | 893 | 926 |
| Italy | 0.00 | −0.00 | 42.34 | 42.34 | 2,301 | 2,312 | 1,521 | 1,527 |
| Total | 0.00 | −0.00 | 42.30 | 42.31 | 9,431 | 9,416 | 6,737 | 6,686 |

*Note: Orig.* presents the original values from Schneeweis et al. (2014) and *Repl.* presents the values from our replication sample.

We also do the analyses using all available waves of the SHARE dataset i.e. waves 1, 2, 4-8 for the main analysis, and some information from wave 3 and wave 7 (SHARELIFE) for other analyses.[2] Including more waves enriches the sample and analyses, especially for the slope analysis. The increase in the sample size should reduce the standard errors, thereby making the estimates more precise ceteris paribus. Using three waves for the slope analyses allows for a maximum of two possible estimates of cognitive decline per individual. With the additional four waves, a maximum of seven observations per individual is possible. This enables us to analyse the effect of education on cognitive decline over a longer period. Finally, the inclusion of multiple waves also allows us to use a single measure of years of education, namely self-reported years of education, so that we do not have to approximate years of education based on information on the highest educational attainment and conversion tables, which is also a difficult task.[3]. Table A2.2 in the Appendix gives details on the different variables for years of education.

[2]See Börsch-Supan (2022a,b,c,d,e,f,g,h); Börsch-Supan et al. (2013). In wave 7, those who did not participate in wave 3, the SHARELIFE wave, were requested to do the SHARELIFE interview along with a condensed set of questions from the regular questionnaire. Those who already participated in wave 3 received a regular panel questionnaire.

[3]For example, some of the compulsory schooling reforms are not visible in SHARE's conversion tables

We, therefore, do a complete analysis with all the countries used in Schneeweis et al. (2014). Summary statistics for the all waves sample are given in Tables A2.3 and A2.4 in the Appendix.

## 2.3 Main Results

### 2.3.1 Narrow replication

Schneeweis et al. (2014) estimate the causal effect of education on the level of cognitive performance (level analysis, $l$) and on cognitive decline (slope analysis, $s$). For the level and slope analyses, they use the following models (presented differently) respectively:

$$Y_{ickt} = X'_{ickt}\beta_l + \rho_l E_{ickt} + \gamma_c + \lambda_k + \mu_c T + \varepsilon_{ickt} \tag{2.1}$$

$$Y_{ickt} - Y_{ickt+r} = X'_{ickt}\beta_s + \rho_s E_{ickt} + \gamma_c + \lambda_k + \mu_c T + \varepsilon_{ickt} - \varepsilon_{ickt+r} \tag{2.2}$$

where $Y_{ickt}$ is the cognitive achievement of individual $i$ in country $c$ of birth cohort $k$ in survey year $t$. $Y_{ickt} - Y_{ickt+r}$ is the change in cognitive performance in survey year $t$ compared with survey year $t + r$. $E_{ickt}$ is the number of years that the individual spent in education, and $X_{ickt}$ is a vector of control variables. $X_{ickt}$ includes a female dummy variable and an indicator variable for whether a person was born abroad and migrated before age 5. In Eq. (2.1), it also contains indicators for the interview year and control variables for the quality of the interview session (the interviewer's perception of whether something may have impaired the respondent's performance on the tests and whether another person was present during the interview). In Eq. (2.2), it also contains an indicator for the first interview year, control variables for the quality of both interview sessions, and the number of months between the two interviews (*Duration*). $\gamma_c$ and $\lambda_k$ refer to country and cohort fixed effects, and $\mu_c T$ captures country-specific linear trends in birth cohorts.

They use Two-Stage Least Squares (2SLS) to estimate Eqs. (2.1) and (2.2), because $\varepsilon_{ickt}$ and $\varepsilon_{ickt+r}$ might be correlated with years of education. They instrument years of education with the compulsory years of schooling ($Comp_{ck}$) in the respective country and birth cohort. The first-stage which shows the impact of compulsory schooling on years of education is modelled as:

$$E_{ickt} = X'_{ickt}\alpha + \pi Comp_{ck} + \gamma_c + \lambda_k + \mu_c T + \nu_{ickt} \tag{2.3}$$

The compulsory years of schooling are assigned as Schneeweis et al. (2014) did using the information provided in Table 1 from Schneeweis et al. (2014).The information is also available in Table A2.1 in the Appendix. For more details on the reforms, check the Appendix of Schneeweis et al. (2014).

We replicate Fig. 1 in Schneeweis et al. (2014) shown in Figure A2.1 in the Appendix, which depicts the effect of compulsory schooling on actual years of schooling. Compared with the original figure, our graph shifts upward. The upward shift can be attributed to the increased number of years of education found in Denmark and Austria. In spite of this, we also find a jump in the mean years of education at the time of the various reforms indicating an impact of the reforms on years of education.

Replicating Table 4 from Schneeweis et al. (2014) based on Eq. (2.3), our results in Table 2.3 also show an increase in years of education of about one-third of a year on average due to the increase in compulsory schooling years. They also estimate this effect using smaller windows around the pivotal cohort (see Section 2.2). "Smaller windows have the advantage that persons and circumstances before and after the changes in the law are similar but also the disadvantage of producing smaller sample sizes" (Schneeweis et al., 2014). However, it should be noted that a too small estimation window can lead to the situation in which the fixed effects for / trends in birth cohorts can no longer be cleanly distinguished from the effects of the reform due to insufficient overlap of cohorts across countries. We also replicate these in Table 2.3. Our results are not so different from the original results.

The replication results of Table 5 in Schneeweis et al. (2014) are presented in two tables, level analysis in Table 2.4 and slope analysis in Table 2.5. OLS and 2SLS results are based on

Table 2.3: Replication of Table 4 from Schneeweis et al. (2014) - First stage regressions

| | Baseline | | Sample 10 | | Sample 7 | | Sample 5 | |
|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| Compulsory schooling | 0.315** | 0.302** | 0.317** | 0.304** | 0.314** | 0.301** | 0.331** | 0.312** |
| | (0.062) | (0.056) | (0.063) | (0.057) | (0.073) | (0.063) | (0.090) | (0.077) |
| F Statistics | 25.82 | 29.08 | 24.98 | 28.31 | 18.41 | 22.70 | 13.40 | 16.60 |
| Observations | 27,699 | 26,512 | 25,378 | 24,295 | 20,126 | 19,977 | 15,509 | 15,630 |

*Note:* Standard errors clustered at individual level in parenthesis. The sample includes all observations with non-missing immediate memory scores. * $p < 0.05$, ** $p < 0.01$
*Orig.* presents the original results from Schneeweis et al. (2014) and *Repl.* presents the results from our replication sample.

Eq. (2.1) for the level analysis and Eq. (2.2) for the slope analysis.

Just as Schneeweis et al. (2014), our OLS results also show a positive association between education and levels of cognitive functioning. Our results are slightly larger. For 2SLS results in the level analysis, the direction of the results are almost the same for all the cognitive tests except in sample 7 and sample 5 of good orientation. We also do not find statistically significant results for fluency, good Numeracy and good Orientation. For immediate memory and delayed memory, the standard errors are quite similar but the effects are smaller than the original hence, most of them lose some strength in statistical significance. For instance in the baseline results, Schneeweis et al. (2014) find that an additional year of schooling increases immediate memory by 0.14 words at 5% significance level but we find an increase of 0.11 words at 10% significance level. Just as Schneeweis et al. (2014), the size of the effects is larger the smaller the sample around the pivotal cohorts with sample 5 having the largest effect. For those in sample 5, the effects for immediate memory and delayed memory remain strong. This indicates that the effect of an additional year of schooling on memory is more evident amongst those closer to the reform.

Table 2.4: Replication of Table 5 (Level Analysis) from Schneeweis et al. (2014) - Baseline results

| dep. var: | Immediate Memory | | Delayed Memory | | Fluency | | Good Numeracy | | Good Orientation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| A. OLS | | | | | | | | | | |
| Baseline | 0.112** | 0.126** | 0.125** | 0.139** | 0.504** | 0.568** | 0.032** | 0.038** | 0.004** | 0.005** |
| | (0.003) | (0.004) | (0.004) | (0.004) | (0.015) | (0.017) | (0.001) | (0.001) | (0.001) | (0.001) |
| Observations | 27,699 | 26,512 | 27,693 | 26,503 | 27,555 | 26,408 | 22,368 | 26,512 | 22,467 | 26,512 |
| B. 2SLS | | | | | | | | | | |
| Baseline | 0.144* | 0.113$^+$ | 0.171* | 0.129 | −0.260 | −0.334 | −0.013 | −0.026 | −0.007 | −0.019 |
| | (0.066) | (0.068) | (0.078) | (0.082) | (0.322) | (0.346) | (0.023) | (0.026) | (0.013) | (0.015) |
| Observations | 27,699 | 26,512 | 27,693 | 26,503 | 27,555 | 26,408 | 22,368 | 26,512 | 22,467 | 26,512 |
| Sample 10 | 0.155* | 0.122$^+$ | 0.184* | 0.139$^+$ | −0.020 | −0.109 | −0.012 | −0.023 | −0.006 | −0.019 |
| | (0.067) | (0.069) | (0.080) | (0.083) | (0.308) | (0.331) | (0.023) | (0.026) | (0.013) | (0.015) |
| Observations | 25,378 | 24,295 | 25,375 | 24,287 | 25,245 | 24,199 | 20,450 | 24,295 | 20,540 | 24,295 |
| Sample 7 | 0.205** | 0.140$^+$ | 0.217* | 0.143 | −0.161 | −0.057 | −0.023 | −0.031 | 0.002 | −0.019 |
| | (0.079) | (0.077) | (0.093) | (0.092) | (0.366) | (0.368) | (0.026) | (0.030) | (0.015) | (0.016) |
| Observations | 20,126 | 19,977 | 20,124 | 19,971 | 20,021 | 19,901 | 16,257 | 19,977 | 16,333 | 19,977 |
| Sample 5 | 0.233* | 0.198* | 0.324** | 0.272* | −0.361 | −0.382 | −0.032 | −0.041 | 0.001 | −0.024 |
| | (0.093) | (0.090) | (0.118) | (0.113) | (0.445) | (0.448) | (0.032) | (0.036) | (0.017) | (0.019) |
| Observations | 15,509 | 15,630 | 15,422 | 15,626 | 15,507 | 15,566 | 12,559 | 15,630 | 12,618 | 15,630 |

*Note:* Standard errors clustered at individual level in parenthesis. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
*Orig.* presents the original results from Schneeweis et al. (2014) and *Repl.* presents the results from our replication sample.

Similar to Schneeweis et al. (2014), we only find a statistically significant association between schooling and cognitive decline in delayed memory. We also do not find statistically significant effects for decline in immediate memory, delayed memory and numeracy. We find larger effects than the original for fluency but they are only statistically significant at 10% significance level. For decline in orientation we find very similar effect sizes as Schneeweis et al. (2014), however, our estimates are more precise and thus turn out to be significant. This might be a result of the larger sample size (our sample also includes wave 4 observations, while the sample used in Schneeweis et al. (2014) does not).

Importantly and other than one would expect, the estimates suggest that education accelerates the decline in orientation. A possible explanation for positive effects of education on decline in orientation is that more education leads to higher orientation, which than leads to a faster decline in orientation in older ages. This explanation is not supported by the results for the level analysis though.

Table 2.5: Replication of Table 5 (Slope Analysis) from Schneeweis et al. (2014) - Baseline results

| dep. var: $\Delta$ in ... | Immediate Memory | | Delayed Memory | | Fluency | | Numeracy | | Orientation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| A. OLS | | | | | | | | | | |
| Baseline | −0.002 | −0.002 | −0.016** | −0.015** | −0.011 | 0.001 | −0.003 | 0.000 | 0.000 | −0.000 |
| | (0.004) | (0.005) | (0.005) | (0.005) | (0.016) | (0.018) | (0.005) | (0.002) | (0.002) | (0.001) |
| Observations | 9,431 | 9,370 | 9,435 | 9,368 | 9,378 | 9,326 | 6,737 | 9,370 | 6,768 | 9,370 |
| B. 2SLS | | | | | | | | | | |
| Baseline | −0.080 | −0.123 | −0.083 | −0.088 | −0.755* | $-0.760^{+}$ | −0.041 | −0.025 | 0.061 | 0.057* |
| | (0.085) | (0.101) | (0.093) | (0.107) | (0.367) | (0.415) | (0.082) | (0.043) | (0.045) | (0.025) |
| Observations | 9,431 | 9,370 | 9,435 | 9,368 | 9,378 | 9,326 | 6,737 | 9,370 | 6,768 | 9,370 |
| Sample 10 | −0.079 | −0.134 | −0.101 | −0.107 | −0.780* | $-0.824^{+}$ | −0.025 | −0.015 | 0.062 | 0.057* |
| | (0.084) | (0.102) | (0.093) | (0.108) | (0.366) | (0.424) | (0.080) | (0.043) | (0.044) | (0.025) |
| Observations | 8,561 | 8,514 | 8,567 | 8,513 | 8,513 | 8,474 | 5,973 | 8,514 | 6,002 | 8,514 |
| Sample 7 | −0.045 | −0.100 | −0.010 | −0.049 | −0.606 | −0.617 | 0.081 | 0.025 | 0.077 | 0.066* |
| | (0.095) | (0.115) | (0.103) | (0.123) | (0.391) | (0.459) | (0.086) | (0.050) | (0.048) | (0.032) |
| Observations | 6,757 | 6,966 | 6,762 | 6,965 | 6,717 | 6,933 | 4,729 | 6,966 | 4,752 | 6,966 |
| Sample 5 | 0.062 | 0.042 | 0.073 | 0.001 | −0.616 | −0.791 | 0.078 | 0.051 | 0.050 | $0.070^{+}$ |
| | (0.104) | (0.124) | (0.116) | (0.135) | (0.422) | (0.520) | (0.095) | (0.056) | (0.049) | (0.036) |
| Observations | 5,154 | 5,364 | 5,157 | 5,363 | 5,117 | 5,334 | 3,605 | 5,364 | 3,627 | 5,364 |

*Note:* Standard errors clustered at individual level in parenthesis. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
*Orig.* presents the original results from Schneeweis et al. (2014) and *Repl.* presents the results from our replication sample.

Schneeweis et al. (2014) also check for whether the effects vary by gender and family background using sample 10. 2SLS regressions are estimated for males and females separately. The measure Schneeweis et al. (2014) wanted to use for family background was education of parents. This is, unfortunately, not available hence they use the number of books an individual had available at home at age 10 as a proxy. Based on the variable for the number of books, they split the sample into two: individuals with few books (0 - 10 or 11-25 books) and individuals with many books (26 - 100, 101 - 200 or more than 200 books).

Looking at the level analysis for gender in panel A of Table 2.6, we only find a significant effect on immediate memory for males at 10% significance level. Just as Schneeweis et al. (2014) , we do not find significant effects for females. The size of the effects is a bit larger in our replication results for men, in general the results are quite similar. The results for the gender gradient in the slope analysis (panel A of Table 2.7) are also rather similar, although we do not find significant effects on the decline in delayed memory and fluency for males like Schneeweis et al. (2014) did, but instead a marginally significant effect on change in orientation for females.

Table 2.6: Replication of Table 6 (Level Analysis) from Schneeweis et al. (2014) - Heterogeneity Analysis

| dep. var: | Immediate Memory | | Delayed Memory | | Fluency | | Good Numeracy | | Good Orientation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| A. By Gender | | | | | | | | | | |
| Male | 0.234* | 0.283+ | 0.205+ | 0.202 | −0.243 | −0.512 | −0.012 | −0.039 | −0.018 | −0.037 |
| | (0.107) | (0.165) | (0.120) | (0.181) | (0.478) | (0.798) | (0.034) | (0.060) | (0.021) | (0.036) |
| Observations | 11,361 | 10,827 | 11,360 | 10,823 | 11,271 | 10,770 | 9,240 | 10,827 | 9,283 | 10,827 |
| Female | 0.090 | 0.045 | 0.167 | 0.105 | 0.173 | 0.071 | −0.009 | −0.016 | 0.005 | −0.012 |
| | (0.097) | (0.079) | (0.116) | (0.094) | (0.441) | (0.358) | (0.032) | (0.029) | (0.018) | (0.015) |
| Observations | 14,017 | 13,468 | 14,015 | 13,464 | 13,974 | 13,429 | 11,210 | 13,468 | 11,257 | 13,468 |
| B. By Family Background | | | | | | | | | | |
| Few books | 0.104 | 0.054 | 0.047 | 0.010 | −0.025 | −0.078 | 0.011 | −0.002 | −0.018 | −0.033 |
| | (0.135) | (0.089) | (0.166) | (0.112) | (0.616) | (0.397) | (0.049) | (0.033) | (0.031) | (0.021) |
| Observations | 7,853 | 8,002 | 7,857 | 8,002 | 7,830 | 7,983 | 5,165 | 8,002 | 5,176 | 8,002 |
| Many books | 0.326+ | 0.923 | 0.464* | 1.354 | 0.641 | 1.987 | −0.058 | −0.222 | 0.058+ | 0.182 |
| | (0.166) | (1.035) | (0.226) | (1.558) | (0.685) | (3.029) | (0.056) | (0.321) | (0.035) | (0.224) |
| Observations | 5,272 | 5,389 | 5,276 | 5,388 | 5,262 | 5,381 | 3,355 | 5,389 | 3,364 | 5,389 |

*Note:* Standard errors clustered at individual level in parenthesis. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
*Orig.* presents the original results from Schneeweis et al. (2014) and *Repl.* presents the results from our replication sample.

With respect to the role of number of books, we do not find any significant effects in either groups in the level analysis as show in panel B of Table 2.6. Results of the slope analysis in panel B of Table 2.7 show a positive and significant effect on orientation for those who had few books. Although insignificant, we find extremely large effects on immediate memory, delayed memory and fluency for those who had many books, especially fluency. In general, the effects by family background are estimated very imprecisely and thus do not seem to be too trustworthy.

Table 2.7: Replication of Table 6 (Slope Analysis) from Schneeweis et al. (2014) - Heterogeneity Analysis

| dep. var: | Immediate Memory | | Delayed Memory | | Fluency | | Numeracy | | Orientation | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* | *Orig.* | *Repl.* |
| A. By Gender | | | | | | | | | | |
| Male | −0.123 | −0.180 | −0.227+ | −0.289 | −0.826 + | −0.978 | −0.022 | 0.004 | 0.073 | 0.060 |
| | (0.113) | (0.199) | (0.135) | (0.235) | (0.486) | (0.859) | (0.090) | (0.080) | (0.055) | (0.054) |
| Observations | 3,768 | 3,731 | 3,774 | 3,732 | 3,738 | 3,708 | 2,645 | 3,748 | 2,654 | 3,748 |
| Female | −0.031 | −0.093 | 0.032 | −0.001 | −0.787 | −0.724 | −0.042 | −0.033 | 0.040 | 0.054* |
| | (0.135) | (0.111) | (0.152) | (0.120) | (0.604) | (0.451) | (0.148) | (0.049) | (0.075) | (0.025) |
| Observations | 4,793 | 4,783 | 4,793 | 4,783 | 4,775 | 4,772 | 3,328 | 4,799 | 3,348 | 4,799 |
| B. By Family Background | | | | | | | | | | |
| Few books | −0.014 | −0.049 | 0.140 | 0.123 | −0.959 | −0.645 | −0.072 | −0.024 | 0.119 | 0.060* |
| | (0.146) | (0.102) | (0.167) | (0.115) | (0.699) | (0.420) | (0.181) | (0.044) | (0.105) | (0.027) |
| Observations | 4,428 | 4,528 | 4,432 | 4,530 | 4,406 | 4,510 | 3,205 | 4,541 | 3,218 | 4,541 |
| Many books | −0.253 | −1.410 | −0.511 | −2.556 | −1.539 | −8.663 | −0.156 | −0.132 | −0.103 | −0.019 |
| | (0.201) | (3.168) | (0.318) | (5.973) | (1.080) | (26.195) | (0.174) | (0.413) | (0.088) | (0.119) |
| Observations | 2,840 | 2,909 | 2,844 | 2,908 | 2,834 | 2,905 | 1,714 | 2,917 | 1,726 | 2,917 |

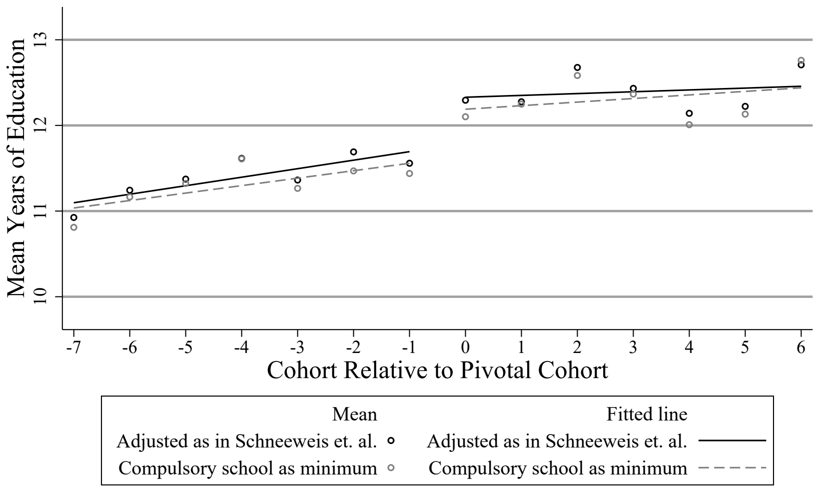
*Note:* Standard errors clustered at individual level in parenthesis. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
*Orig.* presents the original results from Schneeweis et al. (2014) and *Repl.* presents the results from our replication sample.

### 2.3.2 Extension: Wider replication

In the replications so far, we used the years of education variable adjusted in a similar manner as Schneeweis et al. (2014). Given the availability of more information from subsequent waves, we are able to use years of education as provided by the respondents also for wave 1 instead of using the conversion table.[4] The main adjustment we make is using the years of compulsory schooling as the minimum number of years of education. We then check how robust the results are using our adjusted years of education variable. Before we do so in the next section, we check whether we can replicate a first-stage effect using only reported years of education instead of also relying on the conversion tables.

Comparing the two education variables for the Schneeweis et al. (2014) sample, we find only small differences between the means per cohort as depicted in Figure 2.1. Our adjustments are slightly below that of Schneeweis et al. (2014). This implies that the provided information on years of education does not deviate too much from the years of education calculated from the conversion tables. We use our adjusted years of education for the analyses using the all waves sample.

Figure 2.1: Comparing the two education variables - First stage




Now we do a wide replication analysis using all the available waves till date and our adjusted years of education, adding more countries and more reforms, and expanding the range of birth year. We include observations from Belgium [5], Netherlands and Sweden. The reforms we now use are presented in Table 2.8. They include reforms used by Schneeweis et al. (2014) as well as additional reforms from some countries they used and the new countries we consider. Individuals born from 1920 to 1959 are selected for the baseline analysis. Instead of looking at all three sub-samples as Schneeweis et al.

[4]For each individual we use the maximum number of (reported) years of education for all available observations.

[5]Only individuals who went to school in Flanders are considered since the reform only took place in this region.

(2014), we only look at the sample 10. From the baseline data, we consider individuals born up to 10 years before and after the pivotal birth cohort. For countries with multiple reforms, we consider individuals born up to 10 years before the pivotal cohorts of the first reform and 10 years after the pivotal cohorts of the last reform presented in Table 2.8. We only consider the sample 10 sub-sample to allow for enough overlapping of birth cohorts across the different countries. As stated above, the inclusion of the additional data should improve the results especially with respect to precision. Since the panel has been extended with more waves, we include survey year fixed effects in the slope analysis. From the descriptive analysis in Tables A2.3 and A2.4, we gain more than 100,000 observations for the level analysis and over 70,000 observations for the slope analysis.

Table 2.8: Compulsory schooling reforms

| Country | Reform | Increase in Mandatory Years of Schooling | School-Leaving Age | Pivotal Cohort |
|---|---|---|---|---|
| Austria | 1962 | 8 to 9 | 14 to 15 | 1951 |
| Belgium (Flanders) | 1953 | 8 to 9 | 14 to 15 | 1939 |
| Czech Republic | 1948 | 8 to 9 | 14 to 15 | 1934 |
| | 1953 | 9 to 8 | 15 to 14 | 1939 |
| | 1960 | 8 to 9 | 14 to 15 | 1947 |
| Denmark | 1958 | 4 to 7 | 11 to 14 | 1947 |
| France | 1936 | 7 to 8 | 13 to 14 | 1923 |
| | 1959 | 8 to 10 | 14 to 16 | 1953 |
| Germany | | | | |
| Hamburg | 1949 | 8 to 9 | 14 to 15 | 1934 |
| Schleswig-Holstein | 1956 | 8 to 9 | 14 to 15 | 1941 |
| Bremen | 1958 | 8 to 9 | 14 to 15 | 1943 |
| Lower Saxony | 1962 | 8 to 9 | 14 to 15 | 1947 |
| Saarland | 1964 | 8 to 9 | 14 to 15 | 1949 |
| Northrhine-Westphalia | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Hesse | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Rhineland-Palatinate | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Baden-Wuerttemberg | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Bavaria | 1969 | 8 to 9 | 14 to 15 | 1955 |
| Italy | 1963 | 5 to 8 | 11 to 14 | 1949 |
| Netherlands | 1942 | 7 to 8 | 13 to 14 | 1951 |
| | 1947 | 8 to 7 | 14 to 13 | 1951 |
| | 1950 | 7 to 9 | 13 to 15 | 1951 |
| Sweden | 1949 | 6 to 7 | 13 to 14 | 1936 |
| | 1962 | 8 to 9 | 14 to 16 | 1950 |

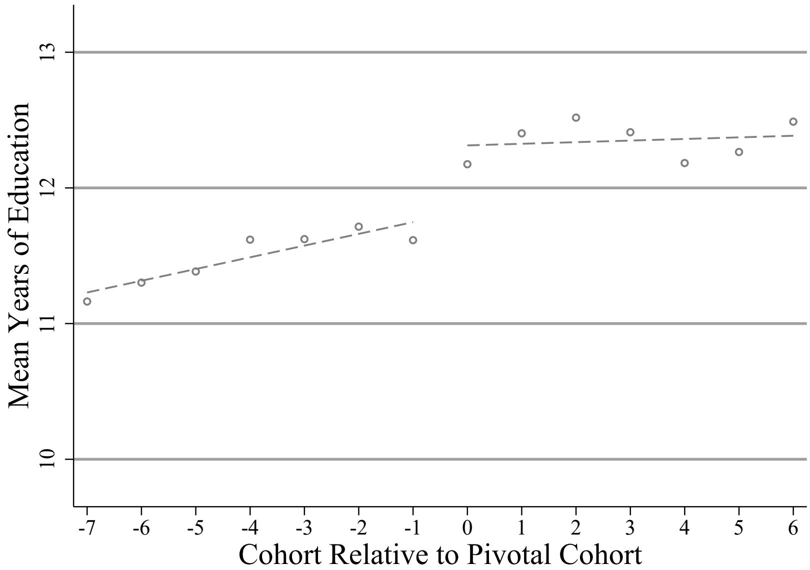
*Source:* Brunello et al. (2016)

Figure 2.2 shows that there is still a jump in years of education as a result of the school reforms even with the inclusion of more waves and reforms.[6] The results in Table 2.9 are

[6]The last reform of countries with multiple reforms, i.e. Czech Republic, France, Netherlands and Sweden, are used for the graph.

slightly smaller and the standard errors are relatively smaller. The first-stage F-statistics are also larger. We find that an additional year of compulsory schooling increases years of schooling on average between one-fourth and one-third of a year.

Figure 2.2: Replication of Fig 1 from Schneeweis et al. (2014) - First stage, all waves sample




*Note:* The last reform of countries with multiple reforms, i.e. Czech Republic, France, Netherlands and Sweden, are used for this graph.

Table 2.9: First stage regressions using all waves

| | Baseline | Sample 10 |
|---|---|---|
| Compulsory schooling | 0.258** | 0.298** |
| | (0.037) | (0.045) |
| F Statistics | 49.04 | 44.61 |
| Observations | 122,906 | 97,632 |

*Note:* The sample includes all observations with non-missing immediate memory scores. Standard errors clustered at individual level in parenthesis. * $p < 0.05$, ** $p < 0.01$

Table 2.10 shows the OLS and 2SLS results for both the level and slope analyses. The positive association between education and levels of cognitive functioning still holds. We also find evidence for a positive effect of education on memory. The estimates suggest that an additional year of education as a result of an additional year of compulsory schooling improves immediate memory by 0.03-0.11 and delayed memory by 0.11-0.15 words on average and are thus in line with the results presented so far. In the slope analysis we find negative associations between education and cognitive decline, indicating that higher education is associated with a slower decline in cognitive abilities. The 2SLS estimates, however, provide no clear evidence that education causally slows down cognitive decline. Compared to the results from the narrow replication presented in Table 2.5, the size of the 2SLS estimates for the change in fluency have reduced considerably and lost their statistical significance. Interestingly, the somewhat suspicious positive effect for the decline in orientation also shrank considerably in the larger sample.

Table 2.10: Baseline results, all waves

| dep. var: | Level Immediate Memory | Delayed Memory | Fluency | Good Numeracy | Good Orientation | Slope (Δ in ..) Immediate Memory | Delayed Memory | Fluency | Good Numeracy | Good Orientation |
|---|---|---|---|---|---|---|---|---|---|---|
| A. OLS | | | | | | | | | | |
| Baseline | 0.115** | 0.132** | 0.494** | 0.037** | 0.005** | −0.000 | −0.005** | −0.010* | −0.001* | −0.001** |
| | (0.002) | (0.003) | (0.009) | (0.001) | (0.000) | (0.001) | (0.001) | (0.005) | (0.000) | (0.000) |
| Observations | 122,906 | 122,790 | 122,552 | 122,906 | 122,906 | 81,375 | 81,266 | 81,092 | 81,375 | 81,375 |
| B. 2SLS | | | | | | | | | | |
| Baseline | 0.027 | $0.107^{+}$ | −0.150 | −0.015 | −0.012 | −0.028 | −0.013 | −0.164 | 0.008 | 0.016 |
| | (0.047) | (0.057) | (0.227) | (0.020) | (0.010) | (0.031) | (0.036) | (0.120) | (0.011) | (0.010) |
| Observations | 122,906 | 122,790 | 122,552 | 122,906 | 122,906 | 81,375 | 81,266 | 81,092 | 81,375 | 81,375 |
| Sample 10 | 0.106* | 0.152* | 0.321 | −0.003 | −0.000 | −0.043 | −0.034 | −0.118 | −0.003 | 0.022* |
| | (0.047) | (0.061) | (0.229) | (0.021) | (0.009) | (0.032) | (0.036) | (0.121) | (0.010) | (0.010) |
| Observations | 97,632 | 97,548 | 97,407 | 97,632 | 97,632 | 65,282 | 65,201 | 65,093 | 65,282 | 65,282 |

*Note:* Standard errors clustered at individual level in parenthesis. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

When we estimate the effects of education in the all waves sample by sex and family background, we find only little evidence for gradients between men and women or between individuals who had few books during childhood and individuals who had many books. Only in the level analysis and with respect to fluency we find larger differences between men and women indicating that education has lasting positive effects on cognitive abilities for women but not for man.

Table 2.11: Heterogeneity Analysis, all waves

| dep. var: | Level Immediate Memory | Delayed Memory | Fluency | Good Numeracy | Good Orientation | Slope (Δ in ..) Immediate Memory | Delayed Memory | Fluency | Numeracy | Orientation |
|---|---|---|---|---|---|---|---|---|---|---|
| A. By Gender | | | | | | | | | | |
| Males | 0.133 | 0.155 | −0.117 | −0.023 | 0.010 | −0.014 | 0.013 | −0.025 | 0.012 | 0.030 |
| | (0.099) | (0.125) | (0.527) | (0.046) | (0.021) | (0.064) | (0.072) | (0.260) | (0.022) | (0.022) |
| Observations | 43,893 | 43,849 | 43,765 | 43,893 | 43,893 | 28,908 | 28,901 | 28,881 | 29,107 | 29,107 |
| Females | 0.082 | 0.139* | $0.473^{+}$ | 0.006 | −0.006 | −0.052 | −0.048 | −0.136 | −0.013 | $0.016^{+}$ |
| | (0.053) | (0.069) | (0.248) | (0.024) | (0.010) | (0.035) | (0.040) | (0.126) | (0.012) | (0.010) |
| Observations | 53,739 | 53,699 | 53,642 | 53,739 | 53,739 | 36,374 | 36,368 | 36,379 | 36,547 | 36,547 |
| B. By Family Background | | | | | | | | | | |
| Few Books | 0.047 | 0.039 | 0.178 | 0.010 | −0.019 | −0.008 | 0.019 | −0.171 | −0.005 | $0.016^{+}$ |
| | (0.056) | (0.072) | (0.254) | (0.024) | (0.012) | (0.030) | (0.036) | (0.119) | (0.011) | (0.010) |
| Observations | 44,088 | 44,051 | 43,991 | 44,088 | 44,088 | 32,146 | 32,144 | 32,129 | 32,324 | 32,324 |
| Many Books | 0.561 | 1.174 | 2.395 | −0.197 | 0.067 | −0.235 | −0.096 | 0.199 | −0.036 | 0.025 |
| | (0.677) | (1.372) | (3.173) | (0.314) | (0.104) | (0.417) | (0.306) | (1.046) | (0.099) | (0.071) |
| Observations | 34,755 | 34,736 | 34,709 | 34,755 | 34,755 | 25,473 | 25,468 | 25,482 | 25,597 | 25,597 |

*Note:* Standard errors clustered at individual level in parenthesis. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

## 2.4 Conclusion

This paper replicates and extends the main results of Schneeweis et al. (2014). We do find similar results in the replication of the main results, although there are some differences in statistical significance. We also replicate the heterogeneity analysis from Schneeweis et al. (2014). The replication shows a gender gradient in the outcome variable recall, as in the original study. Other parts of the replication of the heterogeneity analysis are less conclusive, mainly due to rather noisy estimates in the replication. We then extend the sample used by Schneeweis et al. (2014) by including more reforms as well as subsequent interview waves and re-adjusting the years of education based on the information gained.Here, we also find evidence for positive effects of education on memory.

# Appendix

Table A2.1: Compulsory schooling reforms from Schneeweis et al. (2014)

| Country | Reform | Increase in Mandatory Years of Schooling | School-Leaving Age | Pivotal Cohort |
|---|---|---|---|---|
| Austria | 1962/1966 | 8 to 9 | 14 to 15 | 1951 |
| Czech Republic | 1960 | 8 to 9 | 14 to 15 | 1947 |
| Denmark | 1958 | 4 to 7 | 11 to 14 | 1947 |
| France | 1959/1967 | 8 to 10 | 14 to 16 | 1953 |
| Germany | | | | |
| Hamburg | 1949 | 8 to 9 | 14 to 15 | 1934 |
| Schleswig-Holstein | 1956 | 8 to 9 | 14 to 15 | 1941 |
| Bremen | 1958 | 8 to 9 | 14 to 15 | 1943 |
| Lower Saxony | 1962 | 8 to 9 | 14 to 15 | 1947 |
| Saarland | 1964 | 8 to 9 | 14 to 15 | 1949 |
| Northrhine-Westphalia | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Hesse | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Rhineland-Palatinate | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Baden-Wuerttemberg | 1967 | 8 to 9 | 14 to 15 | 1953 |
| Bavaria | 1969 | 8 to 9 | 14 to 15 | 1955 |
| Italy | 1963 | 5 to 8 | 11 to 14 | 1949 |

*Note:* 1966 is used for calculating the compulsory years of schooling in Austria. The 1967 reform in France is used for calculating the compulsory years of schooling.

Table A2.2: The determinants for the varying years of education variable

| Country | Schneeweis et al. (2014) | Replicated based on Schneeweis et al. (2014) | New variable |
|---|---|---|---|
| *Variables used from SHARE* | | | |
| Austria, Czech, France, Germany and Italy | Corrected reported years of education (dn041_) | Corrected reported years of education (dn041_) | Corrected reported years of education (dn041_) |
| Denmark | Raw reported years of education (dn041_raw) | Corrected reported years of education (dn041_) | Corrected reported years of education (dn041_) |
| *Adjustments* | | | |
| Years of education for Wave 1 based on: | Highest educational attainment and conversion tables | Highest educational attainment and conversion tables | Reported years of education from subsequent waves |
| Further adjustments | Compulsory schooling as the minimum number of years of schooling for wave 1. Computed years of education based on conversion tables as the minimum years if reported years of schooling is less than the compulsory years of schooling | Same as Schneeweis et al. (2014) | Compulsory schooling as the minimum number of years of schooling |

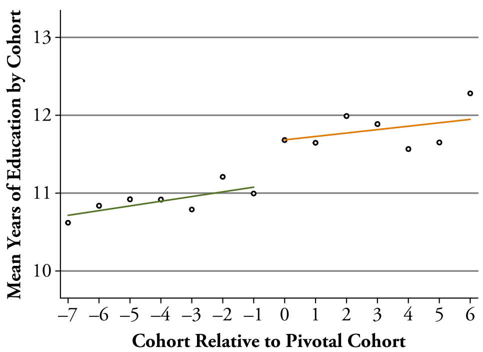
Table A2.3: Replication of Table 2 from Schneeweis et al. (2014) - Descriptive statistics of baseline sample: Level analysis, all waves sample

| | Female | Age | Years of Education | | Immediate Memory | Delayed Memory | Fluency | Good Numeracy | Good Orientation | Observations | Individuals |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Individual | Compulsory | | | | | | | |
| Austria | 0.58 | 68.94 | 10.31 | 8.21 | 5.51 | 4.19 | 22.52 | 0.71 | 0.90 | 13,381 | 4,749 |
| Belgium | 0.53 | 67.91 | 12.33 | 8.68 | 5.15 | 3.71 | 20.15 | 0.49 | 0.85 | 12,202 | 3,664 |
| Czech Republic | 0.58 | 68.33 | 12.27 | 8.56 | 5.35 | 3.73 | 21.90 | 0.60 | 0.87 | 18,290 | 6,999 |
| Denmark | 0.53 | 67.71 | 13.19 | 5.36 | 5.56 | 4.36 | 23.05 | 0.57 | 0.87 | 13,816 | 4,188 |
| France | 0.57 | 68.56 | 11.69 | 8.31 | 5.01 | 3.68 | 19.08 | 0.49 | 0.85 | 17,829 | 5,847 |
| Germany | 0.52 | 67.72 | 12.83 | 8.23 | 5.50 | 4.02 | 21.80 | 0.67 | 0.89 | 11,417 | 4,694 |
| Italy | 0.54 | 68.49 | 8.55 | 5.85 | 4.47 | 2.96 | 14.92 | 0.29 | 0.87 | 18,641 | 6,549 |
| Netherlands | 0.54 | 66.35 | 11.65 | 8.67 | 5.39 | 4.10 | 20.56 | 0.62 | 0.86 | 11,917 | 5,117 |
| Sweden | 0.53 | 69.27 | 11.56 | 7.21 | 5.33 | 4.18 | 23.17 | 0.60 | 0.90 | 16,394 | 5,363 |
| Total | 0.55 | 68.23 | 11.50 | 7.61 | 5.22 | 3.84 | 20.61 | 0.55 | 0.87 | 133,887 | 47,170 |

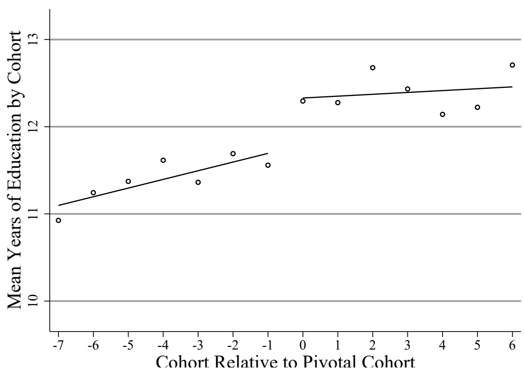
Table A2.4: Replication of Table 3 from Schneeweis et al. (2014) - Descriptive statistics of baseline sample: Slope analysis, all waves sample

| | Immediate Memory | Delayed Memory | Fluency | Numeracy | Orientation | Duration | Observations | Individuals |
|---|---|---|---|---|---|---|---|---|
| Austria | 0.02 | 0.02 | 0.71 | -0.01 | 0.02 | 31.81 | 8,573 | 3,758 |
| Belgium | 0.06 | -0.01 | 0.22 | -0.00 | 0.02 | 30.97 | 8,481 | 2,784 |
| Czech Republic | 0.05 | -0.02 | 0.10 | 0.00 | 0.03 | 32.23 | 11,198 | 4,638 |
| Denmark | 0.14 | 0.13 | 0.42 | -0.01 | 0.03 | 33.66 | 9,595 | 3,242 |
| France | 0.04 | 0.01 | 0.44 | -0.00 | 0.03 | 33.95 | 11,912 | 4,385 |
| Germany | 0.10 | 0.04 | 0.48 | -0.00 | 0.03 | 35.30 | 6,709 | 2,962 |
| Italy | 0.09 | 0.08 | 0.06 | 0.00 | 0.03 | 34.26 | 11,980 | 4,299 |
| Netherlands | 0.13 | 0.13 | 0.37 | -0.01 | 0.02 | 47.14 | 6,793 | 3,262 |
| Sweden | 0.14 | 0.10 | 0.66 | 0.00 | 0.03 | 35.71 | 10,973 | 4,172 |
| Total | 0.08 | 0.05 | 0.37 | -0.00 | 0.03 | 34.60 | 86,214 | 33,502 |

Figure A2.1: Replication of Fig 1 from Schneeweis et al. (2014) - First stage




(a) Original




(b) Replication

# Chapter 3

# Does Education have an Impact on Patience and Risk Willingness?

## 3.1 Introduction

Events in various stages of a person's life may influence the formation of character traits. The early years of a child, known as the formative years, and the adolescent period are examples of stages in a person's life where character traits are formed. For example, the presence or absence of parents and siblings as well as the relationship with them during the formative years are known to influence character traits in children (de Carvalho et al. (2015); Josefsson et al. (2013)). The development process and peer influence during the adolescent period also influence character traits (Blos (1968); Adatto (1980)). Other factors such as gender, age, parental education, wealth and height have also been found to influence certain character traits such as patience and risk attitudes in papers such as Dohmen et al. (2011) and Becker and Mulligan (1997).

This paper, however, explores one of the other possible influences on character traits, which is education. The character traits considered in this paper are patience, also known as time preference, and risk willingness. Until now, few causal analyses have been done to check the validity of the relationships between education and time preference, and between education and risk willingness. Both risk and time preference are of importance in economic analysis such as inter-temporal contexts (Sun and Li, 2010; Kelleher, 2017). The relationships between either of them and other variables have been explored in a number of papers such as Komlos et al. (2004); Anderson and Mellor (2008); Dave and Saffer (2008); Van der Pol (2011) that notwithstanding, it is important to know what affects them as well. Education, indirectly, may be used to shape an individual's character traits through its organisational structure. This may be attributed to the rules and regulations which one has to follow in an academic setting, what is being taught and the attributes of teachers, who indirectly are role models. In relation to this, Berkowitz and Bier (2005) state that, either intentionally or unintentionally, teachers shape the formation of character in students - simply by association - through positive or negative example. Subjects, such as Mathematics, Science, History and language skills, taught in school expose pupils to a lot of information. This may influence certain character traits such as patience and risk willingness.

Patience induces people to invest in modern human capital, but once in school, schooling should make people even more patient (Reyes-Garcia et al., 2007). This is because schooling focuses student's attention on the future and through repeated practice at problem-solving, schooling helps children learn the art of scenario simulation (Becker and Mulligan, 1997). Viscusi and Moore (1989) in their empirical analysis of the wage effects of job risks also find that the rate of time preference (impatience) decreases with education. It is worth noting that there are conflicting theories to the stability of time preference over time. Classic economic theory assumes time preference is a stable parameter, behavioural economists model self control as an exhaustible resource and there is the new psychology view that using self-control may have heterogeneous effects on patience in later decisions (Hoel et al., 2016; Ozdenoren et al., 2012).

Perez-Arce (2017), the first to empirically establish a causal impact of education on time preferences, finds evidence of a causal effect of education on time preferences. He uses the admission lottery implemented by a public college in Mexico City as a natural experiment. Individuals successful in the lottery for a particular academic year can enrol for that academic year while unsuccessful individuals have to wait for the next academic year to enrol. Those who had the delayed admission were more likely to work during the waiting period. Telephone interviews were conducted to ask two sets of questions which were used to measure the degree of patience of the applicants. The questions were about preferences for rewards now or in the future with one set based on monetary rewards and the other based on a trip as the reward. He finds that those who were immediately admitted preferred to give more patient answers to the time preference questions based on trip rewards rather than the questions based on monetary rewards. He notes that monetary rewards do not provide a good measure of patience possibly because respondents are very likely to think of liquidity issues when answering such questions. He further states that if those with immediate admission are more patient than those with delayed admission, it is uncertain if this is because school made the former more patient or work made the latter less patient. Altogether, the literature indicates a positive relationship between education and time preference i.e. education increases patience.

Education would affect productivity not only directly, but also indirectly, by increasing individuals' willingness to take risks (Shaw, 1996). Shaw (1996) states that the correlation between education and risk taking suggests that one characteristic of risk taking is the ability to use information effectively. Education may also enlarge individuals' exposure to the various investment options available to them; one might thus expect more educated individuals to be less risk averse in their asset allocation decisions (Riley Jr and Chow, 1992). Knight et al. (2003) in their paper on education and risk-taking in agriculture in Ethiopia state that, education may increase achievement-orientation and facilitate openness to new ideas and modern practices. This, therefore, increases productivity and income. Grable (2000) also analyses education and financial risk tolerance and finds that respondents with higher attained education were more risk tolerant. He also finds in a second stage of his analysis that a combination of education, financial knowledge, income and occupation explained most of the between-group variability in risk tolerance. Financial knowledge, income and occupation are all in one way or another influenced by education which indicates a direct and indirect relationship between education and financial risk tolerance. Although most of the literature on risk preference and education indicate that education is associated with reducing risk aversion, there are a few that show otherwise. For instance, Barsky et al. (1997) find a U-shaped relationship between years of schooling completed and risk tolerance. They find individuals with exactly twelve years of schooling to be the least risk tolerant.

Jung (2015) is the first to examine the causal effect of education on risk aversion using the British education reform of 1972, which changed the minimum school-leaving age from 15 to 16. She finds a negative correlation between risk aversion and education but a positive causal effect of education on risk aversion. The positive effect is stronger for those with lower education i.e at most a high school diploma since they were mostly affected by the reform. In analysing the heterogeneous effects of the reform across genders for only those with secondary education, she also finds a significant negative correlation and a positive causal effect for the female sample. Therefore, the negative correlations between education and risk aversion found in most literature may not necessarily indicate the direction of the causal effect of education on risk aversion. This shows that the impact of education on risk willingness is not one directional and may depend on the educational system of the observed. Therefore, in this study we focus on Germany which also had a compulsory schooling reform after World War II.

Using the German educational reform as an instrument, this paper aims at a causal analysis of the effect of education on patience and risk willingness. This reform has been used to analyse returns to education in Germany by Pischke and Von Wachter (2008), Kamhöfer and Schmitz (2016) and Cygan-Rehm (2018). It has also been used to analyse the relationship between health and education by Kemptner et al. (2011), the relationship between schooling and citizenship by Siedler (2007) and others. We use a two-stage least squares (2SLS) model as done in studies that used the reform as an instrument. As in previous literature, we find a negative effect of education on risk aversion due to the reform which is larger and significant for the early partakers of the reform. Contrary to the literature, we find a negative effect of education on time preference due to the reform. Our results are quite robust under different specifications.

This paper contributes to the causal analysis of the impact of education on time and risk preferences. It supports research done on returns to education incorporating risk willingness. And it also contributes to the insurance literature where in many papers level of education is used as a proxy for the level of risk aversion.

## 3.2 German Educational Reform Background

The German educational reform made changes to the number of compulsory schooling years by increasing it from eight (8) years to nine (9) years. This occurred from 1946 to 1969 in West Germany after World War II due to lack of labour market opportunities and apprenticeships for school leavers as well as to increase the school leaving age.[1] It was implemented in different years by the various states because in Germany each federal state is responsible for its educational system (see Table A.1 in the Appendix). Decisions and policies concerning the educational system are made at the state level. Pischke and Von Wachter (2008) also mention that although ninth grade students in all the states would typically have the same subjects as in previous years, the content of the curricula for the additional ninth grade differ somewhat between states with focus on either political education, general knowledge, basic skills or others.

Also, in 1966 – 1967, two short school years (SSY) were introduced where the start of the school year moved from Spring to Fall for all states in West Germany except Bavaria. Bavaria already had its start of the school year in Fall. The transition to a fall start of the school year was achieved in most states through two SSY with 24 weeks instead of the regular 37 weeks of instruction each (Pischke, 2007). Therefore, the SSY affected the duration of schooling of those who were affected by it.[2] Pischke (2007) finds that the SSY increased grade repetition and lower track choice. Although the SSY was implemented, the school curriculum did not change during this period and there is the possibility that reading, writing and maths were stressed more to the detriment of other subjects (Pischke, 2007). Hence, there is the probability of this affecting certain character traits as subjects that may influence these character traits may not have been emphasised within these two years.

## 3.3 Data

The German Socio-Economic Panel (SOEP), established in 1984, is a wide-ranging representative longitudinal study of private households which contains yearly information on around 30,000 respondents in nearly 15,000 households (Goebel et al., 2019). We

[1]see Pischke and Von Wachter (2008) for details.
[2]see Pischke (2007) for details.

use observations from the 2018 survey year of the SOEP data from 1984-2019 (SOEP v36eu). Varying sample sizes are used in the analysis with the samples consisting of only German nationals. We initially consider birth cohorts from the years 1930 to 1970. We then consider birth cohorts from the years 1945 to 1970. Similar to Cygan-Rehm (2018) but different from Pischke and Von Wachter (2008), cohorts born before 1945 are excluded to eliminate wartime distortions and temporary extensions of compulsory schooling before the actual implementation of the reform. This also eliminates any possible correlations between the reform and effects of the war that may influence character traits thereby fulfilling the exclusion restriction. We also consider cohorts close to the pivotal birth cohorts in our robustness check. The pivotal birth cohorts are the first birth cohorts affected by the reform. We consider −/+5 years, −/+7 years and −/+10 years around the pivotal birth cohorts. Looking at the cohorts close to the reform is likely to give clearer estimates compared to using a larger sample whose estimates may be influenced by confounding factors. This also aids in checking how robust the estimates from the larger sample are.

The SOEP has information on respondents self-rated character traits on a scale of 0 to 10, where 0 represents no willingness to exhibit the character trait and 10 the highest willingness. The character traits of interest here are patience and risk willingness. In the SOEP, the respondents are asked about their personal patience level and their personal willingness to take risks. For the personal patience level they are asked, "Are you generally an impatient person or someone who always shows great patience?". Possible choices are provided on a scale where 0 is "very impatient" and 10 is "very patient". For personal willingness to take risks they are asked, "Are you generally a person who is willing to take risks or do you try to avoid taking risks?". This has a "willingness to take risks" scale where 0 is "Non" and 10 is "Very". The measures of these two self-rated responses have been validated to be good measures of actual behaviour. Vischer et al. (2013) in a time preference experiment using a representative sub-sample of respondents to the 2006-wave of the SOEP, obtained findings which suggest that the measure of patience in the SOEP represents a meaningful proxy for time preferences by using the typical price list decision format. Dohmen et al. (2011) in a field experiment with a sample of 450 subjects drawn from the adult population in Germany using the same methodology as SOEP find that, responses in the general risk willingness question are a reliable predictor of actual risk behaviour. The subjects in this sample answered the same general risk willingness question asked in the SOEP and also made choices in a real-stakes lottery experiment. They also state that asking questions that include more specific contexts produces measures that are even stronger for that given measure.

The descriptive statistics in table 3.1 are based on the full sample made up of birth cohorts from 1930 to 1970. The mean of patience is about 6.2 for all observations. The average patience level is higher for those who were not affected by the compulsory schooling reform compared to those affected by the compulsory schooling reform. The mean of risk willingness is about 4.2. Those affected by the compulsory schooling reform have a higher mean of risk willingness than those who were not affected in all the samples. This indicates that on average those affected by the reform are less patient and less risk averse compared to those not affected by the reform.

Table 3.1: Descriptive Statistics.

| | All | No compulsory schooling | Compulsory schooling |
|---|---|---|---|
| Patience | 6.16 | 6.26 | 6.12 |
| | (2.345) | (2.241) | (2.392) |
| Risk willingness | 4.21 | 4.01 | 4.31 |
| | (2.364) | (2.375) | (2.354) |
| School years | 10.38 | 9.66 | 10.71 |
| | (1.774) | (1.914) | (1.596) |
| SSY | 0.17 | 0.05 | 0.23 |
| | (0.375) | (0.214) | (0.418) |
| Age | 62.03 | 73.78 | 56.51 |
| | (10.138) | (5.641) | (6.384) |
| Female | 0.52 | 0.51 | 0.53 |
| | (0.499) | (0.500) | (0.499) |
| Observations - Patience | 8,985 | 2,874 | 6,111 |
| Observations - Risk willingness | 8,982 | 2,872 | 6,110 |

*Note:* Own calculations based on SOEP. Standard deviations are in parentheses.

We generated a dummy variable for the compulsory schooling reform. This variable is assigned the value 1 if an individual is born in or after the birth year of the first cohorts affected by the reform, or assigned the value 0 if otherwise. The main analysis is based on the reform years from Leschinsky and Roeder (1980) but we also use the reform years from Pischke and Von Wachter (2008) in our robustness checks (see table A3.1 for the different reform years and initial birth cohorts affected by the reform). The number of years of schooling (school years) was generated based on the type of secondary school attended. Information on the type of secondary school an individual attended is also made available in the SOEP data. Individuals who attended Hauptschule (basic/lowest track) were allocated 8 or 9 years depending on whether they were affected by the compulsory schooling reform or not. Those who attended Realschule (intermediate track) were allocated 10 years, Fachhochschulreife (advanced technical certificate) 12 years and Gymnasium (academic track) 13 years. There is about a year difference in the average years of schooling of those who were not affected by the compulsory schooling reform and those affected by the compulsory schooling. We also generated a dummy variable for SSY; = 1 for those affected or = 0 otherwise. The possible impact of SSY on both character traits and years of schooling emphasises its necessity in the analysis as it deals with any omitted variable bias that could arise with its omission. The SSY mostly affected those who affected by the compulsory schooling reform.

We generate a gender dummy, female, which takes the value 1 if an individual is female or 0 otherwise. There is almost a gender balance in each sample. Those not affected by the compulsory schooling reform are on average older which is to be expected. The values of the covariates in table 3.1 are very similar for both the patience and risk willingness samples. (see table A3.2 in the Appendix for a detailed table.)

## 3.4 Empirical Strategy

In order to deal with issues such as reverse causality related to the endogenous variable, years of schooling, we used the educational reform which serves as a natural experiment. The reform changed the number of years of compulsory schooling for all independent of an individual's personal influences on years of schooling. Using the reform as an instrument for the years of schooling, we exogenously determine the number of years of schooling. By implementing the 2SLS model, we analyse the impact of education on time and risk preferences, by regressing each character trait on years of schooling which is instrumented by Germany's compulsory schooling reform.

The first-stage is estimated by

$$\text{Yed}_{is} = \text{X}_i'\sigma + \pi\text{Reform}_s + \gamma_s + \lambda_t + \mu_s T + \xi_{is} \tag{3.1}$$

where $i$ denotes an individual and $s$ the state of schooling. Yed is an individual's number of years of schooling. X represents the covariates used: SSY and gender. Reform is the dummy variable for the compulsory schooling reform of nine (9) years compulsory schooling instead of eight (8). $\pi$ captures the impact of the reform on the number of years of schooling. $\gamma_s$ denotes a set of state dummies which captures state effects as well as differences amongst the states. $\lambda_t$ denotes a set of year of birth dummies which captures year of birth effects that are common across states. $\mu_s T$ denotes state-specific time trends i.e. interactions of school state dummies with a linear trend in year of birth. As mentioned by Pischke and Von Wachter (2008), other state-specific trends may be correlated with the treatment hence the inclusion of state-specific time trends. This is indeed useful since the federal states implemented the reform at different times.

The reduced-form is estimated by

$$\text{Y}_{is} = \text{X}_i'\tau + \psi\text{Reform}_{st} + \gamma_s + \lambda_t + \mu_s T + \varepsilon_{is} \tag{3.2}$$

where Y is the character trait of interest. $\psi$ captures the impact the reform has on the character trait. The effect of the reform on years of schooling, $\pi$ in equation (3.1), and the effect of the reform on the character trait of interest, $\psi$ in equation (3.2), are estimated through a difference-in-differences approach since state and year of birth effects are controlled for.

The second-stage is then estimated using

$$Y_{is} = \alpha' X_i + \rho \widehat{\text{Yed}}_{is} + \gamma_s + \lambda_t + \mu_s T + \eta_{is} \tag{3.3}$$

where Y is the character trait of interest. $\widehat{\text{Yed}}$ is the estimate obtained from the first-stage which represents the exogenous years of education. $\rho$, which is the coefficient of interest in this analysis, is the causal effect. This provides the impact of education on the character traits.

In order for $\rho$ to be interpreted as a causal effect, certain assumptions about the instrument need to be fulfilled.[3] First of all, the instrument needs to be exogenous which has already been mentioned. Secondly, the instrument needs to have an effect on the endogenous variable that is, the reform should have a significant effect on the number of years of schooling. This is captured by $\pi$ in the first stage. Lastly, the instrument should only affect the dependent variable through its effect on the endogenous variable i.e. through the first stage. This is known as the exclusion restriction. It implies that, the reform should not have any direct effect on character traits but rather an indirect effect through its effect on the number of years of schooling. This also means the reform should not be correlated with unobservable characteristics that influence the character traits i.e. $cov(\text{Reform}_{st}, \eta_{ist}) = 0$. Since one would not expect any direct effect of the reform on character traits, this is also fulfilled. Given that all the assumptions are fulfilled, the effect estimated by the 2SLS captures the local average treatment effect (LATE). In this case, the 2SLS estimates provide information on the effect of schooling on those who had an additional year of schooling because of the reform but would not have had an additional year of schooling otherwise.

## 3.5 Empirical Results

Table 3.2 presents the regression results for patience (upper section) and risk willingness (lower section). The first-stage results in the second column shows that the compulsory schooling reform increased school years by about 0.6 years in both the patience and the risk willingness samples under the full sample when all birth cohorts from 1930 to 1970 are considered. This increases to about 0.7 years when the sample is restricted to those born from 1945 onwards. All the estimates are statistically significant at 1% significance level and the F-statistic is large enough (i.e. greater than 10) in each sample. This shows

[3] See Cygan-Rehm (2018) for in-depth discussion.

that the reform did have an impact on the years of schooling by increasing the years of schooling.

The reduced-form results in the third column show the impact the reform had on the character traits. The reform has a negative impact on patience. This is statistically significant when birth cohort before 1945 are omitted. The size of the effects also increases when we restrict the sample to those born from 1945 onwards. The results indicate that, an increase in compulsory schooling by a year decreased patience by 0.3 points (about 5% of the mean and 13% of the standard deviation of patience). There is a positive impact of the reform on risk willingness which indicates that an additional compulsory schooling year increases risk willingness by 0.17 points (about 4% of the mean and 7% of the standard deviation of risk willingness), but this is statistically insignificant.

The OLS results in the first column show a significant negative association between years of schooling and patience and a significant positive association between years of schooling and risk willingness. The size of the estimates is quite similar regardless of whether the sample is restricted or not. Although statistically significant, the estimates are quite small, accounting for less than 5% of their respective standard deviations, which suggests that there is not much impact of an additional year of schooling on these character traits. The OLS results, however, do not solve the endogeneity problem but are used as a benchmark analysis. Hence, the results do not represent causal effects. For the causal effects, the 2SLS results are considered.

The last column presents the second-stage results of the 2SLS. The size of the effects is larger than the OLS estimates in all sample except for the full sample under risk willingness. The effects from the restricted sample are larger than those from the full sample for both patience and risk willingness. This may be due to bias from the unaccounted effects of the war in the full sample. This bias is stronger for risk willingness which indicates that there could be a strong correlation between wartime effects and risk willingness. The results indicate that, an additional year of schooling as a result of the reform decreases patience level by 0.43 points (about 7% of the mean and 19% of the standard deviation of patience). The effect of schooling on patience is statistically significant for birth cohorts born from 1945 but statistically significant at only 10% significance level for the full sample. This negative effect of education on patience is contrary to the literature which suggests that education should positively influence patience. On the other hand, an additional year of schooling as a result of the reform increases risk willingness level by 0.24 points (about 6% of the mean and 10% of the standard deviation of risk willingness). The size of the effect is close to that of Jung (2015) but in different directions. Our results indicate a decrease in risk aversion as a result of education while that of Jung (2015) indicate an increase in risk aversion as a result of education. This goes to show that there is a possibility the impact of education

Table 3.2: Regression results - main results.

| **Patience** | | | | |
|---|---|---|---|---|
| | OLS | First-stage | Reduced-form | Second-stage |
| *Dependent variable:* | Patience | School years | Patience | Patience |
| ***All birth cohorts*** | | | | |
| School years | $-0.053^{***}$ | – | – | $-0.365^{*}$ |
| | (0.015) | | | (0.207) |
| Reform | – | $0.592^{***}$ | $-0.216^{*}$ | – |
| | | (0.105) | (0.114) | |
| First-stage F-statistic | – | 31.54 | – | – |
| Observations | 8,985 | 8,985 | 8,985 | 8,985 |
| ***Birth cohorts from 1945 onwards*** | | | | |
| School years | $-0.057^{***}$ | – | – | $-0.434^{**}$ |
| | (0.017) | | | (0.209) |
| Reform | – | $0.686^{***}$ | $-0.298^{**}$ | – |
| | | (0.119) | (0.142) | |
| First-stage F-statistic | – | 33.31 | – | – |
| Observations | 7,439 | 7,439 | 7,439 | 7,439 |

| **Risk willingness** | | | | |
|---|---|---|---|---|
| | OLS | First-stage | Reduced-form | Second-stage |
| *Dependent variable:* | Risk willingness | School years | Risk willingness | Risk willingness |
| ***All birth cohorts*** | | | | |
| School years | $0.096^{***}$ | – | – | 0.086 |
| | (0.015) | | | (0.259) |
| Reform | – | $0.591^{***}$ | 0.051 | |
| | | (0.106) | (0.154) | |
| First-stage F-statistic | – | 31.29 | – | – |
| Observations | 8,982 | 8,982 | 8,982 | 8,982 |
| ***Birth cohorts from 1945 onwards*** | | | | |
| School years | $0.081^{***}$ | – | – | 0.240 |
| | (0.017) | | | (0.251) |
| Reform | – | $0.685^{***}$ | 0.165 | – |
| | | (0.120) | (0.178) | |
| First-stage F-statistic | – | 32.63 | – | – |
| Observations | 7,437 | 7,437 | 7,437 | 7,437 |

*Note:* Own calculations based on SOEP. All regressions include a dummy for female, indicators for SSY, year of birth and school state, and state-specific linear trends. Robust standard errors clustered at state$\times$ year of birth in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

on risk willingness (or risk aversion) depends on the educational system since these results are from two different countries with different educational systems. The effects are however statistically insignificant for risk willingness.

Existing literature gives reasoning to why education should increase an individual's patience level but our results show otherwise. Becker and Mulligan (1997) explain that, patience level is not constant but rather it adjusts according to the propinquity of future pleasures. Relating this to schooling, the negative change can be attributed to income generation and delays in personal life goals, such as having a family. Although more years of schooling comes with higher income, there may be some delay in capital accumulation. Less or no income may be accumulated during the years of schooling compared to those in full employment. With the exception of Cygan-Rehm (2018), literature on returns to schooling in Germany shows zero returns to schooling. This means that, an additional year of schooling does not necessarily imply higher income in Germany. According to Pischke and Von Wachter (2008), a possible explanation for the lack of wage returns might be that the wage-setting institutions in Germany prevent the adjustments necessary to reflect any returns [4]. Also, there is quite a difference in wages between the public and private sectors of Germany. This may influence an individual with more years of education thereby being impatient and risk willing to opt for employment in the private sector. The private sector may be considered a "risky" option given that equal opportunity and anti-discrimination policies are implemented more effectively in the public sector and job retention, especially for women after maternity break, is easier in the public sector (Melly, 2005).

## 3.6 Robustness Checks

### 3.6.1 Discontinuity samples

We consider three discontinuity samples, −/+5 years, −/+7 years and −/+10 years around the pivotal birth cohorts. We also check whether the estimates are affected by the war by restricting the samples to those born from 1945 onwards. The results for the restricted samples are provided in the Appendix. Since the reform years from Leschinsky and Roeder (1980) are used, restricting the sample automatically excludes all cohorts who attended school in Hamburg and Schleswig-Holstein. From the first-stage in table 3.3, the effect of the reform on years of schooling ranges from 0.8 to 1 year increase. Those born 5 years around the pivotal birth years have a year increase in schooling as a result of the reform. This reduces as the years around the pivotal birth years increase. This implies that those who were immediately affected by the reform experienced a full year increase while this diminishes for those further away from the reform. This increase is relatively larger than in the full sample. Unlike the full sample, the discontinuity sample looks at a smaller window around the reform years in order to capture the immediate effect of the reform. Indeed, one would expect less than a year increase because of the SSY. However, this was only implemented simultaneously with the reform by only four out of ten West Germany states.

The reform has a negative impact on patience across all samples except for those born 5 years around the pivotal years. However, the impact of the reform is negative across

[4] see Kuhlmann and Röber (2006) for details on wage-setting in Germany

Table 3.3: Regression results - Discontinuity samples.

| Patience | | | | | | |
|---|---|---|---|---|---|---|
| *Sample:* | −/+5 years | | −/+7 years | | −/+10 years | |
| | First-stage | Reduced-form | First-stage | Reduced-form | First-stage | Reduced-form |
| Reform | 1.076*** | 0.120 | 0.812*** | −0.064 | 0.760*** | −0.228 |
| | (0.108) | (0.173) | (0.148) | (0.149) | (0.124) | (0.138) |
| F-statistic | 99.33 | – | 30.16 | – | 37.47 | – |
| Observations | 2,443 | 2,443 | 3,347 | 3,347 | 4,768 | 4,768 |

| Risk willingness | | | | | | |
|---|---|---|---|---|---|---|
| | -/+5 years | | -/+7 years | | -/+10 years | |
| | First-stage | Reduced-form | First-stage | Reduced-form | First-stage | Reduced-form |
| Reform | 1.076*** | 0.448** | 0.807*** | 0.113 | 0.759*** | 0.146 |
| | (0.108) | (0.214) | (0.149) | (0.220) | (0.125) | (0.174) |
| F-statistic | 99.33 | – | 29.27 | – | 36.98 | – |
| Observations | 2,443 | 2,443 | 3,346 | 3,346 | 4,764 | 4,764 |

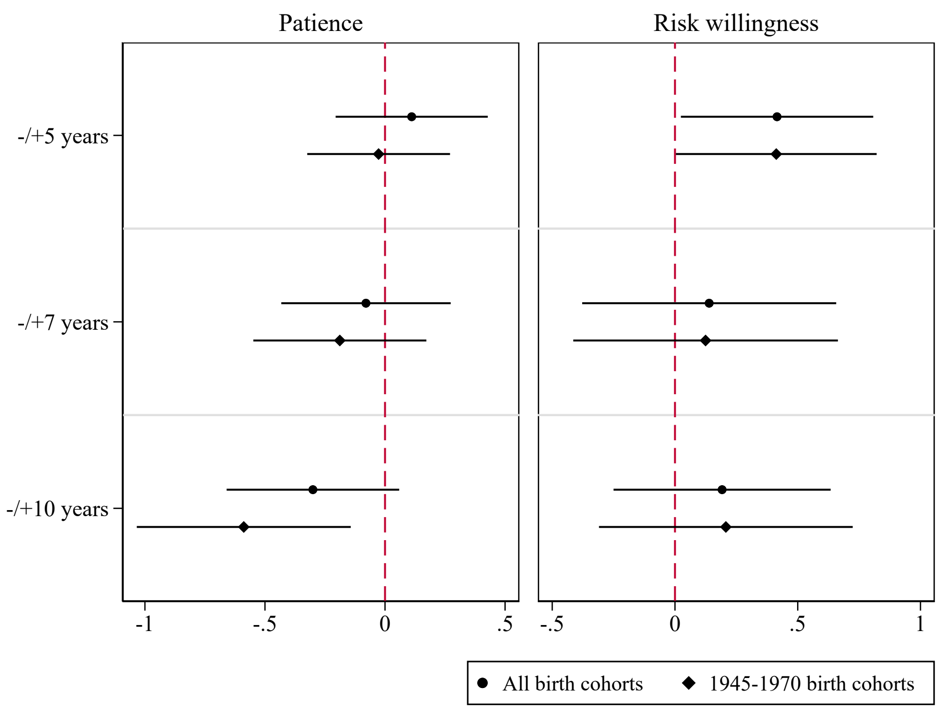
*Note:* Own calculations based on SOEP. All regressions include a dummy for female, indicators for SSY, year of birth and school state, and state-specific linear trends. Robust standard errors clustered at state× year of birth in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

all samples when the samples are restricted to those born from 1945 onwards. The effect of the reform on patience is largest in the −/+10 years sample which is about 4% of the mean (10% of the standard deviation) of patience. These effects are statistically insignificant. The reduced-form effect of patience in the restricted −/+10 years sample is however larger, about 7% of the mean (18% of the standard deviation) of patience, and statistically significant. There is a positive impact of the reform on risk willingness in all samples. The largest effect found in the −/+5 years sample is about 11% of the mean (19% of the standard deviation) of risk willingness and statistically significant. This effect is slightly larger in the restricted −/+5 years sample.

Figure 3.1 depicts the second-stage results of the impact of schooling on patience and risk willingness based on the discontinuity samples, both full and restricted samples. The effect of schooling on patience increases as the years around the pivotal years increase. The effects in the −/+5 years and −/+7 years samples are, however, close to zero. There is a larger effect in the −/+10 years sample. This effect is statistically insignificant in the full sample but statistically significant in the restricted sample. The effect in the restricted −/+10 years sample is larger, about 9% of the mean (25% of the standard deviation) of patience, than the effect found in table 3.3. For risk willingness, the effect is largest in the −/+5 years sample and it is statistically significant. The effects in the full and restricted samples are very similar for risk willingness. The bias in the effects for risk willingness disappear in the discontinuity samples. This could imply that the bias in the effects for risk willingness found in table 3.3 may not be as result of unaccounted war effects but probably due to some measurement error.

With the exception of −/+5 years sample for patience, the direction of the results for the discontinuity samples are consistent with the main results in table 3.3. The direction of our results for the discontinuity samples are similar regardless of whether birth cohorts who were affected by the war are included or not. The differences in the size, and direction in the case of patience, of the effects across the different discontinuity samples may be attributed to the SSY. The birth cohorts very close to the pivotal years in the states, North Rhine-Westphalia, Hesse, Rhineland Palatinate and Baden-Wüttemberg, were affected by the SSY. Therefore, the pivotal birth cohorts and the birth cohorts after them, those closest to the pivotal years, in these states experienced the introduction of the compulsory schooling reform as well as compressed school years simultaneously. This could explain the difference in the results especially, between the $-/+5$ discontinuity sample and the other two discontinuity samples. The birth cohorts farther from the pivotal years, also from these same states, may or may not have experienced SSY. However, those who did probably had time to adjust to the increase in years of schooling. For the other states, some of the birth cohorts farther from the pivotal years were affected by the SSY. The other states that were affected by the SSY had already implemented the compulsory schooling reform years before the introduction of the SSY. The cohorts in those states therefore, did not have to adjust to the reform and SSY at the same time. It can therefore be said that those who experienced both the compulsory schooling reform and the compressed school years are more risk willing (or less risk averse) and more patient.

Figure 3.1: 2SLS results - Discontinuity samples




*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\rho$ based on Eq. (3.3) with 95 % confidence intervals. All regressions include a dummy for female, indicators for SSY, year of birth and school state, and state-specific linear trends. Robust standard errors clustered at state$\times$ year of birth.

### 3.6.2 Various Specifications

We also investigate how robust our results from the full sample are under various different specifications. We however consider only those born from 1945 onwards to eliminate the possible bias from the effects of the war. Results based on all birth cohorts can be found in the Appendix. We re-estimate our model without state-specific trends and with quadratic state-specific trends. Adding quadratic trends help to account for slow-moving trends in each state prior to the reform however, adding quadratic state-specific trends can unintentionally pick up reform induced dynamics and bias the results (Lundborg et al., 2014; Cygan-Rehm, 2018). They may, as a result, partly control for the effect being estimated (Lundborg et al., 2014). In all three specifications, we find a negative effect of schooling on patience regardless of the sample. When trends are not included the effect is zero. We also use years of education provided in the SOEP instead of the computed years of schooling for the estimations. The SOEP provides information on the total number of years of education which is computed based on respondents' educational history which includes secondary vocational education. However, SOEP does not consider the educational reform in the computation hence, we adjust the years of education based on the state of schooling and the reform years. A year is deducted from the years of education of those in the basic track who were not affected by the reform. We also exclude the pivotal birth cohorts and estimate the original models as the second specification. In a regression discontinuity (RD) setting, observations at the thresholds are likely to bias estimates as a result of heaping and manipulation of the running variable as well as from stacking different thresholds (Barreca et al., 2016; Fort et al., 2016). Although we do not use the traditional RD setting, we stack different thresholds as result of the different implementation years of the reform in the different states. We are likely to also face this problem hence the check. Finally, we use the reform years from Pischke and Von Wachter (2008) for our estimations.

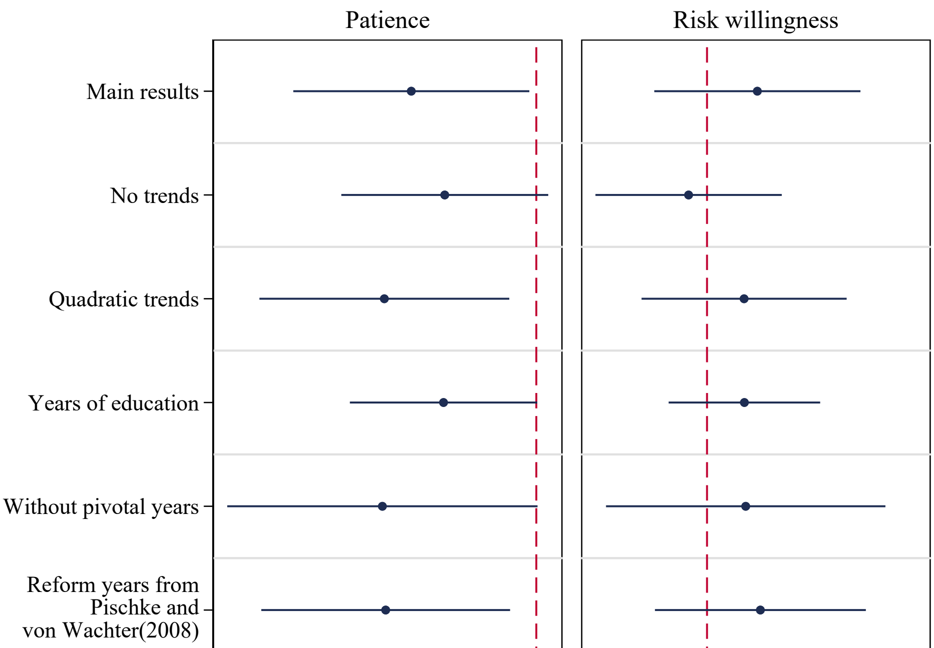
Table 3.4 provides the first stage and reduced-form results for the various specifications. With the exception of the specification with years of education, the others have years of schooling as our variable of interest. The first stage results show that the reform increased years of schooling by 0.62-0.73 years and years of education by almost a year. The reduced-form results for the years of education sample are the same as that for the years of schooling in the restricted sample in table 3.3. The reduced-form results show a consistent negative impact of the reform on patience. The size of the effects in the other specifications, except the no trends specification, is slightly larger than that of the main results. These results are statistically significant across all specifications except the specification without trends which is only statistically significant at 10% significance level. The positive impact of the reform on risk willingness is consistent across all specifications except the no trends specification. The results are not statistically significant.

Table 3.4: Regression results based on various specifications.

| *Sample:* | Patience | | | | Risk willingness | | | |
|---|---|---|---|---|---|---|---|---|
| | First-stage | F-statistic | Reduced -form | Obs. | First-stage | F-statistic | Reduced -form | Obs. |
| No trends | 0.654*** | 54.38 | −0.208* | 7,439 | 0.655*** | 54.42 | −0.058 | 7,437 |
| | (0.089) | | (0.117) | | (0.089) | | (0.148) | |
| With quadratic trends | 0.726*** | 32.11 | −0.383** | 7,439 | 0.723*** | 31.34 | 0.128 | 7,437 |
| | (0.128) | | (0.151) | | (0.129) | | (0.186) | |
| Years of education | 0.925*** | 17.95 | −0.298** | 7,439 | 0.922*** | 17.52 | 0.165 | 7,437 |
| | (0.218) | | (0.142) | | (0.220) | | (0.178) | |
| Without pivotal years | 0.625*** | 23.12 | −0.335** | 7,214 | 0.623*** | 22.63 | 0.115 | 7,212 |
| | (0.130) | | (0.161) | | (0.131) | | (0.220) | |
| Reform years based on Pischke and Von Wachter (2008) | 0.667*** | 31.70 | −0.349** | 7,439 | 0.665*** | 31.04 | 0.170 | 7,437 |
| | (0.118) | | (0.142) | | (0.119) | | (0.177) | |

*Note:* Own calculations based on SOEP. The first-stage and reduced-form columns provide the coefficients of the compulsory schooling reform for the respective models per the specification on each row. All regressions include a dummy for female, indicators for short SSY, year of birth and school state, and state-specific linear trends unless specified otherwise. Robust standard errors clustered at state× year of birth in parentheses. Only birth cohorts from 1945 onwards are considered. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.2: Regression results based on various specifications




*Note:* Own calculations based on SOEP. All regressions include a dummy for female, indicators for SSY, year of birth and school state. The 2SLS regressions include state-specific trends as specified in parentheses. Robust standard errors clustered at state× year of birth are used for the confidence intervals. Only birth cohorts from 1945 onwards are considered.

In figure 3.2, we compare our original 2SLS results with the various 2SLS results. The negative effect of a year increase in schooling on patience is robust across all specifications but the size of the effect varies. The size of the effect for the quadratic trends, without pivotal years and reform years based on Pischke and Von Wachter (2008) specifications are very similar and slightly larger than the main results. The positive effect on risk willingness is also quite robust across all specifications except the specification without trends. The size of the effects is not too different from the main results.

Years of schooling and years of education yield quite similar results for risk willingness than for patience. Factoring in additional years of education beyond high school seems to influence patience as well. Therefore, using either years of schooling or the adjusted years of education provided in the SOEP does not necessarily yield similar results. With patience, the size of the effects without pivotal birth cohorts are a bit larger. This shows some sort of positive bias from the pivotal birth cohorts on the effect of schooling on patience in the main results. Using different sets of reform years changes the dynamics of sample. For instance, the proportion of those affected by the reform varies. Some observations that are in the treatment group in one sample will be in the control group of the other and vice versa. Both sets of reform years yield similar results in the direction and size of the estimates.

Our results seem to be robust under the different specifications especially with regard to direction and size. The results for risk willingness are much more robust under all specifications.

## 3.7 Conclusion

This study investigated whether education has an impact on the character traits, patience (time preference) and risk willingness. We use the SOEP data in the analysis with the increasing years of compulsory schooling due to the German educational reform from eight (8) years to nine (9) years in West Germany after World War II as the instrument. A number of research has been done on the correlation between education and risk with just a few on the causal effect of education on risk. Not much empirical research has been done on the relationship between education and patience. Finding the relationship between education and these character traits raises the problem of endogeneity due to the fact that education is a choice variable, and also the problem of reverse causality. In using the compulsory schooling reform, we exogenously determine the number of years of schooling which help resolve these potential problems. In order to obtain causal effects, we use 2SLS for our analysis.

We find significant positive correlation between education and risk willingness (negative correlation between education and risk aversion) which is in line with the literature. We also find that an additional year of schooling as a result of the reform has a positive effect on risk willingness. This effect is larger and statistically significant for the early partakers of the reform. This negative causal effect of education on risk aversion is contrary to the positive causal effect found by Jung (2015). These results are based on two different countries with different educational systems. This goes to show that different educational systems may affect risk aversion differently.

Contrary to the literature, we find significant negative correlation between education and patience. We also find a negative causal effect of an additional year of schooling as a result of the reform. The effect is small and close to zero for the early partakers but increases when we consider those further. Our results are precise when we exclude birth cohorts who were affected by the war. The negative effect on patience could be due to delays in certain aspects of one's life such as capital accumulation as a result of the years spent in school.

The difference in the results of the sample with those closest to the pivotal years compared to the other samples can be associated with the SSY. Quite a large proportion of those closest to the pivotal years who were affected by the reform were simultaneously affected by the SSY. Those farther from the pivotal years who may have been affected by the SSY did not have to deal with a new reform as well. Being affected by these two changes at the same time may be the reason for the different sizes of the effects for those closest to the pivotal years.

Our results are quite robust under different specifications; excluding state-specific trends, including quadratic state-specific trends, using years of education instead of years of schooling, excluding pivotal birth cohorts and using a different set of reform years. The results are much more robust for risk willingness.

A contributing factor of this paper is that, it provides a causal analysis of the effect of education on patience and risk willingness. It also confirms and contradicts what is hypothesised in the literature. It provides one more factor to aid in understanding the differences in individual risk preferences. It serves as a backing for research done on returns to education that incorporate risk willingness. It also contributes to the insurance literature which uses level of education as a proxy for the level of risk aversion in most papers. This will equip them with more information so as to better their use of education as proxy for risk aversion. Although years of schooling is supposed to make an individual more patience, as suggested in the literature, this does not hold in this case. This shows that the expected positive effect of education on patience may or may not hold for every country. Factors such as educational system and the wage structure of a country may change this positive effect of education on patience and need to be considered in the building of models that sought to relate education with patience (time preference).

# Appendix

Table A3.1: The year of final introduction of the compulsory schooling reform from Leschinsky and Roeder (1980) and Pischke and Von Wachter (2008).

| | **Leschinsky and Roeder (1980)** | | **Pischke and Von Wachter (2008)** | |
|---|---|---|---|---|
| Federal State | Reform Year | Birth Cohort | Reform Year | Birth Cohort |
| Hamburg | 1946 | 1932 | 1949 | 1935 |
| Schleswig-Holstein | 1947 | 1933 | 1956 | 1942 |
| Saarland | 1958 | 1944 | 1964 | 1950 |
| Bremen | 1959 | 1945 | 1958 | 1944 |
| Lower Saxony | 1962 | 1948 | 1962 | 1948 |
| North Rhine-Westphalia | 1967 | 1953 | 1967 | 1953 |
| Hesse | 1967 | 1953 | 1967 | 1953 |
| Rhineland Palatinate | 1967 | 1953 | 1967 | 1953 |
| Baden-Württemberg | 1967 | 1953 | 1967 | 1953 |
| Bavaria | 1969 | 1955 | 1969 | 1955 |

*Source:* Cygan-Rehm (2018)

Table A3.2: Detailed Descriptive Statistics.

| Sample | **Patience** | | | **Risk Willingness** | | |
|---|---|---|---|---|---|---|
| | All | No compulsory schooling | Compulsory schooling | All | No compulsory schooling | Compulsory schooling |
| ***Full sample*** | | | | | | |
| Patience | 6.16 | 6.26 | 6.12 | - | - | - |
| | (2.345) | (2.241) | (2.392) | | | |
| Risk willingness | - | - | - | 4.21 | 4.01 | 4.31 |
| | | | | (2.364) | (2.375) | (2.354) |
| School years | 10.38 | 9.66 | 10.71 | 10.38 | 9.66 | 10.71 |
| | (1.774) | (1.914) | (1.596) | (1.774) | (1.914) | (1.596) |
| SSY | 0.17 | 0.05 | 0.23 | 0.17 | 0.05 | 0.23 |
| | (0.375) | (0.214) | (0.418) | (0.375) | (0.214) | (0.418) |
| Age | 62.03 | 73.78 | 56.51 | 62.03 | 73.78 | 56.51 |
| | (10.138) | (5.641) | (6.384) | (10.137) | (5.641) | (6.385) |
| Female | 0.52 | 0.51 | 0.53 | 0.53 | 0.51 | 0.53 |
| | (0.499) | (0.500) | (0.499) | (0.499) | (0.500) | (0.499) |
| Observations | 8,985 | 2,874 | 6,111 | 8,982 | 2,872 | 6,110 |
| ***-/+5 years*** | | | | | | |
| Patience | 6.23 | 6.34 | 6.15 | - | - | - |
| | (2.371) | (2.295) | (2.425) | | | |
| Risk willingness | - | - | - | 4.22 | 4.17 | 4.25 |
| | | | | (2.364) | (2.379) | (2.353) |
| School years | 10.18 | 9.75 | 10.51 | 10.18 | 9.75 | 10.51 |
| | (1.767) | (1.888) | (1.598) | (1.767) | (1.888) | (1.598) |
| SSY | 0.413 | 0.133 | 0.621 | 0.413 | 0.133 | 0.621 |

Table continues on the next page

| Continuation of Table A3.2 | | | | | | |
|---|---|---|---|---|---|---|
| Sample | **Patience** | | | **Risk Willingness** | | |
| | All | No compulsory schooling | Compulsory schooling | All | No compulsory schooling | Compulsory schooling |
| | (0.492) | (0.339) | (0.485) | (0.492) | (0.339) | (0.485) |
| Female | 0.530 | 0.534 | 0.527 | 0.530 | 0.534 | 0.527 |
| | (0.499) | (0.499) | (0.499) | (0.499) | (0.499) | (0.499) |
| Age | 65.41 | 68.31 | 63.26 | 65.41 | 68.31 | 63.26 |
| | (4.438) | (2.933) | (4.131) | (4.438) | (2.933) | (4.131) |
| Observations | 2,443 | 1,041 | 1,402 | 2,443 | 1,041 | 1,402 |
| ***-/+7 years*** | | | | | | |
| Patience | 6.23 | 6.30 | 6.18 | - | - | - |
| | (2.365) | (2.304) | (2.406) | | | |
| Risk willingness | - | - | - | 4.22 | 4.13 | 4.29 |
| | | | | (2.368) | (2.370) | (2.365) |
| School years | 10.22 | 9.74 | 10.56 | 10.22 | 9.74 | 10.56 |
| | (1.772) | (1.896) | (1.598) | (1.772) | (1.895) | (1.598) |
| SSY | 0.409 | 0.100 | 0.625 | 0.409 | 0.100 | 0.625 |
| | (0.492) | (0.300) | (0.484) | (0.492) | (0.301) | (0.484) |
| Female | 0.517 | 0.517 | 0.516 | 0.517 | 0.517 | 0.516 |
| | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) |
| Age | 65.06 | 69.20 | 62.17 | 65.06 | 69.20 | 62.17 |
| | (5.300) | (3.289) | (4.448) | (5.299) | (3.290) | (4.448) |
| Observations | 3,347 | 1,376 | 1,971 | 3,346 | 1,375 | 1,971 |
| ***-/+10 years*** | | | | | | |
| Patience | 6.23 | 6.26 | 6.21 | - | - | - |
| | (2.344) | (2.287) | (2.379) | | | |
| Risk willingness | - | - | - | 4.23 | 4.14 | 4.29 |
| | | | | (2.363) | (2.355) | (2.367) |
| School years | 10.27 | 9.77 | 10.58 | 10.27 | 9.77 | 10.58 |
| | (1.767) | (1.915) | (1.593) | (1.767) | (1.914) | (1.593) |
| SSY | 0.291 | 0.076 | 0.423 | 0.292 | 0.076 | 0.423 |
| | (0.454) | (0.265) | (0.494) | (0.455) | (0.265) | (0.494) |
| Female | 0.517 | 0.510 | 0.522 | 0.517 | 0.510 | 0.522 |
| | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) | (0.500) |
| Age | 64.25 | 70.40 | 60.49 | 64.25 | 70.39 | 60.49 |
| | (6.714) | (3.733) | (5.186) | (6.710) | (3.728) | (5.186) |
| Observations | 4,768 | 1,812 | 2,956 | 4,764 | 1,810 | 2,954 |

End of Table

Note: Own calculations based on SOEP. Standard deviations are in parentheses.

Table A3.3: Regression results - Discontinuity samples from 1945 onwards.

| | **Patience** | | | | | |
|---|---|---|---|---|---|---|
| *Sample:* | −/+5 years | | −/+7 years | | −/+10 years | |
| | First-stage | Reduced-form | First-stage | Reduced-form | First-stage | Reduced-form |
| Reform | 1.12*** | −0.030 | 0.797*** | −0.150 | 0.715*** | −0.421*** |
| | (0.096) | (0.173) | (0.144) | (0.145) | (0.137) | (0.149) |
| F-statistic | 136.63 | – | 30.74 | – | 27.34 | – |
| Observations | 2,337 | 2,337 | 3,170 | 3,170 | 4,298 | 4,298 |
| | **Risk willingness** | | | | | |
| | -/+5 years | | -/+7 years | | -/+10 years | |
| | First-stage | Reduced-form | First-stage | Reduced-form | First-stage | Reduced-form |
| Reform | 1.12*** | 0.462** | 0.792*** | 0.099 | 0.712*** | 0.148 |
| | (0.096) | (0.227) | (0.145) | (0.223) | (0.138) | (0.193) |
| F-statistic | 136.63 | – | 29.75 | – | 26.54 | – |
| Observations | 2,337 | 2,337 | 3,169 | 3,169 | 4,295 | 4,295 |

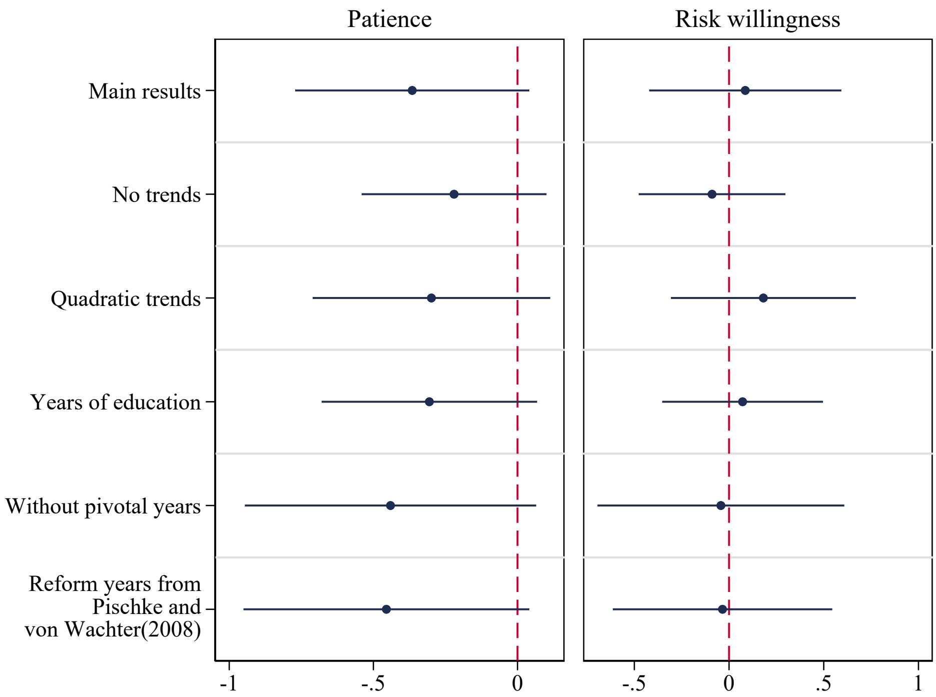
*Note:* Own calculations based on SOEP. All regressions include a dummy for female, indicators for SSY, year of birth and school state, and state-specific linear trends. Robust standard errors clustered at state$\times$ year of birth in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3.4: Regression results - Robustness checks for all birth cohorts (1930 - 1970).

| *Sample:* | **Patience** | | | | **Risk willingness** | | | |
|---|---|---|---|---|---|---|---|---|
| | First-stage | F-statistic | Reduced -form | Obs. | First-stage | F-statistic | Reduced -form | Obs. |
| No trends | 0.671*** | 65.11 | −0.148 | 8,985 | 0.671*** | 65.06 | −0.060 | 8,982 |
| | (0.083) | | (0.108) | | (0.083) | | (0.132) | |
| With quadratic trends | 0.618*** | 30.04 | −0.185 | 8,985 | 0.617*** | 29.82 | 0.112 | 8,982 |
| | (0.113) | | (0.129) | | (0.113) | | (0.157) | |
| Years of education | 0.707*** | 14.52 | −0.216* | 8,985 | 0.706*** | 14.44 | 0.051 | 8,982 |
| | (0.185) | | (0.114) | | (0.185) | | (0.154) | |
| Without pivotal years | 0.556*** | 23.81 | −0.245* | 8,758 | 0.555*** | 23.63 | −0.024 | 8,755 |
| | (0.114) | | (0.127) | | (0.114) | | (0.184) | |
| Reform years based on Pischke and Von Wachter (2008) | 0.498*** | 22.94 | −0.227** | 8,985 | 0.498*** | 22.80 | −0.017 | 8,982 |
| | (0.104) | | (0.115) | | (0.104) | | (0.147) | |

*Note:* Own calculations based on SOEP. The first-stage and reduced-form columns provide the coefficients of the compulsory schooling reform for the respective models per the specification on each row. All regressions include a dummy for female, indicators for short SSY, year of birth and school state, and state-specific linear trends unless specified otherwise. Robust standard errors clustered at state$\times$ year of birth in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A3.1: 2SLS results - Robustness checks for all birth cohorts (1930 - 1970)




*Note:* Own calculations based on SOEP. All regressions include a dummy for female, indicators for SSY, year of birth and school state, and state-specific linear trends unless specified otherwise. Robust standard errors clustered at state× year of birth are used for the confidence intervals.

# Chapter 4

# Life-cycle Health Effects of Compulsory Schooling*

## 4.1 Introduction

Estimating the effects of education on socio-economic outcomes has been an important part of applied microeconometrics in the past three decades. While most of the literature has focused on labor market outcomes, effects on health have been studied as well. In Table 4.1 we list 22 studies that estimate health effects of education and use methods of instrumental variable estimation for identification. More than half of these studies do not find statistically significant effects overall or in relevant subgroups. All these mentioned studies have in common that they aggregate effects over age groups, often over several decades. Yet, this may miss relevant patterns. Kaestner et al. (2020) extend the classic Grossman (1972) model of demand for health and conclude that "it is unlikely that the relationship between education and health will be constant over the life cycle and that education is likely to have little effect on health at younger ages when there is little depreciation of the health stock" (Kaestner et al., 2020). Thus, an estimated small and insignificant effect averaged over younger and older individuals does not necessarily imply that health is not causally affected by education. It may well be that the effect sets in late in life which is blurred, however, by a zero effect for younger individuals.

It is well known that the socio-economic status-health gradient increases over the life-cycle (e.g., Case and Deaton, 2005, Galama and van Kippersluis, 2019). This descriptive pattern has also been shown more specifically for the education-health gradient. As an example, Kaestner et al. (2020) find no differences in mortality by education until

*This chapter is joint work with Hendrik Schmitz and is published as a working paper: Schmitz, H. and Tawiah B. B. (2023). Life-cycle Health Effects of Compulsory Schooling. Ruhr Economic Papers, 1006, RWI.

the age of 60, but afterwards hazard rates diverge by education. In contrast, they find an education-morbidity gradient only for the age group 45-60 but explain this with possible selective mortality. Bijwaard et al. (2015) find an increasing difference in mortality between those with primary education and those with more than primary education mostly after age 60. They find that the differences are mainly due to selection effects (based on cognitive abilities) at early ages, while the role of education increases after age 60. Leopold and Leopold (2018) find differences in self-rated health between higher-educated and lower-educated individuals over ages 30 to 80, which increase from age 50 (for men). Ross and Mirowsky (2010) find a physical impairment gap between the well-educated and poorly educated over the life-cycle which is more pronounced for women. These studies provide a descriptive picture of the education-health gradient over the life-cycle but do not claim causality.

We contribute to the literature on health effects of education by trying to find out whether these effects vary over the life-cycle, thereby going beyond the descriptive analyses. In our study, exogenous variation comes from compulsory schooling reforms in West Germany. Reforms were introduced on federal state level for birth cohorts between 1931 and 1954, depending on the state. Our main data set is the German Socio-Economic Panel study (SOEP), a representative survey running from 1984 until today. We pool these data with the Survey of Health Ageing, and Retirement (SHARE) and the German National Educational Panel Study (NEPS). While our data set clearly has disadvantages compared to administrative data in some aspects, its main advantage is that it covers a 36 year-period and allows to follow the individuals born around the reform periods over many decades and estimate both short-run and long-run effects of education on health within the same framework and data set. This allows us to learn about the point in the life-cycle when potential health effects of education set in. Estimating these health effects on outcomes below the level of mortality would not be possible with any available administrative data set in Germany.

The only two studies we are aware of that also explicitly look at health effects of education over the life-cycle are Clark and Royer (2013) and Gehrsitz and Williams Jr (2022).[1] Clark and Royer (2013) find that two changes in British compulsory schooling laws did not affect mortality as a whole, but also not when focussing on 5-year age groups between 20-24 and 65-69. Gehrsitz and Williams Jr (2022) study effects of a reform in Scotland and report results by age for 30-55 years old individuals. They do not find effects on self-reported health but a reduction in hospitalizations for selected conditions. This mainly holds for men and starts after age 40. In contrast to Clark and Royer (2013), we study life-cycle effects on morbidity and health care utilization and also go beyond age 69 (and age 55 as in Gehrsitz and Williams Jr, 2022).

[1] Bhuller et al. (2017) and Delaney and Devereux (2019) study life-cycle effects of education on earnings.

Table 4.1: Effect of education on health – previous economic literature

| Authors | Country | Type of education | Instrument | Studied age group | Results |
|---|---|---|---|---|---|
| Adams (2002) | USA | Secondary school | Quarter of birth | 51 to 61 | Positive effects |
| Arendt (2005) | Denmark | Middle school | CSR | 25 to 64 | No effects |
| Lleras-Muney (2005) | USA | | | | |
| Oreopoulos (2006) | UK | Secondary school | CSR | 32 to 64 | Positive effects |
| Albouy and Lequien (2009) | France | Secondary school | CSR | 48 to 80 | No effects |
| Silles (2009) | UK | Secondary school | CSR | 25 to 60 | Positive effects |
| Kemptner et al. (2011) | Germany | Secondary school | CSR | 16 to 65 | No effects - women<br>Positive effects - men |
| Braakmann (2011) | UK | Secondary school | Being February-born | 28 to 45 | No effects |
| Lager and Torssander (2012) | Sweden | Various types | CSR | 15 to 64 | No effects |
| Clark and Royer (2013) | UK | Secondary school | CSR | 12 to 74 | No effects |
| Jürges et al. (2013) | UK | Secondary school | CSR | 32-53 + 44-77 | No effects |
| Gathmann et al. (2015) | Europe | Secondary school | CSR | 50+ | No effects - women<br>Positive effects - men |
| Palme and Simeonova (2015) | Sweden | Secondary school | CSR | 28 to 66 | Negative effects |
| Buckles et al. (2016) | USA | College | Vietnam War draft | 28 to 65 | Positive effects |
| Brunello et al. (2016) | Europe | Secondary school | CSR | 50+ | Positive effects |
| Meghir et al. (2018) | Sweden | Secondary school | CSR | 16 to 75 | No effects |
| Davies et al. (2018) | UK | Secondary school | CSR | 37 to 74 | Positive effects |
| Kamhöfer et al. (2019) | Germany | College | Expansions in college availability | 39 to 68 | No effects - mental health<br>Positive effects - physical health |
| Dahmann and Schnitzlein (2019) | Germany | Secondary school | CSR | 50 to 85 | No effects |
| Janke et al. (2020) | UK | Secondary school | CSR | 42 to 60 | No effects (except for diabetes) |
| Fischer et al. (2021) | Sweden | Secondary school | CSR | 18 to 81 | Positive effects |
| Begerow and Jürges (2022) | Germany | Secondary school | CSR | 50 to 79 | No effects |

Notes: Own research of studies without the claim of completeness. CSR stands for compulsory schooling reforms. The age ranges are not always clearly specified in the papers and sometimes deducted by ourselves using information provided on used birth cohorts as well as calendar years when the outcomes are measured. "No effects" usually means no significant effects and abstracts from economic effect sizes which might be non-zero. Brunello et al. (2016) use various European countries.

Our results suggest a positive correlation of health and education which increases over the life-cycle. For example, one more year of schooling goes along with 0.5 more diagnosed conditions for individuals aged 50-54 and 1.5 more diagnoses for individuals aged 75-79. It also goes along with a 0.5 percentage point higher likelihood to report being in poor health for individuals aged 40-44 but 1.5 percentage point higher likelihood for individuals aged 75-79. Thus, we can replicate a common pattern found in the literature. Yet, when looking at the causal relationship, we hardly find any effects for health and health care utilization at all. An exception is obesity, where positive effects of schooling start to be visible around age 60 and become very large in age group 75-79. An ex-post simulated power analysis as suggested by Black et al. (2022) and an analysis of selective panel attrition indicate that attrition and power do not play an important role in our sample until the age of 74. In contrast, the subgroup of 75-79 years old individuals (the oldest in our sample) suffers from small sample size and potential attrition problems. Yet, the point estimates for hospital stays, poor health and diagnoses also point at zero effects in this group. Yet, due to the mentioned problems, this – and the large estimated effect on obesity – should be interpreted with caution.

This paper is structured as follows. In Section 4.2 we present the institutional framework, data, and descriptive statistics. In Section 4.3 we show and discuss the main results:

instrumental variables estimations for different age groups. We also provide robustness checks, carry out a power analysis and inspect panel attrition. In Section 4.4 we study a possible reason for the zero effects. We conclude in Section 4.5.

## 4.2 Institutional framework and Data

### 4.2.1 Institutional framework and sample selection

In Germany, children enter primary school at the age of six. After four years in primary school they attend one of the three secondary school tracks. Secondary schools in Germany can, generally, be differentiated into basic (*Hauptschule*), intermediate (*Realschule*) and high schools (*Gymnasium*). The basic track (up to 8th or 9th grade) prepares students for apprenticeship, the intermediate track (up to 10th grade) qualifies students for apprenticeship or training in white collar jobs, and the high school certificate (up to 12th or 13th) gives access to academic education in colleges or universities. Before the German educational reform, which occurred from 1946 to 1969 in West Germany, basic track schools covered grades five to eight. The reform increased the number of compulsory schooling years from eight years to nine years. Decisions and policies regarding the educational system in Germany are made at the federal state level, hence the reform was implemented in different years by the various states (Tawiah, 2022). Some states introduced a compulsory ninth grade earlier, while the majority of the states only introduced an additional year of schooling due to the Hamburg Accord (*Hamburger Abkommen*) in 1964 (Kamhöfer and Schmitz, 2016). See Table 4.2 for the reform years. The reform was introduced due to a shortage in labor market opportunities and apprenticeships for school leavers, and to also increase the school leaving age (see Pischke and Von Wachter, 2008, for details).

#### Data

We pool data from three sources. The largest one and, thus, our main data source is the German Socio-Economic Panel (SOEP) which is a wide-ranging representative longitudinal study of households in Germany. SOEP, established in 1984, contains yearly information on around 30,000 respondents in nearly 15,000 households (Goebel et al., 2019). For our analysis we use SOEP version 37 containing yearly information from 1984 to 2020 (SOEP, 2022). In order to increase the number of observations, we augment our baseline sample with observations from the Survey of Health Ageing, and Retirement (SHARE) and the German National Educational Panel Study (NEPS): Starting Cohort Adults data (NEPS Network., 2022). SHARE is a representative micro dataset which provides health and socio-economic information of people age 50 and older from 28 European countries and Israel. We consider waves 1, 2 and 4-8 but not wave 3 (SHARELIFE) which considers different topics that are not of interest here (Börsch-Supan, 2020a,b,c,d,e,f,g, 2021; Börsch-Supan et al., 2013). NEPS is a longitudinal dataset that provides information on the acquisition of education in Germany, and educational processes and trajectories across the entire life span (Blossfeld et al., 2011). We consider all 12 waves of the NEPS from 2007 to 2020.

We restrict the sample to individuals born five years before and after the pivotal cohorts – that is, the first birth cohorts that were affected by the reform. Table 4.2 reports the reform years and shows how the age range of individuals we can identify effects for differ by federal states. For instance, for the outcome variables available from 1984 to 2020 in the SOEP (later for the other data sets), the youngest possible age is 25 for a person from Bavaria, born in 1959, observed in 1984. The oldest possible age is 94 for a person from Hamburg, born in 1926, observed in 2020. In our analysis below, we will form 5-year age groups to estimate effects. We restrict the sample to individuals between 30 (starting with age group 30-34) and 79 (for age-group 75-79) years to make sure that effects for certain age groups are not completely driven by individuals from single states. Nevertheless, effects for the age group 75-79, our oldest age group in the sample, will only be identified from individuals in Schleswig-Holstein, Hamburg, Lower Saxony, Saarland and Bremen. We do not consider this a problem of internal validity and, moreover, do not see a clear reason to assume that the effects in these federal states should differ from effects in this age group in the other states. Yet, there may be some concern regarding certain events during the early childhood years of those in this age group, such as malnutrition resulting from the food crisis in Germany from 1944 to 1948 which was severe in 1945, affecting the educational achievement, occupational status and income of individuals born in the winter of 1945/46, that may have long-term effects on health (Jürges, 2013). Such events may drive cohort/federal state effects which may influence the results instead of education. Individuals in age group 75-79 had already been born by 1945, implying that a majority of them were not affected by the food crisis in-utero. None of the individuals from Hamburg are affected and only 5% of observation in this age group were born in 1945. We, therefore, do not expect the food crisis to have a great impact on our results for the oldest age group but, obviously, cannot rule that out.

The data has information on age, gender, the state in which an individual attended school, years of education and the type of school-leaving degree. We use the school-leaving degree to infer years of schooling as our explanatory variable of interest.

### 4.2.2 Outcome variables and descriptive statistics

The health outcomes we consider are hospital stay in the previous year, number of illnesses diagnosed, poor self-rated health and obesity. More specifically, *Hospital stay*

Table 4.2: Reform years, corresponding first birth cohorts and ages

| Federal State | Pivotal birth cohort | Reform year | Youngest age in 1984 | Oldest age in 2020 |
|---|---|---|---|---|
| Schleswig Holstein | April 1932 | April 1947 | 47 | 93 |
| Hamburg | April 1931 | April 1946 | 48 | 94 |
| Lower Saxony | April 1947 | April 1962 | 32 | 78 |
| Bremen | April 1944 | April 1959 | 35 | 81 |
| North Rhine-Westphalia | April 1951 | April 1966 | 28 | 74 |
| Hesse | April 1951 | April 1966 | 28 | 74 |
| Rhineland Palatinate | April 1952 | April 1967 | 27 | 73 |
| Baden-Württemberg | April 1952 | April 1967 | 27 | 73 |
| Bavaria | August 1954 | August 1969 | 25 | 71 |
| Saarland | April 1943 | April 1958 | 36 | 82 |

Source: Begerow and Jürges (2022) for the reform years. Youngest age in 1984 calculated as follows: 1984 - pivotal cohort - 5. Oldest age in 2020 calculated as follows: 2020 - pivotal cohort + 5.

is an indicator variable based on the question whether a person was admitted at a hospital for at least one night the previous year. The number of illnesses diagnosed (called *diagnoses* from now on) is constructed from a question asking if an individual has ever been diagnosed by a doctor of one or more illnesses from a list of illnesses. The 13 illnesses asked are sleep disturbance, diabetes, asthma, heart disease, cancer, stroke, migraine, high blood pressure, depressive psychosis, dementia, joint disorder (also osteoarthritis, rheumatism), chronic back complaints and other illnesses. We count the number of diagnoses. *Poor health* is based on the 5-point scale of self-rated health and equals one if individuals choose the worst category. *Obesity* is a binary variable that indicates a body-mass index larger than 30 (based on self-stated body weight and height).

Table 4.3 reports numbers of observations in the final sample by outcome variable and age group. Next to the number of observations, we show from which data set the observations come. Clearly, SOEP has the most observations. Yet, as SHARE samples older individuals, it helps to increase numbers of observations particularly for the oldest age group. Note that diagnoses and hospital visits are not included in the NEPS data.

Table 4.4 reports descriptive statistics of all outcome variables. Some outcome variables are not available in all waves, hence, the sample size varies for the different outcomes with self-rated health having the largest sample (95,827 observations from 13,618 individuals). The smallest sample has 20,418 observations from 6,799 individuals. 12.4% of the observations stayed at least one night in the hospital the previous year. The maximum number of diagnoses in the sample is 11 out of the 13 options mentioned above. There is an average of about 1.6 illnesses being diagnosed and 21% are obese, while almost 4% state that they are in poor health. The average age is 57 years and the sample is almost gender balanced. The average years of schooling is about 10.4 years.

Table 4.3: Number of observations

| Age group | Hospital Obs | Hospital (% SOEP / %SHARE / % NEPS) | Poor health Obs | Poor health (% SOEP / %SHARE / % NEPS) | Diagnoses Obs | Diagnoses (% SOEP / %SHARE / % NEPS) | Obese Obs | Obese (% SOEP / %SHARE / % NEPS) |
|---|---|---|---|---|---|---|---|---|
| 30 | 4693 | ( 100 / 0 / 0 ) | | | | | | |
| 35 | 6776 | ( 100 / 0 / 0 ) | 1674 | ( 100 / 0 / 0 ) | | | | |
| 40 | 7273 | ( 100 / 0 / 0 ) | 5601 | ( 100 / 0 / 0 ) | | | | |
| 45 | 10859 | ( 100 / 0 / 0 ) | 10756 | ( 100 / 0 / 0 ) | | | 2618 | ( 99 / 1 / 0 ) |
| 50 | 14206 | ( 95 / 5 / 0 ) | 15290 | ( 87 / 5 / 8 ) | 1427 | ( 48 / 52 / 0 ) | 6244 | ( 85 / 12 / 3 ) |
| 55 | 15225 | ( 90 / 10 / 0 ) | 21355 | ( 64 / 7 / 29 ) | 4788 | ( 68 / 32 / 0 ) | 9063 | ( 73 / 17 / 10 ) |
| 60 | 14787 | ( 86 / 14 / 0 ) | 23272 | ( 55 / 9 / 36 ) | 7592 | ( 73 / 27 / 0 ) | 9559 | ( 64 / 22 / 14 ) |
| 65 | 9491 | ( 85 / 15 / 0 ) | 13695 | ( 59 / 11 / 30 ) | 5054 | ( 71 / 29 / 0 ) | 6359 | ( 65 / 23 / 12 ) |
| 70 | 2857 | ( 87 / 13 / 0 ) | 3627 | ( 69 / 10 / 21 ) | 1271 | ( 71 / 29 / 0 ) | 2079 | ( 67 / 17 / 16 ) |
| 75 | 502 | ( 73 / 27 / 0 ) | 507 | ( 72 / 27 / 1 ) | 247 | ( 44 / 56 / 0 ) | 331 | ( 57 / 42 / 1 ) |

*Notes*: Own calculations based on SOEP, SHARE, and NEPS. Age group 30 stands for age group 30-34, age group 35 stands for age group 35-39, and so on. These are the observations in the final selected sample that enter the regressions below.

Table 4.4: Descriptive statistics

| | Mean | SD | Min. | Max. | Observations | Survey years |
|---|---|---|---|---|---|---|
| *Outcome variables* | | | | | | |
| Hospital stay (yes = 1, no = 0) | .124 | .33 | 0 | 1 | 86,669 | 1984 - 2020[a] |
| Poor self-rated health (yes = 1, no = 0) | .036 | .185 | 0 | 1 | 95,827 | 1992 - 2020 |
| # Diagnoses | 1.616 | 1.568 | 0 | 11 | 20,418 | 2009 - 2020[b] |
| Obese (yes = 1, no = 0) | .208 | .406 | 0 | 1 | 36,382 | 2002 - 2020 |
| *Treatment and instrument* | | | | | | |
| Years of schooling | 10.411 | 1.814 | 8 | 13 | 86,669 | |
| Reform | .606 | .489 | 0 | 1 | 86,669 | |
| *Other information* | | | | | | |
| Age | 57.159 | 8.049 | 30 | 79 | 86,669 | |
| Female | .505 | .5 | 0 | 1 | 86,669 | |

*Notes*: Own calculations based on SOEP, SHARE, and NEPS. The statistics for age, female, years of schooling and reform are based on the estimation sample for hospital stay. Hospital stay is an indicator variable for whether an individual was admitted at a hospital for at least one night the previous year. Obese is a binary indicator of having a BMI > 30. Poor self-rated health is a binary indicator of checking the lowest of five possible categories in self-rated health. [a]No data for hospital stay in 1990 and 1993 in the sample. [b]biennial.

### 4.2.3 OLS estimations

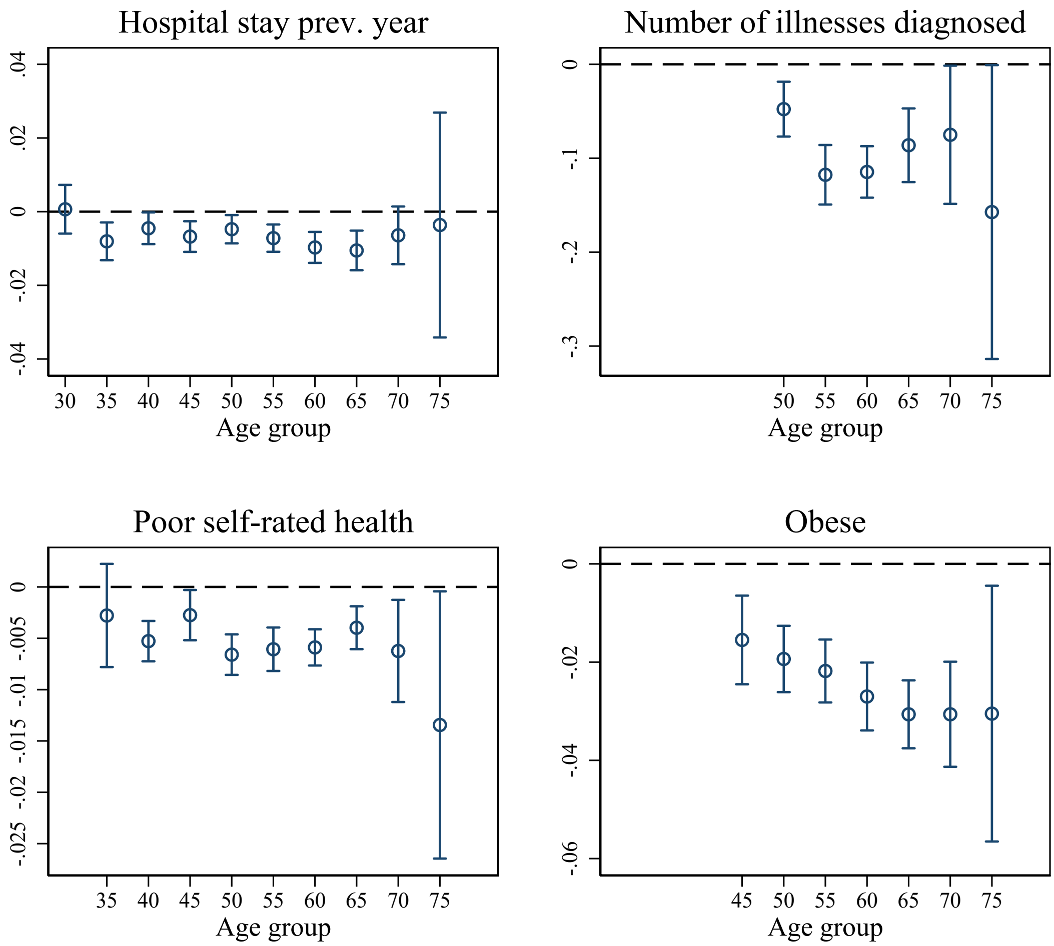
As another descriptive statistic, we present results of OLS regressions of the following form:

$$H_{ist} = \sum_g \beta_g Yed_{is} \times agegroup_{it} + \beta X_{ist} + \varepsilon_{ist} \tag{4.1}$$

where $H_{ist}$ is a health outcome of individual $i$ who attended school in state $s$. $Yed_{is}$ is an individual's number of years of schooling. To flexibly account for the correlation of schooling and health, we define 5-year age brackets, denoted $agegroup_{it}$, as follows: 30-

34, 35-39, 40-45,..., 75-79.[2] The vector $X$ includes a constant, a full set of age dummies (in years), federal state dummies, female indicator, survey as well as interview year dummies, and state-specific time trends, i.e. interactions of school state dummies with a linear trend in year of birth. We cluster standard errors on state×year of birth level. The coefficients of interest are the $\beta_g$, they are reported in Figure 4.1.

Figure 4.1: OLS results




*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on Eq. (4.1) with 95 % confidence intervals. Standard errors clustered at state×year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

The results show a positive relationship between education and health over the life-cycle. With small exceptions, an additional year of schooling is related to better health throughout the life-cycle. Broadly, one more year of education is related to a 10% lower number in the measure of negative health, when we compare the coefficients with the

[2] Age groups start with 35-39 for poor health, with 45-49 for obesity, and with 50-54 for diagnoses because they are only covered later in the data.

sample means in Table 4.4. That is, for example, the coefficient for hospital visits is around -0.01 while the sample mean is around 0.1. For all measures but hospital stay, the health-gap in education widens over the life-cycle, where the estimated coefficients for age group 75-79 are two to three times larger than those for the youngest age groups.

## 4.3 Instrumental variables estimations by age group

### 4.3.1 Empirical Strategy

We run the two-stage-least-squares (2SLS) equivalent to regression equation (4.1) where we instrument $Yed_{is} \times agegroup_{it}$ with $Reform_{is} \times agegroup_{it}$ for the age groups used above in Section 4.2.3. $Reform_{is}$ is an indicator variable for whether an individual was affected by the reform or not. If certain assumptions hold, the estimated coefficients of the instrumented $Yed_{is} \times agegroup_{it}$ identify the local average treatment effect of education on health for the different age groups, that is, the effects for those individuals who increase years of schooling solely because they are forced to do so due to the reform. Given that there are no never-takers of the reform, this group of compliers is composed of individuals at the lowest margin of willingness to take education.

To interpret the coefficients causally, the assumptions for instrumental variables need to be fulfilled. First, the instrument needs to be exogenous. This is fulfilled if all other changes that occur across states prior to reform are uncorrelated with the law change itself and the outcomes given controls. The inclusion of state-specific time trends helps to deal with any factors that affected states over time. Secondly, the exclusion restriction needs to hold. Given how large the reform was, this is not completely obvious. While it is conceivable that the reform may have had effects on health through other channels than education, there is no evidence for this so far and this assumption is standard in the literature.

Third, the instrument needs to be correlated with the endogenous variable, that is, the reform should have a significant effect on the number of years of schooling. While not shown in a separate table, the reform increased the average years of schooling by around 0.5 in our data with an F-statistic on the excluded instrument of around 35. These results are in line with the many other studies that evaluate these reforms in Germany (e.g. Kemptner et al., 2011, or Kamhöfer and Schmitz, 2016). Finally, when allowing for potential effect heterogeneity, the monotonicity assumption needs to hold in order to be able to interpret the results for a well-defined subgroup, namely the compliers. This assumption means that no individual would reduce the years of schooling due to the reform. Again, we follow the vast previous literature and assume that individuals with high educational attainment do not attain less schooling due to an increase in mandatory years in the basic track.

### 4.3.2 Estimation results

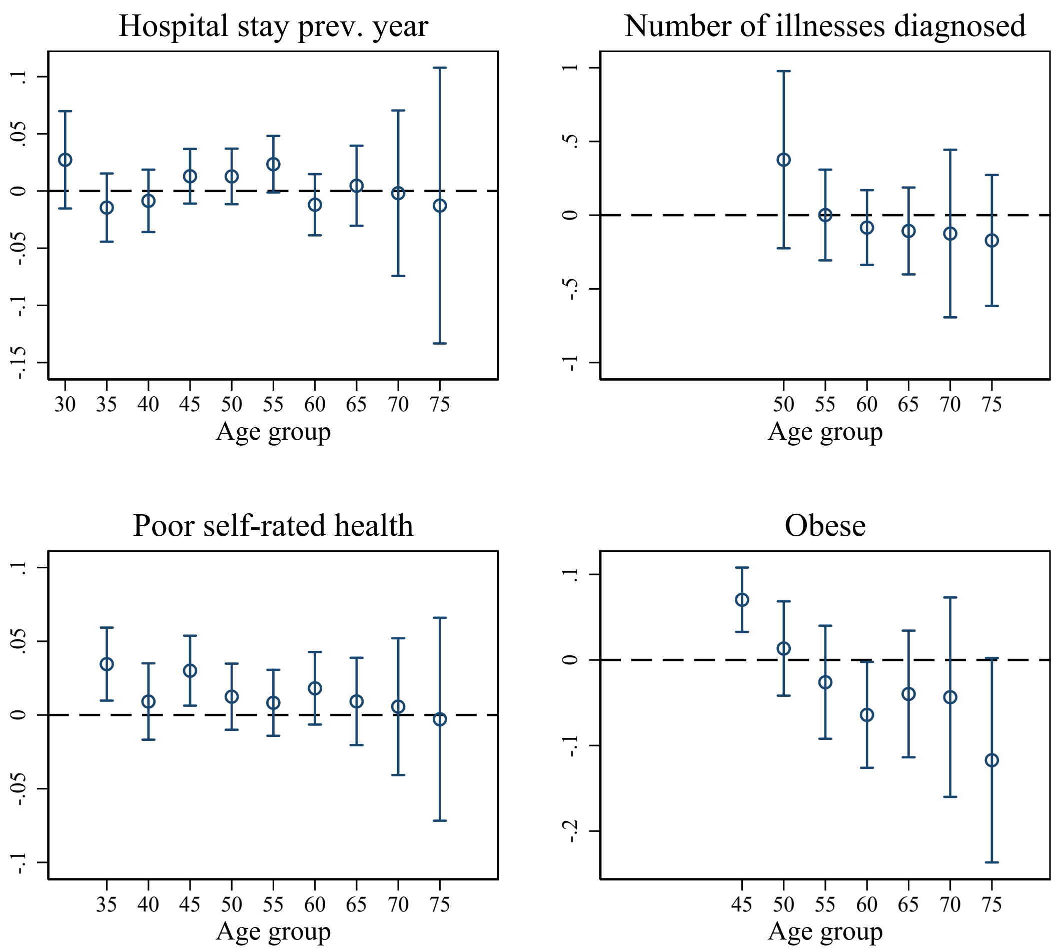
The main results using instrumental variables estimations are shown in Figure 4.2. For the three outcomes hospital stay, poor self-rated health and diagnoses, the estimated age-group-specific effects fluctuate around zero throughout the life-cycle. While being a bit noisy, they clearly provide evidence that the correlation between health and education from Figure 4.1, which even increases over the life-cycle, is unlikely to be due to a causal effect of education. While precision is an obvious issue given our sample size, the coefficients are all close to zero. An exception is obesity, where the effect of schooling is positive in the age group 45-49 with a five percentage point increase in obesity due to more schooling. Over the life-cycle this turns negative, resulting in a 4 to 8 percentage point lower likelihood to be obese for individuals in the range 60-74 (significant in the age group 60-64 only), and a more then ten percentage point lower likelihood to be obese due to an additional year of schooling in the age group 75-79 (only significant at the 10 percent level).

In Figure 4.3 we repeat the analysis separately by gender. We do not find a structural difference in results but note that the negative effect on obesity seems to be driven by women. In summary, the main finding of the paper is the following: individuals with more schooling are in better health and the health gap by education increases over time. However, there is hardly any local average treatment effect of additional schooling for individuals at low education margins. Up to the age of 79, we do not observe an improvement in health due to eduction. In that sense, the results are in line with those of Clark and Royer (2013), although they use different data and outcome variables. They are in contrast, however to some of the findings by Gehrsitz and Williams Jr (2022). An exception is obesity which seems to decrease over the life-cycle due to education. However, this relationship is estimated with low precision only.

**Robustness checks**

We conduct different robustness checks and report their results in the Appendix. In Figure A4.1 we repeat the baseline estimation of Figure 4.2 but do not additionally account for state-specific trends. In Figure A4.2 we account for short school years (SSY). In 1966 – 1967, there was the introduction of two short school years in all states in West Germany except Bavaria. The start of the school year moved from spring to fall but it was already in fall for Bavaria, see Pischke (2007) for details. This was achieved in most states through two SSY with 24 weeks instead of the regular 37 weeks of instruction each. The introduction of the SSY occurred simultaneously with the compulsory schooling reform in some states. Therefore, the SSY is a possible confounding factor indicating our results may be biased with its omission. For the estimation, we include an indicator variable for SSY in the 2SLS regressions.

Figure 4.2: Instrumental variables estimations: Baseline results



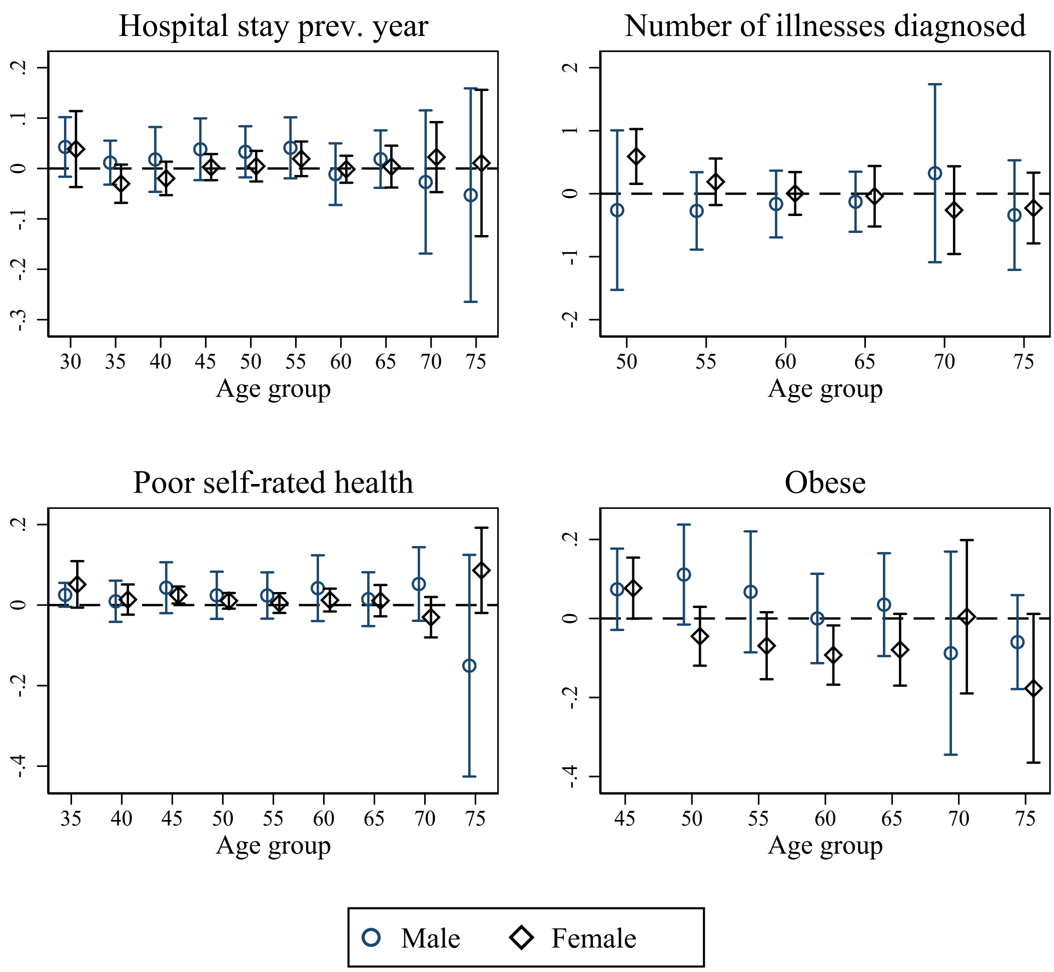

*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state×year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female, and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

In Figure A4.3 we make a different sample selection. Instead of five years around the pivotal cohort in each state we use all birth cohorts from 1930 to 1960. Finally, in Figure A4.4, we only use SOEP as a data source. The results in the robustness checks are fairly similar to those in the baseline specification.

### 4.3.3 A simulated ex-post power analysis

Findings close to zero for a relevant share of the population, together with a comparably small sample size in the age groups and larger standard errors raise the question of

Figure 4.3: Heterogeneous effects by gender




*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state$\times$year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

statistical power. For instance, how likely is it that the true effect is economically large but that we fail to identify it given our sample? That is, how likely are we to make a type-II-error?[3] In order to receive estimates of statistical power and minimum detectable effect sizes (MDE) in our data and application, we follow the simulation-based approach suggested by Black et al. (2022). In the spirit of their approach, we search for the minimum effect size that has 80% power at the 5% significance level, meaning that – if this was indeed the true effect size – in 80% of all cases we would reject the hypothesis of no effect at the 5% significance level.

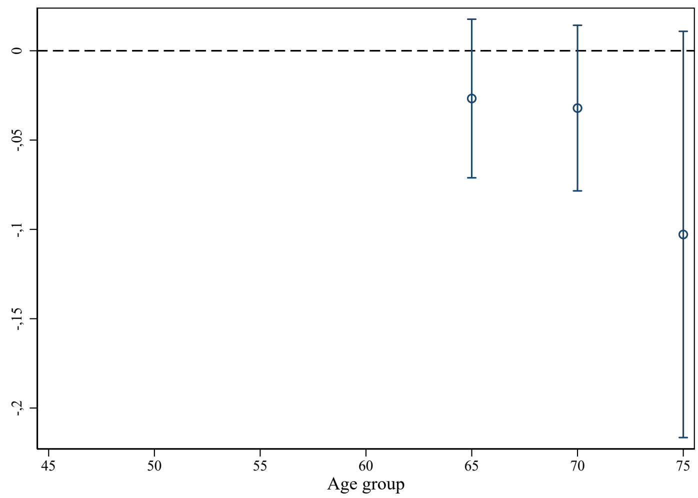
[3]The description of our procedure takes a lot from Freise et al. (2022).

As an exemplary procedure we choose the estimation using the outcome variable *obesity* where small to no effects are found for age groups below 75 and significant effects at the 10 percent level are found for age group 75-79. To simplify matters, instead of estimating the two stage least squares regressions resulting in Figure 4.2, we, in this section, estimate the reduced form-relationship. That is, we estimate the direct effect of being affected by the reform on the indicator of obesity by running this regression:

$$obese_{ist} = \sum_{g} \beta_g reform_{is} \times agegroup_{it} + \beta X_{ist} + \varepsilon_{ist} \tag{4.2}$$

The reason is that, in the simulated power analysis described below, we randomly assign the treatment (being affected by the reform). By only considering the reduced form, we do not need to make further assumptions on how individuals react to the reform, that is, whether they are compliers or always takers. Provided that the first stage coefficient is large, the reduced form results resemble the two stage-least squares results anyway. Finally, we focus on the 8,769 observations in the three age groups 65-69, 70-75, and 75-79. As seen in Figure 4.4, the reduced form results for this group are in line with the 2SLS results from Figure 4.2.

Figure 4.4: Reduced form regression




*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ with 95 % confidence intervals. Standard errors clustered at state×year of birth. Number of observations: 8,769. The regression equation is: $Obese_{ist} = \sum_g \beta_g reform_{is} \times agegroup_{it} + \beta X_{ist} + \varepsilon_{ist}$. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female, and state-specific linear trends. Age 65 in the figure stands for age group 65-69, age 70 stands for age group 70-75, and so on.

Our procedure to get estimates of the three minimum detectable effect sizes for the three age groups is the following:

1. Use the baseline sample and drop all observations that are affected by the reform, that is, all who are younger or equal to the pivotal cohorts. Thus, only use observations not treated to make sure that possible real treatment effects do not affect our results. This means, from the 8,769 observations between 65 and 79 with information on obesity and who are born +-5 years around the pivotal cohort, we drop 3,212 and keep 5,557 observations.

2. Refill the sample to get a sample size of 8,769 such that in each state we have the exact same number of observations per age group as before. We do this in two ways. First, we take data from individuals in the state born between -6 to -10 years relative to the pivotal cohort. For the remaining observations, we oversample untreated individuals.

3. Randomly select individuals of the sample created in step 2 and assign them the treatment ($reform$) such that the share of treated individuals per federal state is the same as in the original sample.

4. Assign the treated individuals a uniform and constant effect of $X$ which is added to their measure of obesity.

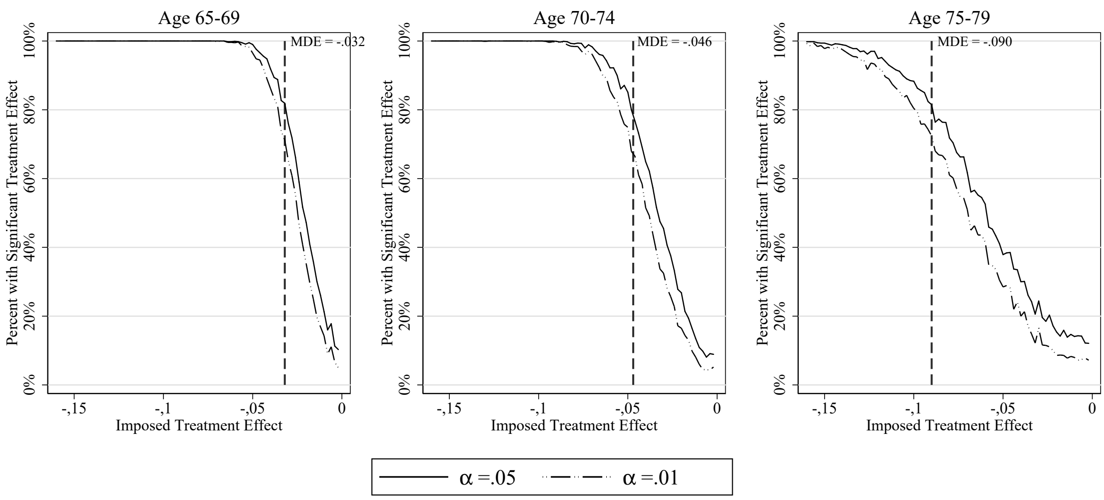
5. Estimate Eq. (4.2) and check whether the estimated coefficients of the treatment effects in the three age groups is significant at the 10% level, that is, whether or not we made a type-II-error (fail to reject the null hypothesis of a zero effect although we know that the true effect is $X \neq 0$).

6. Repeat steps 3 to 5 1,000 times and count the share of significant treatment effect estimates.

7. Repeat step 6 80 times where the imposed treatment effect $X$ is gradually increased from -0.002 to -0.160 in steps of 0.002.

Figure 4.5 reports the results of this exercise where for each of the 80 imposed treatment effects the share of significant estimates in 1,000 repetitions is shown. The figure reports results for two different significance levels. The minimum detectable effect size is defined to be where the 5% significance curve shows 80% power. This is at -0.034 in age group 65-69. This means: if the true effect was -0.032, our data would allow for an analysis with a power of 80%. The minimum detectable effect size for age group 70-74 is -0.046, while it is -0.090 for age group 75-79. The minimum detectable effect sizes for the three age groups are in the same range as our estimated effects. We interpret these findings as evidence that power-problems do not rule out a useful analysis by age groups given our data and that the estimated effects in Figure 4.4 seem to be well-powered. Yet, it should be noted the MDE in the age group 75-79 is very large. This works in the example of obesity, where indeed our point estimate is in this range. However, the MDEs in that age group for the three other outcomes are large, too (not shown). Hence, we cannot rule out that actual health effects on education are considerable in this age group. All in all, we conclude that our analysis yields reliable results for individuals up to age 74 but the results need to be interpreted with caution for the group 75-79.

Figure 4.5: Simulated power analysis



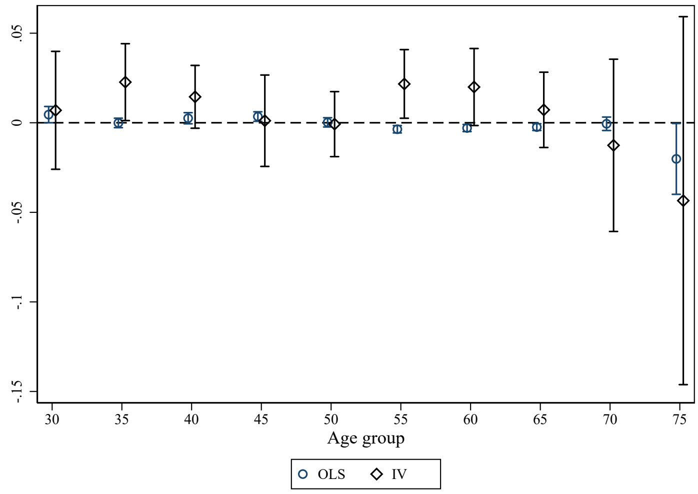

*Notes:* Own calculations, based on the procedure suggested by Black et al. (2022). Detailed simulation procedure described in the text.

### 4.3.4 Attrition

A major concern with longitudinal household surveys, especially those focused on the older population and where the interest is health, is the potential bias from attrition (Banks et al., 2011; Deng et al., 2013; Fichera and Savage, 2015). Attrition hinders a survey from being representative of the target population and, hence, introduces potentially substantial biases to statistical inference (Banks et al., 2011; Deng et al., 2013). Even though the surveys we use are constantly being refreshed, selective attrition (possibly due to mortality) by educational status might be an issue.

We apply two complementary approaches to test for potential attrition problems. In the first approach, we use the data from our working sample and generate a binary indicator *attrition*. This indicator equals one if a person does not appear in the next survey wave and zero if she either appears in the next wave or if it is the last wave (year 2020) in the survey. We generate this indicator before we make the sample selection based on the pivotal cohort. According to this definition, 25 percent of all person-year-observations in our sample drop out between two waves. Next, we use *attrition* as an outcome and run an OLS and IV regressions as before. Figure 4.6 shows the results of this exercise. Until age group 70-74, there is hardly any difference in attrition by education. This is different for age group 75-79 where one more year of education goes along with a marginally significant two percentage point decrease in attrition (OLS, five percent in IV, but not significant). This difference is small, however.

Figure 4.6: Effect of education on attrition over the life-cycle

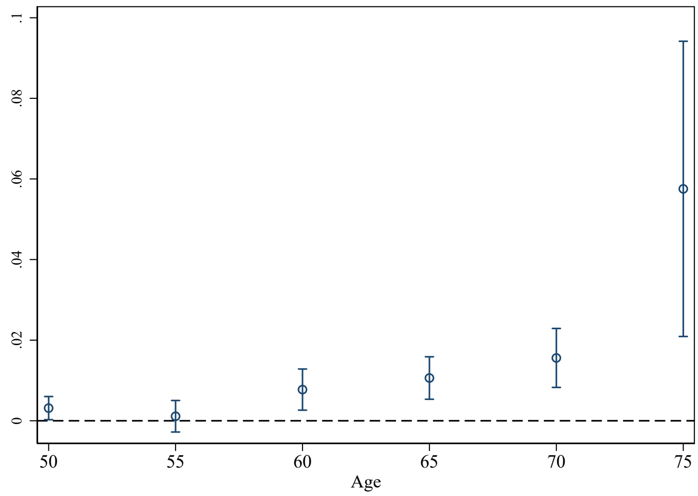



*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on OLS and 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state×year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female, and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

As a second approach, we keep each individual only once in the sample and ask for the likelihood to (still) be in the sample at ages 50-54, 55-59,...75-79. To give an example, consider a person born in 1930 who drops out in 2000. The indicator *before50* takes on the value one for her, while the indicator *before 75* takes on the value zero. An individual born in 1960 would have non-missing values for indicators *before50*, *before55*, and *before60*. As she turns 65 after 2020, the remaining indicators are missing for this person. 86% of all individuals are still in the survey at age 50 (based on 15,094 individuals). This number constantly goes down to 47% who are still in the survey at age 80 (based on 427 individuals). We, then, run separate regressions of all indicators on years of schooling, female, birth year dummies, survey and federal state dummies and state-specific linear trends in birth year. Results are shown in Figure 4.7. We observe a higher likelihood to stay in the sample with more education. The differences are statistically significant but small in economic terms until (and including) age group 70-74. They are somewhat larger, however, for age group 75-79 where education has a stronger relationship with the likelihood to stay in the sample.

We take these results to draw a similar conclusion as with the power analysis in the previous subsection. While attrition does not seem to play an important role for the

Figure 4.7: Effect of education on (still) being in the sample at certain ages




*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of six separate regressions. 50 stands for the coefficient of years of education in a linear regression of *before50* on years of education, female, birth year dummies, survey and federal state dummies and state-specific linear trends in birth year. 95 % confidence intervals. Standard errors clustered at state×year of birth.

results of age groups until 74, the estimated effects for the highest age groups might be subject to selective attrition and, again, should be interpreted cautiously.

## 4.4 Potential mechanisms

A possible reason for only very small health effects of the reform (if any), even in the long-run, might be its institutional setting. Pischke and Von Wachter (2008) already argued that basic skills of the compliers, necessary for the labor market, might already have been settled after eight years of schooling and that the ninth grade did not further improve them. This is at least consistent with the finding of no returns to cognition of that reform (Kamhöfer and Schmitz, 2016). Another hypothesis could possibly be more important for health effects: the reform might not have affected the types of occupation the compliers worked in afterwards. Apart from health behaviors, job types might be the most important channel how education affects health (Marmot, 2004; Erikson, 2006; Burgard and Lin, 2013; Darin-Mattsson et al., 2017)..

To test this, we look at four different classifications of occupations: white-collar vs. blue-collar jobs, physically highly demanding vs. physically less demanding jobs, psychosocially highly demanding vs. psychosocially less demanding jobs and manual vs. non-manual jobs. Occupations are classified as physically (psychosocially) highly demanding if the Overall Physical (Psychosocial) Exposure Index for the occupation derived by Kroll (2011) is larger than five and as less demanding if it is less or equal to five, as done by Mazzonna and Peracchi (2017). We group the occupations into manual and non-manual based on the 11 classes of the Erikson and Goldthorpe (EGP) class schema.[4] EGP classes I, II, III, IVa, IVb and V are classified as non-manual, and classes VI, VII and IVc as manual.[5] We restrict the sample to those within the working age group i.e. 30 - 65 years and to the SOEP due to data availability.

Table 4.5 shows results of eight separate regressions (four times OLS and four times 2SLS) where we regress the four outcome variables explained above on years of schooling and the same control variables as before. Here, however, we do not separate results by age groups. The coefficient of years of schooling is reported in the table. In our sample, 67 per cent have white-collar jobs, 53 per cent have physically less demanding jobs, 49 per cent have psychosocially less demanding jobs and 34 per cent are manual workers. OLS estimates show a significant correlation of education and possibly healthier jobs. 2SLS results, however, have coefficients close to zero which are also not statistically significant. It seems that there is no effect of an additional year of compulsory schooling on healthier jobs. This may be part of the explanation why we do not see effects of this reform in the long-run.

Table 4.5: Potential mechanisms

| | Observations | Sample mean | OLS | 2SLS |
|---|---|---|---|---|
| White collar job | 53,944 | 0.67 | 0.097*** | 0.013 |
| | | | (0.004) | (0.044) |
| Physically low jobs | 53,816 | 0.53 | 0.108*** | 0.022 |
| | | | (0.005) | (0.058) |
| Psychosocially low jobs | 53,373 | 0.49 | 0.030*** | -0.017 |
| | | | (0.007) | (0.070) |
| Manual work | 71,049 | 0.34 | -0.090*** | -0.007 |
| | | | (0.003) | (0.044) |

*Note:* Own calculations based on SOEP and individuals younger than 66. Point estimates of the coefficient of years of schooling from regressing each outcome on years of schooling, female, age fixed effects, state fixed effects, survey and survey year fixed effects, and state-specific linear trends for OLS, and instrumenting years of schooling with the reform for the 2SLS version. Standard errors clustered at state$\times$year of birth.

## 4.5 Conclusion

We study the relationship of education and health over the life-cycle using compulsory schooling reforms in West Germany as exogenous variation. Our main contribution to the literature is to estimate effects for different age groups starting age 30 and up to age 79, several decades after education took place. This allows to scrutinize a pattern that may have been missed in the previous literature: zero aggregate effects, as often

[4]In the SOEP, the EGP is derived from the ISCO-88 classification as well as the information on self-employment and number of employees/supervisory status (SOEP Group, 2022).

[5]See Table A4.1 in the Appendix for details.

found in the literature, might blur potential health effects that only show up late in life. Stronger effects in older ages can be justified theoretically (Kaestner et al., 2020) but may also be expected by descriptive results of an increasing education-health gradient over the life cycle, as found by previous studies (e.g., Case and Deaton, 2005, Galama and van Kippersluis, 2019).

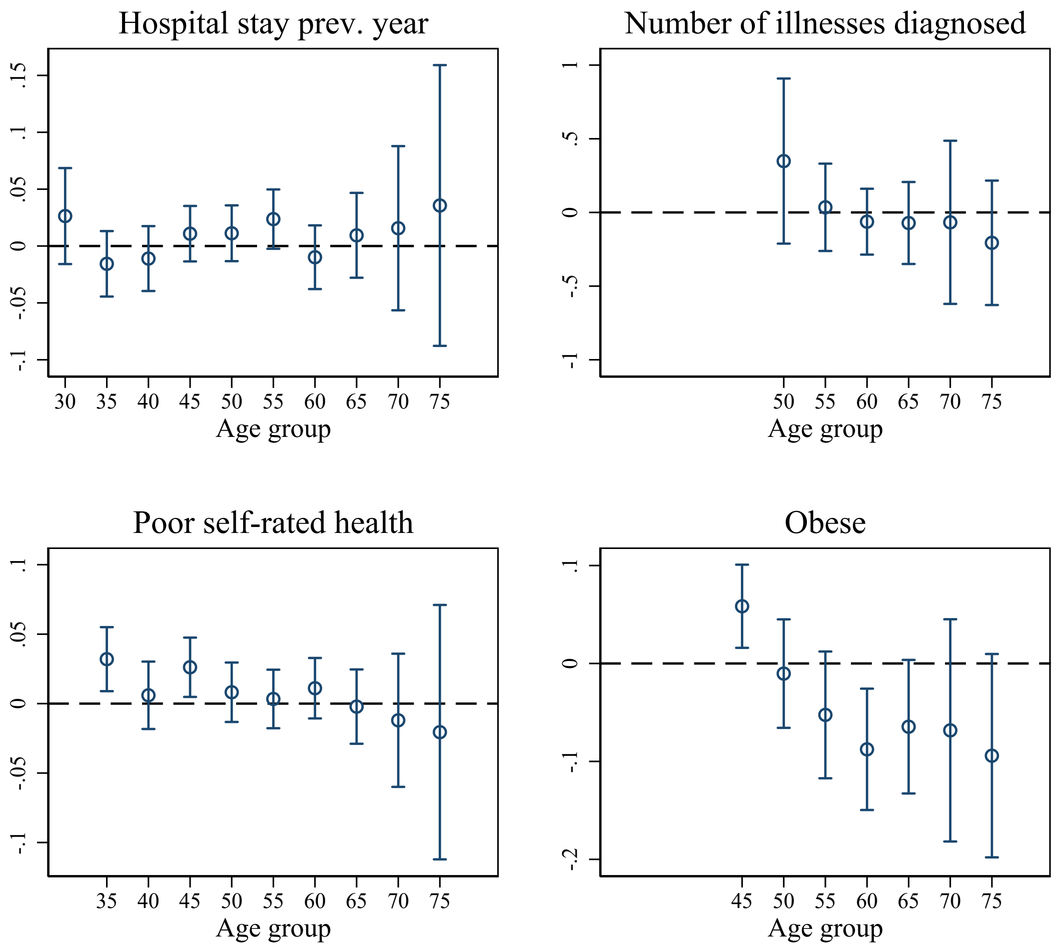
While we find an increase in the health gap by education over the life cycle, we basically do not find causal effects of an additional year of compulsory schooling on health and health care utilization for any age group up to 70-74. We also tend to interpret findings for age group 75-79 as evidence of absence of health effects even though the effects are a bit less clear and we suffer from a small sample here. Yet, most point estimates are basically zero for our oldest age group. Obesity (as a measure of health behavior and not necessarily health status) is the only exception. Here we do find effects starting age 60 and increasing until 79. Of course, we only identified a local average treatment effect, i.e., effects for individuals at the lowest margin of education willingness. A possible reason why there are no long-term health effects of this reform might be its institutional setting. The additional year of compulsory schooling did not bring individuals on a different career path. Yet, the most likely channel of how improved education could affect health is through better (and healthier) jobs.

This might be different for other changes in the German educational system. For instance, the educational expansion in the 1960s to 1980s with a strong increase in the number of universities and high schools (*Gymnasien*) allowed many individuals to get much more education. Kamhöfer et al. (2019) do not only find positive (physical) health effects of this reform for individuals decades later but also that better jobs are a possible mechanism for this effect.

Germany has carried out several reforms of its education system in recent years, also for higher education margins such as university entrance diplomas. While these reforms – most notably the compression of secondary school education from 9 to 8 years, going along with increased instruction times – have been evaluated in terms of short-term health outcomes (e.g., Quis, 2018, Marcus et al., 2020), it cannot be ruled out that larger effects will only show up in some decades. Yet, as these reforms, again, most likely did not have significant effects on individuals' career paths and chosen jobs, the results from this paper at least allow for the prediction that long-run health effects of these reforms might not be substantially larger than the short-term effects.

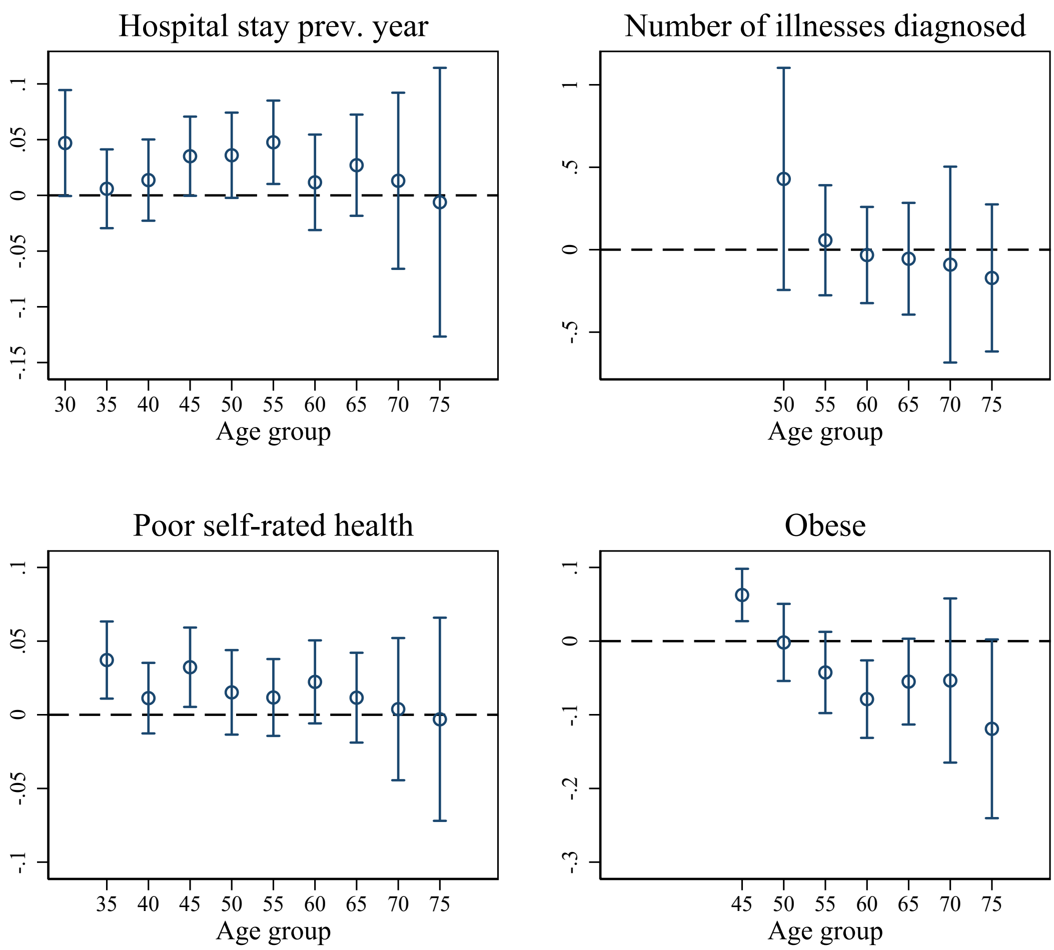
# Appendix

Figure A4.1: No trends



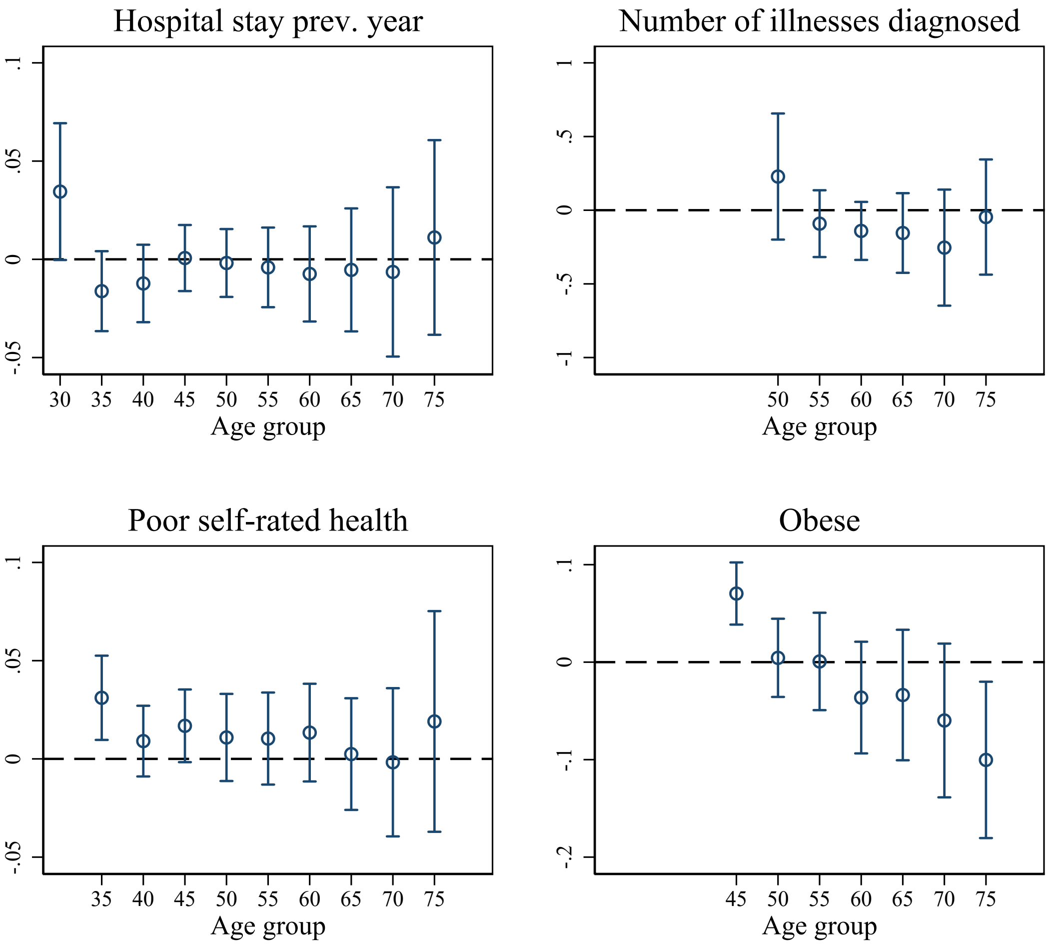

*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state$\times$year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, and female. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

Figure A4.2: Including short school years



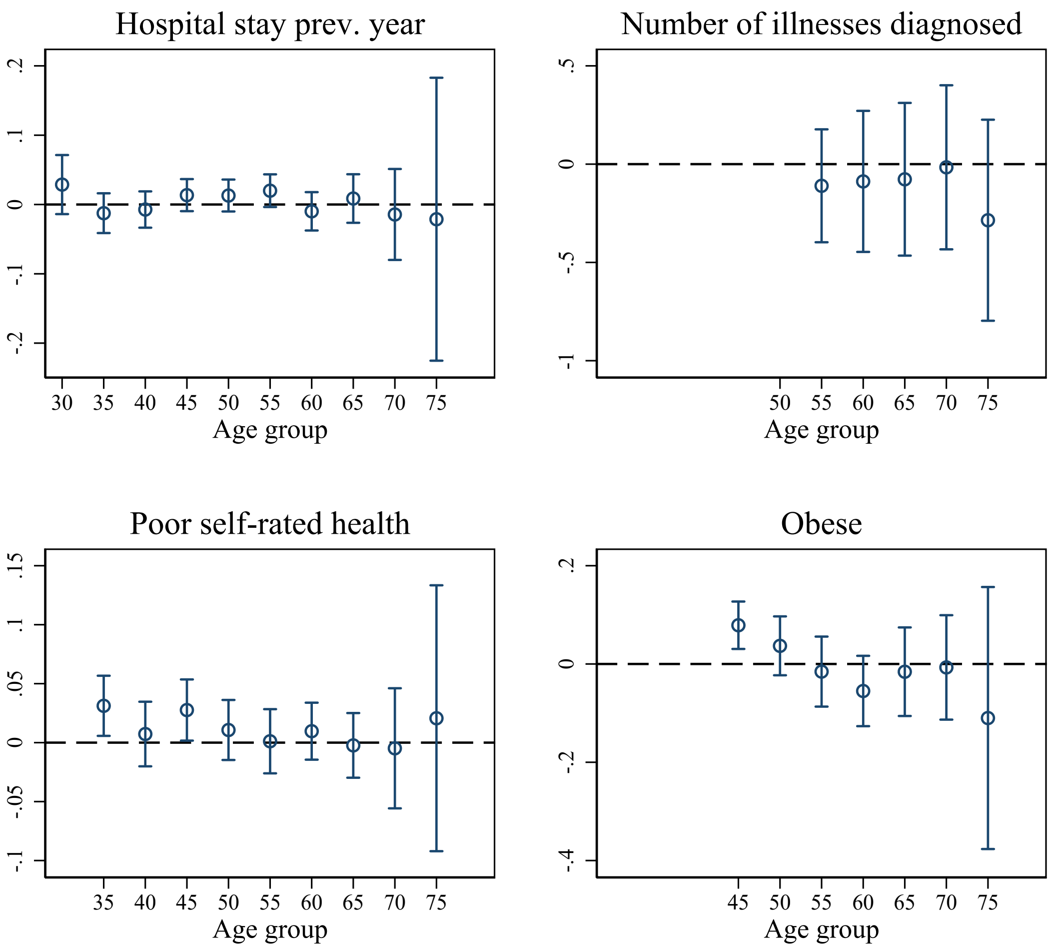

*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state$\times$year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female, state-specific linear trends, and an indicator for short school years.. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

Figure A4.3: Sample 1930 - 1960




*Notes:* Own calculations based on SOEP, SHARE, and NEPS. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state$\times$year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

Figure A4.4: Only SOEP




*Notes:* Own calculations based on SOEP. Point estimates of the coefficients $\beta_g$ based on 2SLS versions of Eq. (4.1), where instruments are interactions of reform dummy (pivotal cohort and older) with the age groups. 95 % confidence intervals. Standard errors clustered at state$\times$year of birth. Controls: age fixed effects, state fixed effects, survey fixed effects, survey year fixed effects, female and state-specific linear trends. Age 30 in the figure stands for age group 30-34, age 35 stands for age group 35-39, and so on.

Table A4.1: Job classifications

| EGP classification | Manual/Non-manual |
|---|---|
| (I) Higher Managerial and Professional Workers | Non-manual worker |
| (II) Lower Managerial and Professional Workers | Non-manual worker |
| (IIIa) Routine Clerical Work | Non-manual worker |
| (IIIb) Routine Service and Sales Work | Non-manual worker |
| (IVa) Small Self-Employed with Employees | Non-manual worker |
| (IVb) Small Self-Employed without Employees | Non-manual worker |
| (V) Manual Supervisors | Non-manual worker |
| (VI) Skilled Manual Workers | Manual worker |
| (VIIa) Semi- and Unskilled Manual Workers | Manual worker |
| (VIIb) Agricultural Labour | Manual worker |
| (IVc) Self-Employed Farmers | Manual worker |

*Source:* SOEP Group (2022). *Notes:* Own grouping for manual/non-manual and calculation based on SOEP and NEPS. Number of observations: 77,887.

# Chapter 5

# The Effect of Children on Health

## 5.1 Introduction

Much research has been done to understand the effect of children on labour supply (Angrist and Evans, 1996; Jacobsen et al., 1999; Vere, 2011; Zhang, 2017; Guo et al., 2018) and on investments in children (Black et al., 2005; Cáceres-Delpiano, 2006; Angrist et al., 2010; Fitzsimons and Malde, 2014; Bhalotra and Clarke, 2020). However, less attention has been given to the effect of children on parental health. Caring for children requires time, energy and financial resources especially at younger ages. Raising children and pursuing a career simultaneously can be challenging, and balancing these two can affect parental health (Bucher-Koenen et al., 2020). This is, therefore, relevant for policy decisions surrounding fertility. Negative health effects of family and work need to be considered when policies targeted at increased fertility rates and/or increased labour supply are being designed (Bucher-Koenen et al., 2020). This paper aims at investigating the long term effects of having children on parental health using the German Socio-Economic Panel (SOEP).

Raising young children can be physically and mentally demanding for parents (Kruk and Reinhold, 2014), and is expected to have more effect on mothers' mental health than on fathers' mental health since mothers are primarily responsible for them (Gove and Geerken, 1977). Having children also has some long term effects on parents' health. Women with children are associated with a higher likelihood of dying from diabetes and circulatory diseases such as hypertension but with decreased risks of breast, ovarian and endometrial cancers than women without children (Beral, 1985). The arrival of children may be stressful and possibly harmful to a father's health but as the family size stabilises, the presence of children can be associated with better paternal health (Bartlett, 2004). Heliövaara and Aromaa (1981) find that the Body Mass Index (BMI) of women increases with age and the number of children. The increase in BMI with respect to the increase in the number of children was strongest in the youngest age group of women aged 25-30. They also found a relationship between the number of children and the prevalence of obesity. The relationship was most pronounced in the youngest age groups and women aged 75-84. Hank (2010) finds that higher parity (number of children birthed) is associated with better self rated health satisfaction among mothers and fathers aged 50 plus in Germany. He finds a negative relationship between having more children and physical health amongst women in East Germany and positive relationship between having more children and mental health amongst women in West Germany.

Most of the findings from analysing the impact of children on parents' health have not been causal. There are just a few causal analyses. Cáceres-Delpiano and Simonsen (2012) analyse the effect of number of children on a mother's health using the US census. They consider mothers in the fertile ages of 20 to 45. Using the Two stage least squares (2SLS) method and multiple birth as an instrument, they find an increase in the likelihood of being obese, smoking and suffering from high blood pressure. Kruk and Reinhold (2014) use the SHARE data to analyse the effect of children on depression in old age using men and women aged 50 to 90 in Europe. They implement a 2SLS using twin birth and same sex of first two children as instruments. They find that women are at a higher risk of being depressed if they have a third child as a result of having multiple birth. They do not find such an effect on men and no effect when the same-sex instrument is used. Bucher-Koenen et al. (2020) analyse the double burden of raising children and working on mortality among Swedish mothers age 55 and older. They find that mothers of twins are more likely to die from lung cancer and chronic obstructive pulmonary disease and, heart attacks and strokes. They also find stronger effects for highly educated mothers and those with above-median pension income.

This paper contributes to the few causal analyses by analysing the aggregate effects of children on health and also introducing life-cycle effects of children on health. I analyse the effect on general physical and mental health, and further look at the impact on various illnesses. Unlike Cáceres-Delpiano and Simonsen (2012) who consider women in their fertile ages, this paper looks at long-term effects on both men and women by considering individuals years after they have had their last birth. Kruk and Reinhold (2014) only focus on depression but in this paper, a measure for the general mental health which does not only account for depression is considered, as well as another mental health illness amongst the various illnesses. I do not only consider women as most of the literature related to fertility but men as well. The analyses are, however, done separately for men and women. The German Socio-Economic Panel (SOEP) data is used for the analyses considering individuals from ages 50 to 90. The SOEP is a representative survey that has been running since 1984. According to Andersen et al. (2007), "the annual SOEP attrition analysis suggests that the health status of respondents has only a marginal effect on their probability of dropping out or remaining in the sample." They further state that this makes the SOEP a good representation of respondents with poor health given its longitudinal structure than any cross-sectional study. The SOEP is, therefore, suitable for health analysis, especially when the focus is on older individuals and attrition is a potential bias (Deng et al., 2013; Fichera and Savage, 2015). Twin birth and same-sex children are used to exogenously determine the number of children. However, same-sex children, that is the first two children have the same sex, is found to be a weak instrument. Even though having two boys first influences the likelihood of having more children for both men and women, it is not a strong instrument to yield unbiased results. Based on the twin birth instrument, I find that having more than two children has negative effects on the health of both men and women. There is an increase in the BMI of women who have more than two children as result of having twins at second birth by 1.88 points (7% relative to the mean of 26.77 and 37% of the standard deviation). For men, there is a decline in the mental health (measured by MCS) and physical health (measured by PCS) by 2.73 points (5% relative to the mean of 52.53 and 28% of the standard deviation) and 2.53 points (5.5% relative to the mean of 46.22 and 26% of the standard deviation) respectively. Over the life-cycle, the BMI of women increases until age 64 but not for men. There is a decline in mental health from age 75 for both men and women, but a significant effect is only found for men in the 80-84 age group. No significant effects on physical health are found, except a negative effect for men in age group 60-64. In order to address issues that threaten the validity of the twin instrument, I *(i)* use same-sex twins instead of twin birth at second birth as instrument, *(ii)* exclude individuals with any birth from 1990 onwards, and *(iii)* use the same-sex twin instrument and exclude individuals with any birth from 1990 onwards. The main results are quite robust under different checks. Illnesses considered in further analyses include diabetes, cardiac disease, cancer, stroke, high blood pressure, depression, dementia, joint disease and back trouble. One would expect that finding a decline in general health as result of having more than two children would reflect in certain illnesses, however, this is not the case. Although I find that the BMI of women increase, I do not find any effect on any of the various illnesses considered except a slight decrease in the likelihood to be diagnosed with dementia. I rather find an increase in the likelihood for men to be diagnosed with stroke, high blood pressure and cardiac disease. Neither the decline in the mental health of men reflects in the likelihood of being diagnosed with depression or dementia nor the decline in physical health reflect in the likelihood of being diagnosed with back trouble or joint disease.

The results are local average treatment effects for those who have more than two children as a result of having twins at second birth. Contrary to Cáceres-Delpiano and Simonsen (2012), I do not find any effect on being diagnosed with high blood pressure. Although Kruk and Reinhold (2014) find a negative aggregate effect on depression for women but not for men, I do not find any effect for either men or women. The increasing BMI amongst women until age 64 may be attributed to pre-retirement conditions and the negative effect on mental health of men at older ages to long-term effects of critical periods experienced in earlier years.

The rest of the paper is structured as follows: Section 5.2 presents the identification strategy. The data and descriptive statistics are provided in Section 5.3. In Section 5.4, the empirical strategy is presented. In Section 5.5, the main results are shown and discussed, validity issues are addressed by running robustness checks in Section 5.6, Section 5.7 looks at the effect on various illnesses, and Section 5.8 looks at attrition. Section 5.9 concludes.

## 5.2 Identification Strategy

Interpreting associations between the number of children and health causally is problematic. Having children or not as well as the number of children an individual or a couple decides to have is mostly by choice. Certain personal factors such as education and wealth, and cultural factors influence the number of children an individual has (Martin, 1995; Kravdal, 2002; Bongaarts, 2010; Colleran et al., 2015; Günther and Harttgen, 2016). This raises the problem of omitted variable bias. Although I am interested in the effect of the number of children on health, there is also the likely effect of health on the number of children. This means there is also the problem of reverse causality. These two problems make it difficult to make any causal statements. In order to deal with these problems, I introduce two instruments that randomly assign children to parents; twin births and same sex of first two children.

Twin births do not occur regularly. The proportion of twin births in the sample used is $\approx 2\%$, which shows how rare twin births are. Twin birth randomly assigns an additional child to a couple that may have been expecting a singleton birth. In the sample of individuals with at least two children, having twin birth at second birth then randomly increases the number of children to three instead of two. With twin births, there are no never-takers. The twins instrument is based on the idea that the event of a twin birth is essentially random and hence, unrelated to potential outcomes or demographic characteristics (Angrist and Fernandez-Val, 2010). In order for the twins instrument to be valid, it should not have any direct effect on the health outcomes but an indirect effect through increasing the number of children. If there is some sort of correlation between the probability of having twin births and unobservable variables, this will make the twins instrument invalid. There is evidence showing a correlation between maternal health and having twin births. Bhalotra and Clarke (2019) find that healthier women are more likely to have twin birth hence, there is positive selection of women into twin birth. However, Farbmacher et al. (2018) have shown that this selection is mainly attributed to dizygotic twins, commonly known as fraternal twins. Strong relationships have been established between having dizygotic twins, and mother's age, height, weight and fertility treatments (Reddy et al., 2005; Farbmacher et al., 2018), which introduce selection bias making it difficult to identify any causal effects. Fertility treatments/Artificial Reproductive technologies (ART) also tend to increase the probability of having twin births (Vitthala et al., 2009), which are likely to be correlated with observable and unobservable characteristics of the parents such as age at birth, level of education and income (Kruk and Reinhold, 2014; Lundborg et al., 2017). Any relation between these unobservable characteristics and health outcomes threaten the validity of the instrument. In order to address these threats, I conduct robustness checks using same-sex twins instead of twins at second birth as the instrument, which corrects the selection bias from dizygotic twins even if the zygosity is not known (Farbmacher et al., 2018), and excluding births from 1990 onwards since these fertility treatments became available from the 1990s onwards (Kruk and Reinhold, 2014).

Parents with their first two children being the same-sex are more likely to have a third child than those with mixed-sex (Angrist and Fernandez-Val, 2010). The sex of a child is equally random since a couple cannot determine the sex of a child at conception. Therefore, having the first two children being the same sex or not is random. Since sex composition is random, one would expect that it affects the health outcomes only by increasing the number of children.

I do not expect to have similar estimates independent of the instrument used even if both instruments are valid. The twins and same-sex instruments consider different compliant sub-populations i.e. those who have three children as a result of having twins at second birth versus those who have three children because their first two births have the same sex. The characteristics of these two sub-populations of compliers may differ. Twin birth presents an unexpected increase in the number of children while having a third child in response to a same-sex composition is a desired and anticipated increase in the number children (Kruk and Reinhold, 2014). Having twin birth implies taking care of two children at the same but not so with having a third child at a delayed time. This should have varying effects on health. I will therefore estimate different local average treatment effects (LATE) for the compliant sub-populations based on the two instruments. This means the estimates will depict the effect of those who have three or more children as a result of having twin birth at second birth (same sex of first two children) who otherwise would have had two children.

## 5.3 Data

The data I use in this analysis is the German Socio-Economic Panel (SOEP) with yearly information from 1984 to 2020.[1] The SOEP, established in 1984, is a wide-ranging representative longitudinal study of private households which contains yearly information on around 30,000 respondents in nearly 15,000 households (Goebel et al., 2019). Starting in 2002 the SOEP health module in the individual questionnaire has been revised and put into a two year replication period (SOEP Group, 2022). I, therefore, consider ten survey years from 2002 to 2020.[2] The health module includes the generated SF-12-Variables[3] and variables on height and weight with imputation flags and a user-friendly longitudinal checked generated variable of the Body Mass Index (BMI) (SOEP Group, 2022). I restrict the sample to individuals from the age of 50 years to 90 years with at least two children. I consider those in this age group because they are either more likely to be done with childbearing or are likely not to have any more children. At age 50 most women are likely to be in the menopausal period. Mean/median age at menopause varies across different countries ranging from 44.6 years to 52 years (Thomas et al., 2001). The median age at menopause in Western populations of women is approximately 51 years however, about 5% of women experience early menopause which occurs at ages 40 - 45 (Santoro, 2003). The average number of years since an individual had a child in the sample is 32 years.

### 5.3.1 Treatment variable and instruments

The main treatment variable is the number of children. The treatment variable is equal to 1 if an individual has at least three children and 0 otherwise. This will be instrumented with having twin birth at second birth and having the same sex of the first two children in separate analyses. Twin births are determined based on the year of birth of the children since direct information on twin births is not available in the SOEP data.[4] I then determine the position of the twins amongst their respective siblings based on the year of birth. Respondents with twins at second birth were assigned 1 for the twin birth instrument and all others were assigned 0. Individuals who had twins at first birth are excluded from this sample. The same-sex instrumental variable is assigned 1 if the

[1] Socio-Economic Panel (SOEP) version 37, published in 2022, doi:10.5684/soep.core.v37eu

[2] There are, however, twelve survey years for MCS and PCS. From the data, less than 1.5% of the respondents have their MCS and PCS values being reported in 2017 and 2019 instead of 2016 and 2018 respectively.

[3] see Andersen et al. (2007) for details on the SOEP version of SF-12v2.

[4] Although SOEP asks for the month of birth, not every parent volunteers this information making it difficult to determine twin birth with both the year and month of birth. However, based on the available information on the month of birth of the children, I am able to detect some non-twin siblings born in the same year but in different months. Though not perfect, this excludes a proportion of non-twin siblings that may be classified as twins based on just the year of birth.

first two children of a respondent have the same sex and 0 otherwise. Individuals with twins at second birth are excluded from this sample since they do not already have two children.

### 5.3.2 Health outcomes

The health measures considered are Body Mass Index (BMI), Mental Component Summary Scale (MCS) which is a measure for mental health and Physical Component Summary Scale (PCS) which is a measure for physical health.

BMI, also known as Quételet's index, is the ratio of a person's weight (in kilogrammes) relative to the square of the person's height (in metres) (Shetty and James, 1994). The World Health Organisation (WHO) classifies BMI $< 18.50$ as underweight, BMI 18.50 - 24.99 as normal range and BMI $\geq 25$ as overweight with sub-classifications of overweight; pre-obese (BMI 25 - 29.99), obese class I (BMI 30 - 34.99), obese class II (BMI 35 - 39.99) and obese class III (BMI $\geq 40$) (WHO, 2000). Their classification is based on the association between BMI and mortality. Pre-obese is also known as overweight and obese class III as extreme obesity (NHLBI, 1998). BMI has been used as a measure for chronic energy deficiency (CED), obesity, nutritional status, mortality and chronic diseases such as coronary heart disease (CHD), diabetes mellitus (Shetty and James (1994); Weisell (2002); Gray et al. (2015)). Those in the underweight category are more susceptible to CED and those in the obese categories are more susceptible to CHD and diabetes mellitus, however, they all are more susceptible to mortality (Shetty and James, 1994; WHO, 1995, 2000; NHLBI, 1998; Katzmarzyk et al., 2001; Flegal et al., 2005; Pischon et al., 2008). Though there are some criticisms with BMI not being a good measure of health, it is still seen as a useful measure of health.

Good mental and physical health are crucial for the health of the whole being. Mental health could be an important determinant of physical health and also, there may be negative spillovers of poor physical health on mental health (Ohrnberger et al., 2017). Studies have also shown the associations between psychological factors, such as anger, anxiety and depression, and other diseases such as coronary heart disease and hypertension (Kubzansky and Kawachi, 2000; Yan et al., 2003; Suls and Bunde, 2005). MCS and PCS, measures for mental health and physical health respectively, are among the generated SF-12-Variables in the SOEP health module. These variables are based on the internationally recognised SF-12v2 questionnaire on health-related quality of life containing 12 questions, which can be grouped into eight sub-scales and also into two superordinate dimensions of mental health and physical health (Andersen et al., 2007). The SOEP version of SF-12v2 deviates from the original SF-12v2 to some degree in the formulation and order of questions and in the general layout (Andersen et al., 2007). Due to the differences between SOEP-SF-12 and the original SF-12v2, Andersen et al.

(2007) developed a computational algorithm based on the empirical results of the 2004 SOEP data. They, however, state that they designed the method used to compute scale values for the eight sub-scales and the two main dimensions as close as possible to the procedure used with the original SF-12v2. The sub-scales are physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health. Four of the sub-scales consist of one item and four of them consist of two items. Using a $z$-transformation, the items are transformed to values that fall between 0 and 100. The eight $z$-transformed sub-scales are grouped under the two superordinate scales, MCS and PCS, using explorative factor analysis (Principal Component Analysis (PCA), varimax rotation). The values of these two dimensions are $z$-transformed, as done for the sub-scales, to have a mean of 50 and standard deviation of 10. Higher points on the scale indicate better health and lower points indicate lower health.

### 5.3.3 Data Description

Other covariates considered are age, migration background, survey year dummy variables and whether a person lives in West Germany or East Germany. It has been shown that women in East Germany tend to have more children compared to those in West Germany (Kreyenfeld, 2004). Associations between parity and health, and timing of first birth and health have also been found to vary between East and West German women (Hank, 2010).

In the sample, the mean age of both women and men is 64 years. Based on the instrument, I have two samples of different sizes. This is as a result of some missing values for the sex of a child and/or year of birth of a child. Table 5.1 provides gender-wise descriptives for those with two children and those with three or more children based on the instrument sample. Those with two children on average give birth a bit later than those with three or more children. The average BMI of both men and women, regardless of the sample, fall in the overweight range. The average MCS (mental health measure) and PCS (physical health measure) of men are higher than that of women in all the samples. For both instrument samples, there is a one point difference between the average BMI of women with two children and that of women with three or more children, with women with three or more children having the higher BMI. A similar statement can be made for the MCS and PCS of women however, the values are higher for women with two children. These indicate that the average health of women with two children is better than that of women with three or more children. On the other hand, the average BMI of men with two children and that of men with three or more children are almost the same. There is a slight difference between the average MCS of men with two children and that of men with three or more children. The average PCS of men with two children is about a point higher than the average PCS of men with three

Table 5.1: Descriptive Statistics

| Sample | *Multiple birth instrument* | | | | *Same-sex instrument* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Two children* | | *Three or more children* | | *Two children* | | *Three or more children* | |
| | Women | Men | Women | Men | Women | Men | Women | Men |
| BMI | 26.432 | 27.554 | 27.319 | 27.471 | 26.429 | 27.560 | 27.314 | 27.456 |
| | (4.874) | (3.965) | (5.293) | (4.244) | (4.885) | (3.967) | (5.278) | (4.224) |
| MCS | 50.701 | 52.661 | 49.840 | 52.307 | 50.707 | 52.680 | 49.787 | 52.379 |
| | (10.235) | (9.505) | (10.766) | (10.085) | (10.246) | (9.494) | (10.761) | (10.094) |
| PCS | 45.036 | 46.492 | 43.193 | 45.769 | 45.055 | 46.508 | 43.194 | 45.767 |
| | (10.161) | (9.664) | (10.691) | (10.204) | (10.177) | (9.669) | (10.673) | (10.225) |
| No. of children | 2.000 | 2.000 | 3.548 | 3.466 | 2.000 | 2.000 | 3.549 | 3.466 |
| | (0.000) | (0.000) | (0.968) | (0.849) | (0.000) | (0.000) | (0.961) | (0.853) |
| Age | 63.651 | 63.345 | 65.044 | 63.861 | 63.596 | 63.305 | 65.026 | 63.972 |
| | (9.555) | (9.269) | (10.425) | (9.962) | (9.550) | (9.269) | (10.394) | (9.953) |
| Age at first birth | 24.804 | 28.172 | 23.159 | 27.146 | 24.851 | 28.244 | 23.150 | 27.105 |
| | (4.512) | (5.261) | (4.092) | (5.213) | (4.548) | (5.315) | (4.089) | (5.150) |
| Migration bkgd* | 0.116 | 0.113 | 0.158 | 0.168 | 0.115 | 0.113 | 0.157 | 0.170 |
| | (0.320) | (0.317) | (0.365) | (0.374) | (0.319) | (0.316) | (0.364) | (0.376) |
| West Germany | 0.709 | 0.734 | 0.751 | 0.815 | 0.710 | 0.735 | 0.751 | 0.816 |
| | (0.454) | (0.442) | (0.432) | (0.388) | (0.454) | (0.441) | (0.433) | (0.387) |
| Twins | 0.000 | 0.000 | 0.045 | 0.055 | | | | |
| | (0.000) | (0.000) | (0.207) | (0.228) | | | | |
| Same-sex | | | | | 0.476 | 0.483 | 0.497 | 0.527 |
| | | | | | (0.499) | (0.500) | (0.500) | (0.499) |
| Observations | 21,818 | 18,628 | 13,562 | 10,579 | 22,011 | 18,865 | 13,458 | 10,404 |

*Note:* Own calculations based on SOEP. Statistics of all variables except MCS and PCS are based on the BMI sample. Standard deviations are in parentheses. * Migration background

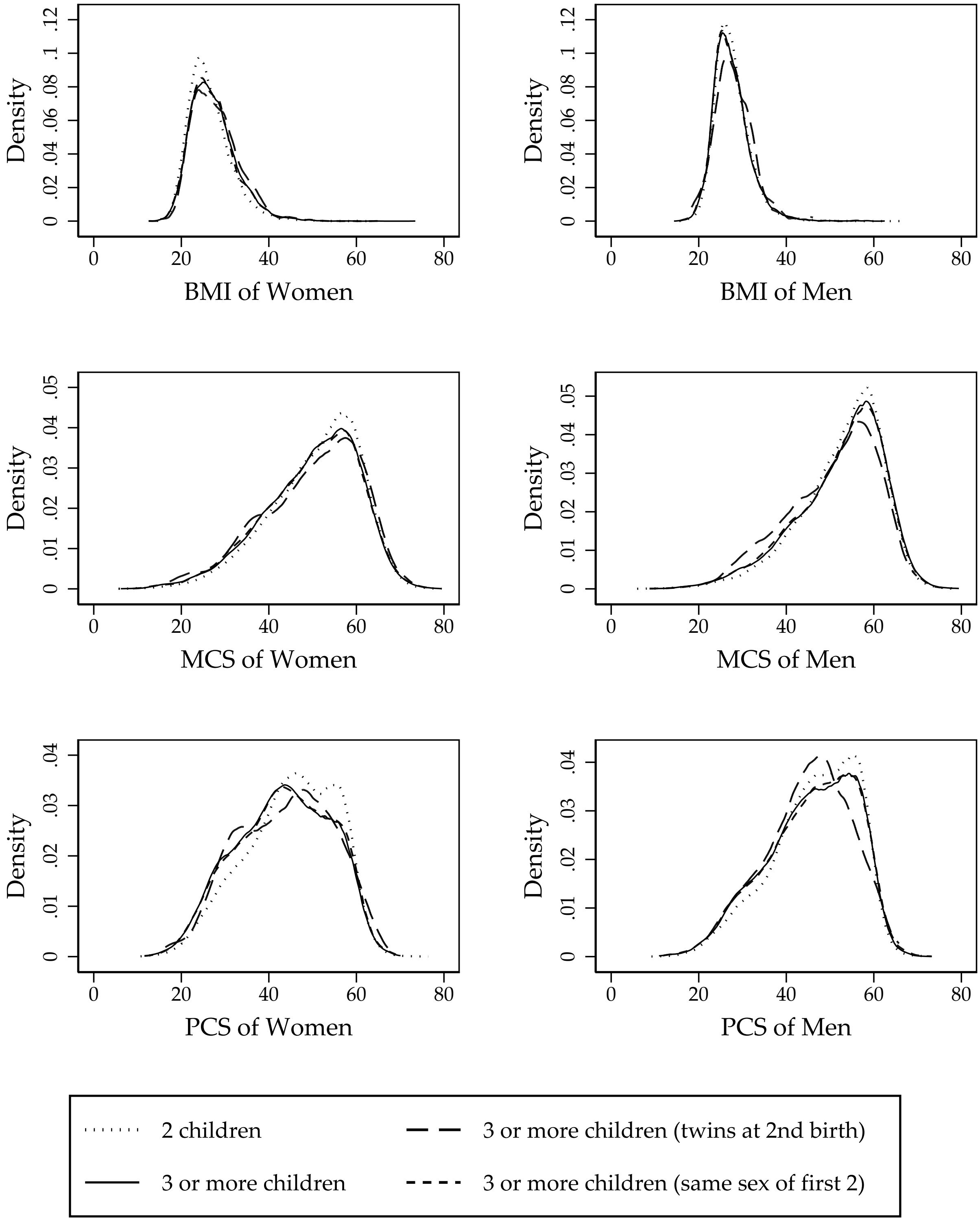
or more children. In summary, women with three or more children on average have a slightly higher BMI, and poorer mental and physical health compared to women with two children, whereas men with two children only have a slightly better mental and physical health than men with three or more children. It should be noted that having twins instead of singletons is likely to have an extra impact on health through various channels such as risks during pregnancy and delivery (Bucher-Koenen et al., 2020), and the stress of raising two children of the same age (Kruk and Reinhold, 2014).

On average, 1.7% of individuals with at least two children have their youngest child being less than 10 years of age.[5] Similar to Kruk and Reinhold (2014), a large proportion of individuals in our sample have been done with childbearing for a long while. Hence, the results depict long-term effects.

I take a look at the distribution of the health outcomes of individuals with two children, individuals with three or more children, individuals with three or more children given that they had twins at second birth and individuals with three or more children given that their first two children are of the same sex. I look at the distributions for men and women separately. Figure 5.1 shows that the distributions for those with three or more children and those with three or more children given that their first two children are

[5] Observations with the last child being born less than 5 years prior to the survey year are excluded from the sample

Figure 5.1: Kernel densities by the treatment status for each gender.




*Source:* Own calculations based on SOEP.

of the same sex are similar. For MCS, there is a flatter curve which spreads to the left for those with twins at second birth in comparison to the others. Those with twins at second birth have relatively higher BMI compared to the others. Women with two children have better BMI compared to the others. The MCS of women and men with twins at second birth is relatively lower. PCS of men with twins at second birth seems not to be better than the others.

## 5.4 Empirical Strategy

I first analyse the impact of children on health by implementing the two-stage least squares (2SLS) method. As stated before, the instruments to be used separately are twin birth and same sex of first two children. These instruments have been used to analyse the effect of children on labour supply and earnings (Angrist and Evans, 1996; Silles, 2016) as well as the effect of children on depression (Kruk and Reinhold, 2014). The variable of interest, having three or more children, will be instrumented.

The first stage is estimated by

$$m2kids_i = \alpha_0 + \alpha_1 instr_i + \alpha_2 age_i^2 + \alpha_3 migback_i + \alpha_4 west_i + \nu_s + \lambda_t + \varepsilon_i \qquad (5.1)$$

where $m2kids$ is a dummy variable for having more than two children: $m2kids = 1$ if an individual has more than two children and $m2kids = 0$ if an individual has two children. $instr$ is an indicator variable for either instruments. For instance, if twin birth is used as the instrument, then $instr$ is equal 1 if an individual has twin birth at second birth and 0 otherwise. $age^2$ is the square of an individual's age at the time of survey, $migback$ is an indicator for whether an individual has a migration background or not and $west$ is an indicator for whether an individual lives in West Germany or not. $\nu_s$ is a set of birth year fixed effects and $\lambda_t$ is a set of survey year fixed effects which account for heterogeneity between birth cohorts and survey years respectively. Due to perfect collinearity with birth year and survey year ($age = surveyyear - birthyear$), linear age is not included but rather age squared. The first-stage effect of the instrument on having three or more children is captured by $\alpha_1$.

The second stage is then estimated using

$$health_i = \beta_0 + \rho \widehat{m2kids}_i + \beta_1 age_i^2 + \beta_2 migback_i + \beta_3 west_i + \nu_s + \lambda_t + u_i \qquad (5.2)$$

where $health$ is the health measure of interest. $\widehat{m2kids}$ is the estimate obtained from the first stage. $\rho$ captures the causal effect of having three or more children on the health measure.

I also analyse the changes in the effect of having more than two children over time. Given the possibility that there may not be variability in the instruments, especially the twin birth instrument, for each individual age, I use age groups instead. I use 5-year age groups ($grp_j$) to capture the progression of the effects over the life-cycle. To do this each age group is interacted with the variable of interest, $m2kids * grp_j$. In order to get an idea of how having more than two children is associated with health outcomes in each group, I use the following OLS regression:

$$health_{it} = \beta_0 + \sum_{j=1}^{n} \phi_j grp_j + \sum_{j=1}^{n} \theta_j m2kids_i \times grp_j + \beta_1 migback_i + \beta_2 west_i + \\ \beta_3 agefirst_i + \beta_4 agefirst_i^2 + \lambda_t + u_i \tag{5.3}$$

where $agefirst$ is the age at which an individual had their first child and $agefirst^2$ is the square of $agefirst$.[6] Instead of using 2SLS to estimate the effects over the life-cycle, I rather estimate the direct effect of having twin birth at second birth on health by substituting $m2kids_i \times grp_j$ with $twinsecond \times grp_j$ in Eq. 5.3.[7] This is referred to as the reduced-form effects. By doing so, mechanisms through which twin birth affects health are captured. In addition to the main mechanism of increasing the number of children, other mechanisms include: *(a)* women who have twin birth, on average, face greater health risks compared to women who have singleton births; *(b)* since twins are extremely closely spaced, twin birth may affect the spacing of additional children, which could have a direct effect on mother's health; *(c)* twin birth may affect the likelihood of divorce, which may act as an additional channel of stress and hence affect one's health; and *(d)* labour market choices may be adjusted with the birth of twins, probably women will decrease their labour force participation and men may take on more workload to earn more income, which may affect their health (Bucher-Koenen et al., 2020). The effect of twin birth on health will, therefore, provide an idea of the general effect of having more children on health.

## 5.5 Empirical Results

### 5.5.1 Aggregate Effects

The first-stage results depicted in Table 5.2 show a statistically significant effect of having twins at second birth on having three or more children for both men and women. The estimates indicate the proportion of those who had a third child as result of having twins at second birth. This implies that ≈40% of women and ≈35-36% of men with

[6] I do not factor age at first birth in the 2SLS model since it is an endogenous variable. The timing for the first birth is mostly by choice. This may therefore be correlated with some unobservable variables which in turn may have an influence on the health outcomes.

[7] I do not estimate the life-cycle effects using the same-sex instrument since it is not a strong instrument here. This is discussed in Section 5.5.

at least two children would have had a third child regardless. The estimates for the twins instrument are quite similar across health outcomes. For the same-sex instrument, the story is different. I find a significant effect for men, where they are 3.9 percentage points more likely to have a third child if their first two children are the same sex, but not for women. This may imply that sex composition is not enough reason to have an additional child for women in this sample. However, looking at having two boys first versus two girls first (Table A5.1 in the Appendix), I find that having boys first does increase the likelihood of having a third child for both men and women.[8] There is no significant effect of having girls first. This indicates that the girl child seems to be more sought after in the case of Germany. Therefore, having two girls does not lead to a third child in order to probably have a boy child, but having two boys does increase the likelihood of having a third child to probably have a girl child. Although significant effects of two boys first as well for same-sex first (for men only) on having a third child are found, these instruments are not strong enough to yield unbiased 2SLS results because the first-stage F-statistics are small ($< 10$).[9] As a result, I do not consider the same-sex instrument in the rest of the analysis.

The lower section of Table 5.2 shows the OLS and the second-stage estimates for having three or more children. The OLS results show an association between poor health and having three or more children. For women, having more than two children is positively correlated with BMI and negatively correlated with mental health (MCS) and physical health (PCS). Having three or more children is not significantly associated (economically or statistically) with the BMI of men but is negatively associated with their MCS (statistically insignificant) and PCS. The 2SLS results show an increase in the BMI of women who had three or more children as a result of having twins at second birth by 1.88 points (7% relative to the mean of 26.77 and 37% of the standard deviation). The average height of a women in the sample is 1.64 metres. An increase in BMI by 1.88 points corresponds to a 5 kilogramme increase in the weight of this woman. The average woman in the sample is in the overweight class, therefore, such an increase moves an average women closer to the obese class. There is no significant effect on the BMI of men. There is a significant decline in the MCS of men by 2.73 points (4.9% relative to the mean of 52.53 and 28% of the standard deviation) and the PCS of men by 2.53 points (5.5% relative to the mean of 46.22 and 26% of the standard deviation). There is no significant effect on the MCS and PCS of women.

[8] 25.7% of the observations in the sample had two boys first and 23.4% had two girls first.

[9] The bias of 2SLS is approximately: $E[\widehat{\beta}_{2SLS} - \beta] \approx \frac{\sigma_{u\varepsilon}}{\sigma_{\varepsilon}^2} \frac{1}{F+1}$ where $F$ is the first-stage F-statistic, $\sigma_{u\varepsilon}$ is the covariance between the error term in the first-stage and the error term in the second-stage and $\sigma_{\varepsilon}^2$ is the variance of the error term in the first-stage. If $F \to 0$, then the bias of 2SLS $\to \frac{\sigma_{u\varepsilon}}{\sigma_{\varepsilon}^2}$ which is close to the bias of OLS.

Table 5.2: Regression results

| Sample | BMI | | MCS & PCS | | BMI | | MCS & PCS | |
|---|---|---|---|---|---|---|---|---|
| | First-stage: *Dependent variable - 3 or more children* | | | | | | | |
| | Women | Men | Women | Men | Women | Men | Women | Men |
| Twins at second birth | 0.604*** | 0.640*** | 0.600*** | 0.629*** | | | | |
| | (0.013) | (0.012) | (0.013) | (0.013) | | | | |
| Same sex | | | | | 0.018 | 0.039*** | 0.017 | 0.039*** |
| | | | | | (0.012) | (0.013) | (0.012) | (0.013) |
| F-statistic | 2236.59 | 2760.25 | 2233.78 | 2362.02 | 2.14 | 8.84 | 1.91 | 9.32 |
| Observations | 35,380 | 29,207 | 35,195 | 29,107 | 35,469 | 29,269 | 35,284 | 29,172 |

| | | OLS | | 2SLS | |
|---|---|---|---|---|---|
| | | Women | Men | Women | Men |
| *Dep. variable* - ***BMI***: | 3 or more children | 0.887*** | −0.006 | 1.876** | 0.769 |
| | | (0.125) | (0.110) | (0.829) | (0.756) |
| | Observations | 35,380 | 29,207 | 35,380 | 29,207 |
| *Dep. variable* - ***MCS***: | 3 or more children | −0.643*** | −0.268 | −0.683 | −2.727** |
| | | (0.198) | (0.208) | (1.593) | (1.307) |
| | Observations | 35,195 | 29,107 | 35,195 | 29,107 |
| *Dep. variable* - ***PCS***: | 3 or more children | −1.288*** | −0.625*** | −0.879 | −2.529** |
| | | (0.210) | (0.219) | (1.445) | (1.210) |
| | Observations | 35,195 | 29,107 | 35,195 | 29,107 |

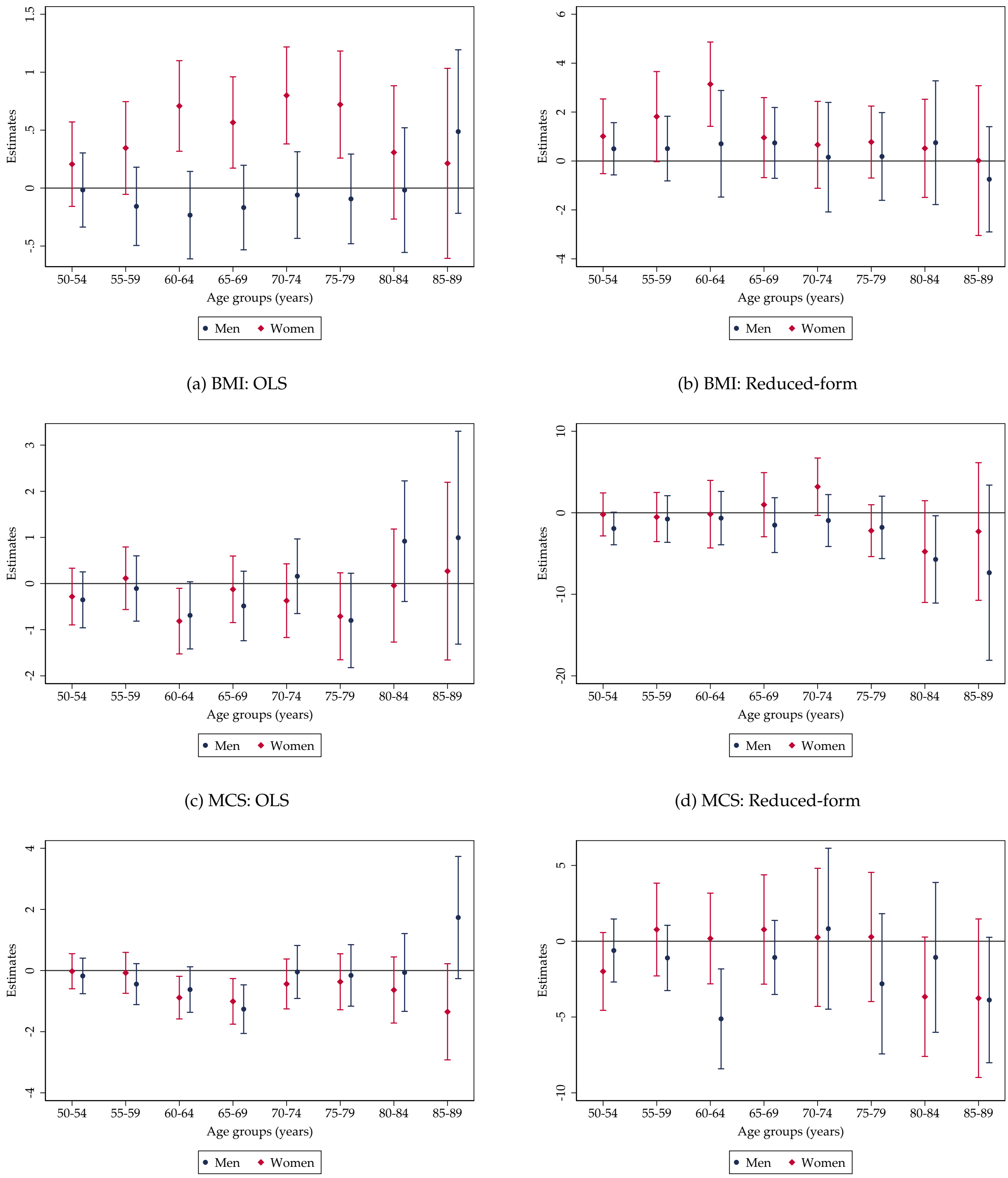
*Note:* Own calculations based on SOEP. All regressions include age squared, and indicators for year of birth, migration background, West Germany and survey years. Robust standard errors clustered at individual level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

### 5.5.2 Life-cycle Effects

I then look at the changes in the associations and effects over different five-year age groups. Figure 5.2 displays the OLS results, which depict the association between having more than two children and the health outcomes over the age groups, in the left panel. The OLS results only provide a descriptive relationship and cannot be interpreted as causal effects. For the causal effect, I consider the reduced-form results in the right panel, which depict the effect of having twin birth at second birth on the health outcomes over the age-groups. The estimates are based on Eq. 5.3 except that, for the reduced-form estimates $m2kids \times grp$ is substituted with $twinsecond \times grp$.

Figure 5.2a indicates that women with more than two children have an increasing BMI compared with women with two children until age 64 when it stabilises and begins to decline from age 80. The associations are only statistically significant for women ages 60 to 79 when the increase is relatively constant. No significant association between having more than two children and BMI is found for men. Looking at the effect of twin birth at second birth on BMI in Figure 5.2b, the increase until age 64 is still evident but starts to decline immediately contrary to the results in Figure 5.2a. The size of the effects until age 64 are larger than the corresponding OLS results. The largest and only statistically significantly increase of 3.14 points (12% relative to the age group mean of 26.92) is in the 60-64 age group. Just as with the OLS results, no effects are found for men. The

Figure 5.2: Regressions over age groups




(a) BMI: OLS

(b) BMI: Reduced-form

(c) MCS: OLS

(d) MCS: Reduced-form

(e) PCS: OLS

(f) PCS: Reduced-form

*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\theta_j$ based on OLS and reduced-form versions of Eq. (5.3). All regressions include quadratic age at first birth, and indicators for the age groups, year of birth, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

increase in BMI until age 64 could be associated with pre-retirement conditions.[10] There is some evidence indicating that health deteriorates faster prior to retirement but slows

[10]The mandatory retirement age in Germany is 65 years.

down after retirement. Westerlund et al. (2009) find that suboptimal self-perceived health increased faster before retirement rather than after, more so for women than men, and van Zon et al. (2016) find that prior to retirement, average limitations in mobility and large muscle functions increased. Therefore, the results give an indication that having to care for more than two children adds an extra toll on the already deteriorating health before retirement, which is further exaggerated by twin birth.

Figure 5.2c shows that the difference in MCS is only statistically significant for women in the 60-64 age group. No significant associations are found for men. In Figure 5.2d, the effect of twin birth on mental health is close to zero until age 70 for women and 75 for men. MCS declines for both men and women from age 75. However, there is only a statistically significant effect found for men in the 80-84 age group of 5.72 (11% relative to the mean of 52.91). A possible explanation for the decline in mental health later in life is critical events experienced earlier in life. Kruk and Reinhold (2014) find some evidence to support their hypothesis that by raising the risk of experiencing critical periods – periods of stress, financial hardships and poor health status – earlier in life, children affect their parents' mental health in old age. Having to care for two children instead of one at the same time, in addition to other children, definitely increases the risk of experiencing such critical periods.

The trend in the association between PCS and having more than two children is similar for both men and women until age 85 as shown in Figure 5.2e. There is almost no difference in the PCS of women with more than two children and women with two children until age 60 when the physical health of women with more than two children decreases. The decrease is constant until age 69 after which it increases, though still negative, remains relatively constant and decreases again at age 85. For men with more than two children, physical health declines until age 69 compared to men with two children. Afterwards, there is almost no difference in physical health until age 85 when the physical health of men with more than two children improves. The associations are, however, only statistically significant for women ages 60 to 69 and men ages 65 to 69. Figure 5.2f indicates that the effect of twin birth on PCS of women is almost non-existent for most age groups. There is a decline in age groups 50-54, 80-84 and 85-89 but they are statistically insignificant. Although the effect on physical health of men fluctuates, it is mostly negative for most age groups. There is only a statistically significant decline in the 60-64 age groups of men. This could also be associated with pre-retirement conditions.

## 5.6 Robustness Checks

To address the threats associated with using multiple births as instruments as mentioned in Section 5.2, I use same-sex twins as the instrument instead of twin birth at second birth and exclude individuals who had any birth from 1990 onwards. I do so in three ways: *(i)* individuals with mixed-sex twins are excluded from the treatment, *(ii)* individuals who had a child from 1990 onwards are excluded from the main sample, and *(iii)* a combination of *(i)* and *(ii)*, where individuals with mixed-sex twins are excluded from the treatment and individuals who had a child from 1990 onwards are excluded from the main sample. Excluding individuals who had a child from 1990 onwards as well as using the same-sex twins instrument should exclude as much non-random selection into twinning as a result of fertility treatments.

Table 5.3: First-stage results

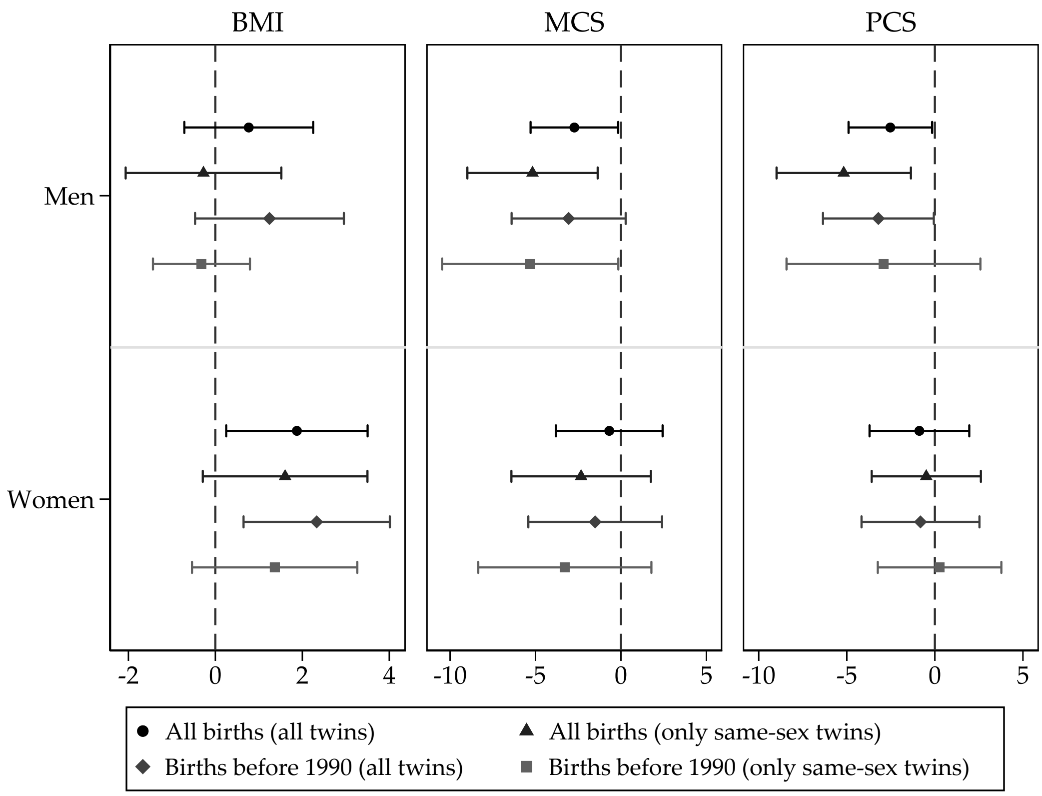
| Sample | BMI | | MCS & PCS | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| *All births* | | | | |
| Same-sex twins | 0.589*** | 0.626*** | 0.587*** | 0.613*** |
| | (0.016) | (0.015) | (0.016) | (0.016) |
| F-statistic | 1344.77 | 1753.05 | 1371.54 | 1488.46 |
| Observations | 35,380 | 29,199 | 35,195 | 29,099 |
| *Births before 1990 - with twins at second birth instrument* | | | | |
| Twins at second birth | 0.627*** | 0.701*** | 0.627*** | 0.705*** |
| | (0.018) | (0.020) | (0.018) | (0.020) |
| F-statistic | 1232.26 | 1243.74 | 1206.38 | 1297.40 |
| Observations | 28,757 | 20,881 | 28,235 | 20,360 |
| *Births before 1990 - with same-sex instrument* | | | | |
| Same-sex twins | 0.608*** | 0.685*** | 0.608*** | 0.690*** |
| | (0.023) | (0.027) | (0.023) | (0.026) |
| F-statistic | 709.26 | 658.50 | 710.88 | 722.24 |
| Observations | 28,757 | 20,873 | 28,235 | 20,352 |

*Note:* Own calculations based on SOEP. All regressions include age squared, and indicators for year of birth, West Germany and survey years. Robust standard errors clustered at individual level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The first-stage results presented in Table 5.3 show that the same-sex twin is a strong instrument regardless of whether the sample is restricted to only births before 1990 or not. Twins at second birth is still strong even in the restricted sample.

In Figure 5.3, the second-stage results of the three robustness checks are presented along with the main results from Table 5.2. The robustness checks still yield no significant effects on the BMI of men, and the mental and physical health of women as in the main results. The direction of the significant effects found in Table 5.2 are found to be robust. The increase in BMI of women is still evident but the effects are not significant and a bit smaller in size when only same-sex twins are used as treatment. As mentioned earlier, excluding dizygotic twinning from the treatment should exclude any selection bias, which could be the reason behind these results. Associations have been found

Figure 5.3: Comparing effects - second-stage results




*Note:* Own calculations based on SOEP. Point estimates of the coefficient $\rho$ of Eq. (5.2), where the instrument for all twins is twin birth at second birth and that for only same-sex twins is same-sex twins at second birth. All regressions include age squared, and indicators for year of birth, migration background, West Germany and survey years. Robust standard errors clustered at individual. 95% confidence intervals.

between mothers who had dizygotic twins and increased BMI before pregnancy but not for mothers who had monozygotic twins in different countries (Reddy et al., 2005; Hoekstra et al., 2010). However, little is known about the association with BMI after birth. If this relation between dizygotic twin mothers and increased BMI persists after birth, this could explain the statistically insignificant and smaller results found when mixed-sex twins are excluded from the treatment. Given that BMI of respondents are only available in SOEP from 2002 and majority of births considered were before this time, I am unable to assess the pre- and post-pregnancy BMI. The decline in the mental health of men also holds but is only statistically insignificant when individuals with births from 1990 onwards are excluded. It is robust when individuals with births from 1990 onwards and the same-sex twins instrument is used. Removing any selection bias resulting from dizygotic twinning increases the size of the effect on mental health. The decline in physical health of men is robust in all checks except when individuals with births from 1990 onwards and individuals with different sex twins are excluded.

The robustness checks are also done for the life-cycle effects which are presented in Figures A5.2, A5.3 and A5.4. The increasing BMI of women until age 64 is robust under all specifications with dizygotic twinning hardly biasing the results. The significant effects found in the 60-64 age group of women for BMI and in the 80-84 age group of men for mental health are very robust in terms of the direction of the effects but the size varies especially for mental health. Some bias from dizygotic twinning is seen in the results for MCS. There are some fluctuations in the mental health of women over the life-cycle with a significant effect showing up in age group 75-79. The size of the significant effect found amongst men in age group 80-84 increases. The significant effect on physical health of men found in age group 60-64 loses its significance when only same-sex twins are used as treatment. There is improved physical health amongst women in age groups 55-59 and 65-69, when as much non-randomness is removed by excluding individuals with births from 1990 onwards and the same-sex twins instrument is used.

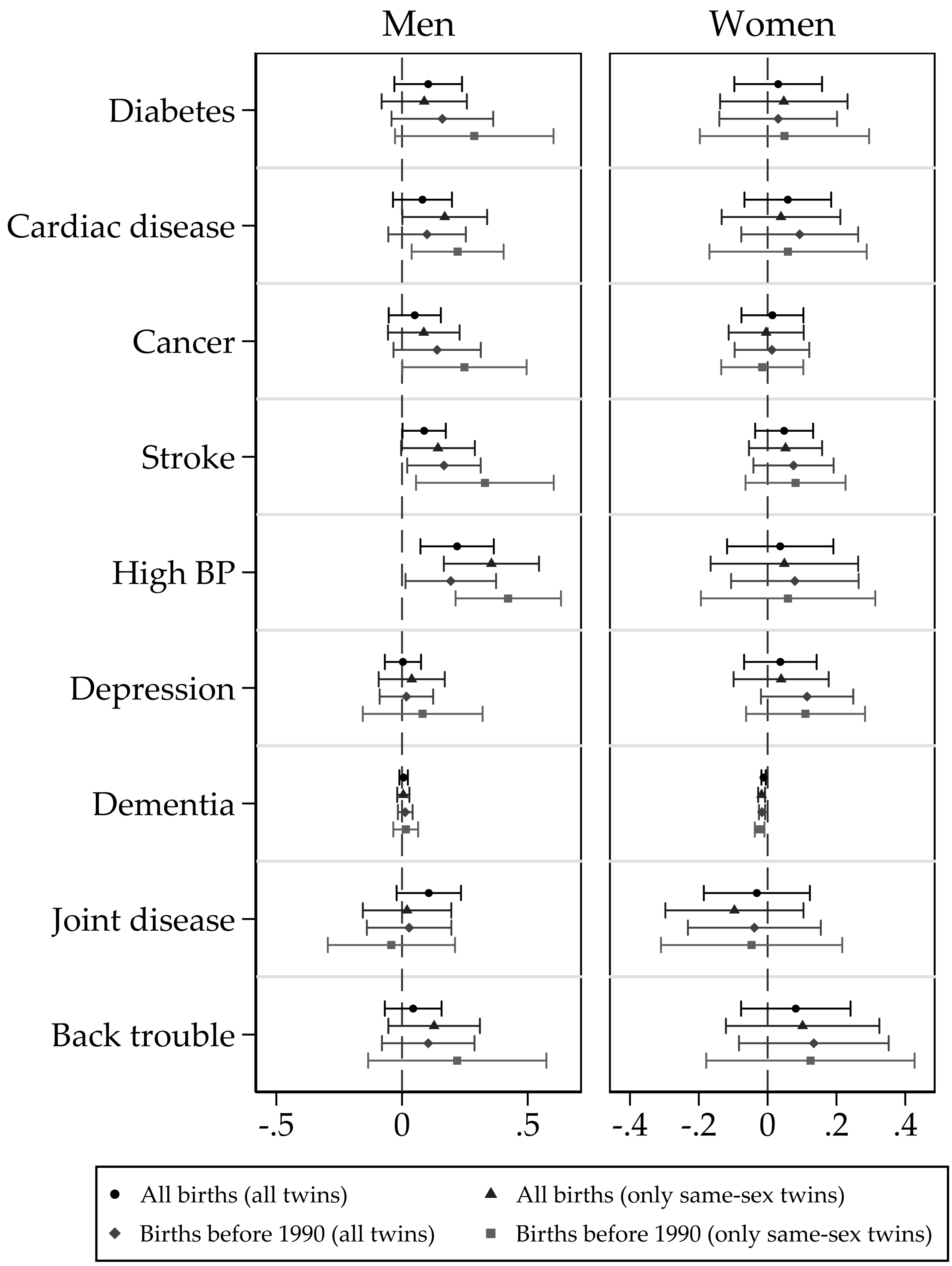
## 5.7 Effect on Various Illnesses

Given that BMI has been associated with diseases such as diabetes and heart related diseases, and poorer mental and physical health could also be associated with certain diseases such as depression and bodily pain, in this section I analyse the effect of having children on various illnesses. In the SOEP, respondents are asked whether they have ever been diagnosed with certain illnesses.[11] The illnesses considered in the analyses are diabetes, cardiac disease, cancer, stroke, high blood pressure, depression, dementia, joint disease and back trouble. Each illness is presented as a dummy variable which takes the value 1 if the specific illness has ever been diagnosed and 0 otherwise. The results are estimated using 2SLS as discussed in Section 5.4 with the illnesses as the outcome variable. For each group (men and women), I run four different analyses for each illness using the twins at second birth and the same-sex twins as instruments (not simultaneously) for the all births sample and births before 1990 sample. The first-stage results are presented in Table A5.7 in the Appendix. The instruments are strong in each sample with an effect size ranging from 0.6 to 0.7 for both instruments.

For women, there is no effect of having more than three children due to having twins at second birth on the likelihood of being diagnosed with any of the illness except for dementia, where there is a slight decrease in the likelihood by less than 5 percentage points. Although there is an effect on BMI for women, this does not necessarily translate to diabetes and heart related diseases as mentioned in Subsection 5.3.2. Contrary to Kruk and Reinhold (2014), I do not find any significant effect of having a third child given that the second birth was twin birth on depression amongst women. Cáceres-Delpiano and Simonsen (2012) find an increase in the likelihood of suffering from high

[11] The question asked in SOEP is "Has a doctor ever diagnosed you to have one or more of the following illnesses?: Sleep disorder, Diabetes, Cardiac disease (also cardiac insufficiency, weak heart), Cancer, Stroke, Migraine, High blood pressure, Depression, Dementia, joint diseases (including arthritis, rheumatism), Chronic back trouble or Other illness"

Figure 5.4: Various illnesses




*Note:* Own calculations based on SOEP. Point estimates of the coefficient $\rho$ of Eq. (5.2), where the instrument for all twins is twin birth at second birth and that for only same-sex twins is same-sex twins at second birth. All regressions include age squared, and indicators for the year of birth, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

blood pressure for younger women ages 20 to 45, but I do not find any such effects for women 50 years and older. It should, however, be noted that the definition of the variable used in Cáceres-Delpiano and Simonsen (2012) is a dummy variable indicating whether a mother experienced high blood pressure during the previous twelve months, which is different from what is used here. In spite of the fact that different age groups are considered, both results cannot be compared given that the variable used in Cáceres-Delpiano and Simonsen (2012) only considers the previous year. The variable does not account for the possibility of a mother experiencing high blood pressure years prior. For men, on the other hand, there are positive significant effects on the likelihood of being diagnosed with stroke and high blood pressure regardless of not finding any significant effects on BMI. Men with same-sex twins also have a higher risk of being diagnosed with cardiac disease. The decline in mental health found amongst men does not seem to lead to mental health diseases such as dementia and depression. This is, however, in line with Kruk and Reinhold (2014) who do not find any effect on depression amongst men. Although a decline in the physical health of men is found, it does not reflect in joint diseases and back trouble.

## 5.8 Attrition

As mentioned earlier, attrition can potentially bias the results of health analysis when using longitudinal household surveys, especially if the focus is on older individuals (Deng et al., 2013; Fichera and Savage, 2015). Although Andersen et al. (2007) indicate that the health status of respondents has only a marginal effect on attrition, I check whether having children could lead to selective attrition, probably due to mortality. Attrition prevents a sample from being a representation of the target population and, thus, presents potentially substantial biases to statistical inferences Deng et al. (2013).

To check whether attrition could be a problem, I use the BMI sample and generate a dummy variable *attrition*, which equals 1 if a respondent does not appear in the next survey year and 0 if the respondent either appears in the next wave or if it is the last survey year (i.e. 2020). The dummy variable is generated prior to the sample selection based on the age. 29% of the respondents in the sample drop out of the survey at some point in time. I use *attrition* as the outcome variable for the OLS and 2SLS regressions as done for the aggregate effects, and also for the OLS and reduced-form regressions as done in the life-cycle effects. The results from the aggregate effects presented in Table A5.8 show that having more than two children does not have any significant relation with attrition. From the life-cycle analyses depicted in Figure A5.5, the OLS results indicate that there is no significant relation between having more than two children and attrition over the life-cycle. However, the reduced-form results show that there is a lower likelihood for women in age group 75-79, who had twins at second birth, to stay in the survey. This implies that the estimated effects for this age group may be biased due to selective attrition and need to be interpreted with caution. For the rest of the results, attrition is not a problem.

## 5.9 Conclusion

This paper contributes to the scarce literature on the causal analysis of the aggregate effect of children on health, and also introduces life-cycle effects. Using the German Socio-Economic Panel data I analyse the effect of children on Body Mass Index (BMI), mental and physical health. Men and women from ages 50 to 90 are considered. I use 2SLS to estimate the aggregate effects and a reduced-form model to estimate effects over the life-cycle. I use twin birth at second birth and same sex of first two births as instruments for having three or more children in a sample with each individual having at least two children. Unlike twin birth at second birth, same sex of first two births is found to be a weak instrument for having more than two children in the case of Germany. Therefore, the same-sex instrument was not used for any further analyses. I find an increased BMI for women who had a third child as result of having twins at second birth. I also find a decrease in mental health and physical health of men. I find an upward trend in the BMI of women until age 64 as a result of twin birth at second birth. This upward trend may be attributed to pre-retirement conditions. Mental health starts to decline from age 75 but a significant effect is only found for men in age group 80-84. This decline can be linked to long-term effects of critical periods experienced earlier in life. Physical health fluctuates over the years especially for men.

Addressing possible validity issues with twin birth instrument, I use same-sex twins at second birth as the treatment instead of twins at second birth, exclude individuals who had any births from 1990 onwards, and a combination of the two i.e. using same-sex twins as instrument and excluding those with births from 1990 onwards. The baseline aggregate results are robust with respect to the direction of the effects. However, there is some variation in the size of the effects and the level of statistical significance. The significant effects found in the life-cycle effects are robust except that of physical health of men. Excluding mixed-sex twins from the treatment eliminates some bias in the results.

Since the health outcomes considered are mainly general measures of health, I check whether the poor health found as a result of having more than two children reflect in the likelihood of being diagnosed with specific illnesses. For women, I do not find any effects on the likelihood of being diagnosed with any of the illnesses except a slight decrease in the likelihood of being diagnosed with dementia. On the other hand, men with more than two children are more likely to be diagnosed with a cardiac disease, stroke and high blood pressure.

In conclusion, having an additional child as a result of twins does have an impact on parental health. Hence, women should be educated on their likelihood of being obese as a result of having more children and how to curb it. Attention should be paid to the mental health of individuals, specifically men, who experience certain stressful periods including poor health, financial crises in younger years while having and raising children, as these periods are very likely to have long-term effects on mental health. Men should also pay attention to the health of their hearts. All in all, both men and women should pay attention to their health especially as the size of the family increases. I would like to note, however, that the results allow for the identification of effects on a small population of compliers. This makes it difficult to draw a general conclusion from the results. Nonetheless, it is still important to draw attention to the possible effects children have on parental health. Since twins add an extra effect due to the unexpected additional child, the effect sizes found are likely to be the upper bounds of the effect of children on parental health.

# Appendix

Table A5.1: First-stage results: Boys vs Girls first

| | BMI | | MCS & PCS | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| Boys first | 0.032** | 0.030** | 0.032** | 0.030** |
| | (0.014) | (0.015) | (0.014) | (0.015) |
| F-statistic | 5.02 | 3.99 | 5.23 | 4.08 |
| Observation | 35,469 | 29,269 | 35,284 | 29,172 |
| Girls first | -0.009 | 0.022 | -0.010 | 0.022 |
| | (0.014) | (0.015) | (0.014) | (0.015) |
| F-statistic | 0.35 | 1.94 | 0.54 | 2.15 |
| Observation | 35,469 | 29,269 | 35,284 | 29,172 |

*Note:* Own calculations based on SOEP. All births are considered. All regressions include age squared and indicators for birth year, migration background, West Germany and survey years. Robust standard errors clustered at individual level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.2: Descriptive statistics for births before 1990 sample.

| | *Two children* | | *Three or more children* | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| BMI | 26.501 | 27.559 | 27.476 | 27.486 |
| | (4.800) | (3.844) | (5.105) | (4.149) |
| MCS | 50.752 | 52.947 | 49.895 | 52.899 |
| | (10.273) | (9.520) | (10.854) | (10.074) |
| PCS | 44.341 | 45.620 | 41.881 | 44.175 |
| | (10.128) | (9.604) | (10.422) | (9.992) |
| Number of children | 2.000 | 2.000 | 3.535 | 3.409 |
| | (0.000) | (0.000) | (0.943) | (0.762) |
| Age | 65.369 | 65.740 | 68.307 | 68.410 |
| | (9.269) | (8.940) | (9.654) | (9.240) |
| Age at first first birth | 24.057 | 27.024 | 22.499 | 25.861 |
| | (4.031) | (4.532) | (3.546) | (4.426) |
| Migration background | 0.106 | 0.092 | 0.147 | 0.141 |
| | (0.308) | (0.290) | (0.354) | (0.348) |
| West Germany | 0.683 | 0.703 | 0.731 | 0.794 |
| | (0.465) | (0.457) | (0.443) | (0.405) |
| Twins | 0.000 | 0.000 | 0.042 | 0.051 |
| | (0.000) | (0.000) | (0.201) | (0.219) |
| Observations | 18,445 | 14,331 | 10,312 | 6,550 |

*Note:* Own calculations based on SOEP. Statistics of all variables except MCS and PCS are based on the BMI sample. Standard deviations are in parentheses.

Table A5.3: Summary statistics: Life-cycle analysis - All births

***Women***

| Age groups: | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75-79 | 80-84 | 85-89 |
|---|---|---|---|---|---|---|---|---|
| BMI | 26.291 | 26.750 | 26.918 | 27.190 | 27.036 | 26.901 | 26.715 | 25.949 |
| | (5.346) | (5.264) | (5.055) | (4.919) | (4.713) | (4.773) | (4.796) | (4.845) |
| MCS | 49.453 | 49.883 | 50.831 | 51.755 | 51.141 | 50.423 | 49.606 | 48.140 |
| | (10.285) | (10.309) | (10.236) | (10.021) | (10.344) | (10.945) | (11.160) | (12.070) |
| PCS | 48.405 | 46.522 | 44.930 | 43.519 | 41.958 | 40.113 | 38.001 | 36.402 |
| | (9.671) | (10.071) | (9.855) | (9.985) | (10.004) | (10.000) | (9.637) | (9.235) |
| Number of children | 2.558 | 2.510 | 2.516 | 2.561 | 2.635 | 2.748 | 2.826 | 2.835 |
| | (0.900) | (0.881) | (0.903) | (0.944) | (1.008) | (1.077) | (1.133) | (1.140) |
| Age | 51.942 | 56.926 | 61.979 | 66.946 | 71.892 | 76.870 | 81.704 | 86.574 |
| | (1.417) | (1.415) | (1.415) | (1.406) | (1.409) | (1.406) | (1.381) | (1.371) |
| Age at first birth | 24.933 | 24.057 | 23.633 | 23.608 | 23.902 | 24.375 | 24.640 | 25.076 |
| | (4.993) | (4.653) | (4.290) | (4.030) | (4.041) | (4.028) | (3.958) | (4.126) |
| Migration background | 0.172 | 0.161 | 0.141 | 0.121 | 0.094 | 0.091 | 0.077 | 0.078 |
| | (0.378) | (0.368) | (0.348) | (0.326) | (0.292) | (0.288) | (0.266) | (0.269) |
| West Germany | 0.744 | 0.725 | 0.706 | 0.716 | 0.723 | 0.733 | 0.724 | 0.743 |
| | (0.437) | (0.447) | (0.456) | (0.451) | (0.448) | (0.443) | (0.447) | (0.437) |
| Twins | 0.022 | 0.014 | 0.013 | 0.013 | 0.013 | 0.024 | 0.027 | 0.023 |
| | (0.147) | (0.118) | (0.113) | (0.112) | (0.113) | (0.154) | (0.163) | (0.151) |
| Twins at second birth | 0.013 | 0.009 | 0.008 | 0.006 | 0.008 | 0.013 | 0.016 | 0.015 |
| | (0.115) | (0.093) | (0.091) | (0.080) | (0.091) | (0.114) | (0.125) | (0.122) |
| Same-sex twins | 0.008 | 0.004 | 0.004 | 0.004 | 0.005 | 0.010 | 0.014 | 0.014 |
| | (0.091) | (0.063) | (0.066) | (0.064) | (0.072) | (0.098) | (0.117) | (0.118) |
| Observations | 7,376 | 6,211 | 5,734 | 5,327 | 4,445 | 3,322 | 2,010 | 857 |

***Men***

| Age groups: | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75-79 | 80-84 | 85-89 |
|---|---|---|---|---|---|---|---|---|
| BMI | 27.461 | 27.723 | 27.873 | 27.760 | 27.560 | 26.985 | 26.631 | 25.711 |
| | (4.275) | (4.197) | (4.309) | (3.991) | (3.793) | (3.464) | (3.571) | (3.134) |
| MCS | 51.076 | 51.221 | 52.677 | 54.359 | 54.225 | 53.404 | 52.912 | 51.017 |
| | (9.563) | (9.892) | (9.721) | (9.212) | (9.182) | (9.828) | (9.967) | (11.567) |
| PCS | 49.362 | 47.665 | 46.213 | 45.432 | 44.294 | 43.205 | 41.313 | 38.183 |
| | (9.273) | (9.676) | (9.701) | (9.548) | (9.721) | (9.664) | (9.468) | (9.310) |
| Number of children | 2.536 | 2.515 | 2.500 | 2.508 | 2.522 | 2.576 | 2.614 | 2.755 |
| | (0.847) | (0.878) | (0.889) | (0.885) | (0.851) | (0.875) | (0.857) | (0.897) |
| Age | 51.971 | 56.932 | 61.974 | 66.961 | 71.890 | 76.827 | 81.660 | 86.533 |
| | (1.404) | (1.418) | (1.410) | (1.408) | (1.405) | (1.402) | (1.355) | (1.387) |
| Age at first birth | 28.292 | 27.805 | 27.457 | 27.420 | 27.543 | 27.798 | 28.264 | 28.858 |
| | (5.412) | (5.583) | (5.353) | (5.063) | (4.817) | (4.950) | (5.030) | (5.437) |
| Migration background | 0.174 | 0.163 | 0.145 | 0.120 | 0.091 | 0.073 | 0.071 | 0.074 |
| | (0.379) | (0.369) | (0.352) | (0.325) | (0.288) | (0.260) | (0.257) | (0.262) |
| West Germany | 0.768 | 0.746 | 0.736 | 0.751 | 0.776 | 0.795 | 0.819 | 0.846 |
| | (0.422) | (0.435) | (0.441) | (0.433) | (0.417) | (0.404) | (0.385) | (0.362) |
| Twins | 0.025 | 0.021 | 0.016 | 0.016 | 0.014 | 0.023 | 0.030 | 0.029 |
| | (0.156) | (0.142) | (0.126) | (0.126) | (0.116) | (0.150) | (0.171) | (0.167) |
| Twins at second birth | 0.018 | 0.014 | 0.011 | 0.010 | 0.007 | 0.012 | 0.014 | 0.014 |
| | (0.133) | (0.119) | (0.103) | (0.100) | (0.084) | (0.107) | (0.117) | (0.119) |
| Same-sex twins | 0.012 | 0.008 | 0.005 | 0.005 | 0.004 | 0.005 | 0.008 | 0.012 |
| | (0.109) | (0.086) | (0.067) | (0.070) | (0.062) | (0.071) | (0.090) | (0.111) |
| Observations | 6,347 | 5,455 | 4,864 | 4,431 | 3,624 | 2,602 | 1,360 | 486 |

*Note:* Own calculations based on SOEP. Statistics of all variables except MCS and PCS are based on the BMI sample. Standard deviations are in parentheses.

Table A5.4: Summary statistics: Life-cycle analysis - Births before 1990

***Women***

| Age groups: | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75-79 | 80-84 | 85-89 |
|---|---|---|---|---|---|---|---|---|
| BMI | 26.381 | 26.755 | 26.956 | 27.181 | 27.033 | 26.900 | 26.721 | 25.962 |
| | (5.126) | (5.064) | (5.040) | (4.901) | (4.713) | (4.776) | (4.796) | (4.844) |
| MCS | 49.019 | 49.732 | 50.856 | 51.750 | 51.148 | 50.425 | 49.594 | 48.115 |
| | (10.290) | (10.293) | (10.233) | (10.046) | (10.323) | (10.948) | (11.160) | (12.075) |
| PCS | 48.084 | 46.222 | 44.873 | 43.499 | 41.969 | 40.084 | 37.977 | 36.360 |
| | (9.508) | (9.966) | (9.744) | (9.976) | (10.000) | (9.985) | (9.625) | (9.200) |
| Number of children | 2.364 | 2.385 | 2.459 | 2.542 | 2.632 | 2.750 | 2.828 | 2.837 |
| | (0.693) | (0.745) | (0.837) | (0.918) | (0.998) | (1.078) | (1.134) | (1.140) |
| Age | 52.134 | 57.057 | 62.045 | 66.961 | 71.892 | 76.869 | 81.705 | 86.575 |
| | (1.419) | (1.412) | (1.413) | (1.405) | (1.409) | (1.406) | (1.381) | (1.372) |
| Age at first birth | 22.432 | 22.739 | 23.164 | 23.535 | 23.902 | 24.374 | 24.640 | 25.076 |
| | (3.490) | (3.687) | (3.852) | (3.938) | (4.035) | (4.027) | (3.958) | (4.126) |
| Migration background | 0.157 | 0.157 | 0.136 | 0.120 | 0.094 | 0.091 | 0.077 | 0.078 |
| | (0.364) | (0.363) | (0.343) | (0.325) | (0.292) | (0.288) | (0.266) | (0.269) |
| West Germany | 0.651 | 0.675 | 0.686 | 0.713 | 0.723 | 0.732 | 0.723 | 0.743 |
| | (0.477) | (0.468) | (0.464) | (0.452) | (0.448) | (0.443) | (0.447) | (0.437) |
| Twins | 0.016 | 0.009 | 0.011 | 0.012 | 0.013 | 0.024 | 0.027 | 0.023 |
| | (0.126) | (0.096) | (0.105) | (0.109) | (0.114) | (0.154) | (0.163) | (0.151) |
| Twins at second birth | 0.012 | 0.007 | 0.007 | 0.006 | 0.008 | 0.013 | 0.016 | 0.015 |
| | (0.110) | (0.082) | (0.085) | (0.078) | (0.091) | (0.114) | (0.125) | (0.122) |
| Same-sex twins | 0.007 | 0.003 | 0.003 | 0.004 | 0.005 | 0.010 | 0.014 | 0.014 |
| | (0.085) | (0.057) | (0.058) | (0.062) | (0.072) | (0.098) | (0.117) | (0.118) |
| Observations | 3,450 | 4,311 | 5,076 | 5,210 | 4,437 | 3,314 | 2,006 | 855 |

***Men***

| Age groups: | 50-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75-79 | 80-84 | 85-89 |
|---|---|---|---|---|---|---|---|---|
| BMI | 27.503 | 27.771 | 27.983 | 27.809 | 27.565 | 27.000 | 26.655 | 25.718 |
| | (3.994) | (4.027) | (4.320) | (3.961) | (3.785) | (3.472) | (3.581) | (3.146) |
| MCS | 50.723 | 51.114 | 52.806 | 54.484 | 54.227 | 53.415 | 52.875 | 50.941 |
| PCS | 48.557 | 47.050 | 45.935 | 45.370 | 44.269 | 43.178 | 41.284 | 38.147 |
| Number of children | 2.287 | 2.324 | 2.382 | 2.447 | 2.504 | 2.564 | 2.606 | 2.749 |
| | (0.602) | (0.656) | (0.731) | (0.809) | (0.841) | (0.865) | (0.854) | (0.897) |
| Age | 52.197 | 57.098 | 62.080 | 66.995 | 71.906 | 76.833 | 81.666 | 86.533 |
| | (1.385) | (1.408) | (1.408) | (1.404) | (1.406) | (1.401) | (1.356) | (1.389) |
| Age at first birth | 24.863 | 25.436 | 26.236 | 26.896 | 27.347 | 27.731 | 28.225 | 28.799 |
| | (3.779) | (4.034) | (4.264) | (4.441) | (4.524) | (4.810) | (4.970) | (5.343) |
| Migration background | 0.126 | 0.135 | 0.127 | 0.114 | 0.089 | 0.072 | 0.071 | 0.075 |
| | (0.331) | (0.342) | (0.333) | (0.317) | (0.285) | (0.258) | (0.258) | (0.263) |
| West Germany | 0.627 | 0.673 | 0.703 | 0.739 | 0.773 | 0.792 | 0.819 | 0.844 |
| | (0.484) | (0.469) | (0.457) | (0.439) | (0.419) | (0.406) | (0.385) | (0.363) |
| Twins | 0.016 | 0.012 | 0.011 | 0.016 | 0.013 | 0.022 | 0.030 | 0.029 |
| | (0.126) | (0.107) | (0.106) | (0.126) | (0.112) | (0.147) | (0.170) | (0.168) |
| Twins at second birth | 0.014 | 0.011 | 0.009 | 0.010 | 0.007 | 0.012 | 0.014 | 0.015 |
| | (0.119) | (0.102) | (0.095) | (0.100) | (0.081) | (0.108) | (0.118) | (0.120) |
| Same-sex twins | 0.009 | 0.006 | 0.003 | 0.005 | 0.003 | 0.005 | 0.008 | 0.012 |
| | (0.097) | (0.075) | (0.059) | (0.067) | (0.059) | (0.071) | (0.090) | (0.111) |
| Observations | 2,223 | 3,026 | 3,721 | 3,988 | 3,495 | 2,563 | 1,344 | 482 |

*Note:* Own calculations based on SOEP. Statistics of all variables except MCS and PCS are based on the BMI sample. Standard deviations are in parentheses.

Table A5.5: Descriptive Statistics: Various illnesses - All births

| | *All* | | *Two children* | | *Three or more children* | |
|---|---|---|---|---|---|---|
| | Women | Men | Women | Men | Women | Men |
| Diabetes | 0.118 | 0.149 | 0.099 | 0.141 | 0.150 | 0.165 |
| | (0.323) | (0.357) | (0.299) | (0.348) | (0.357) | (0.371) |
| Cardiac disease | 0.135 | 0.184 | 0.124 | 0.175 | 0.154 | 0.200 |
| | (0.342) | (0.387) | (0.329) | (0.380) | (0.361) | (0.400) |
| Cancer | 0.081 | 0.079 | 0.079 | 0.075 | 0.083 | 0.085 |
| | (0.272) | (0.269) | (0.270) | (0.264) | (0.276) | (0.278) |
| Stroke | 0.032 | 0.039 | 0.028 | 0.038 | 0.038 | 0.042 |
| | (0.175) | (0.194) | (0.164) | (0.190) | (0.191) | (0.200) |
| High Blood Pressure | 0.404 | 0.415 | 0.396 | 0.418 | 0.416 | 0.409 |
| | (0.491) | (0.493) | (0.489) | (0.493) | (0.493) | (0.492) |
| Depression | 0.097 | 0.055 | 0.094 | 0.052 | 0.103 | 0.061 |
| | (0.296) | (0.228) | (0.292) | (0.221) | (0.304) | (0.239) |
| Dementia | 0.006 | 0.004 | 0.005 | 0.003 | 0.007 | 0.007 |
| | (0.074) | (0.064) | (0.068) | (0.052) | (0.083) | (0.081) |
| Joint disease | 0.363 | 0.250 | 0.350 | 0.251 | 0.385 | 0.249 |
| | (0.481) | (0.433) | (0.477) | (0.434) | (0.487) | (0.432) |
| Back trouble | 0.246 | 0.210 | 0.236 | 0.206 | 0.262 | 0.219 |
| | (0.431) | (0.408) | (0.425) | (0.404) | (0.440) | (0.414) |
| No. of children | 2.558 | 2.506 | 2.000 | 2.000 | 3.502 | 3.430 |
| | (0.911) | (0.826) | (0.000) | (0.000) | (0.904) | (0.780) |
| Age | 64.354 | 64.089 | 63.958 | 63.933 | 65.024 | 64.375 |
| | (10.050) | (9.775) | (9.619) | (9.446) | (10.706) | (10.344) |
| Migration background | 0.118 | 0.120 | 0.107 | 0.107 | 0.136 | 0.144 |
| | (0.323) | (0.325) | (0.309) | (0.309) | (0.343) | (0.351) |
| West Germany | 0.725 | 0.755 | 0.710 | 0.727 | 0.752 | 0.805 |
| | (0.446) | (0.430) | (0.454) | (0.445) | (0.432) | (0.396) |
| Twins | 0.017 | 0.020 | 0.000 | 0.000 | 0.046 | 0.057 |
| | (0.129) | (0.140) | (0.000) | (0.000) | (0.209) | (0.231) |
| Twins at second birth | 0.010 | 0.013 | | | | |
| | (0.100) | (0.113) | | | | |
| Same-sex twins | 0.006 | 0.007 | | | | |
| | (0.080) | (0.082) | | | | |
| Three or more children | 0.371 | 0.354 | | | | |
| | (0.483) | (0.478) | | | | |
| Observations | 23,095 | 19,441 | 14,517 | 12,567 | 8,578 | 6,874 |
| Observations - Joint disease & Back trouble | 19,689 | 16,764 | 12,405 | 10,800 | 7,284 | 5,964 |

*Note:* Own calculations based on SOEP. Standard deviations are in parentheses.

Table A5.6: Descriptive Statistics: Various illnesses - Births before 1990

| | *All* | | *Two children* | | *Three or more children* | |
|---|---|---|---|---|---|---|
| | Women | Men | Women | Men | Women | Men |
| Diabetes | 0.138 | 0.180 | 0.114 | 0.169 | 0.186 | 0.208 |
| | (0.345) | (0.384) | (0.317) | (0.374) | (0.390) | (0.406) |
| Cardiac disease | 0.163 | 0.229 | 0.144 | 0.212 | 0.201 | 0.268 |
| | (0.370) | (0.420) | (0.351) | (0.409) | (0.401) | (0.443) |
| Cancer | 0.088 | 0.100 | 0.085 | 0.091 | 0.093 | 0.123 |
| | (0.283) | (0.301) | (0.279) | (0.287) | (0.290) | (0.329) |
| Stroke | 0.038 | 0.046 | 0.032 | 0.043 | 0.048 | 0.055 |
| | (0.190) | (0.210) | (0.177) | (0.202) | (0.214) | (0.227) |
| High Blood Pressure | 0.454 | 0.467 | 0.440 | 0.463 | 0.483 | 0.475 |
| | (0.498) | (0.499) | (0.496) | (0.499) | (0.500) | (0.499) |
| Depression | 0.091 | 0.044 | 0.090 | 0.041 | 0.093 | 0.051 |
| | (0.288) | (0.205) | (0.286) | (0.198) | (0.290) | (0.221) |
| Dementia | 0.007 | 0.006 | 0.006 | 0.004 | 0.009 | 0.010 |
| | (0.083) | (0.074) | (0.076) | (0.060) | (0.095) | (0.100) |
| Joint disease | 0.408 | 0.289 | 0.387 | 0.284 | 0.449 | 0.301 |
| | (0.491) | (0.453) | (0.487) | (0.451) | (0.497) | (0.459) |
| Back trouble | 0.268 | 0.226 | 0.251 | 0.224 | 0.303 | 0.232 |
| | (0.443) | (0.418) | (0.434) | (0.417) | (0.460) | (0.422) |
| No. of children | 2.498 | 2.407 | 2.000 | 2.000 | 3.477 | 3.370 |
| | (0.857) | (0.729) | (0.000) | (0.000) | (0.857) | (0.686) |
| Age | 67.532 | 68.446 | 66.427 | 67.446 | 69.704 | 70.816 |
| | (9.288) | (8.753) | (9.017) | (8.585) | (9.430) | (8.692) |
| Migration background | 0.105 | 0.094 | 0.096 | 0.085 | 0.123 | 0.116 |
| | (0.306) | (0.292) | (0.294) | (0.279) | (0.329) | (0.320) |
| West Germany | 0.690 | 0.710 | 0.674 | 0.682 | 0.722 | 0.775 |
| | (0.462) | (0.454) | (0.469) | (0.466) | (0.448) | (0.418) |
| Twins | 0.014 | 0.016 | 0.000 | 0.000 | 0.042 | 0.054 |
| | (0.118) | (0.125) | (0.000) | (0.000) | (0.201) | (0.225) |
| Twins at second birth | 0.009 | 0.010 | | | | |
| | (0.095) | (0.100) | | | | |
| Same-sex twins | 0.006 | 0.004 | | | | |
| | (0.076) | (0.067) | | | | |
| Three or more children | 0.337 | 0.297 | | | | |
| | (0.473) | (0.457) | | | | |
| Observations | 17,512 | 12,699 | 11,607 | 8,931 | 5,905 | 3,768 |
| Observations - Joint disease & Back trouble | 14,374 | 10,451 | 9,617 | 7,384 | 4,757 | 3,067 |

*Note:* Own calculations based on SOEP. Standard deviations are in parentheses.

Table A5.7: First-stage results: Various illnesses

| | All births | | Births before 1990 | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| ***Using twins at second birth instrument.*** | | | | |
| Twins at second birth | 0.605*** | 0.647*** | 0.636*** | 0.716*** |
| | (0.013) | (0.012) | (0.020) | (0.022) |
| F-statistic | 2159.63 | 2782.61 | 1047.69 | 1013.12 |
| Observations | 23,095 | 19,441 | 17,512 | 12,699 |
| ***Using same-sex twins instrument.*** | | | | |
| Same-sex twins | 0.585*** | 0.630*** | 0.605*** | 0.701*** |
| | (0.017) | (0.016) | (0.026) | (0.037) |
| F-statistic | 1246.09 | 1544.94 | 550.88 | 359.49 |
| Observations | 23,095 | 19,436 | 17,512 | 12,694 |

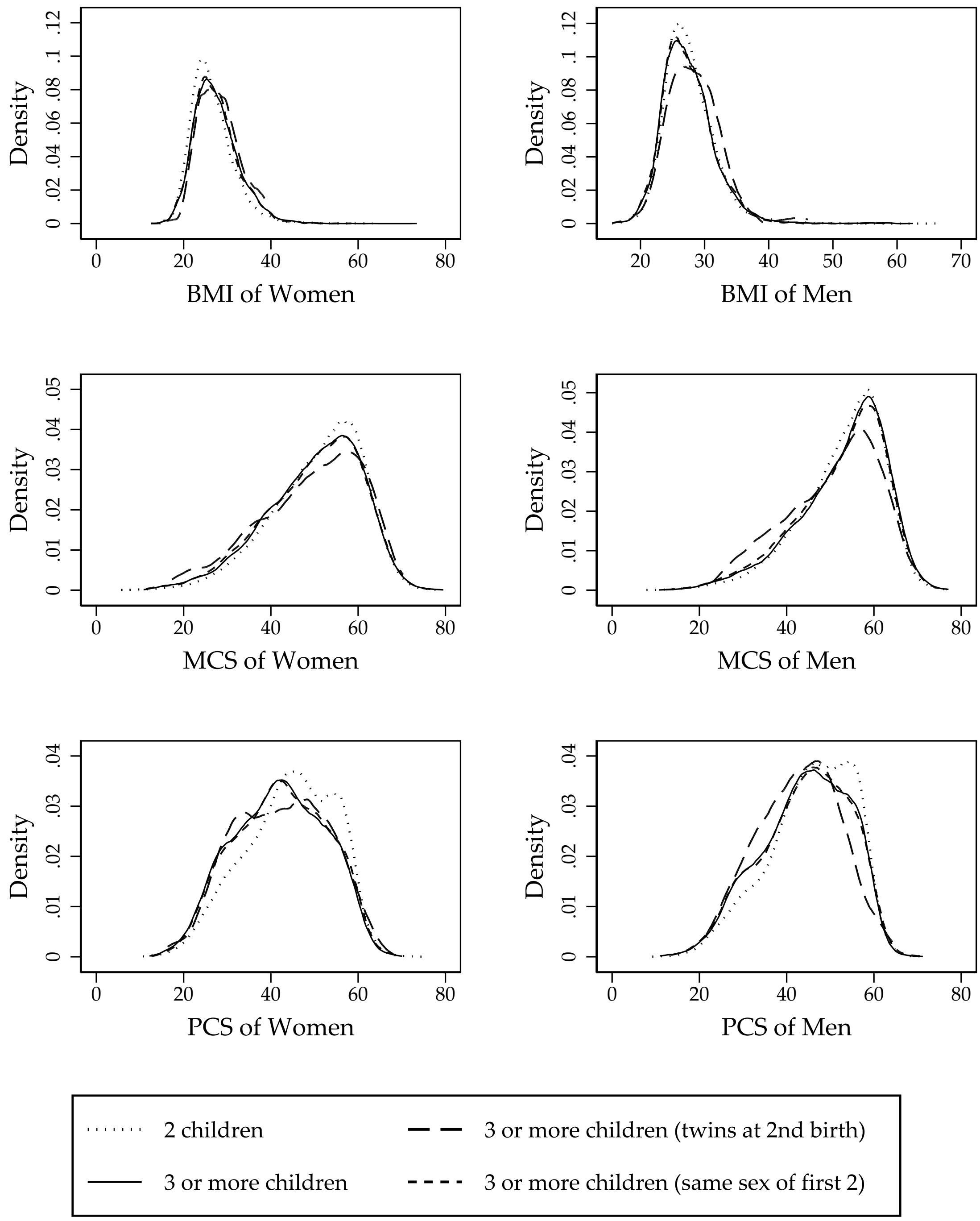
*Note:* Own calculations based on SOEP. All regressions include age squared and indicators for year of birth, migration background, West Germany and survey years. Robust standard errors clustered at individual level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5.8: Attrition - Aggregate results

| | OLS | | 2SLS | |
|---|---|---|---|---|
| | Women | Men | Women | Men |
| Three or more children | 0.004 | 0.006* | 0.001 | -0.002 |
| | (0.003) | (0.003) | (0.023) | (0.023) |
| Observations | 35,380 | 29,207 | 35,380 | 29,207 |

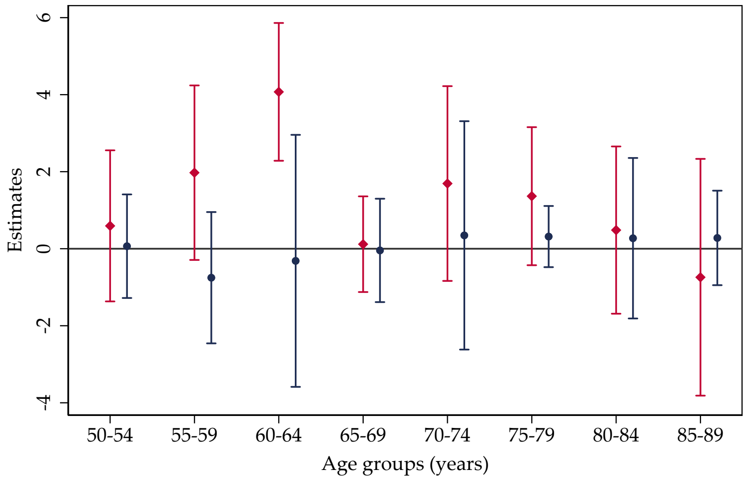
*Note:* Own calculations based on SOEP. All births are considered. This is based on the BMI sample. All regressions include age squared and indicators for birth year, migration background, West Germany and survey years. Robust standard errors clustered at individual level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A5.1: Kernel densities by the treatment status for each gender - only births before 1990.

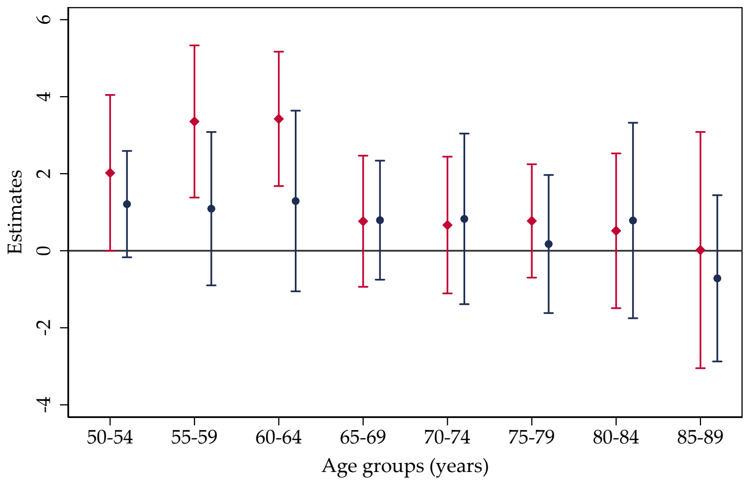



*Source:* Own calculations based on SOEP.

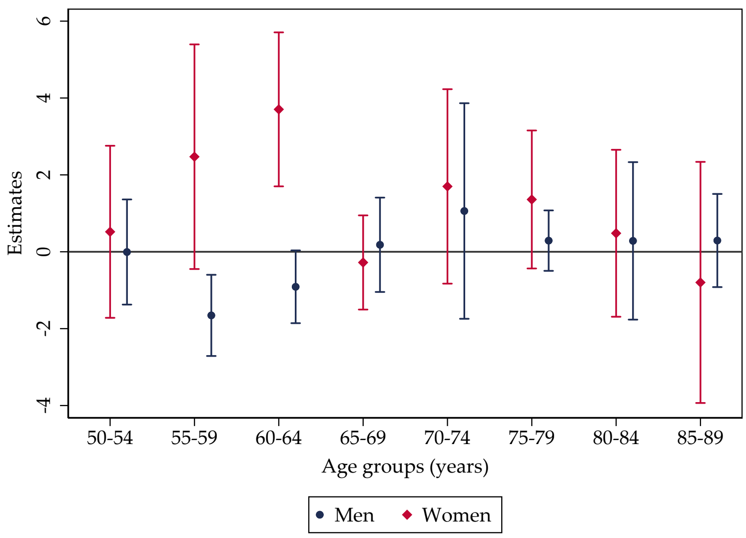
Figure A5.2: Effect over age groups - BMI




All births excluding mixed-sex twins from the treatment



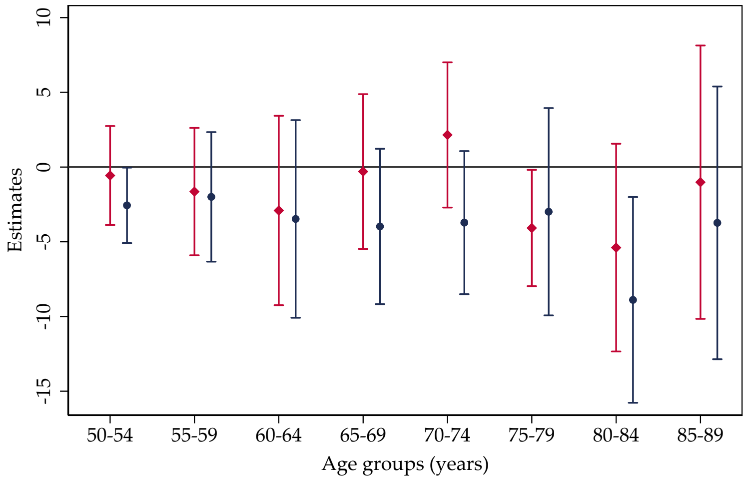

Births before 1990



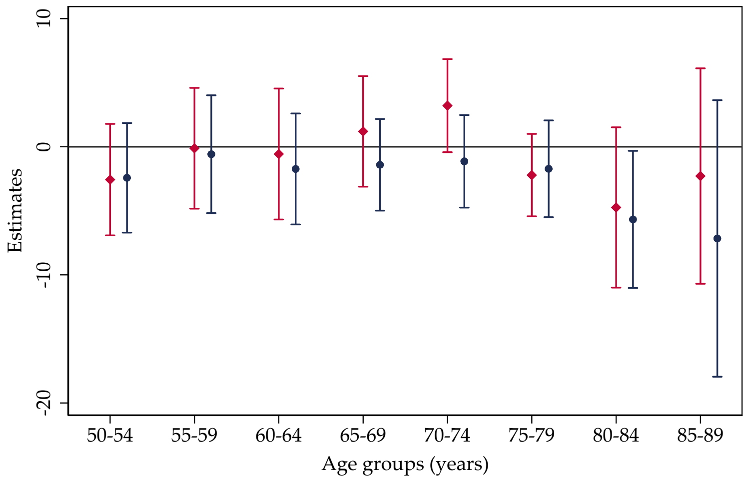

Births before 1990 excluding mixed-sex twins from the treatment

*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\theta_j$ based on OLS and reduced-form versions of Eq. (5.3). All regressions include quadratic age at first birth, and indicators for the age groups, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

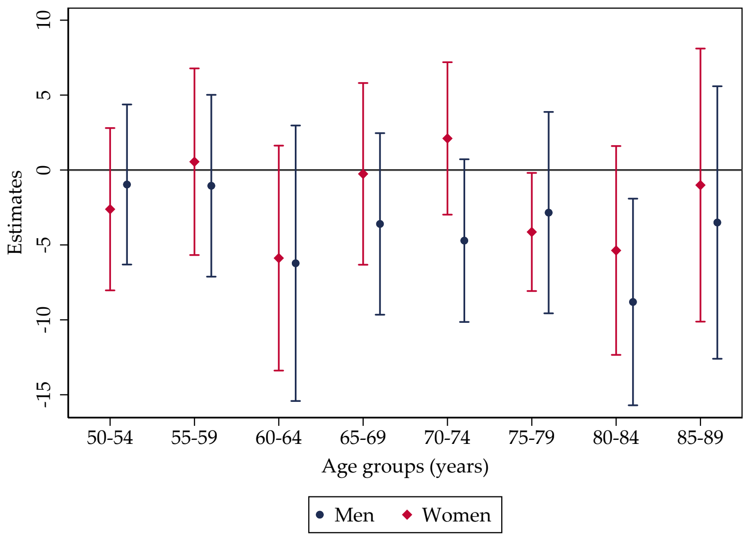
Figure A5.3: Effect over age groups - MCS




All births excluding mixed-sex twins from the treatment



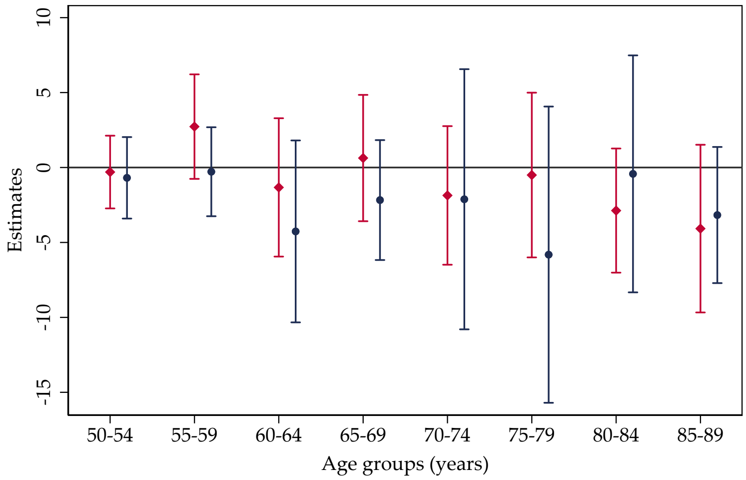

Births before 1990



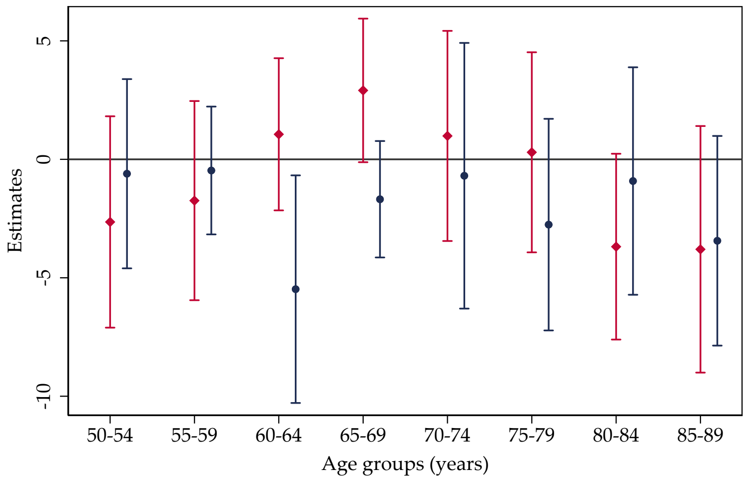

Births before 1990 excluding mixed-sex twins from the treatment

*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\theta_j$ based on OLS and reduced-form versions of Eq. (5.3). All regressions include quadratic age at first birth, and indicators for the age groups, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

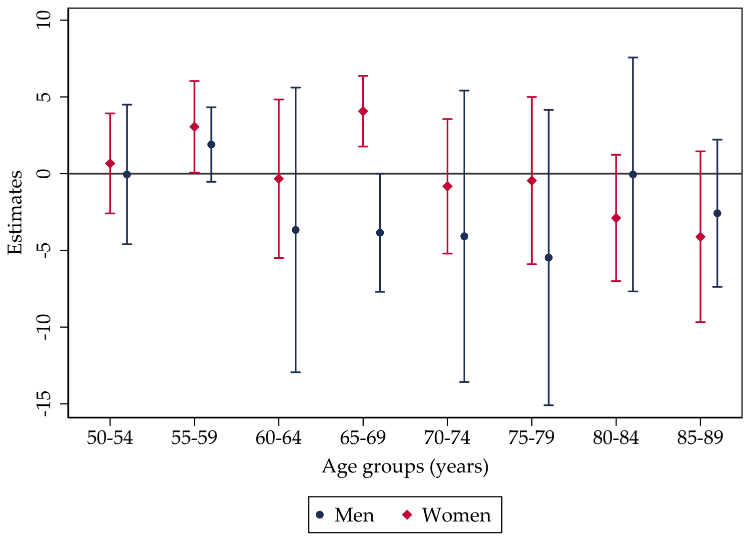
## Figure A5.4: Effect over age groups - PCS




All births excluding mixed-sex twins from the treatment



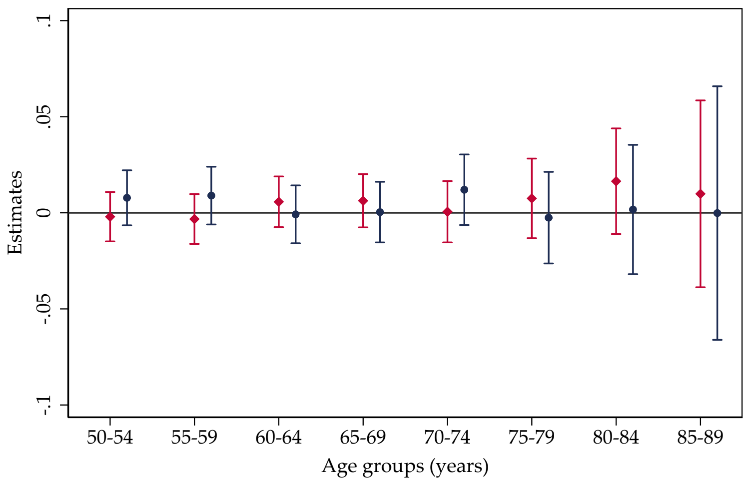

Births before 1990



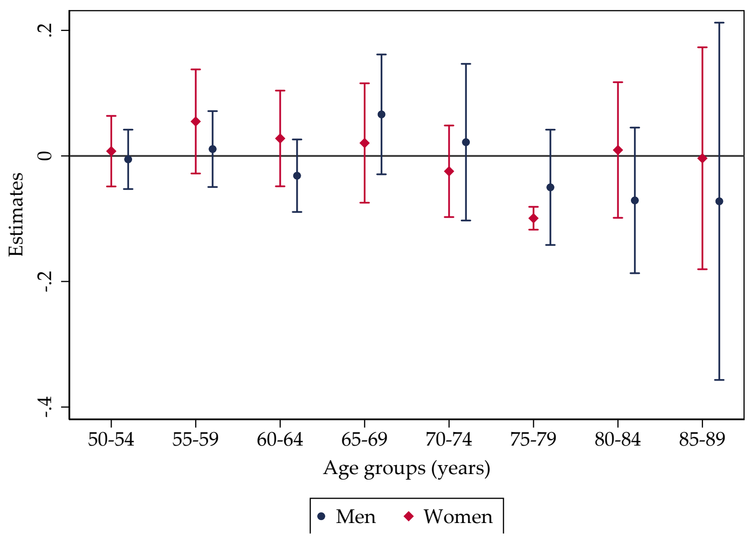

Births before 1990 excluding mixed-sex twins from the treatment

*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\theta_j$ based on OLS and reduced-form versions of Eq. (5.3). All regressions include quadratic age at first birth, and indicators for the age groups, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

Figure A5.5: Effect of having children on attrition over the life-cycle




OLS




(a) Reduced-form

*Note:* Own calculations based on SOEP. Point estimates of the coefficients $\theta_j$ based on OLS and reduced-form versions of Eq. (5.3) with attrition as the dependent variable. All regressions include quadratic age at first birth, and indicators for the age groups, migration background, West Germany and survey years. Robust standard errors clustered at individual level. 95% confidence intervals.

# Bibliography

Adams, S. J. (2002). Educational attainment and health: Evidence from a sample of older adults. *Education Economics*, 10(1):97–109.

Adatto, C. P. (1980). Late adolescence to early adulthood. *The course of life: Psychoanalytic contributions towards understanding personality development*, 2:463–476.

Albouy, V. and Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics*, 28(1):155–168.

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. G. (2007). Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Schmollers Jahrbuch*, 127(1):171–182.

Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, 27(5):1260–1274.

Angrist, J. and Fernandez-Val, I. (2010). Extrapolate-ing: External validity and overidentification in the late framework. Technical report.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824.

Angrist, J. D. and Evans, W. N. (1996). Children and their parents' labor supply: Evidence from exogenous variation in family size. Technical report.

Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, 24(2):149–160.

Banks, J. and Mazzonna, F. (2012). The effect of education on old age cognitive abilities: Evidence from a regression discontinuity design. *The Economic Journal*, 122(560):418–448.

Banks, J., Muriel, A., and Smith, J. P. (2011). Attrition and health in ageing studies: Evidence from ELSA and HRS. *Longitudinal and life course studies*, 2(2).

Banks, J. and Oldfield, Z. (2007). Understanding pensions: Cognitive function, numerical ability and retirement saving. *Fiscal studies*, 28(2):143–170.

Barreca, A. I., Lindo, J. M., and Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic inquiry*, 54(1):268–293.

Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2):537–579.

Bartlett, E. E. (2004). The effects of fatherhood on the health of men: A review of the literature. *Journal of Men's Health and Gender*, 1(2-3):159–169.

Becker, G. S. and Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3):729–758.

Begerow, T. and Jürges, H. (2022). Does compulsory schooling affect health? Evidence from ambulatory claims data. *The European Journal of Health Economics*, 23(6):953–968.

Beral, V. (1985). Long term effects of childbearing on health. *Journal of Epidemiology & Community Health*, 39(4):343–346.

Berkowitz, M. W. and Bier, M. C. (2005). What works in character education: A research-driven guide for educators. *Washington, DC: Character Education Partnership*.

Bhalotra, S. and Clarke, D. (2019). Twin birth and maternal condition. *The Review of Economics and Statistics*, 101(5):853–864.

Bhalotra, S. and Clarke, D. (2020). The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association*, 18(6):3090–3139.

Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030.

Bijwaard, G. E., van Kippersluis, H., and Veenman, J. (2015). Education and health: The role of cognitive ability. *Journal of Health Economics*, 42:29–43.

Black, B., Hollingsworth, A., Nunes, L., and Simon, K. (2022). Simulated power analyses for observational studies: An application to the Affordable Care Act Medicaid expansion. *Journal of Public Economics*, 213:104713.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2):669–700.

Blos, P. (1968). Character formation in adolescence. *The Psychoanalytic Study of the Child*, 23(1):245–263.

Blossfeld, H.-P., Von Maurice, J., and Schneider, T. (2011). The National Educational Panel Study: need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft*, 14(2):5–17.

Bongaarts, J. (2010). The causes of educational differences in fertility in Sub-Saharan Africa. *Vienna yearbook of population research*, pages 31–50.

Börsch-Supan, A. (2020a). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.710. Data set.

Börsch-Supan, A. (2020b). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.710. Data set.

Börsch-Supan, A. (2020c). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 - SHARELIFE. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.710. Data set.

Börsch-Supan, A. (2020d). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.710. Data set.

Börsch-Supan, A. (2020e). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.710. Data set.

Börsch-Supan, A. (2020f). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 7.1.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.710. Data set.

Börsch-Supan, A. (2020g). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 7.1.1. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w7.711. Data set.

Börsch-Supan, A. (2021). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 1.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8.100. Data set.

Börsch-Supan, A. (2022a). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.800. Data set.

Börsch-Supan, A. (2022b). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w2.800. Data set.

Börsch-Supan, A. (2022c). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 3 - SHARELIFE. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w3.800. Data set.

Börsch-Supan, A. (2022d). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.800. Data set.

Börsch-Supan, A. (2022e). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.800. Data set.

Börsch-Supan, A. (2022f). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w6.800. Data set.

Börsch-Supan, A. (2022g). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w7.800. Data set.

Börsch-Supan, A. (2022h). Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w8.800. Data set.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, Sabrina, o. b. o. t. S. C. C. T. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42(4):992–1001.

Braakmann, N. (2011). The causal relationship between education, health and health related behaviour: Evidence from a natural experiment in England. *Journal of Health Economics*, 30(4):753–763.

Brinch, C. N. and Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, 109(2):425–430.

Brunello, G., Fort, M., Schneeweis, N., and Winter-Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors? *Health economics*, 25(3):314–336.

Bucher-Koenen, T., Farbmacher, H., Guber, R., and Vikström, J. (2020). Double trouble: The burden of child-rearing and working on maternal mortality. *Demography*, 57(2):559–576.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M., and Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics*, 50:99–114.

Burgard, S. A. and Lin, K. Y. (2013). Bad jobs, bad health? How work and working conditions contribute to health disparities. *American Behavioral Scientist*, 57(8):1105–1127.

Cáceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources*, 41(4):738–754.

Cáceres-Delpiano, J. and Simonsen, M. (2012). The toll of fertility on mothers' wellbeing. *Journal of health economics*, 31(5):752–766.

Case, A. and Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, NBER Chapters, pages 185–212. National Bureau of Economic Research, Inc.

Clark, D. and Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6):2087–2120.

Colleran, H., Jasienska, G., Nenko, I., Galbarczyk, A., and Mace, R. (2015). Fertility decline and the changing dynamics of wealth, status and inequality. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806):20150287.

Crespo, L., López-Noval, B., and Mira, P. (2014). Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, 43:36–46.

Cygan-Rehm, K. (2018). Is additional schooling worthless? Revising the zero returns to compulsory schooling in Germany. CESifo Working Paper 7191, Munich.

Dahmann, S. C. and Schnitzlein, D. D. (2019). No evidence for a protective effect of education on mental health. *Social Science & Medicine*, 241:112584.

Darin-Mattsson, A., Fors, S., and Kåreholt, I. (2017). Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *International journal for equity in health*, 16(1):1–11.

Dave, D. and Saffer, H. (2008). Alcohol demand and risk preference. *Journal of Economic Psychology*, 29(6):810–831.

Davies, N. M., Dickson, M., Davey Smith, G., Van Den Berg, G. J., and Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature human behaviour*, 2(2):117–125.

de Carvalho, H. W., Pereira, R., Frozi, J., Bisol, L. W., Ottoni, G. L., and Lara, D. R. (2015). Childhood trauma is associated with maladaptive personality traits. *Child abuse & neglect*, 44:18–25.

Delaney, J. M. and Devereux, P. J. (2019). More education, less volatility? The effect of education on earnings volatility over the life cycle. *Journal of Labor Economics*, 37(1):101–137.

Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y., and Zheng, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. *Statistical Science*, 28(2):238–256.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Erikson, R. (2006). Social class assignment and mortality in Sweden. *Social Science & Medicine*, 62(9):2151–2160.

Farbmacher, H., Guber, R., and VikstrÃ¶m, J. (2018). Increasing the credibility of the twin birth instrument. *Journal of Applied Econometrics*, 33(3):457–472.

Fichera, E. and Savage, D. (2015). Income and health in Tanzania. An instrumental variable approach. *World Development*, 66:500–515.

Fischer, M., Gerdtham, U.-G., Heckley, G., Karlsson, M., Kjellsson, G., and Nilsson, T. (2021). Education and health: long-run effects of peers, tracking and years. *Economic Policy*, 36(105):3–49.

Fitzsimons, E. and Malde, B. (2014). Empirically probing the quantity–quality model. *Journal of Population Economics*, 27:33–68.

Flegal, K. M., Graubard, B. I., Williamson, D. F., and Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. *Jama*, 293(15):1861–1867.

Fort, M., Ichino, A., and Zanella, G. (2016). Cognitive and non-cognitive costs of daycare 0–2 for girls. IZA Discussion Papers No. 9756, Bonn.

Freise, D., Schmitz, H., and Westphal, M. (2022). Late-career unemployment and cognitive abilities. *Journal of Health Economics*, 86:102689.

Galama, T. J. and van Kippersluis, H. (2019). A theory of socio-economic disparities in health over the life cycle. *The Economic Journal*, 129(617):338–374.

Gathmann, C., JÃŒrges, H., and Reinhold, S. (2015). Compulsory schooling reforms, education and mortality in twentieth century Europe. *Social Science & Medicine*, 127:74–82. Special Issue: Educational Attainment and Adult Health: Contextualizing Causality.

Gehrsitz, M. and Williams Jr, M. C. (2022). The effects of compulsory schooling on health and hospitalization over the life-cycle. *Working Paper*.

Glymour, M. M., Kawachi, I., Jencks, C. S., and Berkman, L. F. (2008). Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*, 62(6):532–537.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360.

Gove, W. R. and Geerken, M. R. (1977). The effect of children and employment on the mental health of married men and women. *Social Forces*, 56(1):66–76.

Grable, J. E. (2000). Financial risk tolerance and additional factors that affect risk taking in everyday money matters. *Journal of business and psychology*, 14(4):625–630.

Gray, N., Picone, G., Sloan, F., and Yashkin, A. (2015). The relationship between BMI and onset of diabetes mellitus and its complications. *Southern medical journal*, 108(1):29.

Grossman, M. (1972). *The Demand for Health: A Theoretical and Empirical Investigation*. National Bureau of Economic Research.

Günther, I. and Harttgen, K. (2016). Desired fertility and number of children born across time and space. *Demography*, 53(1):55–83.

Guo, R., Li, H., Yi, J., and Zhang, J. (2018). Fertility, household structure, and parental labor supply: Evidence from China. *Journal of Comparative Economics*, 46(1):145–156.

Hank, K. (2010). Childbearing history, later-life health, and mortality in Germany. *Population studies*, 64(3):275–291.

Heliövaara, M. and Aromaa, A. (1981). Parity and obesity. *Journal of Epidemiology & Community Health*, 35(3):197–199.

Hoekstra, C., Willemsen, G., van Beijsterveldt, C. T., Lambalk, C. B., Montgomery, G. W., and Boomsma, D. I. (2010). Body composition, smoking, and spontaneous dizygotic twinning. *Fertility and Sterility*, 93(3):885–893.

Hoel, J. B., Schwab, B., and Hoddinott, J. (2016). Self-control exertion and the expression of time preference: Experimental results from Ethiopia. *Journal of Economic Psychology*, 52:136–146.

Jacobsen, J. P., Pearce, J. W., and Rosenbloom, J. L. (1999). The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. *The Journal of Human Resources*, 34(3):449–474.

Janke, K., Johnston, D. W., Propper, C., and Shields, M. A. (2020). The causal effect of education on chronic health conditions in the UK. *Journal of Health Economics*, 70:102252.

Josefsson, K., Jokela, M., Hintsanen, M., Cloninger, C. R., Pulkki-Råback, L., Merjonen, P., Hutri-Kähönen, N., and Keltikangas-Järvinen, L. (2013). Parental care-giving and home environment predicting offspring's temperament and character traits after 18 years. *Psychiatry Research*, 209(3):643–651.

Jung, S. (2015). Does education affect risk aversion? Evidence from the British education reform. *Applied Economics*, 47(28):2924–2938.

Jürges, H. (2013). Collateral damage: The German food crisis, educational attainment and labor market outcomes of German post-war cohorts. *Journal of Health Economics*, 32(1):286–303.

Jürges, H., Kruk, E., and Reinhold, S. (2013). The effect of compulsory schooling on health – evidence from biomarkers. *Journal of Population Economics*, 26(2):645–672.

Kaestner, R., Schiman, C., and Ward, J. (2020). Education and health over the life cycle. *Economics of Education Review*, 76:101982.

Kamhöfer, D. A. and Schmitz, H. (2016). Reanalyzing zero returns to education in Germany. *Journal of Applied Econometrics*, 31(5):912–919.

Kamhöfer, D. A., Schmitz, H., and Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association*, 17(1):205–244.

Katzmarzyk, P. T., Craig, C. L., and Bouchard, C. (2001). Original article underweight, overweight and obesity: Relationships with mortality in the 13-year follow-up of the Canada Fitness Survey. *Journal of clinical epidemiology*, 54(9):916–920.

Kelleher, J. P. (2017). Pure time preference in intertemporal welfare economics. *Economics & Philosophy*, 33(3):441–473.

Kemptner, D., JÃŒrges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, 30(2):340–354.

Knight, J., Weir, S., and Woldehanna, T. (2003). The role of education in facilitating risk-taking and innovation in agriculture. *The Journal of Development Studies*, 39(6):1–22.

Komlos, J., Smith, P. K., and Bogin, B. (2004). Obesity and the rate of time preference: Is there a connection? *Journal of biosocial science*, 36(2):209–219.

Kravdal, Ø. (2002). Education and fertility in sub-Saharan Africa: Individual and community effects. *Demography*, 39(2):233–250.

Kreyenfeld, M. (2004). Fertility decisions in the FRG and GDR: An analysis with data from the German Fertility and Family Survey. *Demographic Research*, 3:275–318.

Kroll, L. E. (2011). Construction and validation of a general index for job demands in occupations based on ISCO-88 and KldB-92. *methods, data, analyses*, 5(1):28.

Kruk, K. E. and Reinhold, S. (2014). The effect of children on depression in old age. *Social Science & Medicine*, 100:1–11.

Kubzansky, L. D. and Kawachi, I. (2000). Going to the heart of the matter: Do negative emotions cause coronary heart disease? *Journal of psychosomatic research*, 48(4-5):323–337.

Kuhlmann, S. and Röber, M. (2006). Civil service in Germany: Between cutback management and modernization. In *State and Local Government Reforms in France and Germany*, pages 89–109. Springer.

Lager, A. C. J. and Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences*, 109(22):8461–8466.

Leopold, L. and Leopold, T. (2018). Education and health across lives and cohorts: A study of cumulative (dis)advantage and its rising importance in Germany. *Journal of Health and Social Behavior*, 59(1):94–112. PMID: 29337605.

Leschinsky, A. and Roeder, P. M. (1980). Didaktik und Unterricht in der Sekundarschule I seit 1950-Entwicklung der Rahmenbedingungen. *PB Max-Planck-Institut für Bildungsforschung (ed.), Bildung in der Bundesrepublik Deutschland-Daten und Analysen*, 1:283–392.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *The Review of Economic Studies*, 72(1):189–221.

Lundborg, P., Nilsson, A., and Rooth, D.-O. (2014). Parental education and offspring outcomes: Evidence from the Swedish Compulsory School Reform. *American Economic Journal: Applied Economics*, 6(1):253–78.

Lundborg, P., Plug, E., and Rasmussen, A. W. (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review*, 107(6):1611–37.

Marcus, J., Reif, S., Wuppermann, A., and Rouche, A. (2020). Increased instruction time and stress-related health problems among school children. *Journal of Health Economics*, 70(C).

Marmot, M. (2004). Status syndrome. *Significance*, 1(4):150–154.

Martin, T. C. (1995). Women's education and fertility: Results from 26 Demographic and Health Surveys. *Studies in family planning*, pages 187–202.

Mazzonna, F. and Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, 52(1):128–151.

Meghir, C., Palme, M., and Simeonova, E. (2018). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics*, 10(2):234–56.

Melly, B. (2005). Public-private sector wage differentials in Germany: Evidence from quantile regression. *Empirical Economics*, 30(2):505–520.

NEPS Network. (2022). National Educational Panel Study, Scientific Use File of Starting Cohort Adults.

NHLBI (1998). Clinical guidelines for the identification, evaluation, and treatment of overweight and obesity in adults: The evidence report. Technical Report 98-4083. NHLBI (National Heart, Lung, and Blood Institute) Obesity Education Initiative Expert Panel on the Identification, Evaluation, and Treatment of Obesity in Adults (US).

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., et al. (2015). Promoting an open research culture. *Science*, 348(6242):1422–1425.

Ohrnberger, J., Fichera, E., and Sutton, M. (2017). The dynamics of physical and mental health in the older population. *The Journal of the Economics of Ageing*, 9:52–62.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, 96(1):152–175.

Ozdenoren, E., Salant, S. W., and Silverman, D. (2012). Willpower and the optimal control of visceral urges. *Journal of the European Economic Association*, 10(2):342–368.

Palme, M. and Simeonova, E. (2015). Does women's education affect breast cancer risk and survival? Evidence from a population based social experiment in education. *Journal of Health Economics*, 42:115–124.

Perez-Arce, F. (2017). The effect of education on time preferences. *Economics of Education Review*, 56:52–64.

Pesaran, H. (2003). Introducing a replication section. *Journal of Applied Econometrics*, 18(1):111.

Pischke, J.-S. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *The Economic Journal*, 117(523):1216–1242.

Pischke, J.-S. and Von Wachter, T. (2008). Zero returns to compulsory schooling in Germany: Evidence and interpretation. *The Review of Economics and Statistics*, 90(3):592–598.

Pischon, T., Boeing, H., Hoffmann, K., Bergmann, M., Schulze, M. B., Overvad, K., Van der Schouw, Y., Spencer, E., Moons, K., Tjønneland, A., et al. (2008). General and abdominal adiposity and risk of death in Europe. *New England Journal of Medicine*, 359(20):2105–2120.

Quis, J. S. (2018). Does compressing high school duration affect students' stress and mental health? Evidence from the National Educational Panel Study. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 238(5):441–476.

Reddy, U. M., Branum, A. M., and Klebanoff, M. A. (2005). Relationship of maternal body mass index and height to twinning. *Obstetrics & Gynecology*, 105(3):593–597.

Reyes-Garcia, V., Godoy, R., Huanca, T., Leonard, W. R., McDade, T., Tanner, S., and Vadez, V. (2007). The origins of monetary income inequality: Patience, human capital, and division of labor. *Evolution and Human Behavior*, 28(1):37–47.

Riley Jr, W. B. and Chow, K. V. (1992). Asset allocation and individual risk aversion. *Financial Analysts Journal*, 48(6):32–37.

Ross, C. E. and Mirowsky, J. (2010). Gender and the health benefits of education. *The Sociological Quarterly*, 51(1):1–19.

Santoro, N. (2003). Mechanisms of premature ovarian failure. In *Annales d'endocrinologie*, volume 64, pages 87–92.

Schmitz, H. and Tawiah, B. B. (2023). Life-cycle health effects of compulsory schooling. Ruhr Economic Papers 1006, RWI.

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

Shaw, K. L. (1996). An empirical analysis of risk aversion and income growth. *Journal of Labor Economics*, 14(4):626–653.

Shetty, P. and James, W. (1994). Body mass index: a measure of chronic energy deficiency in adult. Rome: Food and Agricultural Organization. *FAO Food and Nutrition paper*, (56).

Siedler, T. (2007). Schooling and citizenship: Evidence from compulsory schooling reforms. ISER Working Paper Series No. 2007-02, Colchester.

Silles, M. A. (2009). The causal effect of education on health: Evidence from the United Kingdom. *Economics of Education Review*, 28(1):122–128.

Silles, M. A. (2016). The impact of children on women's labour supply and earnings in the UK: Evidence using twin births. *Oxford Economic Papers*, 68(1):197–216.

SOEP (2022). Socio-Economic Panel (SOEP), data for years 1984-2020, SOEP-Core v37, EU Edition.

SOEP Group (2022). SOEP-Core v37 – PGEN: Person-Related Status and Generated Variables. *SOEP Survey Papers 1186: Series D – Variable Descriptions and Coding.*

Suls, J. and Bunde, J. (2005). Anger, anxiety, and depression as risk factors for cardiovascular disease: The problems and implications of overlapping affective dispositions. *Psychological bulletin*, 131(2):260.

Sun, Y. and Li, S. (2010). The effect of risk on intertemporal choice. *Journal of Risk Research*, 13(6):805–820.

Tawiah, B. B. (2022). Does education have an impact on patience and risk willingness? *Applied Economics*, 54(58):6687–6702.

Tawiah, B. B. (2023). The effect of children on health. Working Papers Dissertations 103, Paderborn University, Faculty of Business Administration and Economics.

Tawiah, B. B. and Schiele, V. (2023). Does education improve cognitive performance four decades after school completion? a replication study of nicole schneeweis, vegard skirbekk and rudolf winter-ebmer (demography, 2014). Working Papers Dissertations 102, Paderborn University, Faculty of Business Administration and Economics.

Thomas, F., Renaud, F., Benefice, E., De Meeüs, T., and Guegan, J.-F. (2001). International variability of ages at menarche and menopause: Patterns and main determinants. *Human biology*, pages 271–290.

Van der Pol, M. (2011). Health, education and time preference. *Health economics*, 20(8):917–929.

van Zon, S. K., BÃŒltmann, U., Reijneveld, S. A., and de Leon, C. F. M. (2016). Functional health decline before and after retirement: A longitudinal analysis of the Health and Retirement Study. *Social Science & Medicine*, 170:26–34.

Vere, J. P. (2011). Fertility and parents' labour supply: new evidence from US census data: Winner of the OEP prize for best paper on Women and Work . *Oxford Economic Papers*, 63(2):211–231.

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., and Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, 120(2):142–145.

Viscusi, W. K. and Moore, M. J. (1989). Rates of time preference and valuations of the duration of life. *Journal of public economics*, 38(3):297–317.

Vitthala, S., Gelbaya, T., Brison, D., Fitzgerald, C., and Nardo, L. (2009). The risk of monozygotic twins after assisted reproductive technology: A systematic review and meta-analysis. *Human Reproduction Update*, 15(1):45–55.

Weisell, R. C. (2002). Body mass index as an indicator of obesity. *Asia Pacific journal of clinical nutrition*, 11:S681–S684.

Westerlund, H., KivimÃ€ki, M., Singh-Manoux, A., Melchior, M., Ferrie, J. E., Pentti, J., Jokela, M., Leineweber, C., Goldberg, M., Zins, M., and Vahtera, J. (2009). Self-rated health before and after retirement in France (GAZEL): a cohort study. *The Lancet*, 374(9705):1889–1896.

WHO (1995). Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee. techreport 854:1-452.

WHO (2000). Obesity: preventing and managing the global epidemic: Report of a WHO consultation. techreport 894.

Yan, L. L., Liu, K., Matthews, K. A., Daviglus, M. L., Ferguson, T. F., and Kiefe, C. I. (2003). Psychosocial factors and risk of hypertension: the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Jama*, 290(16):2138–2148.

Zhang, J. (2017). A dilemma of fertility and female labor supply: Identification using Taiwanese twins. *China Economic Review*, 43:47–63.

# Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Datum Unterschrift