Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

RIGOROUS THESIS

Effect of education on health: The Czech Republic case

Author: **Ing. Mgr. Marek Pazitka**

Supervisor: **PhDr. Julie Chytilová, Ph.D. (Diploma thesis)**

Academic Year: **2014/2015**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, February 8, 2015

Signature

Acknowledgments

The author is grateful especially to PhDr. Julie Chytilová, Ph.D. for her inspiring insights and thoughtful guidance. I feel extremely grateful for her time and energy. I am also thankful to Ing. Tomaš Mlčoch for the encouragement and the valuable comments. At last but not least, I want to thank my girlfriend and family for their continuous support.

This paper uses data from SHARE wave 4 release 1.1.1, as of March 28^{th} 2013 and SHARE wave 2 release 2.6.0, as of November 29^{th} 2013. The SHARE data collection in 2004-2007 was primarily funded by the European Commission through its 5^{th}, 6^{th} and 7^{th} framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the U.S. National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGHA 04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see http://www.share-project.org for a full list of funding institutions).

Abstract

Previous research has uncovered a large, positive and causal link between education and health. This paper is devoted to examining the topic in the former Czechoslovakia. My analysis is conducted on a data set pooled from the Survey of Health, Ageing and Retirement in Europe (SHARE). I utilize a continuum of ages at school entry, caused by the use of a single school cut-off, to identify the effect of education on health, which is uniquely created from the PCA method and using 30 questions of the SHARE. Therefore, I apply instrumental variable approach with a month of birth as an instrument for education. The results from the first-stage suggest that the instrument is not valid, since a correlation between the instrumental ($Month\ of\ birth$) and the instrumented variable (education) is very low and insignificant. The results remain insignificant even after adjusting for different measures of education, health, institutional changes or heterogeneous effects. As the most probable cause, I state the inability to control for non-compliers in my instrumental variable regressions. As a consequence, all the results regarding the link between education and health are inconclusive.

JEL Classification I12, I21, I28, J13
Keywords health, education, birth month, relative maturity effect

Author's e-mail marek.pazitka@gmail.com
Supervisor's e-mail chytilova@fsv.cuni.cz

Abstrakt

Velká, pozitivní a kauzalní souvislost mezi vzděláním a zdravím člověka byla odhalena předchozím výzkumem. Tato studie se věnuje analýze tohoto vztahu v bývalém Československu. V mojí práci jsem využil soubory dat získané z průzkumu Survey of Health, Ageing, and Retirement in Europe (SHARE). Užívání jednoho ročního termínu pro nástup dětí do školy způsobuje continuum věku v této době, což lze využit pro identifikaci vlivu vzdělání na zdraví, které je v mé praci unikátně vytvořeno pomocí metody principálních komponentů z 30 různých otázek z SHARE. Při analýze jsem použil metodu instrumentalní proměnné, která je v mém případě měsíc narození. Nicméně, výsledky z prvního kola této metody naznačují, že instrumentalní proměnná není platná, protože korelace mezi měsícem narození a vzděláním je velmi nízká a nevýznamná. Toto platí i po změně proměnné pro vzdělaní i zdraví nebo po kontrole institutionalních změn, případně po povolení heterogeních efektů. Jako nejpravděpodobnější příčina se jeví fakt, že jsem s použitými daty nebyl schopen kontrolovat pro tkz. “non-compliers”. Důsledkem toho je, že všechny výsledky týkající se souvislosti mezi vzděláním a zdravím jsou neprůkazné.

Klasifikace JEL	I12, I21, I28, J13
Klíčová slova	zdraví, vzělaní, měsíc narození, efekt relativní vyspělosti

E-mail autora	marek.pazitka@gmail.com
E-mail vedoucího práce	chytilova@fsv.cuni.cz

Contents

List of Tables

List of Figures

Acronyms

EU European Union

IV Instrumental variable

NHL The National Hockey League

OECD Organisation for Economic Co-operation and Development

PCA Principal component analysis

SHARE Survey of Health, Ageing and Retirement in Europe

TSLS Two-stage least squares

UK United Kingdom

Rigorous Thesis Introduction

The presented rigorous thesis is an author's diploma thesis defended at the Institute of Economic Studies (Faculty of Social Studies, Charles University in Prague) in September 2014.

Introduction

Previous literature shows large and significant correlation of education and health (Grossman and Kaestner, 1997). This correlation is robust even after controlling for different measures of socio-economic status (income, race, etc.) or using different measures of health (morbidity rates, self-reported health, etc.). This fact is very important for a lot of questions. Returns to education are usually interpreted in terms of increases of market productivity or better job opportunities, but if education causes health, we should add it to returns to education, otherwise, the returns would be underestimated. The underestimation could lead to wrong public policy.[1] Since, the goal of many governments is to increase number of college graduates - see, for example, Europe 2020 (the EU's growth strategy for the coming decade), which has a goal of at least 40% of 30-34-year-olds with third-level education (or equivalent), literature should verify the effectiveness of a greater allocation of public resources into education.

In addition, the effect of education on health, if causal, can be viewed as an another attribute into the decision about redshirting a child.[2] In the Czech Republic (or previously in former Czechoslovakia) laws state that all children who are six years old and born no later than 31$^{\text{th}}$ of August have to enter the first year of primary school, so those born in the summer are usually the youngest in the class, unless they are redshirted. Redshirting might be an effective way to improve an educational outcome, if the relatively older children perform better throughout high school and beyond, but this practice also bears some costs, because a child loses one year of his productive live.

[1]For example, Carneiro, Heckman, and Vytlacil (2011) estimate the marginal returns to college and they calculate that some marginal policy changes induce students into college, producing very low returns. However, in calculation of returns they are using just earnings increase (as many others), so their result could substantially underestimate the true returns to education.

[2]Terminology originally from United States college sports, in which redshirting means purposeful deferral of beginning of training (in our case education). To redshirt a child is to let him stay at home for one or more years longer than needed by eligibility rules.

However, if education affects health, the costs of redshirting are reduced - with better health, one could work longer and also life expectancy could increase.

The research on the topic of the relative maturity effect[3] started in the 1960s,[4] so there is an enormous amount of literature. Bedard and Dhuey (2006), using data from all countries of OECD report that the youngest members of each cohort score 4-12 percentiles lower than the oldest members in the fourth grade, and 2-9 percentiles lower in the eighth grade. Their study also includes the Czech Republic, but in a quite different time period than my work. Du, Gao, and Levi (2012) find that the number of summer-born CEOs is disproportionately small relative to the number of CEOs born in the other seasons on data set of S&P 500 companies between 1992 and 2009. Further, Crawford et al. (2011) state that the month of birth not only affects their test scores, but it also affects how they feel about their own ability and the degree to which they believe they can influence their future. Education choices are likely to have potentially far-reaching consequences for various parts of life, so it seems plausible that the month of birth could affect one's choices, experiences or achievements during adulthood.

Although, most of the papers concern the relative maturity effect and investigate its impact on educational attainment, wages or career, it is also vital to know whether the relative age effect has a lasting impact on different measures of well-being such as health. I am not the first to consider this issue - Crawford et al. (2013) investigate the relative maturity effect in the UK. They estimate the effect on the probability of employment, occupation, earnings for adults and health (self-perceived and mental), but even through significant differences were observed in educational attainment, they did not find widespread differences in adulthood. But I am first in considering the maturity age effect on health in the Czech Republic (or generally in Czechoslovakia) and compared to Crawford et al. (2013) I also created a unique health variable containing information from 30 different questions related to health. I possess data on people older than 50 years, which is convenient due to the fact that differences in health between higher and lower educated individuals are increasing with age (Leopold and Engelhartdt, 2011). Therefore, for determining the long-term effect on health, it is better to use older people, since the improvement in health caused by later enrollment in school can be slowly increasing.

[3]In my work I use this term for an overall effect. In Section 1.2 I present and describe sub-effect, but it is not the aim of this work to measure them separately.

[4]As a pioneer in this field of study we can count the work of King(1955).

For the identification of the causal effect of education on health, I am using an instrumental variable approach. The dependent variable in the first stage is the years of schooling (alternatively, the acquired level of education) and independents are the birth month relative to the school cut-off date (instrument) plus vector of controls. The birth month manipulates with treatment (education), because children are relatively and absolutely older than their peers and can therefore gain more skills before starting school, as well as being more mature. These advantages can project into a higher educational attainment. The relative maturity effect is, however, not correlated with the labor market outcomes or health (except for mechanism through education), since the absolute age is an important determinant for performance and the initial gap in age disappears as children get older (Kawaguchi, 2011).

The Identification strategy of this paper relies on a few assumptions, in the following lines the most important are presented. Firstly, time of birth is random through the year. Buckles and Hungerman (2012) show that children born at different quarters have different socioeconomic characteristics (for example, children born in the winter are disproportionately likely to be born to women who are teenagers, who are unmarried, and who lack a high school degree). Due to this concern, I applied a different identification strategy, which uses only some months (one or two) around the cut-off point. Secondly, school cut-off dates only determine relative age if the rules are strictly followed. The probability of non-compliance is increasing, if we are closer to the cut-off, so in the last identification strategy I exclude one month directly next to the cut-off from each side in a sample of two months around each side of the cut-off. Furthermore, a comparison of estimates from the different samples - a full sample, sample with a one and two month bandwidth around cut-off and a July and October bandwidth which can serve as a robustness check. A similar procedure use, for example, Puhani and Weber (2006).

The main research question to be addressed is thus whether an individual's age relative to academic year cut-off continues to affect health observed in the elderly. With my data set drawn from the SHARE survey. I did not find a statistically significant effect of the relative maturity on education. Insignificant results violate one of the assumptions of the instrumental variable approach - a strong first-stage (measured by F-statistic above 10), so I cannot state any conclusion about the relationship between education and health. First-stage regressions are insignificant even after adjusting for different measures of education (years vs. levels of schooling), different measures of health (physical health, mental health etc.), institutional changes or heterogeneous effects (man vs. woman).

I identified two hypotheses explaining weak relationship between the birth month and education. The first one, a huge proportion of non-compliers, which can be described as people not complying with the cut-off, for example, a child born in August and despite this postponing the compulsory school attainment by one year. I cannot control non-compliance and since its presence is causing downward bias in the resulting coefficients, it is not unlikely that the non-compliance effect (non-complying agents born before September would be the oldest in a class) wiped out the relative maturity effect. Secondly, we cannot reject the hypothesis that the relative maturity effect is simply not significant at the time of high school entrance tests, because relative differences between absolute ages are already too small and without strong streaming, the differences in skills would not prevail.

The balance of my work proceeds as follows. In Chapter 1, I present a theoretical framework for my thesis, which contains three parts - an institutional background, the effect of the relative maturity on education and finally the effect of education on health. In Chapter 2, I introduce my data set, the creation of my dependent variables and descriptive statistics. Chapter 3 presents the instrumental variable estimation methodology and discusses necessary adjustments for proper estimation. In Chapter 4, I present results from the ordinary least square and the instrumental variable approach. Chapter 4.3.3 concludes and discusses potential fields of future research.

Part I

Theoretical part

Chapter 1

Theoretical Background

1.1 Institutional background

This chapter presents the institutional background. The institutional background is included due to the fact that the institutional stability is important for the hypothesis and any major changes in the Czechoslovak schooling system should be reflected in the analysis. It is also possible to split sample in order to check for the robustness of the estimates across the whole period of my interest, since some changes can also change the relative maturity effect's size. This analysis will be presented in Subsection 4.3.2.

The Czechoslovak compulsory starting age for school was set at six years old. This fact remain unchanged since 1869, when so called "Hasner's Law" came into force. Children with a birthday before the cut-off on the 1st of September are obliged to go to school and children with a birthday after the cut-off have to wait another year.[5] When we compare this fact internationally, we find that it is equal to the median and the mode value of selected countries displayed in Table 1.1.[6] This comparison could be useful, when we take into account external validity (extent to which the results can be generalized) of the study.

During the time of my interest, the Czechoslovak schooling system did not have any preschool programs. Before the compulsory school attendance, children usually attended kindergarten, where they could play with others. For this

[5]Assuming that the rules are strictly followed.

[6]I should note that these statistics are from the year 2012, meanwhile, the agents in my data set attended school in quite a different time. However, changes in compulsory starting age are rather rare. In 1971, the average compulsory starting age was 6.21 with standard deviation of 0.63 among 199 countries and in 2010 it was 6.11 with standard deviation of 0.57 (World Development Indicators, The World Bank), so the table contains relevant information.

Table 1.1: The compulsory school starting age by country

Age 5	Age 6	Age 7
Australia	Argentina	Brazil
United Kingdom	Austria	Switzerland
Ireland	Belgium	China
New Zealand	Canada	Denmark
	Chile	Finland
	Czech Republic	Croatia
	France	Poland
	Germany	
	Mexico	
	Netherlands	
	Norway	
	Slovak Republic	
	Spain	

Note: In Switzerland entry age differs by region.
Source: World Development Indicators, The World Bank.

reason, entering a compulsory school in Czechoslovakia meant moving from a playground to school desks, where a usual day took approximately around five to six hours (approximately from eight o'clock in the morning to one or two o'clock in the afternoon). The usual length of kindergarten was three years, from three to six years of children's life.

In my sample, I can expect most of the people born in the range of approximately 1920 to 1960, but due to the fact that Czechoslovak universities were not functional from the years 1939 to 1945, I will have to restrict my sample (more detailed description will be provided in the Section 2.1). Another important event in the development of Czechoslovak schooling system was the law number 95/1948 about the basic form of uniform education from year 1948. The law issued by communists abolished eight year-long gymnasiums. In Table 1.2 you can see the structure of the Czechoslovak schooling system between 1948 and 1953. In 1953, the law number 31/1953 changed the Czechoslovak schooling system again. The compulsory school attendance shortened by one year, from nine to eight and was followed by a three year-long higher non-compulsory education, which served as university preparation.

During the whole above mentioned period, university education was tuition-free, but there is a possibility that it was not available to all people for political reasons. However, I lack the data to control this feature, but there is no reason

Table 1.2: The Czechoslovak schooling system between 1948 and 1953

School	Length	Age
National school	5	6-11
Secondary school	4	11-15
Gymnasia for four years, as well as SOS and SOU	4	15-19

Note: Higher education is excluded. SOS and SOU are types of high schools.
Source: The law number 95/1948 about basic form of uniform education.

to think that these restrictions are correlated with birth month. So, even through I cannot control for these restrictions in education, it will not bias my results under the condition that they do not change based on the birth month.

1.2 The effect of the relative maturity on education

1.2.1 The basic model of the relative maturity effect

Motivated by the research of Dixon et al. (2011), the effect of relative maturity can be described in the following way. Firstly, there is the selection of individuals based on ability. Once those individuals are selected, they are placed into different streams (e.g., gifted or competitive) and these different streams provide discriminate opportunities for instruction, contact time, and competition. As described in the previous chapter, the Czechoslovak schooling system meets such conditions; in fact, practically every schooling system satisfies those conditions, because there is almost always an entrance exam (based on ability or talent) for continuing at a higher level of education and there is practically always one single cut-off.

Similarly, Allen and Barnsley (1993) claim that the selection characterizes all educational and training systems. But, most of the selection processes are not error-free. Individuals are usually selected based on ability, but it is often impossible to observe ability independent of the maturity. For this reason, the relative maturity effect is an error made during the selection process amongst individuals who were situated in the same selection period (usually a year).

Allen and Barnsley (1993) also provide a simple, and yet informative model describing the relative maturity effect. Although, the model is described with an example of Canadian hockey, which I will adjust for the case of academia.

As an input, let's assume that the Czechoslovak educational system can be generalized to a model, which is characterized by the following three features. Firstly, it has a finite number of discrete periods of learning. Secondly, entry to the first period occurs during childhood (when observing ability independently on maturity is practically impossible).[7] Thirdly, different training is provided to different streams and the selection to the streams is based on the observed ability (further stated skills).

The skills (observed ability) of an individual are composed from three things - ability (constant and innate), maturity (increasing function of age) and accumulated knowledge. As an age goes to adulthood, the differences in skill are given by the combination of the ability and the accumulated knowledge.

As another simplifying assumption, let's consider that people can be divided into two groups based on ability: group 1 and group 2. Further, there are two streams: stream A and stream B. The selection is based on the skills (function of ability, maturity and accumulated knowledge). In Table 1.3 we can see that the relative maturity effect bears death weight losses, since the selection into streams is not optimal due to maturity.

Table 1.3: Possible errors in the relative maturity model

	A	B
1	✓	Type II error
2	Type I error	✓

Note: Rows depict groups of people with different levels of ability. Columns depict different kind of streams.
Source: Model is based on work of Allen and Barnsley (1993).

There are two effects in the model, which determine the amount of errors made. The first one is a training effect - different streams have different rates of learning, otherwise, the whole concept of streaming would be meaningless. This effect implies a bigger persistence of errors. The second is the ability effect. This effect corrects initial errors - as individuals age, the relative differences in maturity are getting lower, because initial ability becomes more important.

When we confront the model with the real world, we can conclude that systems that stream more have greater persistence of the selection errors. I already mentioned that the authors of the model used Canadian hockey as an

[7] For example, Musch and Grondin (2001) note that skill is as much the product of maturity as it is ability and it is practically impossible to distinguish ability from maturity during childhood.

example and the Canadian hockey is a system with an extreme persistence of errors. In the NHL (National Hockey League) the relative maturity effect is greatest at the end of the system - in adulthood.[8]

When we applied the previously described model to our case, the essential question for this paper is if the differences caused by relative maturity at the end of a compulsory education are still high enough to affect the result of the entrance exams.[9] In the case that this assumption would hold, there is still one concern - the rate of accumulation of knowledge has to be higher for different high schools. When those two requirements are fulfilled, we should observe the training effect - the persistent relative maturity effect.

For example, Bedard and Dhuey (2006) use data from OECD to show that the youngest members of each cohort score 4-12 percent lower than the oldest members in the fourth grade, and 2-9 percent lower in the eighth grade. *Further, they show that in Canada and the United States, the youngest members of each cohort are less likely to attend a university.* Their result has one important lesson - the power of the effect is decreasing over time, but even at the end of the system (college), there is some effect. In terms of our model, the training effect is quite high.

1.2.2 The division of the relative maturity into sub-effects

The effect of the relative maturity in school can be split into several different sub-effects.

The absolute age effect The age in which children wrote their tests. If children sit some exams (entrance exam or graduation exam) on the same day (or a very similar one), those born later in the academic year will be younger than their peers.

The school starting age effect "Readiness for school" - on the one hand, it is probable that children will learn more in a school than in kindergarten or at home, but on the other hand, it could be that they are just not ready

[8]For example, Gladwell (2008) states that 40% of the Canadian professional hockey players were born in the first quarter (January to March), 30% in second (April to June), 20% in third (July to September) and only 10% in fourth (October to December).

[9]There are two possible mechanisms. First, at the time of the entrance exam (usual age of child is around 14-15 years), the relative maturity effect is still significant, in other words, it is still too difficult to distinguish ability (innate) from skills (function of ability and maturity). And secondly, there are streams (selection into classes based on observed skills) in compulsory education, which could increase the persistence of errors.

and they need more time to develop a certain set of emotional, behavioral, and cognitive skills needed to learn, work, and function successfully in school.

The length of schooling In the former Czechoslovakia, the length of schooling in time of an entrance test for high school was very similar for candidates and deviations of just a few days cannot play a role in this. There is one exception - children, who have to repeat a grade.[10] There is one exception - children, which have to repeat a grade.[11]

The relative age effect Children after the cut-off are relatively older than their peers and at the beginning of the compulsory school attainment, a one year difference could be a lot.[12] Moreover, the oldest children in a class are stronger and more mature, and this relative standing in class may have an effect on self-esteem, aspirations and the child's social development (Solli, 2011).

The different streams in schools The curriculum could differ based on ability - the more skilled will be placed into more challenging programs (streams) and if the relative maturity affects skills, then it could at least partially affect such a placement.

This said, I will not try to uncover how strong these particular effects are, I will however refer to previous literature to show the expected mechanism. Actually, most of the studies find that the absolute age is the strongest factor. Cahan and Cohen (1989) show on the sample of Jerusalem's Hebrew-speaking, state-controlled, elementary school students that age significantly affects intelligence (measured by various types of intelligence tests).

Further evidence provided by Crawford et al. (2007) also claims that the major reason why children before the cut-off perform significantly worse than after the cut-off is simply that they are almost a year younger. Fredriksson and

[10] Of course, here I refer to the Czechoslovak schooling system. For example, in the USA schooling system, the situation is different (Angrist and Krueger, 1991).

[11] In fact, relatively younger children are more likely to repeat a grade. Without more detailed data it is impossible to determine the final effect of repetition. Student have change to go again through the lectures and they are even more mature than any other non-repeating student. On the other hand, it lowers self-confidence, which could play role in a school, and also if there are some entrance requirements in terms of grades, one year repetition could significantly lower the odds. Besides, my estimate will not be bias upwards by this feature, it will be lower or same that the actual.

[12] I refer here to Table 1.1, an average starting age was 6.11 years among 22 countries in 2010.

Öckert (2005) investigate the same topic on the data set from Sweden. Thanks to the nature of their data, they could exploit it within school variation in order to find out what drives the above mentioned mechanism. They show that the effect is caused by the absolute age rather than the relative age in the class.

The above mentioned papers were unable to distinguish between the effect of school starting age and a direct age-at-test effect, as they are perfectly collinear. Black et al. (2011) separate these two with a Norwegian data set. This separation is possible thanks to unique policy in Norway, where all men must take an IQ test at around 18 years old during a time of compulsory military enrollment. They find that the absolute age effect (or the effect of age at test) is again significant and robust, on the contrary to the school starting age, which appears to have a very small effect on educational attainment.

It is a question, if there is really something like "readiness for school". For example, Cunha et al. (2006) argue that skills accumulated in early childhood are complementary to later learning, which implies that the rate of learning could be different with a different school starting age. To put it simply - early schooling would not necessarily be more productive. Heckman (2006) summarizes evidence on the effects of early environments on disadvantaged child, adolescent, and adult achievement. Based on a large amount of literature, he states that the rate of return from early interventions is high, and the return from later interventions is lower.[13] This said, disadvantaged children are a different problem, but we can expect some external validity. Similar results are provided by Elder and Lubotsky (2009), which find that the association between achievement test scores and entrance age appears to decline sharply and so they cast doubt against the effectiveness of raising the kindergarten entrance age as a mean to raise an achievement. Of course, the problem is what is school, for example, pre-school is not generally taken as school, but it is very different from kindergarten and certainly plays a significant role in a child's development.

[13] To be more specific, the rate of return from early interventions is estimated around 8% per annum, compared to later interventions, which barely paid themselves or has negative rate of return.

1.3 The effect of education on health

The correlation between health and education is a well-established empirical fact[14] and it is robust across a variety of health measures (for example, morbidity rates, self-reported health status or other measures of health). The correlation is even larger than the one with the access to health-care insurance in USA (Newhouse, 1993). Moreover, Auster, Leveson and Sarachek (1969) suggest that education could be a cost-effective way of achieving better health. However, for such a claim we need to know the causal relationship between the two variables and ideally the mechanism behind this relationship.

An example from recent literature, which uncovers causal education-health relationships, is Lleras-Muney (2005). This paper, using data from the USA, shows an education with compulsory education laws from 1915 to 1939 and finds that education has a causal impact on mortality (proxy for health). A similar paper of Silles (2009), based on data from the United Kingdom, also finds a causal relationship running from more schooling to better health. The paper uses the instrumental variable approach with changes in compulsory schooling laws in the United Kingdom as an instrument to test the hypothesis. Both papers concludes that the effect is even larger than standard regression estimates suggest.

Cutler and Lleras-Muney (2010) go further and examine the possible explanations for the relationship between education and health. They report that income, health insurance, and family background can account for about 30 percent of the effect, knowledge and measures of cognitive ability explain an additional 30 percent, social networks account for 10 percent and discounting, risk aversion, or the value of future (measured by proxies - mostly question about future, risk ect.) do not account for any of the effect.

[14] For example, see Grossman and Kaestner (1997).

Part II

Empirical part

Chapter 2

Data

2.1 Data set description

I pool data from the Survey on Household Health, Ageing and Retirement in Europe, or SHARE,[15] for the years 2006 and 2010. I am using data only for the Czech and Slovak Republic (former Czechoslovakia) and only data from a second (2006) and a fourth wave (2010), because in a first wave there are no Czech or Slovak respondents and a third one has a different concept (SHARE-LIFE, which is the name of the third wave, focuses on people's life history, it is a retrospective data collection). So, the starting number of observations is 7 643, 2 830 from wave 2 and 4 813 from wave 4.

I cannot really be sure that a particular person studied in Czechoslovakia, I can however increase this probability by taking only those persons, which are Czech or Slovak citizens and are born in Czechoslovakia. Of course, there is a small probability that this person was born in Czechoslovakia, emigrated and then returned, but even in the case that this would be true, there is no indication that such behaviour is correlated with a birth month. To conclude, because of this restriction I have deleted 177 cases.

Lastly, I restricted my sample based on year of birth. Although, the SHARE data should be exclusively for people over 50 years old, I found certain people, who were much younger at the time of the interview. Some of them were not even old enough to successfully complete university, so I restricted the samples

[15] The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of more than 85,000 individuals (approximately 150,000 interviews) from 19 European countries (+Israel) aged 50 or over. The SHARE data set was introduced in Borsch-Supan et al. (2006) and methodological details are in Borsch-Supan and Jurges (2005).

in a way that people are at least 35 years old at the time of interview. As a result of this restriction, 20 cases were deleted. The final data set has 7446 observations.

In Section 1.1 I described several institutional changes in the Czechoslovak schooling systems. Because my methodology heavily lies on the schooling system (for example, different streams for gifted or non-gifted children as described in Section 1.2), I will have to take into account these institutional changes. For this reason, I created not one, but four different data sets and I will test my hypothesis on all of them.

The first one contains all the observations and this is the benchmark data set, in which no institutional changes are taken into account. The second one will be adjusted by the biggest disruption in the data - the Nazi's shutdown of the universities in years 1939 to 1945. Despite the fact that I cannot know for sure, if a person was affected or not, I can significantly lower the probability that the affected person will not be in the data set, if I restrict the samples by deleting years 1916 to 1926. This procedure assumes that a student started their studies at the age of nineteen.

Ideally, I would only need observations from the time period in which all the people attained the very same schooling system. Hoverer, there is trade-off, since I am losing observations. From this reason, I will use two more data sets - the first one will start with the people born in the year 1935, so for students starting their studies at universities in 1954 or later and the second one will start with the people born in the year 1945, so for students starting their studies at universities in 1963 or later. The last data set contains only 3680, so it is half as large, which could have an influence on the significance of the coefficients in the model.

Even though the last data set has a much smaller number of observations, it is worth analyzing, because it has two crucial advantages. The first is already mentioned as institutional stability and the second is documented in the work of Doblhammer, van den Berg, and Fritze (2013). They investigate the determinants of cognitive ability among the elderly and show that economic conditions at birth significantly influence cognitive functioning later in life in various domains. Based on their research we can say that people born during World War II could be affected by such an effect and if the distribution of babies between months during the war was different to usual, which could be the case, because health could differ between months (Ueda et al., 2013) and distribution among month could change, because of the War. For example, there is evidence that exposure of pregnant women to negative events influences the health of newborns (Glynn et al. 2001; Simeonova, 2011; Xiong et al., 2008). Moreover, this effect is not random - it affect the most vulnerable, which are usually boys, so we cannot rule out that distributions among month were affected. A comparison of estimates from this restricted and from the benchmark data set could suggest, if the institutional changes or the War significantly affected the results, in other words, if the restriction is justified.

2.2 Variables measuring health

In order to perform my analysis, I need a variable measuring health. The SHARE data set contains 32 different variables concerned with health, which can be divided into five groups - cognitive function, mental health, physical health, behavioral risk (smoking, drinking or physical inactivity) and the measured values - grip strength and walking speed. It would be counterproductive to use all of these variables in my analysis, so I need to create a comprehensive health measure.

An ideal method for creating such a variable seems the principal component analysis (further stated PCA). In the SHARE data set, in the case of the health variables, lower values are usually 22) "good" (for example, "0" is best and "5" is worst), but variables like walking speed or grip strength are in meters per second or kilograms, in this case, naturally, more are better here. For this reason, I will recode all the variables to make sure that they all have the same interpretation - higher values are "good" and lower are "bad" (for example,"5" is best and "0" is worst). This recoding will enable easier interpretation of principal components.

2.2.1 Construction of health variables

In this chapter, I will explain the mechanism of the PCA. The intuition underlying the PCA is to capture most of the information in our observed variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]$ with a lesser number of new variables called principal components. This is done by finding a linear combination of original variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]$ with maximum variance. For the PCA we need standardized $\boldsymbol{X}$, because some variables have a bigger variance then others and this difference could affect the results.[16] Let me denote the linear combination by vector $\boldsymbol{u}_i = (u_1, u_2, \ldots, u_p)$, then we can express the PCA goal for the first component as maximizing the variance of the elements of $\boldsymbol{z}_1 = \boldsymbol{X}\boldsymbol{u}_1$. The second component will then capture the largest amount of variance, which is not already captured by the first component and these new variables are orthogonal (uncorrelated). For other components, if there are any, the PCA will do exactly the same thing.

For deriving the PCA, we will maximize $\boldsymbol{Z}$, which could be written as:

$$var(\boldsymbol{z}_i) = \frac{1}{n-1}\boldsymbol{u}_i'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{u}_i = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i \tag{2.1}$$

where $\boldsymbol{R}$ is equal to $\frac{1}{n-1}\boldsymbol{X}'\boldsymbol{X}$, which is the sample correlation matrix. Equation 2.1 now has a trivial solution - choose very large $\boldsymbol{u}_i$. For avoiding such a trivial solution we impose a constrain of unit length on the unit on the vector $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$. Such constrain optimization problems could be solved by the method of Lagrange multipliers, such as:

$$\mathcal{L} = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i - \lambda_i(\boldsymbol{u}_i'\boldsymbol{u}_i - 1) \tag{2.2}$$

where λ_i is called the Lagrange multiplier. The multiplier makes sure to penalize the objective function, if the equality constraint $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$ is not met. Taking the derivative of $\mathcal{L}$ with respect to the elements of $\boldsymbol{u}_i$ yields:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}_i} = 2\boldsymbol{R}\boldsymbol{u}_i - 2\lambda_i\boldsymbol{u}_i \tag{2.3}$$

setting Equation 2.3 to zero and solving yields:

$$\boldsymbol{R}\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i \tag{2.4}$$

This equation has a special structure: it is known as an eigenvalue - eigenvector problem, where $\boldsymbol{u}_i$ is called the eigenvector and λ_i the eigenvalue. Provided that the correlation matrix is full rank, the solution will consist of p positive eigenvalues and associated eigenvectors, which have a special relationship with

[16]Think, for example, on variable wage and number of schooling years. I expect that in most cases variable wage will have a much bigger variance and that would cause the PCA to focus more on such a variable. Variables with relatively large variances could dominate results in the PCA and that is the reason for the standardization.

the variance of the principal components:

$$var(\boldsymbol{z}_i) = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i = \boldsymbol{u}_i'\lambda_i\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i'\boldsymbol{u}_i = \lambda_i \tag{2.5}$$

where we are using fact that $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$. Thus, the eigenvalue λ_i is exactly the variance of associated principal component $\boldsymbol{z}_i$. Let $\boldsymbol{D}$ denote the diagonal covariance matrix of the principal components:

$$trace(\boldsymbol{D}) = trace(\boldsymbol{U'DU}) = trace(\boldsymbol{DU'U}) = trace(\boldsymbol{RI}) = trace(\boldsymbol{R}) = p \tag{2.6}$$

where we know that $trace(\boldsymbol{R})$ is simply the sum of the ones along the diagonal of the correlation matrix and p is the number of variables. Such property is useful for expressing a fraction of the total amount of variation accounted for by some subset of the principal components. Another useful by-product is a correlation matrix of the principal component scores $\boldsymbol{Z}$ with the original data $\boldsymbol{X}$. The correlation matrix is given by the following expression:

$$corr(\boldsymbol{X},\boldsymbol{Z}) = \frac{1}{n-1}\boldsymbol{X'Z}_s = \frac{1}{n-1}\boldsymbol{X'XUD}^{-\frac{1}{2}} = (\boldsymbol{UDU'})\boldsymbol{UD}^{-\frac{1}{2}} = \boldsymbol{UD}^{\frac{1}{2}} = \boldsymbol{F} \tag{2.7}$$

where $\boldsymbol{F}$ is referred to as component loadings, which are especially useful in interpretation of the PCA results.

So far I have assumed that my data is appropriate for using the PCA. For assessing in such a manner, we could use the Bartlett's sphericity test. This test directly addresses the question if correlation matrix should be factored. The test is an approximate chi-squared test:

$$\chi^2[\frac{(p^2-p)}{2}] = -[(n-1) - \frac{(2p+5)}{6}]ln|\boldsymbol{R}| \tag{2.8}$$

where ln $|\boldsymbol{R}|$ is a natural logarithm of the correlation matrix, $\frac{(p^2-p)}{2}$ is the number of degrees of freedom associated with the chi-square test statistic, p is the number of variables and n is the number of observations. A null hypothesis of the test is that the true correlation matrix of the underlying population is an identity matrix, so if we are unable to reject the null hypothesis, we can conclude that dimension reduction is inappropriate.

The important question in the PCA is how many components we should choose. Kaiser's rule advises to take all eigenvalues bigger than 1. Intuition behind the rule reflects the common sense notion that any principal component should account for at least as much variation as any of the original variables in $\boldsymbol{X}$. The second method is a graphical one called a scree plot. The approach involves plotting the eigenvalues for each principal component in order from the largest to the lowest ones. Then we look for an "elbow" in the curve, in which is a point after which the remaining eigenvalues decline in approximately linear fashion.

Despite all the benefits, the PCA method has one crucial assumption - it can only work on variables with ratio variables.[17] For example, the PCA cannot work on categorical data.[18] So, I will only select those variables, which meet the above mentioned conditions.

2.2.2 Included variables in the principal component analysis

As I mentioned above, the SHARE data contains 32 health related variables. However, not all of them are ideal for my analysis and also not all of them are mutually exclusive, so I will only use some variables, which are described below.

Physical well-being

Maximum of grip strength measure : Two grip strength measurements on each hand were recorded using a dynamometer

Number of chronic diseases : Number of selected chronic diseases

Self-perceived health - US version : 1 = poor, 5 = excellent

Mental health

Appetite : 0 = change in appetite, 1 = no change in appetite

CASP: quality of life and well-being index : The CASP score measures quality of life and is based on four subscales on control, autonomy, pleasure and self-realization. The CASP score is the sum of these four subscales and ranges from 12 to 48

[17] A ratio variable, has a clear definition of zero and the difference between a value of 100 and 90 is the same difference as between 90 and 80. Examples of ratio variables are height, weight or enzyme activity.

[18] A categorical variable, also called a nominal variable, is for mutually exclusive, but not ordered, categories. For example, you can code regions of some country.

Concentration : Information on difficulties with the concentration on a television program, film, radio program or reading (0 = yes)

Depression : 0 = depressed recently, 1 = not depressed recently

Depression scale : The depression scale measures the current depression as a composite index of twelve items: depressed mood, pessimism, suicidally, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. The scale ranges from 0 "very depressed" to 12 "not depressed at all". Variable is constructed as sum of variables Appetite, Concentration, Depression, Enjoyment, Fatigue, Guilt, Interest, Irritability, Pessimism, Sleep, Suicidally, Tearfulness

Enjoyment : 0 = no enjoyment recently, 1 = some enjoyment recently

Fatigue : Respondent had too little energy to do the things she/he wanted to do in the previous month (0 = yes)

Guilt : 0 = selected, 1 = not selected

Interest : 0 = changes in general interest, 1 = no change

Irritability : 0 = selected, 1 = not selected

Numeracy score - mathematical performance : Gives information on the Mathematical performance of the respondents - ranges from 1 (bad) to 5 (good)

Orientation to date, month, year and day of week : Orientation to date, month, year and day of week - ranges from 0 (bad) to 4 (good)

Pessimism : 0 = pessimistic, 1 = optimistic

Recall of words, first trial : Contains the number of words recalled in the first trial of the word recall task - ranges from 0 to 10

Recall of words, delayed : Contains the number of words recalled in the delayed word recall task - ranges from 0 to 10

Sleep : 0= problems with sleep, 1 = no problems with sleep

Suicidality : 0 = selected, 1 = not selected

Tearfulness : Did the respondent cry at all in the last month? (0 = yes)

Functional limits

Activities of the daily living index : The sum of the five tasks: dressing, bathing or showering, eating, cutting up food, walking across a room and getting in or out of bed. The higher the index is, the less difficulties with these activities and the higher the respondent's mobility. The index ranges from 0 to 5

Instrumental activities of the daily living index : The sum of telephone calls, taking medications, managing money, shopping for groceries and preparing a hot meal. The index ranges from 0 to 5 (low: has difficulties)

Fine motor skills index : The sum of picking up a small coin, eating/cutting up food and dressing. The index ranges from 0 to 3 (low: has difficulties)

Gross motor skills index : The sum of walking 100 meters, walking across a room, climbing one flight of stairs and bathing or showering. The index ranges from 0 to 4 (low: has difficulties)

Large muscle index : The sum of sitting two hours, getting up from chair, stooping, kneeling, crouching and pulling or pushing large objects. The index ranges from 0 to 4 (low = has difficulties)

Mobility index : The sum of walking 100 meters, walking across a room, climbing several flights of stairs and climbing one flight of stairs. The index ranges from 0 to 4 (low = has difficulties)

Behavioral Risks

Ever smoked daily : Indicates if a respondent ever smoked daily - cigarettes, cigars, cigarillos or pipe (0 = yes)

Days a week consumed alcohol in the last 3 months : Categorical, number of alcoholic drinks during the last six months (beer, cider, wine, spirits or cocktails) - ranges from 1 (often) to 7 (never)

Smoke at the present time : Indicates whether the respondents smoke cigarettes, cigars, cigarillos or a pipe at present (0 = yes)

Sports or activities that are vigorous : Gives information on the frequency of doing vigorous activities such as sports, heavy housework, or a job that involves physical labor - ranges from 1 (never) to 4 (often)

Not all aspects of health have to be affected by higher education in the same way, so I will create more than one health variable. The first one will capture all relevant variables; the second one will be created only from those concerned by physical health and functional limits; the third and the fourth one will, on the other hand, be concerned by mental health variables (I will present a restricted and extended version of a mental health variable). Then I will focus on correlations between particulars extraction from the PCA. For example, we can expect that the first variable from the first extraction will be mostly about physical health and the second one will be more about mental health or vice versa - that is my reason for exploring the correlations between principal components from different extractions, I want to know the similarities for better interpretation of the final results.

Note that in comparison with the SHARE data, I omitted one whole segment of variable - health care (doctor visits, hospital visits, stayed in nursing home, private care and so on). Those variables are rather endogenous in terms of health care. In most of the cases they are improving health (otherwise it would not make sense to undergo them), but people are not randomly visiting doctors/hospitals. Usually, sick people (people with lower health) are visiting doctors/hospitals or more generally using health care, so the interpretation of such variables would be rather difficult. Of course, in the case of uneven access to health care, this factor could be important, but in the Czech Republic (or former Czechoslovakia) the health care is financed through the public sector and it is open for most of the people without any mentionable barriers.

2.2.3 Principal component analysis - extractions

In this chapter, I will present individual extractions from the PCA. The fist one, physical health, will be presented in more detail for demonstration of the used methodology. The rest will not be as detailed.

Physical health Variable physical health is constructed from physical well-being (Self-perceived health - US version, Number of chronic diseases, Maximum of grip strength measure) and functional limits (Activities of daily living index, Instrumental activities of daily living index, Mobility index, Large muscle index, Gross motor skills index, Fine motor skills index). I am using nine variables in my analysis, so the Bartlett's test χ^2 has 36 degrees of freedom. In such case the critical value at a 99 percent level of confidence (i.e., $\alpha = 0.01$)

is approximately 59. The resulted statistic is approximately 30 000, which is much higher than the critical one. According to the standard tables the p-value is practically zero, so I am rejecting the H_0 - the dimension reduction is justified.

Table 2.1: Eigenvalues and total variance explained - Physical health

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	4.227	46.964	46.964
2	1.244	13.827	60.791
3	.867	9.633	70.424
4	.746	8.285	78.709
5	.600	6.663	85.371
6	.521	5.790	91.161
7	.443	4.923	96.084
8	.251	2.786	98.871
9	.102	1.129	100.000

Source: Authors' calculations based on SHARE data set.

In the PCA the number of variables is equal to the sum of eigenvalues. In the Table 2.1 we can see that the first component is able to explain almost 47% of the variation in the original set of nine variables. According to the scree plot located in Figure 2.1, it is clear that just the first components should be extracted. According to Kaiser's rule we should retain two principal components, however the second component is only slightly above the threshold of the eigenvalue equal one. In my case, I can use the third way of picking up a number of components from the PCA - interpretability of the extracted variable. For such a task, I can use component loadings, which are presented in Table 2.2. Component loadings are correlation between the original variables and the factors, moreover, their squares indicate what percentage of the variance in an original variable is explained by a factor. As you can see the first component is relatively easy to interpret. All of the original variables have the same scale and the bigger a value of a variable, the healthier a person is (except for the case of the number of chronic diseases, in which naturally more is worse). When we look at components loadings, we can see that all are positively correlated, except the number of chronic diseases, which is negatively correlated, in other words, the bigger value of the factor means a healthier person (or at least healthier in terms of the original variables). Furthermore, all values except maximum of grip strength measure are above 0.5,[19] which ensures that a big amount of the variance in the original variables are explained by the first factor. On the contrary, the second component is much harder to interpret, it probably contains less important information. To sum up, the first factor is ideal for extracting, the second is not, even though its eigenvalue is above one.

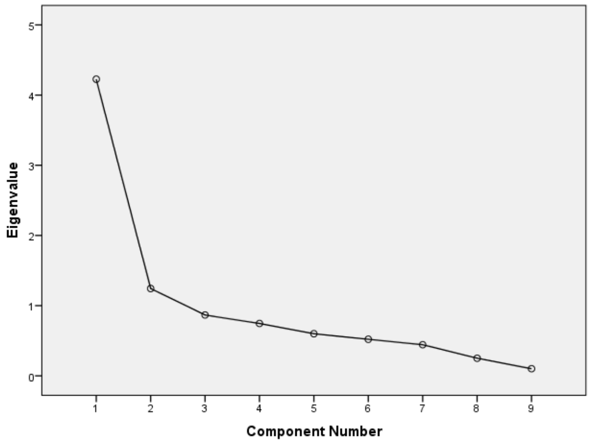

Figure 2.1: The scree Plot

Table 2.2: Component matrix and communalities

Component Loadings	1	2	Communalities
Self-perceived health - us version	.595	-.498	.602
Number of chronic diseases	-.540	.531	.574
Maximum of grip strength measure	.393	-.339	.269
Activities of daily living index	.779	.458	.816
Instrumental activities of daily living index	.628	.359	.523
Mobility index	.834	-.116	.708
Large muscle index	.756	-.234	.627
Gross motor skills index	.852	.164	.753
Fine motor skills index	.658	.407	.599

Source: Authors' calculations based on SHARE data set.

The first component is extracted by the method of Bartlett scores. This method considers the PCA equation as a system of regression equations, in which the original variables are the dependent ones, the factor loadings are the explanatory ones and the factor scores are the unknown parameters. Estimation

[19]When we look at communalities placed in Table 2.2, we can see, why maximum of grip strength measure is not so correlated in the first factor - the total amount of variance of the maximum of grip strength measure shared with all others variables included in analysis is much lower compared to others.

of the scores is obtained by the weighted least squares in order to account for heteroscedasticity. This method is used for other extractions as well.

Factors from all relevant variables (further stated Index of health) Now I will present a factor from all relevant variables, which are presented in Subsection 2.2.2. Overall I am using 30 variables in this analysis, so the Bartlett's test χ^2 has 435 degrees of freedom. The resulted statistic again has a huge value, approximately 55 000, which again means that p-value is practically zero and dimension reduction is justified.

In table Table 2.3 we can see the result from the PCA. Based on a similar procedure as described in the previous paragraph, I extracted the first two components.

Table 2.3: Eigenvalues and total variance explained - All variables

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	6.126	20.421	20.421
2	2.265	7.550	27.971
3	1.823	6.078	34.049
4	1.747	5.825	39.874
5	1.315	4.385	44.259
6	1.144	3.814	48.073
7	1.111	3.704	51.777
8	1.029	3.429	55.206
9	.895	2.984	58.190
10	.874	2.912	61.102

Note: Whole analysis contain 30 variables, but for sake of readability I here present only first ten.
Source: Authors' calculations based on SHARE data set.

Mental health For extraction of the factor mental health I used 17 variables,[20] which means that the Bartlett's test χ^2 has 136 degrees of freedom. The resulted statistic is approximately 20 000, so p-value is nearly zero and dimension reduction is justified. In Table 2.4 you can see details from the PCA. I extracted only the first component, which has all component loading around

[20] Appetite; CASP; Concentration; Depression; Enjoyment; Fatigue; Guilt; Interest; Irritability; Numeracy Score - mathematical performance; Orientation to date, month, year and day of week; Pessimism; Recall of words, first trial; Recall of words, delayed; Sleep; Suicidality; Tearfulness.

0.5, so interpretation is simple. The component has an eigenvalue of 4 and it explains around 23.5 percent of the variation in original 17 variables.

Table 2.4: Eigenvalues and total variance explained - Mental health

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	4.002	23.540	23.540
2	1.755	10.323	33.863
3	1.151	6.771	40.634
4	1.007	5.921	46.555
5	.920	5.413	51.968
6	.834	4.909	56.877
7	.825	4.856	61.732
8	.780	4.587	66.320
9	.762	4.483	70.802
10	.737	4.333	75.136

Whole analysis contains 17 variables, but for the sake of readability I here present only first ten.
Source: Authors' calculations based on SHARE data set.

Mental health - restricted Because in the first PCA used to extract the mental health factor I included 17 variables, which is quite a lot, I also created one factor called 'mental health restricted'. In this analysis I am using only four variables - recall of words (first trial), recall of words (delayed), orientation to date, month, year and day of week, numeracy score - mathematical performance. These variables are heavily correlated, so the Bartlett's test has a p-value of almost zero and dimension reduction is justified. The resulting factor is more focused on the brain's performance compared to the previous one, which focused on psychological state and the brain's performance. As you can see from the results, those four variables are ideal for dimension reduction, because the first eigenvalue is bigger than two, this result means that the first component contains more than 55 percent of the original variation. Moreover, the component loadings have all the same sign and all are in interval of 0.55 to 0.85.

Correlation between different components In Table 2.6 you can see correlations between extracted components. First and second component from the PCA with all relevant variables are uncorrelated, which is property of the

Table 2.5: Eigenvalues and total variance explained - Mental health, restricted

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	2.221	55.524	55.524
2	.811	20.273	75.797
3	.646	16.143	91.940
4	.322	8.060	100.000

Source: Authors' calculations based on SHARE data set.

PCA.[21] The rest of the results suggest that the first component is very good in terms of representability. It contains the majority of the variation of physical health, functional limits and mental health. The correlations with physical health and functional limits and unrestricted mental health components are especially very large (0,891 and 0,832). The correlation with the restricted mental health component is smaller (0.548). These results can be interpreted as follows: the first component mainly represents physical health, functional limits and psychological state, furthermore, it is also correlated with the brain's performance, but this relationship is a bit weaker. Overall the first component is positively correlated with all health related variables used in the analysis, except behavioral risk (since those negatively affect health).

Table 2.6: The correlation between extracted components

	Index of h. 1.	Index of h. 2.	PH and FL	MH - res.	MH
Index of h. 1	1	.000	.891**	.548**	.832**
Index of h. 2	.000	1	-.396**	.114**	.508**
PH and FL	.891**	-.396**	1	.353**	.499**
MH - res.	.548**	.114**	.353**	1	.644**
MH	.832**	.508**	.499**	.644**	1

Note: ** Correlation is significant at the 0.01 level (2-tailed). Whole 1 stands for the first component from the analysis performed on all 30 relevant variables. Similarly, Whole 2 is the second component, PH - physical health, FL - functional limits and MH - mental health. Source: Authors' calculations based on SHARE data set.

The second component, on the other hand, is negatively correlated with physical health and functional limits and positively correlated with both mental health components. Such a component is rather difficult to interpret in a

[21]Even through, as an extraction method I used Bartlett scores and this method does not guarantee that the factor scores will be uncorrelated even in the case of orthogonal solution.

regression analysis - for example, in the case of a positive relationship with some variables, we cannot know, if such a relationship is caused by a decrease of physical health and functional limits or an increase of mental health. For this reason, I will not use this component in a subsequent analysis.

To conclude this chapter, I created several components, which will be used as dependent variables representing health in a regression analysis. The first one is merely focused on physical health and functional limits. The second one is positively correlated with all my health related variables, except behavioral risk, since those are usually considered harmful to health (all other variables are coded in a way that a bigger value means bigger health, besides the number of chronic diseases). This component will be used as all-embracing *Index of health*.[22] The third and fourth are focused on mental health and one is specialized just on the brain's performance and the second is a compound from the physiological state and brain's performance.

2.3 Descriptive statistic

In Table 2.7, I present the descriptive statistics of the selected variables. The scales of many variables can be found in Subsection 2.2.2. You can observe that all four extracted components (*Index of health*, *Mental health component*, *Mental health component - restricted*, *PH and FL component*) have zero mean and standard deviation equals to one. These results are a direct cause of the PCA extraction.

Table 2.7 contains two variables concerned with education - *Levels of education* and *Years of education*. *Levels of education* are coded in following way:

0 : None

1 : Primary education or first stage of basic education

2 : Lower secondary or second stage of basic education

3 : (Upper) secondary education

4 : Post-secondary non-tertiary education

5 : First or second stage of tertiary education

[22]In the upcoming chapters, I am using exclusively this variable as a dependent, because the results using other variables are almost identical.

In most of the cases the level of education is what matters, it is important if a person completed high school or university, not how many years he studied. However, Lleras-Muney (2005) notes in relation to changes of compulsory school attendance that years of schooling matter too in the case of the education effect on health, so I am using both of them in my analysis.

Marital status with range from one to six is defined in following way:

1 : Married and living together with spouse

2 : Registered partnership

3 : Married, living separated from spouse

4 : Never married

5 : Divorced

6 : Widowed

In some cases I use marital status as a series of dummy variables - $Married$, for the sake of simplicity $Married$ include first three categories; $Nevermarried$; $Divorced$; $Widowed$.

When we look at the variable $Month\ of\ birth$, we can see that its average is 6.26, which is quite close to the expected value of 6.5 under the assumption of proper randomization. For the purpose of my analysis, I have recoded the $Month\ of\ birth$. In the recoded $Month\ of\ birth$, which I will use in the analysis, is August coded as 1 and September as 12. This change is convenient for an interpretation of the first-stage in the instrumental variable method. The variable basically indicate the effect of the $Month\ of\ birth$ relative to a school cut-off.[23] In Section 1.2 I described the relative age effect, so we already know that the people born in September have the highest advantage, hence the maximal value (assuming that the rules are strictly followed). So the hypothesis is that the people with the higher values of the $Months\ of\ birth$ (those born right after the cut-off day) will have more educations in terms of $Levels\ of\ education$ or $Years\ of\ education$. This hypothesis is used as a basic principle for the identification strategy presented in next chapter.

[23] Although, the usual way how to code a month of birth, when the goal is to estimate the relative age effect, is to calculate the average/median age in the cohort and then calculate the variable, with help of an age of the people, relatively to the average/median age in the cohort. This procedure in not applicable to my case, since in some cohorts there are just a few observations, so the deviation from a population average/median could be quite large.

Table 2.7: Descriptive statistics of the selected variables

Variables	N	Mean	Std. Dev.
Activities of daily living index	7414	4.83	0.64
Appetite	7302	0.93	0.26
CASP: quality of life and well-being index	7000	34.83	5.86
Concentration	7272	0.85	0.36
Days a week consumed alcohol last 3 months	7371	5.12	2.00
Depression	7282	0.63	0.48
Depression scale	7214	9.83	2.25
Enjoyment	7287	0.92	0.27
Ever smoked daily	7377	0.57	0.50
Fatigue	7277	0.69	0.46
Fine motor skills index	7414	2.90	0.38
Gender (female=1)	7446	0.57	0.49
Gross motor skills index	7414	3.73	0.75
Guilt	7275	0.94	0.24
Index of health (the PCA component)	6485	0.00	1.00
Instrumental activities of daily living index	7414	4.86	0.62
Interest	7291	0.92	0.27
Irritability	7284	0.77	0.42
Large muscle index	7417	3.25	1.13
Levels of education	7360	2.60	1.15
Marital status	7429	2.49	2.11
Maximum of grip strength	6857	35.41	11.70
Mental health component	6899	0.00	1.00
Mental health component - restricted	7220	0.00	1.00
Mobility index	7414	3.42	0.92
Month of birth	7431	6.26	3.46
Number of chronic diseases	7407	1.47	1.42
Numeracy score - mathematical	7370	3.49	1.13
Orientation to date, month, year and day of week	7418	3.76	0.73
Pessimism	7274	0.80	0.40
PH and FL component	6849	0.00	1.00
Recall of words, delayed	7253	3.53	2.01
Recall of words, first trial	7257	5.26	1.70
Self-perceived health - US version	7415	2.66	1.01
Sleep	7294	0.66	0.48
Smoke at the present time	7376	0.78	0.42
Sports or activities that are vigorous	7372	2.79	1.27
Suicidality	7275	0.91	0.28
Tearfulness	7291	0.80	0.40
Years of schooling	7415	12.00	3.06

Note: Variables with sign * is dummy with values ranging from 0 to 1.
Source: Authors' calculations based on SHARE data set.

Chapter 3

Identification Strategy

In order to estimate the causal effect of education on health, I need a situation in which I will have the very same two groups of people with different levels of education. With a simple OLS estimate we cannot tell if our relationship is causal or not. For instance, education can increase health because people could make better decisions or they have more information about health. On the other hand, poor health can cause a reduced education. Moreover, maybe there is a different variable affecting both, for example, discount rates or a genetic characteristic. In conclusion, we need some exogenous variation which will enable us to explore the relationship.

3.1 Instrumental variable methodology

For an estimation of the effect, I adopt the instrumental variable approach or in other words, the two-stage least squares estimator (TSLS).[24] The instrumental variable approach is designed for situations in which the treatment effect is affected by unobserved variables, for which we cannot control. An instrument only has an effect on a treatment - it manipulates it without fully controlling it.

More formally, an instrument has to have two basic properties. The first one is no correlation of the instrumental variable with ϵ - $Cov(\boldsymbol{Z}, \boldsymbol{\epsilon})$, so it has to be uncorrelated with unobserved factors that might influence a treatment. The second is correlation with the explanatory variable $\boldsymbol{X}$ - $Cov(\boldsymbol{X}, \boldsymbol{Z})$, because in

[24]For the sake of precision, the two stage least squares is an instrumental variables estimation method with more instruments than endogenous explanatory variables in the model.

case of an uncorrelated instrument, we cannot expect any significant results. I will test these two assumptions in the result section.

The instrumental variable method is a two step method, in which the first step is the education variable $\boldsymbol{X}$ regressed on the instrumental variable:

$$\boldsymbol{X} = \delta \boldsymbol{Z} + \boldsymbol{v} \tag{3.1}$$

where $\boldsymbol{Z}$ is the instrumental variable fulfilling the above mentioned conditions and $\boldsymbol{v}$ is an error term from the first stage. δ is then defined as:

$$\delta = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} \tag{3.2}$$

From the first stage we get predicted values $\hat{\boldsymbol{X}}$ defined as:

$$\hat{\boldsymbol{X}} = \hat{\delta}\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} \tag{3.3}$$

In order to avoid endogenity, these predicted values are used in the second stage instead of variable $\boldsymbol{X}$:

$$\boldsymbol{y} = \beta\hat{\boldsymbol{X}} + \boldsymbol{u} \tag{3.4}$$

where $\boldsymbol{y}$ is an endogenous variable and $\boldsymbol{u}$ is a random error. The instrumental variable estimates then take the following form:

$$\hat{\beta}_{2SLS} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \tag{3.5}$$

Intuitively we can say that we project the endogenous variable $\boldsymbol{X}$ onto the instrument $\boldsymbol{Z}$ and then this projection is used as an explanatory variable instead of $\boldsymbol{X}$.

You can observe that there are no control variables in either stage. The justification for such a procedure is simple. If the instrument is truly exogenous (there is no correlation with unobserved variables), then no additional controls are required in order to estimate the effect of the instrument on the treatment or in the second stage, the effect of the treatment on the dependent variable. Despite the fact that under given assumptions the coefficients are consistent, I will include some control variables, because control variables can improve the efficiency of the estimates.

3.2 Full sample instrumental variable approach

My first identification strategy is based on the use of the whole sample. As I will describe later, there are indications that my instrument is not exogenous, when the full sample is used, so I will also develop a second identification strategy, which will serve as a check for the first one and generally a comparison of the result can serve as a robustness check. The second identification strategy is methodologically similar to the first one.

Generally, my regression specification will look as the following:

$$Health_i = \beta_0 + \beta_1 Edu\hat{c}ation_i + \boldsymbol{X'}_i + \epsilon_i \tag{3.6}$$

$$Edu\hat{c}ation_i = \delta_0 + \delta_1 Monthofbirth_i + \boldsymbol{X'}_i + v_i \tag{3.7}$$

where in the first step I estimate education based on the $Month\ of\ birth$ and control variables,[25] and in the second step I estimate health based on fitted values of education (variables $Years\ of\ education$ or $Levels\ of\ education$). As described in the theoretical section, there is some justification for the effect of the birth month on education attainment. Despite that, it is important to check if the instrument is weak (small correlation between instrument and explanatory variable) or not. In the case of a weak instrument, the IV estimator δ has a high standard of error and inference using asymptotic approximations for the standard errors is not reliable. Also, with a weak instrument, even a very small correlation between the instrument and the error term may lead to significant inconsistencies (Bound, Jaeger, and Baker 1995). For a better illustration look at Equation 3.8, as the $Cov(\boldsymbol{X}, \boldsymbol{Z})$ is closing towards zero, the bias is increasing rapidly (the bias with the weak instrument could be potentially much bigger than with a OLS estimation). Furthermore, even though the instrumental variable estimates are consistent (assuming a valid instrument), they are always bias in a finite sample. When the instrumental variable is weak, this bias can be large, even in very large samples (Murray, 2006).[26] To put it simply, the instrumental variable approach with a weak instrument is not an optimal strategy. An easy rule of thumb states that an F-statistic below approximately 10 is indicative of a weak instrument problem (Staiger and Stock, 1997; Stock, Wright and Yogo, 2002).

[25] $Month\ of\ birth$ is coded as described in Section 2.3.

[26] This evidence is important for the interpretation of the instrumental variable estimates performed on small samples.

$$plim\hat{\beta} = \frac{Cov(\boldsymbol{Z}, \boldsymbol{\epsilon})}{Cov(\boldsymbol{X}, \boldsymbol{Z})} \frac{\sigma_{\epsilon}}{\sigma_x} \tag{3.8}$$

3.3 Partial sample instrumental variable approach

Based on recent evidence, the second assumption of no correlation between the instrument and unobservable variables may cause some problems. This said, I will test the correlation between the instrument and observable variables in the results section. I have to admit that even in the case of very low correlation, I cannot be sure if the instrument is strictly exogenous. Furthermore, motivated by previous literature, there is some evidence that the birth month is not strictly exogenous.

Firstly, Buckles and Hungerman (2013) explore the seasonality of births and consider a new variable - maternal characteristics.[27] They document large changes in the maternal characteristics of children born through the year, for example, winter births are often realized by teenagers and the unmarried. Their main result is that the season of birth is associated with later health and professional outcomes. If the characteristics of the parents are not random through the year, then the assumption of the instrumental variable approach could be violated. Oreopoulos et al. (2006) show that there is a causal effect of a parent's education on a child's education. Similar results are provided by Sewell and Shah (1968).

Secondly, there is some evidence that certain health indicators are correlated with birth month. For example, Ueda et al. (2013) shows association of the month of birth and risk of mortality in ages above 50 years on data from Sweden. Abel and Krueger (2010) investigate the association between the birth month and longevity for Major League Baseball players. They report that players born in the month of November had the greatest longevities while those born in June had the shortest one. Unfortunately, because this effect is indistinguishable from coefficient of interest δ, I cannot tell what drives my results in case of a positive coefficient. One solution could be to compare my results to the original study - Abel and Krueger (2010), but there are problems connected to such a procedure. They use a data set of Caucasian players from Major League Baseball, which is surely not a random sample of the population.

[27]Before this study, research on seasonality focuses on conditions at conception like climate (Seiver, 1985) or temperature (Lam, Miron and Riley, 1994). But those variables leave too much variation in the seasonality of births unexplained.

Moreover, even with random samples there is no guarantee that their results have some external validity for the Czech Republic (or former Czechoslovakia). In conclusion, I am not aware of any study, which investigates the effect of birth month on longevity in the Czech Republic (or former Czechoslovakia) and I cannot be sure about the external validity of the studies from different states. As a consequence, I need an additional identification strategy, which will not suffer from the seasonal effect, in order to check the robustness of my results from the first identification strategy.

The second strategy is not so different. It uses the same procedure, but a different data set. Instead of using people which are born during the whole year, I am going to use only those close enough to the cut-off. We know that in Abel and Krueger (2010) the biggest difference in longevity was between June and November, so perhaps, when we have the cut-off on the 1st September, the three months bandwidth on both sides is already too much. Therefore, I will use two basic bandwidths - with one and two months around the cut-off.

A possible problem with the identification strategy is non-compliers. I lack the data to identify the ratio of non-compliers in the period of my interest. But, for instance, Dobkin and Ferreira (2010) report that parents from higher socioeconomic classes are more likely to postpone the entrance of their children into school, using data from USA states California and Texas. Even through these results are not a good indicator for the identification strategy, I doubt that parents had such power over the school starting age in the Czechoslovak at that given time. It is more likely, that people were admitted or not based on the capacity possibilities of schools and parents did not significantly affect the school starting age.

I hypothesize that, if there were some non-compliers, in many cases they were just around the cut-off. Let me motivate this claim with a simple model - assume that skills (determining study results) are functions of maturity and ability. Ability is normally distributed between people and people are randomly born during a whole year.[28] Furthermore, the costs of non-compliance are nonzero and randomly distributed in size across a year.[29] In this setup, it is

[28]Of course, I already discussed that this assumption is not valid, but the nonrandom component is quite small, so it will not significantly affect this theoretical model.

[29]Nonzero costs are a fair assumption - you have to submit an application or at least discuss the matter with a particular representative of a school. Random distribution of costs in size throughout a year is a required assumption. If a person knows a representative of the authority responsible for delay applications, the cost would be lower. However, the prediction of the model is unaffected, if these changes in cost are uncorrelated with birth month and there is no reason to think otherwise.

clear that children right before cut-off will have biggest benefits from the delay, because the cost are on average same and benefit are increasing, if we are moving toward the cut-off (since ability is random and maturity is lowering).[30]

As a consequence, I will perform a further sample restriction in order to avoid most of the problem. Because of that I will add estimates, in which I exclude one month on both sides of the cut-off. The estimate could still be bias, but this bias will favor the persons before the cut-off, therefore in the case of positive and significant coefficients, I can state that the coefficient is lower than the actual one, in other words, in reality the effect of education on health is the same or even higher. This methodology is rather complement then substitute for the previously described methodologies using data around cut-off.

3.4 Overview

For better clarity I will present again every identification strategy with a small description and also approximate numbers of available observations, since actual regressions can have lower number of observation, because some variables have missing observations.

Whole sample : This specification uses all possible observations - 7431.

Two month bandwidth around cut-off : Due to indication that births are not distributed randomly through a year, only two months, with 2357 observations, around the cut-off are used.

One month bandwidth around cut-off : Same motivations as for the two month bandwidth applies for the one month bandwidth. In this case, 1149 observations are available.

Bandwidth of July and October : Because bandwidth just around cut-off are especially vulnerable to non-compilers, two month bandwidth excluding August and September is also used. It contains 1208 observations.

[30]For example, the study of the Department for Education (2010) from the United Kingdom is partly validating the model. Data shows that the majority of the reception admissions during spring and summer terms are summer-born students. In the UK there are three different terms for admission

Chapter 4

Results

4.1 Ordinary least squares results

Regressing health on educational outcomes by ordinary least squares regression (further stated OLS) must be expected to yield biased estimates rather than the causal effect of educational results on health. However, such analysis can be useful in at least two ways. Firstly, we can compare results from the OLS with results from the IV strategy and secondly we can establish, that there is significant relationship between education and health.

Table 4.1: The regression results from the ordinary least squares method

	Dependent variable			
	Index of health	PH and FL	MH	MH - rest.
Years of Schooling	0.069	0.041	0.081	0.103
	(0.004)***	(0.004)***	(0.004)***	(0.004)***
N	5416	5695	5748	5996
R^2	0.237	0.187	0.175	0.255
F	187***	145***	135***	228
Levels of Education	0.213	0.125	0.231	0.31
	(0.011)***	(0.011)***	(0.011)***	(0.010)***
N	5373	5646	5705	5949
R^2	0.25	0.192	0.183	0.279
F	199***	148***	141***	254***

Note: Standard errors are in parentheses; ***, **, * significance at the 1, 5 and 10% level. All models contains three groups of control variables: personal characteristic (gender and age), family background (household size, marital status, number of children and grandchildren) and parental status (mother and father still alive). Source: Authors' calculations based on SHARE data set.

Table 4.1 reports the estimated coefficients. The linear regressions look as following:

$$Health_i = \beta_0 + \beta_1 Education_i + \boldsymbol{X'}_i + \epsilon_i \tag{4.1}$$

where as a dependent variables I am using my extracted components representing health - $Index\ of\ health$ (generated from 30 variables), $Physical\ health\ and\ Functional\ limits$ (generated from nine variables), $Mental\ health$ (generated from 17 variables) and $Mental\ health\ restricted$ (generated from four variables). As Education variables I am using $Years\ of\ schooling$ in the first set of regressions and $Level\ of\ education$ in the second one. Furthermore, all models contains vector of control variables $\boldsymbol{X'}_i$, which can be divided into three parts. The first one is personal characteristic, in which gender and age are included, family background with variables household size, marital status, number of children and grandchildren is second one and parental status (mother and father still alive) is the third one.

When we look on the results, we can see that the education variables are significant in all specifications independently on how we define it ($Levels\ of\ education$ or $Years\ of\ schooling$), so education has an effect on all parts of health, but a mental health is much more affected. This result is not surprising, especially in case of the dependent variable $Mental\ health\ restricted$, which is primary focused on brain performance. With these particular results we probably face quite extensive selection bias - more educated people would have probably bigger brain performance even without their education. Nevertheless, the results indicate that there is strong correlation between health and education, as described in Section 1.3.

Before proceeding to the IV results, we should understand to the differences between the nature of the IV estimates and a true causal parameter. The causal parameter, also called an average treatment effect (ATE), is an measure of the average effect of **X** on **Y** compared with no treatment for a random draw from a population. Meanwhile, the IV estimates are called a local average treatment effect (LATE), which can be interpret as the local average effect of treatment on outcome compared with no treatment for a random draw from a subpopulation of compliers.

4.2 Validity of the instrument

In Chapter 3 I introduced two main assumptions for a valid instrument. For the valid instrument, it has to be both correlated with education and uncorrelated with unobserved factors influencing health. In other words, the validity of the instrument depends on the assumption that the month of birth has no direct effect on health or at least month of birth has no direct effect on health when we go sufficiently close to cut-off. Such assumption cannot be tested directly, but I can test whether it is correlated with observed variables that I believe might influence health.

Table 4.2: The correlation between the instruments and the observables

Sampling window	Whole	August, September	July - October	July, October
Index of health	-.015	.000	.029	-.025
PH and FL	.002	.005	.028	-.012
MH	-.009	-.022	.009	-.052
MH - rest.	-.003	-.045*	-.006	-.082**
Gender (female=1)	-.002	.002	-.030	.033
Age at interview (in years)	.010	-.008	-.064*	.044
Household size	.005	.032	.074*	-.009
Number of children	.022	-.015	.009	-.038
Number of grandchildren	-.009	-.034	-.060	-.010
Is natural mother still alive	.015	.004	.005	.003
Is natural father still alive	-.004	-.025	.004	-.051
Single	-.024*	-.017	-.022	-.012
Married	-.006	.015	.042	-.011
Divorced	.007	-.002	-.019	.014
Widowed	.013	-.009	-.027	.007
N	7446	2357	1149	1208

Note: ** Correlation is significant at the 0.01 level (2-tailed).* Correlation is significant at the 0.05 level (2-tailed).The set of variables is partly determined by data availability in the SHARE data set.
Source: Authors' calculations based on SHARE data set.

In Table 4.2 is displayed the simple correlations between the instrument and the full set of control variables for different sampling windows. Significant correlations are marked with one (five percent level) or two asterisks (one percent level). As the table shows, the maximum correlation equals 0.08 in absolute value (correlation with MH component in sampling window of July, October), this correlation is also the only one, which is significant on the one percent level. Except few, all other values are not significant and those that are significant are very close to zero. The instrument (driven by the month of birth) is thus unrelated to most of the variables, for example, to gender, number of children, number of grandchildren, parental status and to most of the marital status.

Table 4.3: Means of Years of schooling conditionally on birth month

Sampling window	Whole	1	2	3
January	11.91	11.94	12.05	12.20
February	12.09	12.09	12.23	12.31
March	11.96	12.05	12.19	12.35
April	12.02	12.11	12.25	12.55
May	12.13	12.22	12.34	12.55
June	12.00	12.10	12.24	12.17
July	12.19	12.27	12.27	12.54
August	11.80	11.86	11.97	12.15
September	11.92	12.03	12.10	12.31
October	11.85	11.97	12.10	12.28
November	11.94	11.91	11.95	12.19
December	12.12	12.23	12.39	12.75
Total	12.00	12.07	12.17	12.36
N	7405	7026	6004	3996

Note: In the first sampling window, years 1916 - 1926 are excluded, the second one is ranging from 1935 and the third one is ranging from 1945.
Source: Authors' calculations based on SHARE data set.

In Table 4.3 and Table 4.4 I present means of $Years\ of\ schooling$ and $Levels\ of\ education$ conditionally on birth month. The means are presented in four columns and each of them represents a different time period. The first one contains all observations, in the second one years 1916 - 1926 are excluded, the third and fourth one are ranging from 1935 and 1945, respectively.

Ideally, we would like to see discontinuity between August and September. Although, almost in all cases values representing September are bigger that August, the differences are not so large, moreover, values for September are not among largest. Those result does not seem optimistic, but I will proceed in my analysis. It is possible that the relative maturity effect is not so important for all periods, as I am discussing in Section 1.1, or it is not so important for all subgroups.

Table 4.4: Means of Levels of education conditionally on birth month

Sampling window	Whole	1	2	3
January	2.56	2.57	2.61	2.63
February	2.64	2.64	2.70	2.68
March	2.60	2.63	2.65	2.65
April	2.64	2.66	2.67	2.71
May	2.61	2.64	2.68	2.75
June	2.60	2.62	2.68	2.59
July	2.65	2.66	2.68	2.71
August	2.54	2.55	2.61	2.58
September	2.58	2.61	2.65	2.70
October	2.56	2.59	2.61	2.63
November	2.58	2.57	2.59	2.61
December	2.63	2.66	2.68	2.76
Total	2.60	2.62	2.65	2.67
N	7348	6967	5960	3965

Note: In the first sampling window, years 1916 - 1926 are excluded, the second one is ranging from 1935 and the third one is ranging from 1945.
Source: Authors' calculations based on SHARE data set.

4.3 Instrumental variable results

In Table 4.5 we can see results from the first stage, where as a dependent I used $Years\ of\ schooling$. The table presents four different specifications. The first specification is just the instrument, personal characteristic (gender and age) are added in the second one, family background in the third one (household size, marital status, number of children and grandchildren) and parental status (mother and father still alive) in the last one.

For assessing the question about weak instrument, we should look on the first specification. Unfortunately, the results are not optimistic, since the biggest value of a F-statistic is 2.3.[31] Based on the results we can safely state that the instrument birth month is not valid for any sampling window. When, we add some other exogenous variables (specification two, three and four), we can see that in two cases the birth month is statistically significant, but it has different sign than expected (people born after cut-off has less education on average). There are two points, which should be discussed. Firstly, it is not unlikely that we are facing a type I error.[32] Both coefficients are significant at 10% level and the table contains 16 different regressions, so we can expect one

[31]The threshold we are seeking is 10.
[32]The type I error occurs when a null hypothesis is rejected, but it is in fact true; that is, the null hypothesis is wrongly rejected.

Table 4.5: The first-stage results, Years of schooling as a dependent

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Birth month	-0.00351	-0.0813	0.152	-0.292
	(-0.31)	(-0.60)	(0.80)	(-1.52)
N	6472	2062	989	1073
F	0.0982	0.360	0.634	2.307
R^2	0.0000152	0.000175	0.000642	0.00215
Specification 2				
Birth month	-0.00456	-0.0992	0.0780	-0.251
	(-0.41)	(-0.74)	(0.41)	(-1.33)
N	6472	2062	989	1073
F	76.40	25.35	10.38	15.42
R^2	0.0342	0.0356	0.0306	0.0415
Specification 3				
Birth month	-0.00629	-0.140	0.0953	-0.340*
	(-0.56)	(-1.04)	(0.51)	(-1.77)
N	6046	1949	933	1016
F	49.45	17.58	11.57	8.282
R^2	0.0687	0.0755	0.101	0.0690
Specification 4				
Birth month	-0.00765	-0.137	0.149	-0.383*
	(-0.65)	(-0.96)	(0.76)	(-1.86)
N	5416	1731	833	898
F	36.64	13.84	10.11	5.844
R^2	0.0694	0.0814	0.119	0.0676

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The specification 1 contains only the instrument, in the second specification I add a personal characteristic (gender and age), a family background is added in the third one (household size, marital status, number of children and grandchildren) and a parental status is the fourth supplement (mother and father still alive).
Source: Authors' calculations based on SHARE data set.

Table 4.6: The second-stage results, Index of health as a dependent, The first stage - Years of schooling

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Years of schooling	-0.296	0.0211	0.331	0.182
	(-0.19)	(0.04)	(0.71)	(0.82)
N	6472	2062	989	1073
F	0.0344	0.00167	0.505	0.677
Specification 2				
Years of schooling	-0.0163	0.244	-0.147	0.136
	(-0.02)	(0.54)	(-0.18)	(0.59)
N	6472	2062	989	1073
F	436.6	109.2	44.30	75.45
Specification 3				
Years of schooling	0.0172	0.263	-0.277	0.130
	(0.03)	(0.77)	(-0.33)	(0.74)
N	6046	1949	933	1016
F	157.7	34.24	9.521	26.28
Specification 4				
Years of schooling	-0.173	0.192	-0.210	0.0604
	(-0.29)	(0.58)	(-0.41)	(0.37)
N	5416	1731	833	898
F	76.64	32.57	9.890	20.87

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The specification 1 contains only the instrumented variable (Years of schooling), in the second specification I add a personal characteristic (gender and age), a family background is added in the third one (household size, marital status, number of children and grandchildren) and a parental status is the fourth supplement (mother and father still alive).
Source: Authors' calculations based on SHARE data set.

or two instances of the type I error. Secondly, all exogenous variables should be tested by an overidentifying restrictions test. For results from this test, see Subsection 4.3.1.

Before we will proceed towards the possible causes and potential solutions of the not-significant relationship between the instrumental and the instrumented variable, we should pay attention to Table 4.6, which reports the second-stage of the instrumental variable results. Since in the instrumental variable method, the birth month manipulates with education and affects health only indirectly through these manipulations, the fact that all coefficients, which stand for education (the $Years\ of\ schooling$ in this case), are insignificant. These results are expected. The last note, which applies to all the following second-stage results presented in this study, is about R^2. This statistic is not presented in the tables, because in case of the instrumental variable estimator it is not interpretable, for example, it is not an exception, when R^2 is negative in the instrumental variable method.[33]

There are six possible causes of the non-significant results in the first-stages, which can be divided into two subgroups - those, which cannot be solved in my analysis and those, which can be. It is not unlikely that in reality there will be some combination of the below mentioned effects.

Let's start with three, which cannot be repaired or identified on the current data set.

The relative maturity effect does not exist at a time of a high school entrance test

It is not unlikely that the Czechoslovak schooling system was designed in a way, that the birth month did not play an important role. As was mentioned in the context of Bedard and Dhuey (2006), the power of the effect is decreasing over time. In order to have a significant birth month effect, we need different performance streams. In the Czechoslovak schooling system, the high schools can be taken as those performance streams, but in this time it could already be too late - we cannot exclude the possibility that the relative maturity effect is not significant at a time of a high school entrance exam.[34] A Counterargument for this hypothesis can be the result of Bedard and Dhuey (2006), because as I mentioned,

[33] In the instrumental variable approach, the residual sum of squares (RSS) does not have to be smaller than the total sum of squares (TSS), because the model's residuals are computed over a set of regressors different from those used to fit the model.

[34] For more references I can suggest Elder and Lubotsky (2009). They argue that the relative maturity effect is a demonstration of skills that were acquired prior to start of compulsory school and have no effect on the rate of learning.

they perform their analysis on the Czech Republic too. They find quite a strong effect even in eight grade, but we need to know, that their data is from a different time, so the comparison could be misleading.

Too many non-compliers The data set 'SHARE' does not contain information about the actual starting age, so, for example, if many people born before September do not comply with the rule (they delay their start in school), then they will be the oldest in a class and results will have a downward bias. This hypothesis is supported by the fact that in many regresions the $Month\ of\ birth$ has a negative sign (non-compliers will have downward bias results). High schools in the Czechoslovakia cannot be considered as streams: This hypothesis basically states that a rate of learning was the same at each high school in Czechoslovakia. There was no significant advantage of going to a particular high school (for example, a gymnasium) in terms of maximization of education attainment. This hypothesis is rather unlikely - even in the time of the communist era, people select themselves into high schools, because some provide them a higher change for college (gymnasiums).

Remaining issues are testable with my data set, so the rest of the paper will be focused on those.

Many persons repeat grade People born during the summer are the youngest in a class, so they have the biggest probability of a repeating grade, which would increase the $Years\ of\ schooling$. However, this hypothesis is not supported by evidence in Table 4.4, where we can see that the pattern of the $Level\ of\ education$ is very similar to the pattern in the $Years\ of\ schooling$ (in both cases the value for September is bigger than the value for August, but lower than the value for July). For the sake of completeness, in Table A.1 and Table A.2 you can also see the result from the first and second stage, where the $Level\ of\ education$ is used as an instrumented variable. The results are similar to those where the $Years\ of\ schooling$ is used.

Changes in the institutional background Here I refer to Section 1.1. The institutional background (types of high school, open access to universities, length of basic schooling and so on) can have an influence on the existence of the relative maturity effect. This hypothesis is tested in Subsection 4.3.2.

Heterogeneity in the relative maturity effect The hypothesis that the relative maturity effect is present only in some subgroups of the population is motivated by a work of Zhong (2012). She found that the male-specific relative age effect becomes insignificant in the fifth grade, but the female relative age effect will outlast this up until the eighth grade.[35] Furthermore, she found out, that the effect differs in terms of parents' education. Elder and Lubotsky (2007). Children of more educated parents are more likely to seize their extra time, because their parents can provide them indispensable guidance. Unfortunately, I do not possess information about parents' education. In chapter Subsection 4.3.3, I present estimates for both genders separately and for respondents with father or mother alive.

4.3.1 Selection of an optimal specification

In this section I am using the test of overidentifying restrictions to determine an ideal specification. In the previous section I presented results from the instrumental variable approach and I used four different specifications in which I am adding additional exogenous (subject of the test, if true) controls to the first and also the second-stage.[36] As I already discussed, additional controls can improve the efficiency of the estimate, but it is crucial to use only truly exogenous ones. This matter can be solved with help of Sargan's (1958) and Basmann's (1960) tests of overidentifying restrictions.

In order to perform this test we need more instruments than regressors, so we need an overidentified equation. A joint null hypothesis that the instruments are valid, i.e., uncorrelated with an error term. A rejection, of course, casts doubt on a validity of the instruments.

Resulting statistics reject the null hypothesis in all specifications (except the first one, in which the test cannot be performed). I perform the test further on several combinations of variables. Those additional tests show indications that a few instruments are valid - birth month, marital status or household size. For the other variables, in many cases it is not surprising. For example, the variable that indicates if parents are alive - longevity is correlated with education and

[35] She is able to distinguish between the different effects described in chapter Subsection 1.2.2. For that reason I changed the terminology from the relative maturity effect (my label for the effects described in chapter Subsection 1.2.2) to the relative age effect (part of the previously mentioned).

[36] Study of Puhani and Weber (2006) use the same procedure.

educated parents are just more likely to have birth at a particular time of the year.

As a result, I will not use any controls in regressions in the following chapters.

4.3.2 Inclusion of the changes in an institutional background

In chapter Section 1.1 I explained the main institutional changes in the period of my interest. Those changes can influence a streaming process. In some cases, it is practically certain - for example, the closing of Czechoslovak universities during years 1939 to 1945. In Table 4.7 you can see the second-stage results of the instrumental variable approach, and in Table A.3 you can see the results from the first stage. In both tables, there is not even one statistically significant regression. Due to insignificant results, we can safely decline the hypothesis that in some time periods the relative maturity effect was statistically significant.

4.3.3 Heterogeneity in the relative maturity effect

In the text above I already presented the motivation behind the hypothesis of the heterogeneous effect of the relative maturity effect. Motivated by Zhong (2012) I present the results for the samples of girls and boys. As a proxy for parents education I am using if a mother or father is alive. Conveniently, those variables are not correlated with age as one could expect (0.35 for father alive and 0.45 for mother alive).

In Table 4.8 you can see the second stage of the instrumental variable approach and in Table A.4 you can see the first stage. Again, not even one statistically significant result.

Table 4.7: The second-stage results, Index of health as a dependent, The specifications are based on the institutional background

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Years of schooling	-0.296	0.0211	0.331	0.182
	(-0.19)	(0.04)	(0.71)	(0.82)
N	6472	2062	989	1073
F	0.0344	0.00167	0.505	0.677
Specification 2				
Years of schooling	-0.310	-0.398	0.385	-0.0028
	(-0.37)	(-0.36)	(0.526)	(0.23)
N	6205	1967	938	1029
F	0.138	0.129	0.46	0
Specification 3				
Years of schooling	-0.0274	0.332	-0.326	0.168
	(-0.03)	(0.68)	(-0.19)	(0.69)
N	960	323	154	169
F	0.000817	0.459	0.0355	0.473
Specification 4				
Years of schooling	-0.291	-0.0620	8.002	0.0262
	(-0.55)	(-0.20)	(0.03)	(0.12)
N	2635	887	443	444
F	0.299	0.0380	0.000673	0.0137

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The first specification is based on a regular data set, the specification 2 includes the people born before 1916 and after 1926, the third specification contains the people born after 1935 and the fourth those, which are born after 1945.
Source: Authors' calculations based on SHARE data set.

Table 4.8: The second-stage results, Index of health as a dependent, Heterogeneous effects

Sampling window	Whole	August, September	July - October	July, October
Females				
Years of schooling	0.0978	-0.0960	0.340	0.0583
	(0.18)	(-0.10)	(0.29)	(0.14)
N	3672	1214	582	632
F	0.0341	0.00948	0.0818	0.0196
Males				
Years of schooling	1.063	0.0983	0.331	0.257
	(0.19)	(0.15)	(0.65)	(0.81)
N	2800	848	407	441
F	0.0344	0.0213	0.421	0.653
Mother alive				
Years of schooling	-0.0135	0.632	-0.0735	0.0979
	(-0.03)	(0.16)	(-0.18)	(0.33)
N	1428	471	234	237
F	0.000653	0.0248	0.0318	0.105
Father alive				
Years of schooling	8.031	0.358	0.238	0.293
	(0.02)	(0.29)	(0.81)	(1.15)
N	585	196	88	108
F	0.000389	0.0848	0.639	1.307

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level.
Source: Authors' calculations based on SHARE data set.

Conclusion

This paper investigates the causal impact of education on health. I applied an instrumental variable approach, in which as an instrument I used a variability in the birth month to predict education attainment. Despite my best attempts, I was unsuccessful in my search for significant results.

I identified six hypothesis focused on the reason of the insignificant results. I was able to reject three of them and since one is rather unlikely (High schools do not differ in quality - they cannot be considered as streams) only two remain. The first one, a huge proportion of non-compliers is something which could be repaired with better data and a second one - the birth month effect is not significant in a time of a high school entrance exam.

Just to show the importance of the research question, let's for a moment assume that the last hypothesis is true for schooling system, in which streaming is active. Such a setup could answer one of the main question connected with the economics of schooling: do schools serve as a screening system (signaling model - Spence, 1973) or do they actually increase human capital (human capital theory - Becker, 1975). To sum up, if the relative maturity effect diminishes quickly, a system is based on a signaling effect, rather than on human capital accumulation.

Future research should primarily address improving an estimation methodology (mainly the above mentioned proportion of non-compliers). Economic theory can still benefit from the relative maturity effect research (for example, correcting returns on education or signaling vs. human capital production model) and similarly public policy makers can use this research in order to minimize inefficiencies connected with the selection processes. The second area for the future research is an implementing the quality of education (measured by the reputable school ratings, for example) into the relationship of education and health.

In terms of the causal effect of education on health it is essential to increase our knowledge about the mechanisms through which education operates on health.

Bibliography

Abel, Ernest L., and Michael L. Kruger. 2010. "Birth Month Affects Longevity." *Death Studies* 34 (8): 757-763.

Allen, Jeremiah, and Barnsley, Roger. 1993. "The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection." *The Journal of Human Resources* 28 (3): 649-659.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4): 979-1014.

Auster, Richard, Irving Leveson, and Deborah Sarachek. 1969. "The Production of Health, an Exploratory Study." *Journal of Human Resources* 4 (4): 411-436.

Basmann, R. L. 1960. "On finite sample distributions of generalized classical linear identifiability test statistics." *Journal of the American Statisical Association* 55 (292): 650-659.

Becker, G. S. *Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education.* 2d ed. New York: Columbia University Press for NBER, 1975.

Bedard, Kelly, and Elizabeth Dhuey. 2006. "The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects." *The Quarterly Journal of Economics* 121 (4): 1437-1472.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *The Review of Economics and Statistics* 93 (2): 455-467.

Börsch-Supan, Axel, and Hendrik Jürges. 2005. "The Survey of Health, Aging and Retirement in Europe-Methodology." *Mannheim : Mannheim Research Institute for the Economics of Ageing.*

Börsch-Supan, Axel, and Hendrik Jürges. 2006. "Early Retirement, Social Security and Well-being in Germany". *National Bureau of Economic Research.*

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90 (430): 443-450.

Buckles, Kasey S., and Daniel M. Hungerman. 2012. "Season of Birth and Later Outcomes: Old Questions, New Answers." *Review of Economics and Statistics* 95 (3) (July 17): 711-724.

Cahan, Sorel, and Nora Cohen. 1989. "Age Versus Schooling Effects on Intelligence Development." *Child Development*: 1239-1249.

Carneiro, P., J. Heckman, and E. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6): 2754-81.

Crawford, Claire, Lorraine Dearden, and Costas Meghir. 2007. "When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England." *Centre for the Economics of Education*, London School of Economics and Political Science.

Crawford, Claire, Lorraine Dearden, and Ellen Greaves. 2011. "Does When You Are Born Matter? The Impact of Month of Birth on Children's." *Institute for Fiscal Studies.*

Crawford, Claire, Lorraine Dearden, and Ellen Greaves. 2013. "The Impact of Age within Academic Year on Adult Outcomes." *IFS Working Papers.*

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." *Handbook of the Economics of Education* 1: 697-812.

Cutler, David M., and Adriana Lleras-Muney. 2010. "Understanding Differences in Health Behaviors by Education." *Journal of Health Economics* 29 (1): 1-28.

Department of Education. *Month of Birth and Education : Schools Analysis and Research Division.* 2010.

Dixon, Jess, Sean Horton, and Patricia Weir. 2011. "Relative Age Effects: Implications for Leadership Development." *International Journal of*

Sport & Society 2 (2): 1-15.

Dobkin, Carlos, and Fernando Ferreira. 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review* 29 (1): 40-54.

Doblhammer, Gabriele, Gerard J. van den Berg, and Thomas Fritze. 2013. "Economic Conditions at the Time of Birth and Cognitive Abilities Late in Life: Evidence from Ten European Countries." *PloS One* 8 (9).

Du, Qianqian, Huasheng Gao, and Maurice D. Levi. 2012. "The Relative-Age Effect and Career Success: Evidence from Corporate CEOs." *Economics Letters* 117 (3): 660-62.

Elder, Todd E., and Darren H. Lubotsky. 2009. "Kindergarten Entrance Age and Children's Achievement Impacts of State Policies, Family Background, and Peers." *Journal of Human Resources* 44 (3): 641-683.

Fredriksson, Peter, and Bjorn Ockert. 2005. "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance."

Gladwell, M. *Outliers : The Story of Success*, Little, Brown and Company, 2008.

Glynn, Laura M., Pathik D. Wadhwa, Christine Dunkel-Schetter, Aleksandra Chicz-DeMet, and Curt A. Sandman. 2001. "When Stress Happens Matters: Effects of Earthquake Timing on Stress Responsivity in Pregnancy." *American Journal of Obstetrics and Gynecology* 184 (4): 637-42.

Grossman, Michael, and Robert Kaestner. 1997. "Effects of Education on Health." *The Social Benefits of Education* (Ann Arbor: University of Michigan Press).

Heckman, James J. 2006. "Skill Formation and the Economics of Investing in Disadvantaged Children." *Science* 312 (5782): 1900-1902.

King, Inez B. 1955. "Effect of Age of Entrance into Grade I upon Achievement in Elementary School." *The Elementary School Journal*, 331-36.

Lam, David A., Jeffrey A. Miron, and Ann Riley. 1994. "Modeling Seasonality in Fecundability, Conceptions, and Births." *Demography* 31 (2): 321-346.

Leopold, Liliya, and Henriette Engelhartdt. 2013. "Education and Physical Health Trajectories in Old Age. Evidence from the Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Public Health* 58 (1): 23-31.

Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *The Review of Economic Studies* 72 (1): 189-221.

Musch, Jochen, and Simon Grondin. 2001. "Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport." *Developmental Review* 21 (2): 147-167.

Murray, Michael P. 2006. "Avoiding Invalid Instruments and Coping with Weak Instruments." *The Journal of Economic Perspectives* 20 (4): 111-32.

Newhouse, Joseph P., and Rand Corporation Insurance Experiment Group. 1993. "Free for All?: Lessons from the RAND Health Insurance Experiment." *Harvard University Press.*

Oreopoulos, Philip, Marianne E. Page, and Ann Huff Stevens. 2006. "The Intergenerational Effects of Compulsory Schooling." *Journal of Labor Economics* 24 (4): 729-760.

Puhani, Patrick A., Weber, Andrea M. 2006. "Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany," *University of St. Gallen Department of Economics working paper series* 2006: 2006-02.

Sargan, J.D. 1958. "The estimation of economic relationships using instrumental variables." *Econometrica* 26: 393-415.

Seiver, Daniel A. 1985. "Trend and Variation in the Seasonality of US Fertility, 1947-1976." *Demography* 22 (1): 89-100.

Sewell, William H., and Vimal P. Shah. 1968. "Parents' Education and Children's Educational Aspirations and Achievements." *American Sociological Review*: 191-209.

Silles, Mary A. 2009. "The Causal Effect of Education on Health: Evidence from the United Kingdom." *Economics of Education Review* 28 (1): 122-28.

Simeonova, Emilia. 2011. "Out of Sight, Out of Mind? Natural Disasters and Pregnancy Outcomes in the USA." *CESifo Economic Studies* 57 (3): 403-31.

Spence, M. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87 (3): 355-374.

Solli, Ingeborg. 2011. "Left Behind by Birth Month." *University of Stavanger*, *Mimeo.*

Staiger, Douglas O., and James H. Stock. 1994. "Instrumental Variables Regression with Weak Instruments." *National Bureau of Economic Research Cambridge*, Mass., USA.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business* & *Economic Statistics* 20 (4).

Ueda, Peter, Anna-Karin Edstedt Bonamy, Fredrik Granath, and Sven Cnattingius. 2013. "Month of Birth and Mortality in Sweden: A Nation-Wide Population-Based Cohort Study." *PloS One* 8 (2).

World Bank. 2013. *World Development Indicators 2013.* Washington, DC: World Bank.

Xiong, X. U., Emily W. Harville, Donald R. Mattison, Karen Elkind-Hirsch, Gabriella Pridjian, and Pierre Buekens. 2008. "Exposure to Hurricane Katrina, Post-Traumatic Stress Disorder and Birth Outcomes." *The American Journal of the Medical Sciences* 336 (2): 111.

Zhong, S., Y., 2012. "The Effects of Relative Age on Early Childhood Academic Achievement: How they Differ Between Gender and Change Across Time." *Bachelor's Thesis*, Stanford University.

Norberg, Karen. 2004. â€śPartnership Status and the Human Sex Ratio at Birth.â€ť Proceedings of the Royal Society of London. Series B: Biological Sciences 271 (1555): 2403â€“10.

Appendix

Table A.1: The first-stage results, Levels of schooling as a dependent

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Birth month	-0.000184	0.0599	-0.00943	-0.0725
	(-0.04)	(0.83)	(-0.18)	(-1.01)
N	6419	2045	983	1062
F	0.00190	0.0342	0.689	1.017
R^2	0.000000296	0.0000167	0.000702	0.000958
Specification 2				
Birth month	-0.000352	0.0383	-0.0124	-0.0602
	(-0.08)	(0.53)	(-0.25)	(-0.85)
N	6419	2045	983	1062
F	37.36	12.74	5.868	7.906
R^2	0.0172	0.0184	0.0177	0.0219
Specification 3				
Birth month	-0.00124	0.0462	-0.0373	-0.111
	(-0.29)	(0.65)	(-0.74)	(-1.56)
N	5999	1933	927	1006
F	51.04	19.94	12.10	9.953
R^2	0.0712	0.0854	0.106	0.0825
Specification 4				
Birth month	-0.00171	0.0437	-0.0425	-0.126*
	(-0.39)	(0.59)	(-0.81)	(-1.69)
N	5373	1717	829	888
F	38.24	15.49	8.958	7.843
R^2	0.0728	0.0908	0.108	0.0897

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The specification 1 contains only the instrument, in the second specification I add a personal characteristic (gender and age), a family background is added in the third one (household size, marital status, number of children and grandchildren) and a parental status is the fourth supplement (mother and father still alive).

Source: Authors' calculations based on SHARE data set.

Table A.2: The second-stage results, Index of health as a dependent, First stage - levels of schooling

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Levels of schooling	-5.791	0.163	0.833	0.725
	(-0.04)	(0.04)	(0.73)	(0.74)
N	6419	2045	983	1062
F	0.00171	0.00139	0.537	0.553
Specification 2				
Levels of schooling	0.201	2.094	-0.377	0.576
	(0.02)	(0.25)	(-0.22)	(0.56)
N	6419	2045	983	1062
F	485.7	21.20	44.21	69.21
Specification 3				
Levels of schooling	0.0915	1.023	-0.581	0.415
	(0.03)	(0.67)	(-0.36)	(0.78)
N	5999	1933	927	1006
F	161.5	24.75	11.61	27.25
Specification 4				
Levels of schooling	-0.683	0.696	-0.801	0.200
	(-0.22)	(0.61)	(-0.39)	(0.41)
N	5373	1717	829	888
F	57.21	28.93	6.999	21.85

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The specification 1 contains only the instrumented variable (Levels of education), in the second specification I add a personal characteristic (gender and age), a family background is added in the third one (household size, marital status, number of children and grandchildren) and a parental status is the fourth supplement (mother and father still alive).
Source: Authors' calculations based on SHARE data set.

Table A.3: The first-stage results, Years of schooling as a dependent, Specifications based on institutional background

Sampling window	Whole	August, September	July - October	July, October
Specification 1				
Birth month	-0.00351	-0.0813	0.152	-0.292
	(-0.31)	(-0.60)	(0.80)	(-1.52)
N	6472	2062	989	1073
F	0.0982	0.360	0.634	2.307
R^2	0.0000152	0.000175	0.000642	0.00215
Specification 2				
Birth month	-0.00678	-0.0683	0.144	-0.255
	(-0.60)	(-0.50)	(0.195)	(0.193)
N	6205	1967	938	1029
F	0.357	0.246	0.55	1.74
R^2	0.0000576	0.000125	0.0006	0.0017
Specification 3				
Birth month	0.0125	-0.300	0.151	-0.768
	(0.39)	(-0.82)	(0.32)	(-1.38)
N	960	323	154	169
F	0.156	0.666	0.103	1.906
R^2	0.000163	0.00207	0.000678	0.0113
Specification 4				
Birth month	-0.0173	-0.239	0.00726	-0.494
	(-0.95)	(-1.15)	(0.03)	(-1.61)
N	2635	887	443	444
F	0.905	1.313	0.000659	2.589
R^2	0.000344	0.00148	0.00000149	0.00582

Note: t-statistic are in parentheses; ***, **, * significance at the 1, 5 and 10% level. The first specification is based on a regular data set, the specification 2 includes the people born before 1916 and after 1926, the third specification contains the people born after 1935 and the fourth those, which are born after 1945.
Source: Authors' calculations based on SHARE data set.

Table A.4: The first-stage results, Years of schooling as a dependent, Heterogeneous effects

Sampling window	Whole	August, September	July - October	July, October
Females				
Birth month	-0.00882	-0.0652	0.0784	-0.198
	(-0.63)	(-0.38)	(0.32)	(-0.81)
N	6472	2062	989	1073
F	0.0982	0.360	0.634	2.307
R^2	0.0000152	0.000175	0.000642	0.00215
Males				
Birth month	0.00325	-0.0902	0.218	-0.349
	(0.18)	(-0.42)	(0.72)	(-1.15)
N	2800	848	407	441
F	0.0335	0.176	0.515	1.322
R^2	0.0000120	0.000208	0.00127	0.00300
Mother alive				
Birth month	0.0122	-0.0415	0.265	-0.344
	(0.55)	(-0.16)	(0.68)	(-0.93)
N	1428	471	234	237
F	0.307	0.0242	0.465	0.873
R^2	0.000215	0.0000515	0.00200	0.00370
Father alive				
Birth month	0.000596	-0.116	0.572	-0.700
	(0.02)	(-0.34)	(1.08)	(-1.56)
N	585	196	88	108
F	0.000383	0.115	1.174	2.427
R^2	0.000000656	0.000593	0.0135	0.0224

Note: t-statistic are in parentheses; ***, **, * significance at the 1, 5 and 10% level.

Source: Authors' calculations based on SHARE data set.