

Can We Trust Older People's Statements on Their Childhood Circumstances? Evidence from SHARELIFE

Enkelejda Havari · Fabrizio Mazzonna

Received: 6 June 2013 / Accepted: 31 October 2014 / Published online: 30 July 2015
© Springer Science+Business Media Dordrecht 2015

Abstract This paper analyzes the quality of subjective assessments related to childhood circumstances when provided by old-age individuals. Early life events are important for social scientists to predict individual outcomes later in life and because of data restrictions, retrospective assessments are often used. Nevertheless, there is widespread skepticism on the ability of old-age respondents to recall with good accuracy events occurred many years ago. Using data from the survey of health, aging and retirement in Europe (SHARE), we assess the internal and external consistency of some measures of childhood health and socio-economic status. Our study suggests that overall respondents seem to remember fairly well their health status and their living conditions between age 0–15. Applying a cross-country comparison (13 European countries), we analyse within survey responses with external historical data (e.g., GDP per capita in period 1926–1956) at a country and cohort level. Our results should mitigate some of the doubts on retrospective data collection and promote their use for research purposes.

Keywords Childhood · Europe · Health · Retrospective data · SHARE survey

1 Introduction

Nowadays, researchers are becoming increasingly interested in analyzing the long-term impact of childhood events on individual outcomes later in life. This is due to a

E. Havari
European Commission - Joint Research Centre, Ispra, VA, Italy

F. Mazzonna (✉)
Università della Svizzera Italiana (USI), Lugano, Switzerland
e-mail: fabrizio.mazzonna@usi.ch

F. Mazzonna
Munich Center for the Economic of Ageing (MEA), Munich, Germany

large literature on child development documenting that early life events can be considered good predictors for health status and labor market outcomes in adulthood. Almond and Currie (2011) argue that children's characteristics (e.g., cognitive abilities) when measured in early life (age 0–5) can predict up to 20 % of the variation in wages at age 33. Childhood health—often proxied by birth-weight or height—seems to play an important role for economic and health outcomes in late life (Case et al. 2005; Case and Paxson 2009; Currie 2009). Another body of research in developmental epidemiology arrives to similar conclusions focusing on the role of adverse early environment, especially in utero, on adult health status (Barker 1998; Gluckman and Hanson 2005).

To assess whether childhood circumstances affect adult health and socio-economic status (SES) directly or indirectly through human capital accumulation (e.g., education), it is crucial to have detailed data on individuals' lives from birth to adulthood. Unfortunately, this is not always feasible due to high data requirements. Nationally representative surveys such as the 1958 British Cohort study, that follow individuals for a long time are very expensive and available for a limited number of countries (e.g., US, UK) and cohorts. Other studies use macroeconomic or contextual variables such as business cycle fluctuations, natural disasters, pandemics, or other type of shocks (e.g., Almond and Mazumder 2005; Case and Paxson 2009; van den Berg et al. 2009), to evaluate the long lasting effects of adverse conditions early in life. However, the exposure of a given cohort to a drop in national income or to famine around birth is not equivalent to the exposure of a single individual to an income shock or to nutritional shortages. During such events, many households are not directly exposed to adverse conditions, because for example they belong to a wealthier class. This means that comparing the outcomes of individuals born during adverse macroeconomic events with the outcomes of those born before or later will likely underestimate the long term effect of adverse conditions early in life (what economists call a reduced form effect).

Therefore, it is increasingly common to collect data retrospectively, by asking individuals to provide subjective assessments about their health status and living conditions in childhood, as well as information about their health, education, employment, life satisfaction in later stages of their life. According to Börsch-Supan and Schröder (2011, p. 5) this method is “*faster, ..., less costly ... and the risk that respondents drop out of the study is much lower than in a longitudinal survey*”. However, recall errors may arise when individuals do not remember precisely when and how an event took place in the past (recall failure), and when they answer to questions about their past based on their current SES and living conditions (this second type of recall error is known as coloring). More generally, recall errors may lead to bias when the accuracy of recall regarding prior exposures is different for cases and controls. For instance, this happens when recall errors are systematically related to some specific individual characteristics (as in the case of coloring). However, if the probability of recall failure is the same across groups under investigation this will lead to a downward bias due to a classical measurement error problem (Raphael 1987). Therefore, the presence of bias might be study specific and depends on the research question under investigation.

Despite these potential concerns, surveys—such as the US Health and Retirement Survey (HRS) and the English Longitudinal Study on Aging (ELSA) in the UK—have collected retrospective information on individuals' life histories (including childhood circumstances). The main purpose is to have a picture of their experiences before the baseline year of the survey and to study patterns in adult health, earnings and labor supply (e.g., Blackwell et al. 2001; Smith 2009a, b). As argued by Smith (2009a), information on pre-baseline health histories is important since it helps relaxing the assumption that health status in the baseline wave summarizes the health history of each respondent. The same strategy has been used for the Survey of Health, Aging and Retirement in Europe (SHARE), a longitudinal survey that collects micro-data on health and SES for individuals aged 50+ across 15 European countries. The third wave of SHARE, named SHARELIFE, provides rich retrospective data on health, accommodation and employment history, childhood circumstances and other general questions on well-being for respondents aged 50+ and their spouses independent of age. Since respondents are often asked to recall events that took place 50 years ago or longer, the assessment of data quality in SHARELIFE is an important issue. This is especially true when information on early life circumstances is used to uncover a causal relationship between childhood events and outcomes in adulthood, which can be used for policy purposes.

Havari and Peracchi (2012) and van den Berg et al. (2012) show that having suffered hunger during childhood has a negative impact on education, health and other dimensions of SES later in life. Using the same data we also show that there exists a gradient between health and SES in childhood and health and SES later in life for SHARE respondents (see Sect. 3.2). Figure 1 reveals a strong relationship between childhood and adult health using questions on self-reported health (SRH).¹ At age 60, those reporting good health in childhood (between age 0 and 15) are approximately 15 percentage points more likely to report good health as adults compared to those in poor health as children, and this difference is statistically significant at a 5 % level at all ages.

However, Fig. 1 does not reveal the whole story because retrospective data may not be of good quality. For instance, the large correlation between self-reported childhood and adult health could be a sign of upward bias due to the presence of coloring. Similarly, the fact that this association declines with age might be interpreted as an increase in age-related recall errors. If this is true, measurement errors increase with age and the observed association between child and old-age health might be biased towards zero. Thus, before analyzing possible relationships between childhood and adult health, it is important to evaluate the accuracy of retrospective records of childhood health conditions.

SHARE is rapidly gaining popularity among researchers (from 2005 to 2013 there are 3400 new registered scientific users), so there is a growing need for validating retrospective information related to the childhood period. So far, there are only a few studies that look at the accuracy of retrospective assessments in SHARE

¹ Self-reported health is measured on a 1–5 scale with 1 for excellent health and 5 for poor health and we use data from SHARELIFE (2008). For illustration purposes we recode it and define good health as “very good or excellent” and bad health as “good, fair or poor”.

(SHARELIFE was released in November 2010). Garrouste and Paccagnella (2010) provide an overall assessment of recall errors in SHARE looking at respondents' coherence in answering to the same questions in different waves of the survey (test-retest methodology). Yet, there are no studies that carefully look at individual responses on childhood characteristics and circumstances available in SHARELIFE.

Our paper contributes to the literature in three ways: (a) we study the importance and predictive power that childhood circumstances may have for outcomes later in life; (b) we investigate the quality of retrospective health data in SHARE extending previous work carried out for similar surveys such as the PSID (Panel Study of Income Dynamics) and the HRS (e.g., Haas 2007; Smith 2009a); (c) we go beyond health measures and we also investigate the quality of information on SES (e.g., number of books at home, experience of hunger, etc.). Until now, the validation studies carried out for similar surveys have only used childhood health measures such as self-reported health status (SRH) and number of chronic diseases. However, as stressed by Smith (2009a, p. 402): "... other domains during childhood, in addition to childhood health, may be critical for later life. These might include childhood exposures to adverse physical or social environments, substance abuse, or conditions in the home that lead to chronic stress". For this reason, we assess the

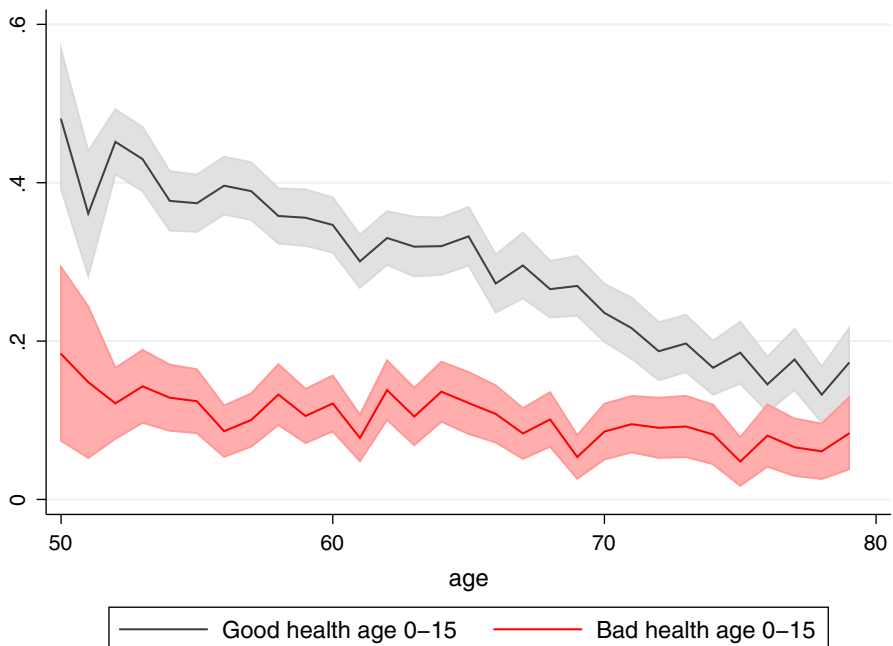


Fig. 1 Percentage of adults in good health at a given age, by their health status during childhood (between age 0 and 15). Self-reported health in childhood and adulthood is on a 1–5 scale with 1 for “excellent”, 2 “very good”, 3 “good”, 4 “fair” and 5 for “poor”. For illustration purposes, we define good health as “excellent or very good” and bad health as “good, fair or poor”. Data are from SHARELIFE (wave 3). The shaded area delimits the 95 % confidence interval. (Color figure online)

quality of a variety of measures that go beyond childhood health, to include indicators of family socio-economic background (e.g., number of books in the house at age 10, accommodation features, etc.), as well as hardship episodes (i.e., episodes of hunger) that could directly affect outcomes later in life.

In the absence of administrative records or more objective data to compare with individual assessments, it is difficult to make a sharp evaluation of childhood retrospective measures. We try to overcome this problem by looking at the consistency of answers within wave (e.g., comparing responses on childhood SRH and specific childhood diseases), and by exploiting the large cohort and geographical variation in SHARE. In particular, we verify whether responses on childhood SES, hunger and financial hardship episodes are consistent with evidence from historical data (e.g., GDP per capita, mortality rates) at a country and cohort level (external consistency). Additionally, in the case of childhood health we are able to distinguish between different types of recall errors and so different potential sources of bias. Since the presence of recall bias might be study specific, it is important for researchers who want to use SHARELIFE data to be aware of all the potential biases that might arise in their study. Our results show a good level of internal and external consistency of most of the self-reported childhood measures. In particular, in the case of childhood health our investigation rejects the presence of memory related to recall failure but shows some evidence of coloring for women.

The remainder of this paper is organized as follows. The next section presents a brief review of the literature on validation studies using survey data. Then in Sect. 3, we describe the data and the methods used for this study and in Sect. 4 we present our results on the validation of the retrospective childhood information in SHARELIFE. Finally, Sect. 5 concludes by discussing our main findings.

2 Background

In the past 20 years, there has been a growing interest in assessing the quality of retrospective information in survey data. This is strictly linked to the diffusion of household surveys that cover different dimensions of individuals' lives (e.g., lifetime earnings, health status, health care) and the growing research on the importance of childhood circumstances in explaining differences in adult outcomes. For this reason, many surveys (e.g., HRS, SHARE, ELSA, etc.) started to collect individual data at different points in time in a longitudinal framework by asking respondents to recall life events before the baseline year of the survey.

However, there is a trade-off between collecting retrospective information over a long period of an individual's life and the desirability of having high quality data (Bound et al. 2001). This holds especially when respondents are asked to remember events that occurred over 50 years ago (as in SHARELIFE). The presence of recall errors may bias the estimates of the long-run effects of childhood events on later-life outcomes.

Social scientists (Dex 1995; Beckett et al. 2001; Smith and Thomas 2003; Auriat 1991) state that the individuals' ability to accurately recall certain events depends on its attributes (e.g., timing, frequency with which it has occurred, etc.),

respondents' characteristics (e.g., age, schooling) and the survey design (e.g., the use of temporal landmarks derived by the most important events during life). Intuitively, an event occurred decades ago would be recalled less precisely compared to an event that took place last week or last year. So, the more salient an event the more accurately it will be remembered.

The literature suggests two ways to deal with recall bias: (a) find ways to minimize recall errors in a survey (ex-ante approach); (b) assess the quality of recall data (ex-post approach). Bound et al. (2001) and Beckett et al. (2001) conveniently summarize the ex-ante approach.

Scientists have tried to improve the design of retrospective surveys by incorporating knowledge of how memory works. There are two main intuitive facts that are taken into account. First, events that are more important are easier to remember. Second, memory is organized in a hierarchical format, namely moving from a general structure to specific one (e.g., Belli 1998). One practical way to improve the recollection process is to use the so-called Life Calendar Methods (LCM), that are based on a bounded recall framework. They exploit the presence of memorable events—such as marriage, birth of a child, the first job, etc.—to recall events that are less unique (e.g., unemployment spells, accommodation transitions, etc.). Nowadays, it is possible to opt for more advanced versions of LCM, where respondents can see the placement of the events on a computer screen while they answer. Belli (2005) discussing the usefulness of this practice within the PSID survey, argue that graphical devices in general and LCM approaches in particular can improve respondents' ability to recall.

On the other hand, the strand of the literature that focuses on the ex-post approach evaluates the quality of retrospective assessments once the data have been collected. One way to do this is to validate retrospective information with more objective assessments. Such practice is followed especially by epidemiologists when dealing with retrospective childhood illnesses. Krall et al. (1988) validated self-reports of respondents' childhood diseases, accidents, hospitalizations at age 30, 40 and 50 by using physician assessments collected during their infancy in a longitudinal prospective study. Similarly, Berney and Blane (1998) reveal that a majority of subjects can recall with good accuracy simple socio-demographic information (e.g., father's occupation and accommodation characteristics), even after 50 years. Francesconi (2005) evaluates retrospective data on family structure (e.g., living) during childhood using the sixth wave of the British Household Panel Survey (BHPS). He compares subjective data with both external sources and internal variables in the BHPS itself, and finds that the BHPS data overestimate the proportion of people who report an experience of life in a non-intact family during childhood by about 10 %. Despite such discrepancy, the author finds a substantial degree of similarity between the family structure information that was collected in the sixth wave of the BHPS and data from external sources.

Smith (2009a) validates the quality of responses on child diseases (infectious diseases, asthma, allergies etc.) by matching disease prevalence rates in the HRS and PSID retrospective surveys with more objective data from the National Health Examination Survey (NHES), that provides detailed health information on American children born between 1949 and 1958 (closer cohorts to those present

in PSID and HRS). Garrouste and Paccagnella (2010) provide a validation of the SHARE data by comparing responses on employment status, presence of cohabiting spouse, number of children alive across the three waves of SHARE. Their main findings are: (i) a strong consistency across waves (less than 10 % of recall errors), (ii) gender, age and family status are the main determinants of recall ability, (iii) education plays a significant role in recollecting employment characteristics and transitions. However, along with the enhancement of household surveys other methods have been applied. Using the HRS and PSID, Haas (2007) and Haas and Bishop (2010) show that the retrospective measure for childhood health is reliably reported over time (the correlation of identical responses across waves ranges between 0.6 and 0.9). Haas (2007) evidences that the quality of the assessments did not vary much by gender or age but was substantially higher among more educated individuals.

Despite the positive results discussed above one should be cautious when using retrospective data of events that happen more frequently (e.g., unemployment spells). Jürges (2007) compares self-reported unemployment in a specific month with the retrospective calendar data on unemployment in that same month reported one year later. He uses data from the German Socio-Economic Panel and finds that 13 % of all unemployment spells are not reported one year later, and another 7 % are misreported. One interesting result is that the ratio of retrospective to current unemployment is related to salience aspects such as loss of life satisfaction. Generally, those with weak labor force attachments such as women with children or individuals close to retirement are more prone to under-report unemployment retrospectively.

3 Data and Methods

3.1 SHARE and SHARELIFE

SHARE is a multidisciplinary, cross-national household survey that started in 2004. It collects data on health, SES, social and family networks for nationally representative samples of elderly people (aged 50+) in the participating countries. The survey was designed to be comparable across countries and harmonized with the HRS and ELSA. Until now there are four waves available (2004, 2006, 2008, 2011) covering 19 European countries, plus Israel.

We use data from 13 countries where individuals participated in both the second and the third wave of the survey (2006, 2008). These countries represent different European regions: (a) Scandinavia (Denmark, Sweden); (b) Central Europe (Austria, Belgium, France, Germany, the Netherlands, Switzerland); (c) Mediterranean countries (Greece, Italy, Spain); (d) Eastern Europe (Poland, Czech Republic). The target population consists of individuals aged 50 or more, who speak the official language of each country and do not live abroad or in an institution, plus their spouses or partners independent of age (see also Börsch-Supan and Jürges 2005).

The third wave of SHARE, named SHARELIFE, collects retrospective information on the life history of each respondent before the baseline year of the survey (2004). According to Schröder (2011), SHARELIFE was designed to analyze the role played by the European welfare states in shaping individual's life and behaviors. The main idea is to compare individuals' decisions and experiences across time and countries and matching them with different institutional reforms that Europeans might have faced over time (e.g., pension reforms, health insurance, etc.).

Furthermore, face-to-face interviews used for waves 1, 2 and 4 of SHARE will continue to take place (wave 5 is released in Spring 2015). SHARELIFE is designed to follow some of the methods and principles reported in the literature (see Background) on how to improve respondents' recall ability.

First, SHARELIFE arranges the different interview modules based on what is usually most important for the respondent and hence remembered most accurately (Schröder 2011). This order is based on pilot studies conducted for the ELSA survey; however, it allows each respondent to change the order of the modules according to what is deemed most important.

Second, the interview is supported by a multidimensional life grid which is a computerized version of the LCM that serves as the basis for the SHARELIFE interview. "*Life events are recorded into a large grid, where sets of topics such as children, partners, or work are combined with the time dimension, which is usually on the horizontal*" (Schröder and Börsch-Supan 2008, p. 7). The life grid represents an improvement with respect to a simple timeline, which records events on an axis and places other events around it. Indeed, the life grid is a multidimensional version of the timeline because respondents can watch on the screen important events of different areas (children, health, job etc.) in parallel (from early childhood through adulthood). In particular, the information provided by the respondent appears immediately on a computer screen (it can be viewed by both respondents and interviewers), and helps respondents to link events in a easier way. For instance the interview starts with the children section, followed by a module on partners. Here respondents are asked questions about marriages, cohabitation, and other important partnerships. The starting and ending dates of each relationship is displayed on the screen to facilitate the recall of more detailed information on each partner (e.g., year of birth, education, etc.). The module in which respondents are asked to remember where they lived throughout their life (and related accommodation transitions) is placed immediately after the children and partners sections. In this way interviewers can easily link other detailed information about personal events and ask questions such as: "Did you move in the new accommodation after the first child was born?" or "Were you still married with X when you moved?". Schröder and Börsch-Supan (2008) argue that this type of anchoring has proved to be very useful to respondents in SHARELIFE (see their paper for more details).

We now briefly discuss problems of non-response and attrition in SHARE and their consequences when making inference. De Luca and Peracchi (2005, 2012) showed that the response rate in SHARE is quite high compared to similar European surveys. However, there are concerns for panel attrition as one third of the original sample is lost between the first and the second wave of the survey and about one

fourth between the second and third. The representativeness of the second wave was ensured by adding a refreshment sample. However, retrospective information on life-course events (collected in wave 3) is not available for all participants in wave 2 due to quite relevant attrition between waves 2 and 3. Ignoring attrition may lead to invalid inferences if this selectivity effect is driven by observable and unobservable factors related to the research question under investigation. For this reason, we deal with this issue by using the calibrated longitudinal weights provided by the public release of SHARE whenever we use the panel sample (see De Luca and Rossetti 2013). These weights are meant to make the distribution of the sample closer to the target population (individuals aged 50+ in 2006 and alive in 2008), by gender, age group and geographical area (at the NUTS1-level). Mazzonna (2014) shows that this weighting procedure leads to similar results with respect to alternative approaches that account for selection on unobservables by relaxing some of the underlying assumptions when using weights.

3.2 Measures of Childhood Circumstances

In our analysis we focus on three modules of the questionnaire that contain valuable childhood information: health conditions, SES and general life questions that allow to establish if a respondent experienced a period of hardship in a given year and country. The health section starts by asking respondents to rate their overall health from birth up to age 15, choosing one of the five categories: (*excellent, very good, good, fair, poor*). Note that the question is formulated similarly to self-reported health in old-age. Other questions look at episodes of absence from school due to health problems and specific diseases or health limitations from birth up to age 15, namely *infectious diseases, polio, asthma, other respiratory problems, allergies, severe diarrhea, meningitis, chronic ear problems, speech impairment, difficulty seeing, severe epilepsy/seizures, emotional, nervous, or psychological problems, fractures, appendicitis, diabetes, heart problems, leukemia, cancer or other not listed*. It is also possible to know—for each of the diseases or limitations—the age interval in which they were encountered (age 0–5, 6–10, 11–15) and if the disease lasted for more than a year.

The childhood SES section asks respondents to provide information on their living conditions and family background characteristics when they were 10 years old. We consider the number of rooms (excluding bathrooms and kitchens) and people in the house, features in accommodation (fixed bath, cold and hot running water supply, inside toilet and central heating), the number of books at home (from none to 2 or more bookcases), the occupation of the main breadwinner (10 categories ranging from manager to elementary occupation), and the self-reported relative position in mathematics (math) and country's language compared to their own peers in the class (much better, better, about the same, worse, much worse).

Interestingly, a strong gradient exists also for other dimensions of SES such as income and education. Figure 2 highlights that individuals from better off families—with higher values of rooms per capita, books at home, accommodation facilities and with a breadwinner being a white collar—have on average a higher income when they are old.

Similarly, Fig. 3 shows that being good in mathematics or language at the age of 10 exerts a positive effect on years of completed education. Interestingly, there are remarkable differences by gender. For instance, differences in years of education by math level are less marked for females than males. When considering other measures of health or income we find similar results when relating them to childhood circumstances.

The last module, General Life (GL), collects information on striking experiences during life such as hunger, financial hardship, stress, and poor health. We consider hunger episodes because they are particularly relevant when studying the long-term consequences of severe childhood circumstances on adult outcomes (van den Berg et al. 2012; Havari and Peracchi 2012). Data have been collected based on the following questions: (i) *Looking back on your life was there a period in which you suffered from hunger?* Then, if the event is remembered and the answer is “Yes”, two other questions followed: “When did this period start (year)?”, and “When did this period stop? (year)”. Figure 4 shows the proportion of respondents reporting to be in excellent or very good health in old-age conditional on experiencing hunger (red bars for those experiencing hardships and blue bars for those who don’t). As expected, the fraction of individuals who report to be in good health when old is lower for those who experienced hunger. Although not causal, this strong relationship confirms the importance of childhood events in predicting long-term

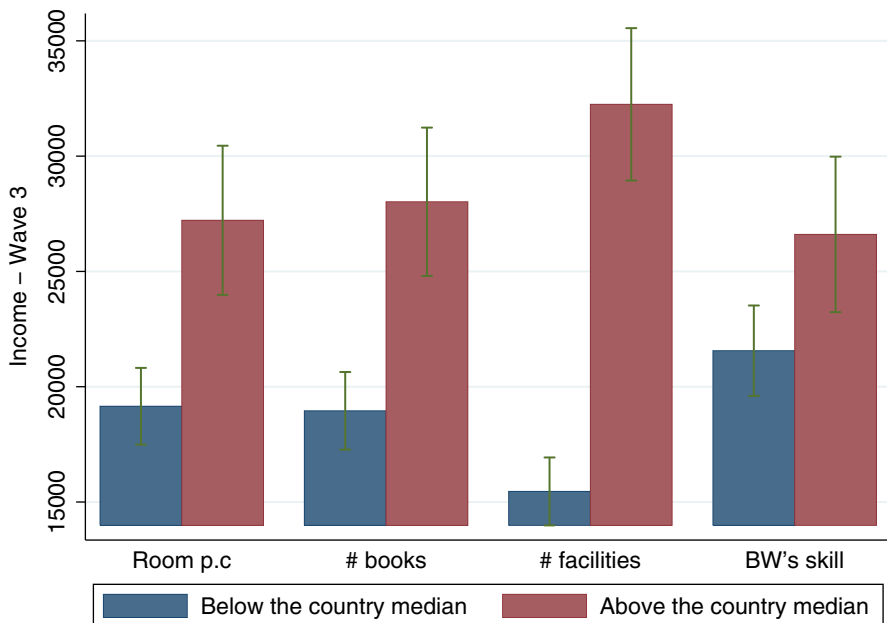


Fig. 2 Childhood SES proxies and adult per capita income. The *blue* and the *red* bars show the average per capita income in adulthood (after age 50) for individuals below and above the country median value of each childhood SES proxy: (a) number of rooms per capita at age 10, (b) number of books in the house at age 10, (c) number of facilities in accommodation (fixed bath inside, hot water, etc.), (d) low-skilled breadwinner (*blue collar* or elementary occupation). (Color figure online)

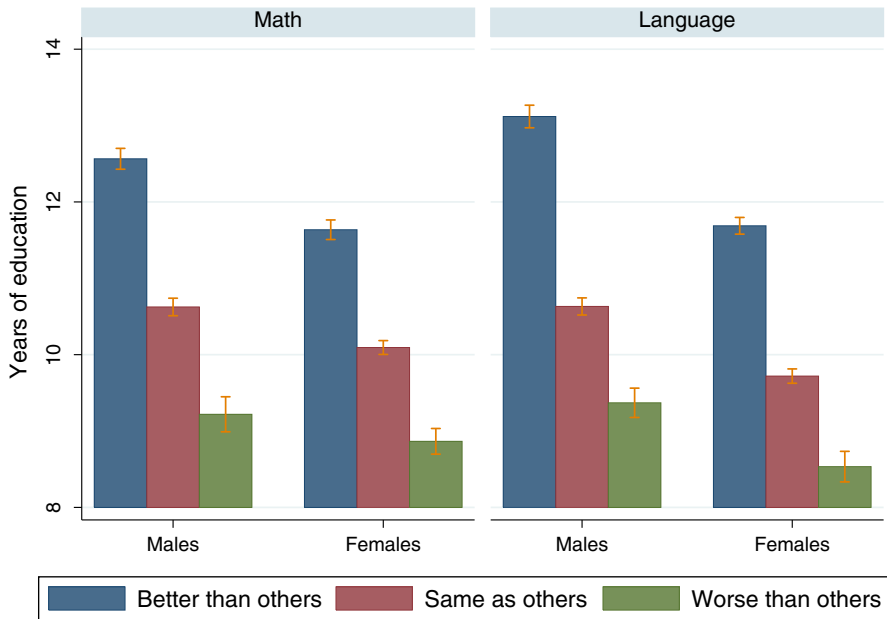


Fig. 3 Average years of completed schooling and school performance at age 10 (math and language), by gender. The relative position in math and language is expressed in three categories: (a) better than others (blue bars), (b) same as others (red bars), (c) worse than others (green bars). (Color figure online)

health outcomes. In Sect. 4 we will present some extended analysis concerning these specific childhood circumstances.

4 Results: Quality of Childhood Measures

As previously discussed, different solutions have been proposed to validate retrospective health assessments: (a) using external medical records (Krall et al. 1988); (b) testing the consistency of responses on the same question over time (Haas 2007); (c) comparing individual responses to aggregate data on prevalence disease rates for a given country and year (Smith 2009a). The first two methods cannot be implemented here because, by now, SHARELIFE is not yet linked to medical records and we have only one wave of retrospective questions. The third method could be feasible in the near future but for the moment we lack data on disease prevalence rates in Europe for our cohorts of interest. Nevertheless, the common approach we follow to evaluate the retrospective information on both childhood health and SES is somewhat similar to that described in (c). In particular, we evaluate the consistency of the information provided in SHARELIFE by matching individual data with aggregate external data on the socio-economic environment for each country and cohort. As already mentioned, this strategy does not allow to make a sharp evaluation of the quality of the retrospective data—as would be the case

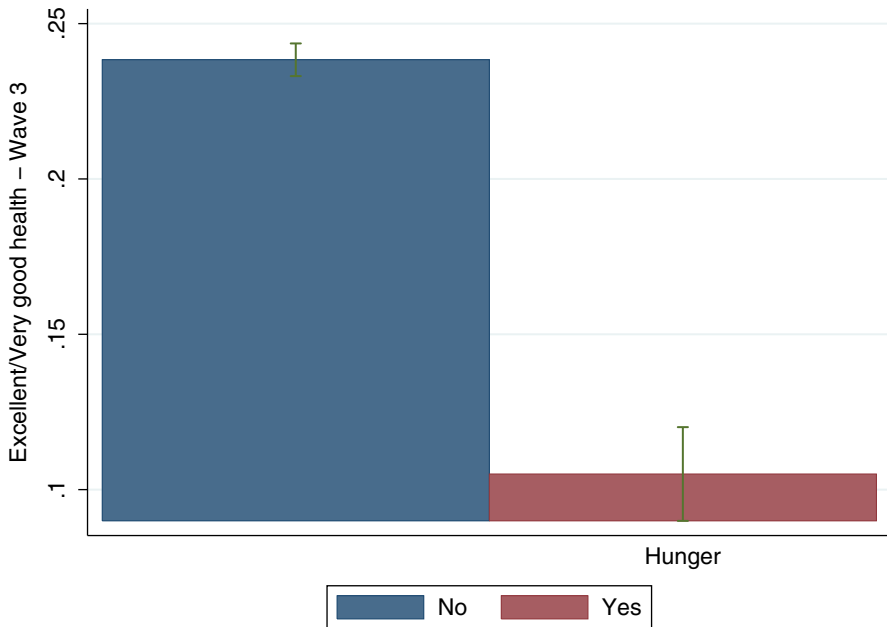


Fig. 4 Childhood hunger and adult health using Wave 3. We show the percentage of respondents reporting excellent or very good health as adults, conditional on having suffered (*red bars*) or not (*blue bars*) from hunger during childhood. (Color figure online)

when matching with administrative records. However, in the case of childhood health (as described in Sect. 4.1) the joint use of external data and large set of information provided in SHARE allows us to implement specific tests about the presence (and sources) of recall bias. In the case of childhood SES (as described in Sect. 4.2) we cannot implement similar tests but we can still evaluate the consistency of the self-reported retrospective information with external socio-economic data or with historical events, as in the case of hunger. In particular, the cohorts present in SHARE were undoubtedly exposed to episodes of famines related to the Second World War (e.g., Dutch Hunger Winter 1944–1945) which can be easily compared with the self-reported episodes of hunger in SHARELIFE.

4.1 Childhood Health

As in Smith (2009a), we start investigating the quality of childhood SRH by looking at how coherent individuals are in reporting information on selected diseases and periods in which they miss school because of health problems (internal consistency). Table 1 shows the marginal effects from a logit regression of childhood SRH being either *excellent* or *very good* on a set of self-reported childhood health conditions: (a) specific illnesses and (b) indicators of absence from school due to health problems.

It is quite standard in the literature to check if individuals are coherent in answering to different questions about health. In both cases, in column (A) we

control for country fixed effects (baseline regression) while in column (B) we control for a cubic polynomial in age and a dummy for female. From the table we notice that reporting illnesses like asthma, respiratory problems, heart problems, diarrhea, etc. reduces the probability of reporting good health during childhood. The more severe the disease, the greater is the effect. All coefficients have a negative sign and are statistically significant at the 1 % level with the exception of short-term diseases like infectious diseases and fractures which may not be recalled with good precision or does not reduce the general childhood health. Having missed school for more than one month, having spent time in bed or in hospital, reduces the probability of reporting good health during childhood. The results reported in Table 1 suggest that individuals are coherent in answering different health questions pertaining to the same age-interval (childhood). However, they do not provide any

Table 1 Internal consistency: Logit estimates of childhood SRH (excellent, very good)

	(A)	(B)	(A)	(B)
Infectious diseases	-0.003	0.001		
Polio	-0.450***	-0.460***		
Asthma	-0.340***	-0.337***		
Resp. problems	-0.235***	-0.236***		
Allergies	-0.218***	-0.227***		
Diarrhea	-0.233***	-0.233***		
Ear problems	-0.182***	-0.207***		
Meningitis	-0.182***	-0.173***		
Difficulties seeing	-0.193***	-0.187***		
Speech impairment	-0.233***	-0.240***		
Headaches or migraines	-0.158***	-0.150***		
Epilepsy or seizures	-0.370***	-0.380***		
Depression	-0.179***	-0.166***		
Fractures	0.019	0.007		
Appendicite	-0.069***	-0.067***		
Diabetes	-0.559***	-0.563***		
Heart problems	-0.326***	-0.320***		
Cancer	-0.624***	-0.622***		
Miss school > 1 month			-0.116***	-0.119***
Miss school > 1 month (bed)			-0.227***	-0.227***
In hospital 3 times a year			-0.254***	-0.250***
Demographics	NO	YES	NO	YES
Country FE	YES	YES	YES	YES
N	23485	22445	22255	21274
Pseudo-R ²	0.0783	0.0822	0.0570	0.0607

Logit estimates (marginal effects) for childhood SRH: 1 “excellent or very good” and 0 “good, fair or poor”. Demographic controls include a cubic polynomial in age and a dummy for female. We add country fixed effects (Italy is the reference country)

Significance *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are robust to heteroskedasticity

information on the presence of recall bias. As already mentioned, it is important to remark that not all recall errors lead to recall bias. Apparently, coloring (or anchoring) seems to be a more serious source of recall bias since it leads to a spurious correlation between child and old-age health. However, also other types of recall errors, such as simple recall failures, if systematically related to some individual characteristics under investigation might lead to recall bias. Therefore, in the analysis that follows we try to analyze all types of recall errors trying to identify their source.

We start looking at the presence of recall errors by focusing on disease rates by age within SHARELIFE. The idea is to see whether we observe a systematic association between age and disease rates. In particular, Table 2 shows the percentage of respondents that declare to have suffered from a specific disease in their childhood (between age 0 and 15), by cohort. It is noticeable how recalled disease prevalence in youth declines with age in the cross section, or rises with the year of birth. For instance, only 76 % of individuals aged 71+ (born between 1918–1937) report to have had infectious diseases in their childhood, whereas the

Table 2 Child diseases (%), by age

Age in 2008	50–60		61–70		71+	
Year of birth	1948–1958		1938–1947		1918–1937	
Age 15 ^a	1963–1973		1953–1962		1933–1952	
Child diseases	Mean	SE	Mean	SE	Mean	SE
Infectious diseases	83.23	(0.40)	82.17	(0.41)	76.95	(0.47)
Meningitis	0.96	(0.10)	0.92	(0.10)	0.49	(0.08)
Asthma	2.04	(0.15)	1.90	(0.15)	1.76	(0.15)
Diarrhea	0.93	(0.10)	1.18	(0.12)	0.82	(0.10)
Respiratory problems	3.36	(0.20)	2.87	(0.18)	2.87	(0.19)
Allergies	4.17	(0.21)	2.87	(0.18)	2.16	(0.16)
Speech impairment	0.84	(0.10)	0.48	(0.07)	0.30	(0.06)
Heart trouble	0.65	(0.09)	0.78	(0.09)	0.65	(0.09)
Ear problem	2.93	(0.18)	2.41	(0.16)	2.11	(0.16)
Headaches or migraines	4.19	(0.21)	4.23	(0.21)	3.56	(0.20)
Fractures	9.82	(0.32)	7.38	(0.28)	4.98	(0.24)
Appendicite	9.46	(0.31)	9.73	(0.31)	6.65	(0.28)
Depression	1.24	(0.12)	1.06	(0.11)	0.56	(0.08)
Diabetes	0.11	(0.04)	0.08	(0.03)	0.05	(0.02)
Cancer	0.03	(0.02)	0.02	(0.02)	0.02	(0.02)
Epilepsy or seizures	0.64	(0.08)	0.38	(0.07)	0.17	(0.05)
Leukemia	0.01	(0.01)	0.07	(0.03)	0.00	(0.00)
N	8881		8914		8240	

We show the percentage of respondents reporting a given disease during childhood (standard errors in parenthesis)

^a The year of birth and the year of the 15th birthday are shown on the top of the table to highlight the childhood period (age 0–15)

percentage of respondents aged 50–60 (born in 1948–1958) claiming to have had infectious diseases is much higher, 83 %.

There can be three (not mutually exclusive) interpretations for this pattern. First, it can be attributed to recall failure since for older individuals might be more difficult to recall selected diseases than it is for younger ones (recall failure hypothesis). Second, lower rates for the oldest cohort can be explained by the lack of diagnostic tools and medical knowledge when they were young and by the absence of disease prevention in most of the countries under consideration (lack of diagnosis hypothesis). Finally, we cannot exclude the presence of a survivorship bias effect (selection hypothesis); in fact, individuals belonging to the earlier cohort that are still alive might be healthier. The distinction is important because each hypothesis differently affects and eventually biases the inference that can be made using these data and there are different ways in which researchers can eventually control for it. For instance, as already discussed, under the recall failure hypothesis the declining age profile in the correlation between child and adult health, shown in Fig. 1, might be explained by the fact that recall failure increases as people age. This implies an increase in measurement errors that leads the correlation towards zero. A downward bias in that correlation is also introduced under the selection hypothesis. Indeed, among those in bad health during childhood only the healthier survive. Instead, in the case of the lack of diagnosis, only the prevalence rates of diseases requiring a medical diagnosis will be underestimated. All in all, only the recall failure hypothesis is a problem associated with the retrospective nature of the data. In fact, in the other two cases even administrative data would suffer of some form of measurement problem or bias.

In the analysis that follows we implement a series of estimations aimed to differentiate among these hypotheses. We start from Table 3 where we exploit differences in disease rates by geographical areas to see whether we find a similar pattern across countries. Countries are grouped according to their geographical location, but this aggregation also reflects different levels of economic development in Europe (e.g., we analyze cohorts born before 1954). For this reason, in the last row we report the average GDP per capita for each macro-area, based on data collected for the years 1926–1956 (GDP per capita is expressed in thousands of 1990 international Geary-Khamis dollars, Maddison 2010). We find a large variation of disease response rates across countries, with low values for Mediterranean countries and Eastern Europe and higher ones for Central Europe and Scandinavia. This result seems to be more consistent with the second interpretation. Individuals born in poorer countries (Italy, Spain, Greece, Czech Republic, Poland)—characterized by limited access to medical services—show lower self-reported disease rates for most diseases (e.g., infectious diseases, asthma, etc.). Nonetheless, we cannot reject the selective mortality hypothesis according to which only the healthiest survive in poorer countries.

A further check on this problem is shown in Table 4. We regress the number of chronic diseases (by summing up responses on most relevant diseases) on a proxy for memory capacity that is built summing up the outcomes of two cognitive ability tests from wave 2 (2006/2007). These two tests consist of a verbal registration and recall of a list of ten items that each respondent hears only once. The constructed

Table 3 Child diseases (%), by geographic area

	Central Europe	Eastern Europe	Mediterranean	Scandinavian
Infectious diseases	83.60	66.52	76.74	93.93
Meningitis	0.72	1.22	0.52	1.12
Asthma	2.01	0.55	1.14	2.89
Diarrhea	1.61	0.71	0.49	1.19
Respiratory problems	3.09	2.44	2.70	1.84
Allergies	2.99	0.79	2.26	5.70
Speech impairment	0.62	0.41	0.24	1.22
Heart trouble	1.36	1.36	0.28	0.68
Ear problem	2.77	2.21	1.75	5.02
Headaches or migraines	5.27	2.82	2.84	5.95
Fractures	8.39	7.46	4.86	9.35
Appendicite	11.44	4.38	7.43	6.69
Depression	1.33	0.64	0.55	1.93
Diabetes	0.06	0.02	0.06	0.02
Cancer	0.01	0.00	0.10	0.02
Epilepsy or seizures	0.42	0.19	0.26	1.14
Leukemia	0.06	0.00	0.00	0.00
GDP ^a per capita at age 10				
Mean	6446	3782	3339	7503
Sd	(2.387)	(0.969)	(1.473)	(1.770)
N	11,128	3716	7274	3917

We report disease rates (%) using SHARELIFE data, by geographic area: Mediterranean (Greece, Italy, Spain), Scandinavia (Denmark, Sweden), Central Europe (Austria, Belgium, Germany, France, Netherlands, Switzerland), Eastern Europe (Czech Republic, Poland)

^a Data on GDP are from Maddison tables and we merge them with the SHARE data by country and year/cohort

memory variable has a large support (from 0 to 20) and it has been extensively used in the literature to capture memory capacity and cognitive abilities in old age (e.g., McArdle et al. 2011; Mazzonna and Peracchi 2012). In the baseline specification we also include a cubic polynomial in age, a dummy for gender, a dummy for high school or college graduate and country dummies. In the second column we add a dummy for living in an urban area during childhood and in the last column we control for indicators of childhood SES (e.g., accommodation features, number of books in the house, etc.).

It is evident from the first column that respondents' memory capacity seems to be not important in predicting the number of chronic diseases, whereas age and

Table 4 Ordered logit estimates of the number of chronic diseases during childhood (self-reported)

Age	−0.017*** (0.006)	−0.017*** (0.006)	−0.010 (0.006)
Age ²	−0.035 (0.079)	−0.035 (0.079)	−0.050 (0.078)
Age ³	−0.002 (0.005)	−0.002 (0.005)	−0.003 (0.005)
Female	0.144*** (0.052)	0.145*** (0.052)	0.137*** (0.052)
Education	0.222*** (0.063)	0.193*** (0.063)	0.057 (0.066)
Memory	0.010 (0.008)	0.007 (0.008)	−0.001 (0.008)
Urban area		0.261*** (0.063)	0.159** (0.064)
Other controls			
Female	YES	YES	YES
Country f.e.	YES	YES	YES
Childhood SES proxies	NO	NO	YES
<i>N</i>	20,257	20,257	20,257

Number of chronic diseases is computed by summing up all the selected diseases. Memory is a proxy for cognitive ability and it is constructed using the outcomes from two cognitive tests: recall first and recall delayed

Significance *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are robust to heteroskedasticity

education matter.² This result is consistent with the lack of diagnosis or selective mortality hypotheses rather than with recall failures. The positive and significant effect of growing up in an urban area confirms this intuition. Living in an urban area might lead to better access to medical care and diagnostic tools (Pampel and Pillai 1986). Another interpretation could be that those grown up in urban areas might be more exposed to pollution, unhealthy environment or lifestyle, leading to a positive selection effect. However, when we use child or adult SRH as a dependent variable we find either no effect or positive effects from growing up in an urban area on self-reported health (results available upon request). It is though unlikely that a selection process working through the number of child chronic diseases and not other measures of child or adult health is taking place. Finally, as we include proxies for childhood SES, education becomes no longer significant. Similar considerations arise when we account for age effects. Given the evidence reported so far, we can exclude the presence of recall failures for self-reported health diseases because age

² An alternative explanation might be that the effect of memory disappears because memory is a function of age and education. However, as aforementioned, this variable has a very large support and we found evidence of substantial variation in the memory test also within age and education groups. In particular, we regressed the variable memory on a cubic polynomial in age, education and other baseline controls including country fixed effects (results are available upon request). As expected, we found that this large set of controls is able to explain less than 30 % of the total variation in the memory test.

and memory are not significantly associated with the number of recalled diseases. However, we still cannot completely distinguish between the selection and the lack of diagnosis hypothesis.

In Table 5 we better analyze the selection hypothesis by looking at the relationship between selective mortality and self-reported measures of childhood health. We restrict our sample to eight countries (BE, CH, ES, FR, IT, DK, NL, SW), for which cohort and gender specific mortality rates are available from the Human Mortality Database (2011). Mortality rates are linked to the corresponding cohorts in SHARELIFE.

In the first column, we show that cohorts affected by higher cumulative mortality rates (CMR) up to 2006 (it is equivalent 1-cohort survivorship to the year 2006) report a significantly lower number of childhood chronic conditions, even after accounting for a cubic polynomial in age and a female dummy. However, CMR is not only a result of the selective mortality due to the aging process but also the consequence of the infant mortality rate (IMR). This distinction is important because IMR is strongly related to the quality of the environment in which individuals were born and to public health in general (Pampel and Pillai 1986)—e.g., access to medical care and infectious disease prevention. Therefore, this distinction is consistent with the lack of diagnosis hypothesis. For this reason, in the second column, we separately evaluate the effect of IMR and CMR. Not surprisingly—since the aging effect should be captured by the cubic polynomial in age—only the IMR has a strong and significant effect on the number of childhood chronic conditions. The lack of diagnosis hypothesis is also confirmed by the results in the last two columns where IMR shows either no effect on childhood SRH or a negative effect on adult SRH. In presence of selection we should have observed a significant effect of IMR also on childhood SRH. However, we only found an effect

Table 5 Infant mortality rate (IMR) as proxy of public health

Dependent variable	(a) N. Child chronic diseases		(b) Childhood SRH		(c) Adult SRH
CMR	−0.032*** (0.008)				
IMR		−0.081*** (0.011)	0.041 (0.026)	0.006 (0.018)	−0.095*** (0.010)
Conditional CMR		0.010 (0.010)	−0.005 (0.011)	0.030 (0.019)	−0.003 (0.010)
Other controls					
Cubic in age	YES	YES	YES	YES	YES
Female	YES	YES	YES	YES	YES
Country FE	NO	NO	YES	NO	NO
<i>N</i>	12,913	12,913	12,913	12,913	12,913

The table shows the effects of Cumulated Mortality Rate (CMR), Infant Mortality Rate (IMR) and CMR conditional on IMR (Cond. CMR) on 3 different dependent variables: (a) the first three columns show ordered logit estimates for childhood self-reported number of chronic diseases; (b) the fourth column shows a logit model for childhood SRH; (c) the fifth column shows a logit model for adult SRH

Significance *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are robust to heteroskedasticity

of IMR on the number of reported childhood chronic conditions. This suggests that there is no effect on the “true” childhood health (as this would have been captured also by childhood SRH) but respondents do not know whether they suffered from a specific disease. In addition, the negative effect on adult SRH seems to suggest a selection problem that goes in the opposite direction; the cohorts in worse health today are those that experienced higher IMR. This last result is consistent with findings in Bozzoli et al. (2009). They argue that high IMR mark the health status of survivors making them more prone to late life diseases (the so-called scarring effect), which usually dominates the selection effect in many developed countries. Lastly, it is important to remark that as we control for country fixed effects—as in the third column—any mortality effect disappears.

Finally, we closely examine issues related to coloring in childhood SRH. As thoroughly discussed, our measures could be subject to coloring because questions on health status in adulthood and childhood are asked in the same section of SHARELIFE. Since data on childhood SRH are available only in SHARELIFE, we cannot check for the consistency of answers across waves (test-retest method) as is done by other researchers (Haas 2007). Therefore, we adopt a different and new strategy which exploits the availability of adult SRH in wave 2 and 3 (years 2006, 2008) instead of childhood SRH. Basically, we assume that the relationship between childhood SRH and adult SRH measured in wave 2 is the same as the relationship between childhood SRH and adult SRH measured in wave 3. This assumption, although untestable, is based on the idea that the actual relationship between child and adult health status after 50 or 52 years should be (on average) the same. The only possible discrepancy could arise if we assume a (very persistent) autoregressive health process. In this case, we could expect a stronger relationship between adult and childhood health in wave 2 with respect to that in wave 3. Therefore, if we observe a stronger relationship between childhood and adult SRH in wave 3 compared to wave 2, we can suspect the presence of coloring, probably due to simultaneous reporting of child and adult health status in wave 3. In this test we consider respondents participating both in wave 2 and 3 and use calibrated cross-sectional weights provided by SHARE to control for presence of attrition between waves. We regress separately adult SRH reported in wave 2 (W2) and wave 3 (W3) on childhood SRH. Under the null hypothesis the difference in the coefficients (between wave 2 and 3) of childhood SRH is equal to zero. Our regressions, and hence the tests, are performed separately for men and women controlling for country fixed effects and a quadratic polynomial in age. Table 6 shows the final results. The first row of the table reports the coefficient of a regression of adult SRH on child SRH in each specific wave and separately by gender. The second row reports the Wald test on the equality of coefficients for W2–W3. Consistent with the coloring hypothesis, results show that the absolute value of the coefficient of childhood SRH is higher in wave 3 than in wave 2. However, this difference is statistically significant only for women. In fact, the Wald test for women rejects the null hypothesis of equality of coefficients between wave 2 and 3 but not for men.

Table 6 Testing for coloring in the childhood SRH, by wave and gender

	W2 (male)	W3 (male)	W2 (female)	W3 (female)
Childhood SRHS	0.142*** (0.017)	0.167*** (0.019)	0.135*** (0.013)	0.177*** (0.013)
<i>N</i>	10263	10263	12535	12535
<i>R</i> ²	0.060	0.079	0.099	0.113
W2–W3		2.13		9.54
<i>p</i> value		(0.145)		(0.002)

We control for country fixed effects (10 countries available for the longitudinal panel) and a quadratic polynomial in age

Significance *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are robust to heteroskedasticity

Therefore, we can conclude that there is statistical evidence of coloring when using self-reported information on childhood health only for women.³

4.2 Childhood SES and Circumstances

To assess the quality of SES variables we compare individual data with external data (GDP per capita and average years of schooling) at a country level. Table 7 reports summary statistics for four of these indicators: number of rooms per capita, number of facilities in the accommodation, the share of respondents reporting a low-skill breadwinner, and the share of respondents reporting at least one bookcase of books in their childhood accommodation. All variables refer to the period at which the respondent was 10 years old. In the last two columns we report for each country the average GDP per capita referring to the period 1926–1956 (expressed in thousands of 1990 international Geary-Khamis dollars), and the average years of schooling measured in 1960 (data are from the Barro-Lee, 2013). These data are linked to SHARELIFE at a country and cohort level.

There is substantial variation in all indicators both within and between countries. Individuals in Mediterranean countries and Poland, as expected, report worse conditions in terms of rooms per capita and accommodation facilities. Furthermore, 86 % of Italian respondents report to have lived in a household in which the main breadwinner was employed in a low-skill occupation. Conversely, the Scandinavian countries and Switzerland have the highest fraction of respondents who grew up in more well-off households. The validity of such results is confirmed by looking at the last two columns. Higher levels of GDP per capita are associated with higher average values of the three indicators of living standards.

Regarding the number of books at home we see a lower fraction of respondents declaring to have had at least one bookcase in the Mediterranean countries and

³ The same test has been performed using the whole panel (respondents present in wave 1, 2 and 3). However, we decided to drop it from the analysis since selective attrition might influence the goodness of the test. In any case the Wald test for the equality of coefficients between W1–W2 and W1–W3 lead to the rejection of the null hypothesis for women, while the evidence is mixed for men. Results are available upon request.

Table 7 Childhood background proxies, average GDP per capita and average years of schooling in 1960 (population 25+), by country

Country	Rooms p.c.	<i>N.</i> facilities	Low skill occ (%)	GDP	One bookcase (%)	Years schooling
Austria	0.69 (0.45)	1.49 (1.56)	83	3.682	26	6.71
Belgium	0.99 (0.48)	1.80 (1.70)	76	5.062	35	7.46
Czech Rep.	0.56 (0.27)	2.30 (1.56)	78	3.121	61	9.39
Denmark	0.91 (0.41)	3.00 (1.93)	79	5.804	56	8.95
France	0.82 (0.44)	2.13 (1.78)	73	4.557	34	5.78
Germany	0.78 (0.40)	2.12 (1.57)	75	4.477	44	9.48
Greece	0.54 (0.23)	1.32 (1.40)	88	2.174	13	4.64
Italy	0.56 (0.36)	1.43 (1.57)	86	3.426	12	4.56
Netherlands	0.80 (0.36)	2.37 (1.13)	74	5.441	45	5.27
Poland	0.38 (0.25)	0.62 (1.29)	91	2.338	19	6.74
Spain	0.62 (0.41)	1.25 (1.45)	88	2.414	19	3.64
Sweden	0.77 (0.41)	3.19 (2.00)	75	5.424	58	7.65
Switzerland	0.88 (0.39)	3.14 (1.66)	74	7.449	50	7.30
Correlation						0.69

This table shows summary statistics (mean and s.d) for some childhood indicators by country. We use: (a) the number of rooms per capita; (b) facilities in accommodation (hot water, bathroom inside, etc.); (c) a dummy for the main breadwinner being in low-skilled occupation; (d) a dummy for having had more than one bookcase of books. In order to compare individual data with macro-data, for each country we report the average years of schooling measured in 1960 considering the population of Europeans aged 25+ (Barro and Lee 2013). The correlation coefficient between number of books at home and years of schooling is 0.69

Poland (in Italy only 12 %) and a higher fraction in Czech Republic, Denmark and Sweden. This result is confirmed by looking at external data on the average years of schooling in each country. The association between these two indicators is corroborated by the very high correlation coefficient (0.69).

Finally, we look at subjective assessments of hunger episodes experienced during life, drawn from the General Life Section in SHARELIFE. Among other episodes such as financial hardship, stress and poor health, we concentrate on hunger since it is a striking event with possible long-lasting effects on individuals' health and SES, especially when experienced during childhood. From SHARELIFE it is possible to construct hunger spells to know not only whether an individual has experienced hunger but also when. Periods of hunger or malnutrition are often related to exogenous shocks such as famines, wars or natural disasters. Exploiting the cross-country dimension of SHARE we match individual hunger spells with external data on the occurrence of episodes of famine or war-related events (World War II and Spanish Civil war). For a detailed description we refer to Havari and Peracchi (2012) and van den Berg et al. (2012).

Figure 5 shows the fraction of individuals who experience hunger in a given year, by country. We thus construct a longitudinal data set with annual observations on hunger onset for each individual in the sample. The red lines correspond to the years in which the most salient famine or war-related events occurred, which are

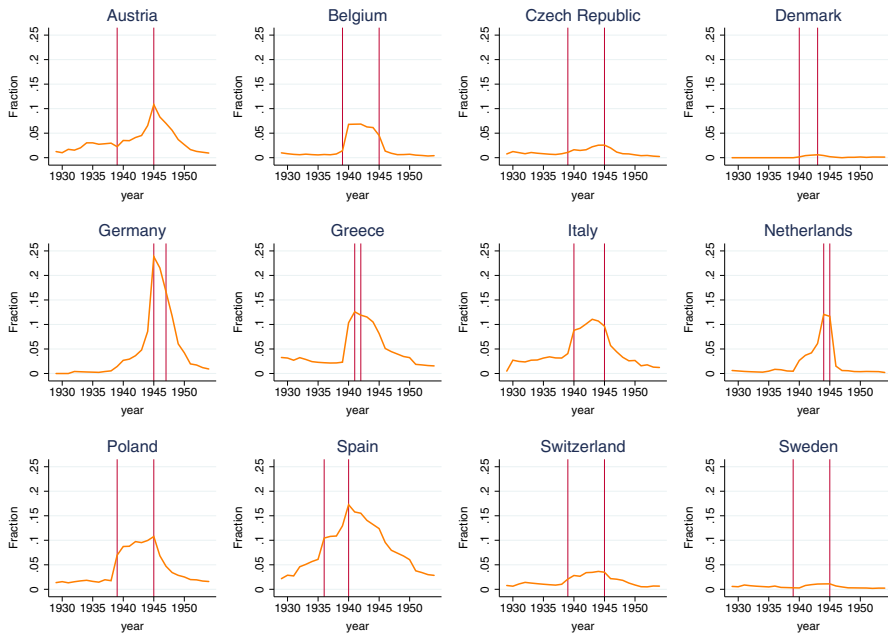


Fig. 5 Fraction of individuals reporting to have suffered hunger in a given year, by country

country-specific. It is evident from the graph that hunger spells are found mostly within the period 1939–1945. The figure clearly shows the two neutral countries in WWII, Switzerland and Sweden. For them we still mark the WWII as reference, since they were mildly affected by the war through trade and naval blockades and accidental bombing. However, it is more interesting to notice the very good correspondence between peaks in reported hunger experience and country specific episodes of war-related famine. Germany experienced at the end of World War II (1944–1945) severe aerial bombing from the Allies which hit a large number of cities, with negative consequences for the industry and the agricultural sector (see Kesternich et al. 2013 for a review). The Allied-occupation and the consequent food rationing led to the experience of a severe famine in the post-war period (1946–1947). Similar is the evidence for Austria, which had been regarded by Nazi Germany as a constituent part of the German state. We also find a good consistency between the self-reported periods of hunger for the Netherlands and Greece and the famine episodes respectively in the years 1944–1945 (so called Dutch Hunger Winter) and in 1941–1942 (see van den Berg et al. (2012) for further details). For Poland we find a high fraction of respondents experiencing hunger from 1939 until 1945. This corresponds to the Nazi occupation that started in September 1939. Spain experienced the civil war between 1936–1939 and although neutral during WWII it also faced a severe food rationing because of the autarky regime imposed by Franco that lasted for several years (del Arco Blanco 2010). These episodes are well represented in the SHARE data by looking at the occurrence of hunger spells in the Spanish sample. Very good correspondence is also found in Belgium where the

1940–1944 peak corresponds to the German invasion. In Italy we observe a high fraction of respondents experiencing hunger throughout WWII. Different is the evidence for Czech Republic, where there are just a few episodes of reported hunger concentrated during the Nazi annexation (1938–1945), and Denmark, that was occupied by the Nazi regime in 1940. However, for these two countries we did not find any evidence in the literature of famine episodes. Overall, with the data at hand we can argue that there is an almost perfect correspondence between individual hunger episodes and historical events like famines and wars in Europe.

5 Conclusions

The importance of childhood circumstances in determining individuals' future health and economic status is well documented in the literature. However, collecting information on childhood circumstances for old-age people is costly if ever possible. For this reason, many old-age surveys, such as HRS, ELSA, and SHARE have started to collect retrospective information by asking old-age people simple information about their childhood health and SES. In this paper we investigate the quality of retrospective childhood information provided in SHARE. Our work aims to support researchers who are interested in studying the causal relationship between childhood events and later life outcomes using retrospective data. Except for the case of the childhood SRH where there is evidence of some coloring, especially for women, our results seem to indicate that self-reported childhood information collected in SHARELIFE shows a good level of internal and external consistency. In particular, we show that the recall of specific diseases during childhood is affected more by scarcer access to medical services around fifty years ago—hence less knowledge about preventive care—than it is by recall failure. Therefore, even administrative data or longitudinal cohort studies would suffer of the same problem that is not associated with the retrospective nature of the data. This also implies that in order to avoid cross-country and cross-cohort comparisons biases need to take into account differences in terms of SES during childhood, which are influenced by differences in public policy interventions (e.g., health care).

All in all, our findings are consistent with the literature on response errors and the quality of recall data. Simple socio-demographic information is relatively accurate when supported by modern survey techniques (e.g., life grid), as they may minimize recall failures (e.g., Berney and Blane 2003). Based on the evidence presented in this paper, we expect that similar results should partly overcome the skepticism about the usage of retrospective data and encourage their collection in longitudinal studies. However, more research is needed before giving a definitive judgment on the reliability of retrospective information on childhood circumstances collected at older ages. In particular, it would be helpful to have repeated subjective assessments of the same childhood information over time in order to verify whether life events and the aging process have an influence on the recall of this information.

Acknowledgments This paper uses data from SHARE release 2.3.1. SHARE data collection in 2004–2007 was primarily funded by the European Commission through its 5th and 6th framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the US National Institute on Aging (Grant No. U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGH04-064; 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).

References

- Auriat, N. (1991). Who forgets? An analysis of memory effects in a retrospective survey on migration history. *European Journal of Population*, 7, 311–342.
- Almond, D., & Currie, J. (2011). Human capital development before age five. *Handbook of Labor Economics*, 4(B), 1315–1486.
- Almond, D., & Mazumder, B. (2005). The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data. *American Economic Review*, 95(2), 258–262.
- Barker, D. J. (1998). In utero programming of chronic diseases. *Clinical Science*, 95, 115–128.
- Barro, R., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, forthcoming. doi:10.1016/j.jdevco.2012.10.001
- Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., & Peterson, C. (2001). The quality of retrospective data: An examination of long-term recall in a developing country. *Journal of Human Resources*, 36(3), 593–625.
- Belli, R. (1998). The structure of autobiographical memory and the history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory*, 6(4), 383–407.
- Belli, R. (2005). The integration of a computer assisted interviewing event history calendar in the panel study of income dynamics. PSID Technical Series Paper #05-01.
- Berney, L., & Blane, D. (1998). Collecting retrospective data: Accuracy of recall after 50 years judged against historical records. *Social Science and Medicine*, 45(10), 1519–1525.
- Berney, L., & Blane, D. (2003). The lifegrid method of collecting retrospective information from people at older ages. *Research Policy and Planning*, 21(2), 13–22.
- Blackwell, D., Hayward, M. D., & Crimmins, E. M. (2001). Does childhood health affect chronic morbidity later in life? *Social Science and Medicine*, 52(8), 1269–1284.
- Börsch-Supan, A., & Jürges, H. (2005). *The survey of health, aging, and retirement in Europe. Methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Börsch-Supan, A., & Schröder, M. (2011). Retrospective data collection in the survey of health, ageing and retirement in Europe. SHARELIFE methodology. In M. Schröder (Ed.), *Retrospective data collection in the survey of health, ageing and retirement in Europe* (pp. 5–10). Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Bozzoli, C., Deaton, A., & Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, 46(4), 647–669.
- Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, 5(59), 3707–3833.
- Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24(2), 365–389.
- Case, A., & Paxson, C. (2009). Early life health and cognitive function in old age. *American Economic Review Papers and Proceedings*, 99(2), 104–109.
- Currie, J. (2009). Healthy, wealthy, and wise? Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47(1), 87–122.
- De Luca, G., & Peracchi, F. (2005). Survey participation in the first wave of SHARE. In A. Börsch-Supan & H. Jürges (Eds.), *The survey of health, aging, and retirement in Europe—methodology*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- De Luca, G., & Peracchi, F. (2012). Estimating Engel curves under unit and item nonresponse. *Journal of Applied Econometrics*, 27(7), 1076–1099.
- De Luca, G., & Rossetti, C. (2013). Weights. In *SHARE release guide 1.1.1, wave 4*.
- del Arco Blanco, M. A. (2010). Hunger and the consolidation of the francoist regime (1939–1951). *European History Quarterly*, 40(3), 458–483.

- Dex, S. (1995). The reliability of recall data: A literature review. *Bulletin de Méthodologie Sociologique*, 49, 58–89.
- Francesconi, M. (2005). An evaluation of the childhood family structure measures from the sixth wave of the British household panel survey. *Journal of the Royal Statistical Society—Series A*, 168, 539–566.
- Garroute, C., & Paccagnella, O. (2010). *Data quality: Three examples of consistency across SHARE and SHARELIFE*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Gluckman, P., & Hanson, M. (2005). The developmental origins of adult disease. *Maternal and Child Nutrition*, 1(3), 130–141.
- Haas, S. (2007). The long-term effects of poor childhood health: An assessment and application of retrospective reports. *Demography*, 44(1), 113–135.
- Haas, S., & Bishop, N. (2010). What do retrospective subjective reports of childhood health capture? Evidence from the WLS and the PSID. *Research on Aging*, 32(6), 698–714.
- Havari, E., & Peracchi, F. (2012). *Childhood circumstances and adult outcomes: Evidence from SHARELIFE*. Working paper 11–15, Einaudi Institute for Economics and Finance (EIEF), Rome.
- Jürges, H. (2007). Unemployment, life satisfaction and retrospective error. *Journal of the Royal Statistical Society—Series A*, 170, 43–61.
- Kesternich, I., Siflinger, B., Smith, J. P., & Winter, J. K. (2013). Individual behavior as a pathway between early-life shocks and adult health. *Economic Journal* (forthcoming)
- Krall, E. A., Valadian, I., Dwyer, J. T., & Gardner, J. (1988). Recall of childhood diseases. *Journal of Clinical Epidemiology*, 41, 1059–1064.
- Maddison, A. (2010). Statistics on world population, GDP and per capita GDP, 1–2008 AD. Retrieved April 17, 2013, from http://www.ggdc.net/MADDISON/Historical_Statistics/horizontal-file_02-2010.xls.
- Mazzonna, F. (2014). The long lasting effect of family background: A European cross-country comparison. *Economics of Education Review*, 40, 25–42.
- Mazzonna, F., & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4), 691–710.
- McArdle, J. J., Smith, J. P., & Willis, R. (2011). Cognition and economic outcomes in the health and retirement survey. In D. Wise (Ed.), *Explorations in the economics of aging*. Chicago: University of Chicago Press.
- Pampel, F., & Pillai, V. (1986). Patterns and determinants of infant mortality in developed nations, 1950–1975. *Demography*, 23, 525–542.
- Raphael, K. (1987). Recall bias: A proposal for assessment and control. *International Journal of Epidemiology*, 16(2), 167–170.
- Schröder, M. (2011). *Retrospective data collection in the survey of health, ageing and retirement in Europe*. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).
- Schröder, M., & Börsch-Supan, A. (2008). *Retrospective data collection in Europe*. Mannheim Research Institute for the Economics of Aging (MEA), Discussion paper no. 172–2008.
- Smith, J. (2009a). Reconstructing childhood health histories. *Demography*, 46(2), 387–403.
- Smith, J. (2009b). The impact of childhood health on adult labor market outcomes. *The Review of Economics and Statistics*, 91(3), 478–489.
- Smith, J., & Thomas, D. (2003). Remembrances of things past: Test–retest reliability of retrospective migration histories. *Journal of the Royal Statistical Society—Series A*, 166, 23–49.
- van den Berg, G. J., Doblhammer-Reiter, G., & Christensen, K. (2009). Exogenous determinants of early-life conditions, and mortality later in life. *Social Science and Medicine*, 68(9), 1591–1598.
- van den Berg, G. J., Pinger, P. R., & Schoch, J. (2012). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. doi:10.1111/ecco.12250.