SHARE ERIC
SURVEY OF HEALTH, AGEING AND RETIREMENT IN EUROPE

SHARE WORKING PAPER SERIES

What lies behind the education gradient in health? New evidence from a distributional perspective

Iryna Kyzyma and María Noel Pi Alperin

Working Paper Series 65-2021
DOI: 10.17617/2.3307418

SHARE-ERIC | Amalienstr. 33 | 80799 Munich | Germany | share-eric.eu

MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging

This project has received funding from the European Union under grant agreements VS/2019/0332, VS/2020/0313 and the European Union's Horizon 2020 research and innovation programme under grant agreements No 870628, No 101015924.

SPONSORED BY THE Federal Ministry of Education and Research

Supported by the NIH National Institute on Aging

About the SHARE Working Paper Series

The series is designed to provide a timely discussion of results based on SHARE data within the SHARE family, i.e., members of the SHARE Country Teams, Area Coordination Teams and other SHARE bodies. The papers are not peer reviewed; the authors are solely responsible for the scientific content and the graphical layout of their submissions. The respective Country Team Leaders and Area Coordinators are encouraged to look over the submissions by their team members.

The publisher (SHARE ERIC) checks working papers in this series for formal issues such as proper acknowledgements to the funders of SHARE. The publisher takes no responsibility for the scientific content of the paper.

Acknowledgements

This paper uses data from SHARE Waves 5 and 6 (DOIs: 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710), see Börsch-Supan et al. (2013) for methodological details. *The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged* (see www.share-project.org).

Iryna Kyzyma also acknowledges funding received from the 'Fonds National de la Recherche Luxembourg' (Grant INTER/NORFACE/16/11333934 – "The impact of childhood circumstances on individual outcomes over the life course" (IMCHILD)).

The authors would like to thank the participants of the SHARE Users Workshop in Luxembourg, the Workshop on Labour Economics in Trier, and iHEA Congress in Basel for numerous helpful comments.

What lies behind the education gradient in health? New evidence from a distributional perspective

Iryna Kyzyma[a,b,*] and **María Noel Pi Alperin**[a]

March, 2021

Abstract: Despite a large number of studies, the evidence on the causal effect of education on health remains inconclusive. Among other things, it might be related to the fact that the impact of education on health in most cases is evaluated at the mean whereas little is known about what happens in other parts of the distribution. Using data from the Survey of Health, Ageing and Retirement in Europe, this paper aims to analyze the education gradient in health and its underlying factors from a distributional perspective. We account for the endogeneity of individual education by exploiting the cross-country cross-time variation in the number of compulsory schooling years induced by policy changes. Our findings show that higher educated individuals indeed enjoy better health than those who are lower educated. The gradient, however, is not constant along the health distribution, with the gap being three times bigger at the top of the distribution than at its bottom. The results further suggest that most of the gradient can be explained by the sub-group differences in individual characteristics, especially those related to employment patterns, health behaviors, and parental background.

Keywords: causality, decomposition, distributional approach, education gradient in health, SHARE.

JEL classification: D30, I10, I20.

[a] Luxembourg Institute of Socio-Economic Research (LISER)
[b] IZA Bonn
* Corresponding author: LISER, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, G.D. Luxembourg, tel.: +352 58 58 55 481, e-mail: iryna.kyzyma@liser.lu. ORCHID: https://orchid.org/0000-0001-8163-2269.

1. Introduction

There is extensive evidence that education is positively correlated with health. The relationship persists across countries and over time, and holds for various health measures (Kunst et al., 2005; Cutler & Lleras-Muney, 2008; Conti et al., 2010; Jürges, 2010). The evidence, however, become much less consistent when one turns to causal estimates. Whereas most of the studies reveal a positive effect of education on health (e.g. Adams, 2002; Oreopoulos, 2006; Mazumder, 2008; Silles, 2009; Schneeweis et al., 2014; Brunello et al., 2016), other studies do not find this effect (Arendt, 2005; Clark and Royer, 2013), or yield heterogeneous results depending on the health measure, gender, or age group (Kemptner et al., 2011; Brunello et al., 2013; Jürges et al., 2013).

One of the potential reasons why this might be the case is that previous research on the relationship between education and health does not go much beyond the mean or focus on the analysis of aggregate measures of health inequality (e.g., a relative risk ratio, between-group differences in the average level of health, and the concentration index). These measures provide information on the level of health inequality in the population, but tell us little about the difference in health outcomes between the lower and higher educated in different parts of the health distribution. For example, would we find the same health gradient at the bottom of the distribution as at the top? Is there a point in the distribution of health at which the education gradient disappears? If this is the case, it might (at least partially) explain the inconclusive evidence on the causal impact of education on health in existing literature.

In this paper, we aim to analyze the education gradient in health and its underlying factors from a distributional perspective. We start by constructing a separate distribution of health for two education sub-groups–lower and higher educated–and then compare the difference in their health levels at each point of the distribution. As a next step, we perform an Oaxaca-Blinder decomposition based on the Recentered Influence Function (RIF) approach of Firpo et al. (2009) to explore the factors underling the education gradient in health at various points of the health distribution.

The analysis relies on data from waves 5 and 6 of the Survey of Health, Ageing and Retirement in Europe, covering multiple European countries and Israel. As a health measure, we use a continuous synthetic indicator of health summarizing the health status of individuals across multiple aspects reflecting their physical and mental health. To overcome the endogeneity of

education, we instrument it using variation in the duration of compulsory schooling induced by the educational reforms, which were passed in different countries in different years.

Our paper contributes to existing literature on education-related health inequalities in two ways. First, rather than relying on differences in the average level of health between education sub-groups, or aggregate health inequality measures, we analyze the gap between the entire distributions of health constructed for individuals with different levels of education while accounting for its endogeneity. This approach allows us to explore whether the education gradient in health remains stable along the distribution of health or varies in its upper and lower parts, where individuals with different levels of health are located.

Second, we perform a decomposition exercise in order to explore which factors underlie the education gradient in health. To date, literature has focused on the decomposition of health inequality indices or identification of particular channels via which education influences health outcomes, using regression techniques. We extend this literature by quantifying the contribution of various individual characteristics to the education gradient in health at different points of the health distribution. Although the analysis is descriptive in nature, as any decomposition exercise, it provides an indicative evidence on the channels via which differences in education translate into differences in health along the entire distribution of health outcomes.

The rest of the paper is structured as follows. Section 2 presents the data and definitions of the constructs (i.e., health and education). Section 3 describes the methodology. Section 4 provides the results and Section 5 concludes.

2. Data and definitions

In this paper, we use data from the Survey of Health, Ageing and Retirement in Europe (SHARE).[1] SHARE is a multidisciplinary, cross-national panel survey, which collects harmonized data on health and health behaviors, socio-economic characteristics (e.g. gender, age, employment status, income, health behavior, life-styles), and family and social networks of individuals aged 50 years or older. The analysis is mainly based on data from wave 6 (2015) to which we matched

[1] For a detailed description of SHARE, see Börsch-Supan et al. (2013).

information on individual's childhood circumstances from wave 5 (2013). The countries covered in these two waves include 13 European countries and Israel.[2]

The main advantage of SHARE for our study is that it contains a wide range of variables referring to both physical and mental health. The richness of this information makes SHARE a unique dataset for studying health-related questions, including socio-economic inequalities in health (O'Donnell, 2009).

2.1. Definition of individual health status

We measure the health status of each individual using the multidimensional approach of Pi Alperin (2016). This approach relies on the aggregation of health items, which reflect different aspects of individual health, in one synthetic health indicator (for a detailed description of the method, see Appendix A). Individual synthetic health scores are calculated as the weighted mean of health items accounting for different degrees of 'deprivation' within each item and the correlation between them following the weighting approach of Betti and Verma (1998). [3] Through the way they are constructed, individual indicators of health are continuous in nature and can take any value on the interval between 0 (healthy in all health items) and 1 (sick in all health items).

The main advantage of the constructed synthetic indicators of health compared to the self-assessed health measure is that they utilize information on multiple questions regarding different aspect of individual health instead of posing a unique question about the state of health in general. These questions refer, among other, to various diagnosed diseases, limitations in daily activities, depression symptoms (used to derive the EURO-D scale –see Prince et al. (1999) for more details), and objective measurements of cognitive abilities of the respondents by interviewers. Although a part of this information is reported, and reporting styles differ substantially by education and countries (Jürges, 2007; Bago d'Uva et al., 2008; O'Donnell, 2009; Bago d'Uva et al., 2011), it has been shown that the inclusion of a larger number of health items in the health indicator allows one to decrease reporting-related bias in the measurement of health (van Doorslaer et al., 2000).

[2] The specificities of SHARE wave 5 and wave 6 are described in Malter and Börsch-Supan (2015, 2017) and Börsch-Supan (2020a, 2020b).

[3] For a sensitivity check, we also used a set of alternative weighting schemes but the conclusions of the paper remained largely unchanged (see Appendix C).

Table 1. Health items included in the synthetic health indicator[4]

Dimensions of health		Health items	List of variables included in each item
Global health	Mental health	Depression	Depression, concentration, guilt, loss of interest, sleep, irritability, appetite, stress, pessimism, suicide, enjoyment, tearfulness
		Memory	Orientation regarding: date, day of the week, month, year
		Cognition	Capacity to memorize a given number of words (first trial and delayed trial)
	Physical health	Long-term illnesses	Heart attack, stroke, cancer, ulcer, cataract, fracture of the femur, other fractures, rheumatism, hypertension, high cholesterol, diabetes, pneumonia, Parkinson's, Alzheimer's, anxiety, arthrosis, renal problems
		Limitation activities 1	Daily activities: dressing, bathing or showering, eating, cutting up food, walking across a room, getting in or out of bed
		Limitation activities 2	Instrumental activities: telephone calls, taking medication, managing money, shopping for groceries, preparing a hot meal
		Limitation activities 3	Mobility: walking 100 meters, walking across a room, climbing several flights of stairs, climbing one flight of stairs
		Eyesight	Farsighted, nearsighted
		Hearing	Hearing difficulties

Note: The selected variables are those available in SHARE.

2.2. Definition of education

As a measure of education, we consider the number of years individuals reportedly spent in schooling. In its original form, this measure is likely to be endogenous (see, among other, Groot and van den Brink, 2007; Cutler and Lleras-Muney, 2008). First, there might be a problem of reverse causality when poor health precludes individuals from attaining higher levels of education. Second, there might be other factors simultaneously affecting educational attainment and health, such as genetic traits or parental background.[5] The extent to which these two groups of factors shape the relationship between education and health is, however, unclear. Cutler and Lleras-Munley (2008), for example, have found that family background accounts for about 38 percent of

[4] See Appendix B for a detailed description of the variables included in the different health items.

[5] For an overview of the literature on the relationship between childhood circumstances and health see Case et al. (2002) and Case et al. (2005)

the correlation between education and health concluding that the true effect of education on health is still relatively large.

In order to account for endogeneity of education, we exploit educational reforms implemented between 1949 and 1970 in a number of countries covered by SHARE.[6] Being passed in different years in different countries, these reforms provide an exogenous variation in the duration of compulsory schooling, which we use to instrument the number of actual years individuals spent in education. This estimation strategy is well established in the literature and was previously applied to SHARE data, among other, by Brunello et al. (2009; 2013; 2017), Gathmann et al. (2015), Schneeweis et al. (2014), Fort et al. (2014).

Table 2 presents the selected compulsory schooling reforms.[7] For each reform, we report the year when it came into force, the change in the number of compulsory schooling years, and the first cohort affected by the reform (pivotal cohort). In general, the described reforms increased the duration of compulsory education by 1 year in Austria, Czech Republic, Germany, Sweden, and Switzerland, by 2 years in France, Israel, Spain, and by 3 years in Denmark and Italy.

In order to avoid comparing individuals born too far from the policy change, we followed the strategy of other researchers (e.g. Brunello et al., 2009, 2013; Fort et al., 2016) and narrowed the pre- and post-treatment samples to individuals born 8 years before / after the reform.[8] Figure 1 plots the average number of years spent in schooling by distance from the reform. It clearly shows that individuals affected by the reforms spend, on average, more time in education as compared to those who are not affected suggesting that the increase in the number of compulsory years of schooling resulted in an increase in educational attainment.

[6] In some countries, which are present in waves 5 and 6 of SHARE, the reforms took place too early or too late for our observation period (i.e. Belgium, Estonia, Luxembourg, and Slovenia). Hence, we had to exclude those countries from the analysis. Following Brunello et al. (2009) we also excluded all East German states since there were no reforms there over the period of interest.

[7] In this paper, we followed the strategy of Brunello et al. (2009) and focused only on one reform per country.

[8] This choice is driven by the trade-off between the necessity to exclude confounding reforms taking place around that time and the need to have a sufficient sample size to perform the estimations.

Table 2. Selected reforms of compulsory schooling

Country	Year of the reform	Change in minimum school leaving age	Change in the number of compulsory years of schooling	Pivotal cohort
Austria	1962	14 → 15	8 → 9	1951
Czech Republic	1960	14 → 15	8 → 9	1947
Denmark	1958	11 → 14	4 → 7	1947
France	1959	14 → 16	8 → 10	1953
Germany:				
Baden-Württemberg	1967	14 → 15	8 → 9	1953
Bayern	1969	14 → 15	8 → 9	1955
Bremen	1958	14 → 15	8 → 9	1943
Hamburg	1949	14 → 15	8 → 9	1934
Hessen	1967	14 → 15	8 → 9	1953
Niedersachsen	1962	14 → 15	8 → 9	1947
Nordrhein-Westfalen	1967	14 → 15	8 → 9	1953
Rheinland-Pfalz	1967	14 → 15	8 → 9	1953
Saarland	1964	14 → 15	8 → 9	1949
Schleswig-Holstein	1956	14 → 15	8 → 9	1941
Israel	1968	13 → 15	8 → 10	1956
Italy	1963	11 → 14	5 → 8	1949
Spain	1970	12 → 14	6 → 8	1957
Sweden	1949	13 → 14	6 → 7	1936
Switzerland	1970	14 → 15	8 → 9	1961

Source: Information on the reforms is taken from Brunello et al. (2017) for Austria, Czech Republic, Denmark, France, Germany, Italy, and Sweden; from Brunello et al. (2013) for Spain; and from Salonen and Pöyliö (2017) for Israel and Switzerland.

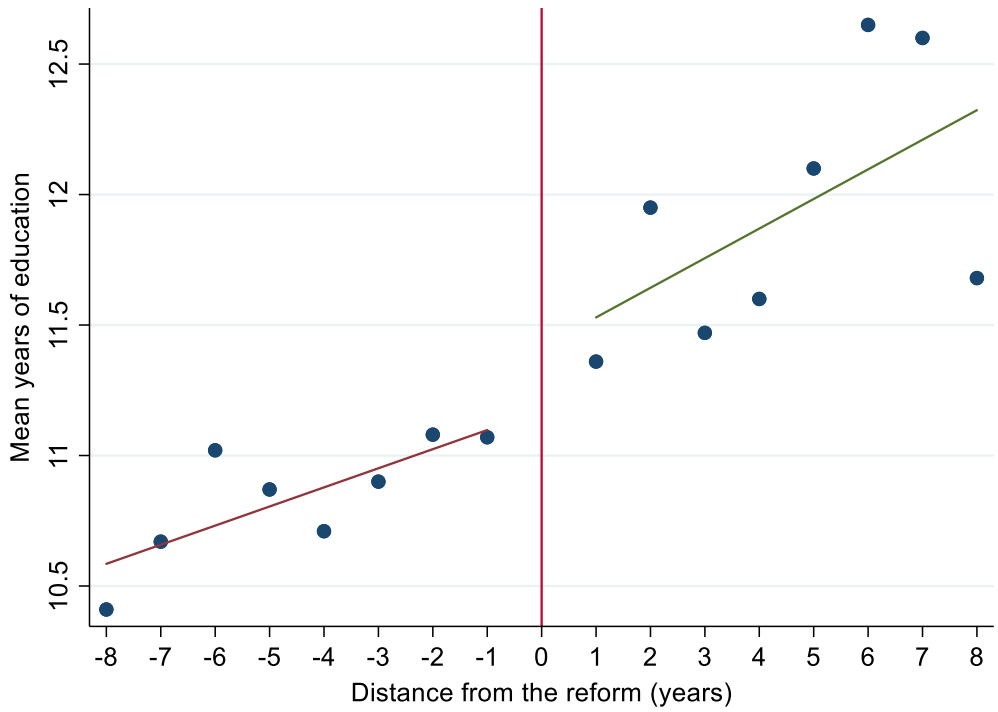

Figure 1. Average number of years spent in education by distance from the reform

Using information on the compulsory schooling reforms described above, we instrument the number of years individuals spent in schooling as follows:

$$Y_{ikc} = \beta_0 + \beta_1 COMP_{kc} + \beta_2 Born_{ikc} + \beta_c D_c + \beta_k D_k + \varepsilon_{ikc}, \quad (1)$$

$$\widehat{Y_{ikc}} = \widehat{\beta_0} + \widehat{\beta_1} COMP_{kc} + \widehat{\beta_2} Born_{ikc} + \widehat{\beta_c} D_c + \widehat{\beta_k} D_k, \quad (2)$$

where Y_{ikc} and $\widehat{Y_{ikc}}$ stand for the actual and instrumented number of years spent in education by individual *i* born in year *k* in country *c*; *COMP* is the number of years of compulsory education required for cohort *k* in country *c*; $Born_{ikc}$ is a dummy capturing whether the person was born in the country of current residence; D_c is a set of country dummies; D_k is a set of year-of-birth dummies; and ε_{ikc} is an individual specific error term. The inclusion of country fixed effects is needed to account for country differences in reporting styles and institutions affecting health of all

individuals living there (for example, the specificities of healthcare systems). Cohort fixed effects aim to account for unobserved characteristics of all individuals born in the same year.[9]

Table 3 presents the estimates from Equation 1. In line with what we have seen in Figure 1, it shows that the compulsory schooling reforms have a significant effect on the number of years individuals spend in education. On average, an additional year of compulsory schooling increases educational attainment by 0.35 years, which corresponds to the findings of previous studies exploiting the variation in schooling laws for instrumenting individual education (see, among other, Brunello et al., 2009, 2013; Fort et al., 2013; Schneeweis et al. 2014).

Table 3. The impact of compulsory schooling reforms on educational attainment

Years of compulsory schooling	0.357*** (0.081)
F-statistics	41.08
R-squared	0.155
Number of observations	15062

Note: The model is estimated in line with Equation (1). Standard errors in parentheses are clustered by country (region for Germany). The presented F-statistics tests the instrumental variable for weakness. The rule of thumb requires this statistics to be greater than 10 (Staiger and Stock, 1997).
*** statistically significant at 0.001 level.

Using the predicted values of education from Equation (2), we split the sample into two groups – those who have attained a lower level of education and those who have attained a higher level of education. As a cut-off point, we use the instrumented value of education at the 75th percentile, which allows us to classify around 25 percent of individuals into the group of higher educated, the value coming close to the portion of the SHARE respondents reporting the completion of studies beyond secondary school.[10]

[9] Initially, we also controlled for the country differences in trends in health, which might be caused by country specific changes in healthcare systems or other factors underlying individual health beyond the reforms. The trend terms (interactions of country dummies with cohort dummies), however, appeared insignificant. Following Reiss (2016), we excluded them from the final specification as such that add extra noise to the model without improving its fit.

[10] As a sensitivity check, we also used the value of education at the 50th percentile of the distribution of the instrumented number of years spent in education. The results, however, remained similar – although the education gradient in health decreased in size (as one would expect), its pattern along the distribution remained relatively unchanged, so as the factors lying behind it (see Appendix D).

2.3. Other variables

In order to investigate how individual characteristics interact with education, contributing to an increase or a decrease in the education gradient in health, we consider a set of additional variables capturing demographics, labor market characteristics, health behavior, and childhood circumstances of the respondents. Table 4 provides a detailed list of these variables and their mean values by education sub-groups.

The lower educated people in our sample are, on average, slightly older than those who are higher educated.[11] The lower educated are also more likely to be retired or inactive on the labor market, and have much lower income than the higher educated. Compared with the higher educated, lower educated individuals are less likely to engage in sports, but they are also less likely to be heavy drinkers or have ever smoked. Finally, lower educated individuals are more likely to have financial difficulties in childhood or to grow up with lower educated parents.

[11] Empirical evidence suggests that people with low levels of education, on average, live a shorter life than people with high levels of education (Hummer & Lariscy, 2011). The fact that in our sample the lower educated are, on average, older than the higher educated can be explained by the expansion of post-secondary education over time, which resulted in a higher level of education among younger cohorts compared with older ones (see Braga et al., 2013).

Table 4. Differences in individual characteristics between the lower and higher educated

Characteristics	Lower educated	Higher educated	Difference
Demographic characteristics			
Age (mean)	63.01	59.2	+ 3.9***
Female (%)	51.6	50.6	+ 1.0
Labor market characteristics			
Total net equivalized household income (mean)	26252	33945	-7693***
Employment status			
Retired (%)	52.2	26.5	+ 25.7***
Employed (%)	28.3	56.9	- 28.6***
Other (%)	19.5	16.6	+ 2.9***
Health behavior characteristics			
Ever smoked (%)	51.4	53.7	- 2.3***
Engaged in sport			
More than once a week (%)	33.6	46.2	- 12.6***
Once a week (%)	13.8	14.3	- 0.5
One to three times a month (%)	9.2	8.1	+1.1*
Hardly any sporting activity (%)	43.3	31.4	+ 11.9***
Drinking habits			
≥ 6 alcoholic drinks at least once per week	5.8	7.7	-1.9***
≥ 6 alcoholic drinks 1–2 times per month	10.2	19.9	-9.7***
Not at all in the last 3 months	83.7	75.8	+11.5***
Childhood circumstances			
Had financial difficulties as a child (%)	26.4	19.1	+7.3***
Education level of parents			
Father had lower education (%)	74.2	36.5	+37.7***
Mother had lower education (%)	84.8	64.4	+20.4***
Number of observations	2942	12120	

Note: All estimates are weighted using individual sample weights. * p=.05; ** p=.01; *** p=.001. The educational categories are defined based on the instrumented values of educational attainment. Parents with lower education are those who have upper-secondary education or any level below it (corresponding to ISCED 1997 categories 0–3). Parents with higher education are those who have any type of post-secondary education (corresponding to ISCED 1997 categories 4–6).

3. Methodology

3.1. Modeling sub-group differences in health

Consider sample S consisting of N individuals. Let h_i be the individual health status, so that $0 \le h_i \le 1$. Then, health status across all individuals in the sample can be summarized with the cumulative distribution function, $F_H(h)$, as follows:

$$F_H(h) = \Pr(H \le h) = \int_0^H f_H(h)dh\,, \tag{3}$$

where $f_H(h)$ is the probability density function of health. The cumulative distribution function of health in Equation (3) summarizes the chances of a randomly chosen individual to have a health score below or equal to a certain level, H.

Consider further that our sample consists of two mutually exclusive educational sub-groups labeled '0' for lower educated and '1' for higher educated. The gap in health scores between these two sub-groups, $\Delta F_H(h)$, can then be expressed as the difference between their cumulative distribution functions of health:

$$\Delta F_H(h) = F_H^0(h) - F_H^1(h) = \int_0^H f_H^0(h)dh - \int_0^H f_H^1(h)dh\,. \tag{4}$$

3.2. Decomposing sub-group differences in health

In order to explore which factors are associated with the health gap identified in Equation (4), we perform an Oaxaca-Blinder type of decomposition, which relies on the estimation of RIF-regressions proposed by Firpo et al. (2009). The key advantage of this approach is that it allows us to go 'beyond the mean' and decompose sub-group differences in the entire unconditional distribution of health. Unlike other methods allowing for a distributional decomposition (e.g. a reweighting approach of DiNardo et al. (1996) or a distribution regression approach of Chernozhukov et al. (2013)), the RIF-based approach is path independent which offers a nice framework for estimating the contributions of individual covariates to the observed differences in health at different points of the distribution.

In practice, the RIF-based decomposition approach relies on the estimation of the RIF, which provides a linear approximation of the unconditional quantiles of the health variable. To derive the RIF, one needs to compute the sample quantiles and estimate the density of the health distribution at those quantiles using kernel density methods.[12] This information is then used to calculate the RIF as follows:

$$RIF(H;Q_\tau) = Q_\tau + \frac{\tau - 1\{H \le Q_\tau\}}{f_H(Q_\tau)} \quad , \tag{5}$$

where *RIF(H; Q_τ)* is the RIF of health variable *H*, Q_τ is the τ-quantile of the unconditional distribution of health, $f_H(Q_\tau)$ is the density of the marginal distribution of health at quantile Q_τ, and $1\{H \le Q_\tau\}$ is an indicator variable taking the value of 1 if an individual has a health score smaller or equal to the quantile Q_τ and taking the value of 0 otherwise.

As soon as the RIF is derived, its conditional expectation can be modelled as a function of covariates using a simple linear regression model:

$$E[RIF(H;Q_\tau) \mid X] = X\beta_\tau + \varepsilon \ , \tag{6}$$

where *X* is a vector of explanatory variables; β_τ is a vector of associated with them parameters estimated by the ordinary least squares (OLS) procedure; and ε stands for an error term with $E(\varepsilon|X) = 0$.

Applying the law of iterated expectations, the unconditional quantile of the health distribution can be derived as follows:

$$Q_\tau = E_x\left[E[RIF(H;Q_\tau) \mid X]\right] = E[X]\hat{\beta}_\tau \ . \tag{7}$$

[12] All densities estimated in the paper are based on the Epanechnikov kernel function with the 'optimal' bandwidth, which minimizes the mean integrated squared error if the true distribution of the data was Gaussian and a Gaussian kernel were used. As a sensitivity check, we re-ran the analysis using the Gaussian kernel function with the bandwidth of 0.06, which was previously used by Firpo et al. (2009) in the wage distribution setting. The results from the sensitivity check are largely in line with the baseline estimates. The main difference is that the contribution of the compositional differences between the lower and higher educated to the health differential increases with the alternative specification at the bottom and in the middle of the distribution and decreases at the 80th percentile. This applies to all three health distributions.

Due to the linear approximation, RIF-regressions can be combined with the Oaxaca-Blinder decomposition approach (Blinder, 1973; Oaxaca, 1973) to analyze sub-group differences in the outcome variable at various points of its distribution (Fortin et al., 2011). In particular, the overall difference in the health scores between the lower and higher educated at quantile Q_τ, Δ_τ^O, can be decomposed as follows:

$$\Delta_\tau^O = [RIF(H^0;Q_\tau^0) - RIF(H^1;Q_\tau^1)] = (\overline{X}^0 - \overline{X}^1)\hat{\beta}_\tau^0 + (\hat{\beta}^0 - \hat{\beta}^1)\overline{X}^1 = \Delta_\tau^X + \Delta_\tau^S, \qquad (8)$$

where $\overline{X}^0$ and $\overline{X}^1$ are the means of covariates X in the sub-samples of the lower and higher educated, and $\hat{\beta}^0$ and $\hat{\beta}^1$ are coefficients from the unconditional quantile regressions estimated for each of these sub-samples.

The first term in Equation (8), Δ_τ^X, captures the difference in health between the lower and higher educated that is attributed to the differences in their observed characteristics (the composition effect). The second term, Δ_τ^S, captures the education gradient in health that is attributed to the sub-group differences in the effects of these characteristics (structural effect).

Decomposition in Equation (8) is expressed from the viewpoint of the higher educated. This implies that the sub-group differences in the observed characteristics are weighted by the coefficients from the RIF-regression estimated on the subsample of higher educated individuals ($\hat{\beta}^0$). We choose the higher educated as a reference group because they are more efficient producers of health and have systematically higher levels of health than the lower educated (Grossman, 1972; Galama and Kippersluis, 2018).

The composition and structure effects from Equation (8) can be further decomposed into contributions linked to individual characteristics:

$$\Delta_\tau^X = \sum_{k=1}^{K} (\overline{X}_k^0 - \overline{X}_k^1)\hat{\beta}_{\tau,k}^0 \qquad (9)$$

$$\Delta_\tau^S = \sum_{k=1}^{K} (\hat{\beta}_k^0 - \hat{\beta}_k^1)\overline{X}_{\tau,k}^1, \qquad (10)$$

where k (with $k = 1, 2, \ldots, K$) is an element of X.

4. Results

4.1. The education gradient along the distribution of health

Figure 2 below plots the differences in the levels of health between the lower and higher educated at various percentiles of the health distribution. Two important messages emerge from this figure. First, higher educated individuals enjoy better health than those who are lower educated along the entire distribution except its bottom 10th percentile, where the healthiest persons are located. Second, the absolute gap in the levels of health between the lower and higher educated is not constant throughout the distribution: it is relatively small if we compare the healthiest group of the lower and higher educated but it is several times bigger if we compare the sickest group.

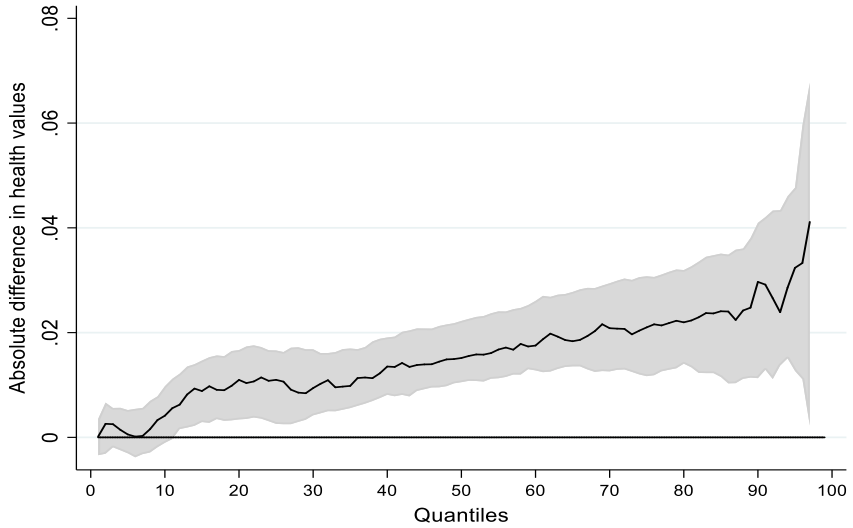

Figure 2. Differences in the levels of health between the lower and higher educated

Note: All estimates are weighted using individual sample weights. 95% bootstrapped confidence intervals (derived from 500 bootstrapped replications).

Table 5 below quantifies the differences in the health scores between the lower and higher educated at various points of the health distribution and at the mean.[13] In line with Figure 2, the gap is not significant at the very bottom of the distribution. At the 25th percentile, however, it constitutes 0.011 points and increases by a factor of three towards the top of the distribution. This evidence implies that the protective effect of education intensifies as we move up the distribution helping to avoid particularly severe health impairments. The fact that the average size of the gap

[13] Recall that health is measured with a composite indicator whose values range between 0 and 1, with most of the values falling between 0 and 0.2.

is only 0.015 points also suggests that without a distributional analysis we would overestimate the size of the gradient at the bottom of the health distribution and underestimate it at the top. Even at the 50th percentile, the difference in health scores between the lower and higher educated is only 0.015 points, implying that the increase in the health gap occurs to a greater extent in the upper part of the distribution than in its lower part.

Table 5. Differences in the levels of health between the lower and higher educated at different points of the health distribution

Educational category	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
			Health values			
Lower educated	0.030*** (0.002)	0.067*** (0.002)	0.111*** (0.002)	0.160*** (0.002)	0.217*** (0.004)	0.122*** (0.001)
Higher educated	0.026*** (0.002)	0.056*** (0.003)	0.096*** (0.002)	0.139*** (0.004)	0.188*** (0.006)	0.106*** (0.003)
	Absolute difference in health values between the lower and higher educated					
Absolute difference	0.004 (0.003)	0.011** (0.004)	0.015*** (0.003)	0.021*** (0.005)	0.030*** (0.007)	0.015*** (0.003)
	Relative difference in health values between the lower and higher educated, in %					
Relative difference	13.3 (8.8)	16.4** (5.3)	13.5*** (2.7)	13.1*** (3.0)	13.8*** (3.3)	+12.3*** (2.5)

Note: All estimates are weighted using individual sample weights. Bootstrapped standard errors in parentheses (derived from 500 bootstrapped replications). * p=.05; ** p=.01; *** p=.001.

Looking at relative differences in the health scores, the lower educated on average have 12.3 percent worse health than the higher educated. The relative difference is the largest at the 25th percentile, where it constitutes 16.4 percent, but it decreases in the middle and upper parts of the distribution falling to 13 - 14 percent. A decline in relative differences along the distribution can be partially explained by higher health values in its upper part as compared to the lower part: when the values of health are small, even a minor absolute difference between the sub-groups might look big in relative terms.

4.2. Decomposition of the education gradient in health

Figure 3 presents the results of the aggregate decomposition of the education-related differential in health at various points of the distribution. The differences are always taken between the lower and higher educated. A positive sign on the compositional component therefore means that characteristics associated with poor health are more prevalent among the lower educated than among the higher educated, contributing to the health differential in favor of the higher educated. Similarly, a positive sign on the structural component means that health returns to individual characteristics are smaller for the lower educated than for higher educated, which increases the health gap between them.

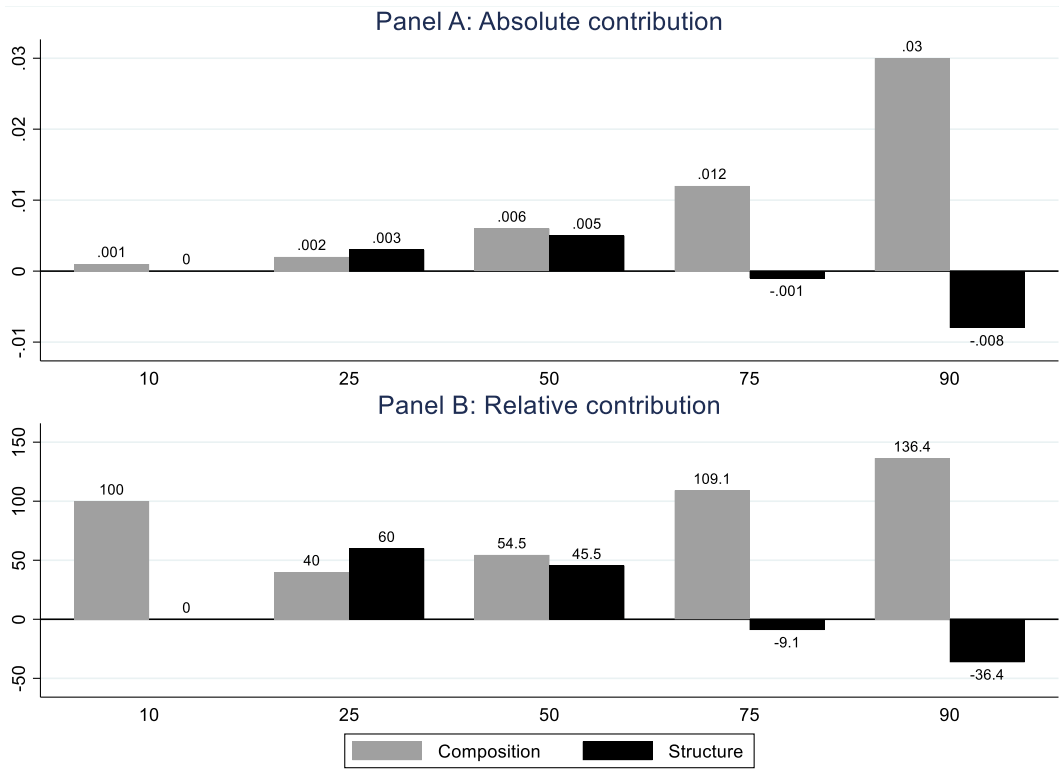

Figure 3. Aggregate decomposition of the education gradient in health

The estimates in Figure 3 suggest that the contribution of the sub-group differences in individual characteristics to the education gradient in health is relatively small at the bottom of the

distribution but steadily increases in size as we move towards its top.[14] At the 25th percentile, for example, the compositional differences between the lower and higher educated account for only 40 percent of the total health gap but their contribution more than triples at the top of the distribution. In contrast, the sub-group differences in the returns to individual characteristics positively contribute to the health gradient only in the middle of the distribution. Towards the top of the distribution, the contribution of the structural component decreases in size and becomes negative implying an inequality reducing effect. This evidence suggests that acquiring higher levels of education influences socio-economic characteristics of individuals, which have a protective impact on their health throughout the life course.

Table 6 below sheds further light on the types of characteristics underlying the health gradient between the lower and higher educated by decomposing the compositional and structural components from Figure 3 into contributions associated with different groups of characteristics. In particular, Panel A shows that differences in demographic characteristics between the lower and higher educated documented in Table 4 do not explain the health gradient at any point of the distribution. Sub-group differences in economic characteristics contribute the most to the explained part of the gradient between the 25th and 75th percentiles and are followed by the differences in health behavior, which are associated with around 20 percent of the education-related health gap. The contribution of economic characteristics, however, decreases and becomes insignificant at the top 10 percentiles of the distribution. In contrast, the role of parental background gains importance in this part of the distribution explaining 81 percent of the health gradient between the lower and higher educated. This evidence suggests that higher educated individuals are not only more likely to come from the families with advantageous social background than the lower educated (see Table 4) but these differences in parental background also translate into differences in health. Furthermore, they become the main factor explaining the health gap between the lower and higher educated at the top of the health distribution encompassing individuals with the most adverse health levels. A potential explanation behind this phenomenon is that parental background brings children a set of advantages (e.g. better nutrition, better educational opportunities, healthier life-styles, better neighborhoods), which convert into health benefits later in life (Currie, 2009; Lundborg et al., 2018).

[14] Recall that the health gradient due to education is very small and not statistically significant at the bottom 10 percentiles of the health distribution.

Table 6. Decomposition results by the groups of individual characteristics

	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Estimated total difference	+0.001 (0.003)	+0.006 (0.004)	+0.010 (0.004)	+0.011 (0.006)	+0.021 (0.009)
Panel A: Due to differences in characteristics (composition)					
Demographic characteristics	-0.003 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.003 (0.003)	+0.002 (0.005)
Economic characteristics	+0.002 (0.001)	+0.003 (0.002)	+0.004 (0.002)	+0.008 (0.003)	+0.005 (0.004)
Health behaviours	+0.001 (0.001)	+0.001 (0.001)	+0.002 (0.001)	+0.002 (0.001)	+0.006 (0.002)
Parental background	+0.001 (0.002)	+0.000 (0.002)	+0.000 (0.002)	+0.004 (0.003)	+0.017 (0.005)
Total	+0.001 (0.003)	+0.002 (0.003)	+0.005 (0.003)	+0.012 (0.005)	+0.030 (0.008)
Panel B: Due to differences in returns to characteristics (structure)					
Demographic characteristics	+0.169 (0.007)	+0.163 (0.080)	+0.119 (0.071)	+0.224 (0.094)	+0.021 (0.164)
Economic characteristics	+0.019 (0.013)	+0.018 (0.013)	+0.004 (0.012)	+0.020 (0.019)	+0.021 (0.034)
Health behaviours	+0.000 (0.005)	+0.005 (0.006)	-0.002 (0.005)	-0.009 (0.007)	-0.021 (0.012)
Parental background	+0.007 (0.008)	+0.010 (0.008)	+0.018 (0.008)	+0.005 (0.010)	-0.010 (0.014)
Constant	-0.195 (0.073)	-0.194 (0.084)	-0.134 (0.075)	-0.241 (0.102)	-0.020 (0.174)
Total	-0.000 (0.004)	+0.003 (0.005)	+0.005 (0.004)	-0.001 (0.007)	-0.009 (0.011)

Note: All estimates are weighted using individual sample weights. Bootstrapped standard errors in parentheses (derived from 500 bootstrapped replications).

Panel B in Table 6 reveals that higher and lower educated have different returns to demographic characteristics and these differences contribute significantly to the health gradient along the entire distribution of health except its top decile. The impact of education on health also varies depending on parental background but only in the middle of the distribution. In particular, we find that education magnifies early advantages of social background around the 50th percentile of the health distribution contributing to an increase in health disparities between the lower and higher educated. In contrast, exercising healthier behaviors give more health benefits to those who are lower educated than to those who are higher educated, at least at the top of the health distribution where the estimates become significant. It might be linked to the fact that lower

educated in that part of the distribution already have much worse levels of health than their higher educated counterparts, which makes healthier behaviors even more important for the former.

Figures 4 and 5 decompose the contributions of the four groups of factors outlined in Table 6 into contributions associated with particular covariates. Figure 4 does it for the part of the gradient associated with the sub-group differences in individual characteristics whereas Figure 5 plots the contributions associated with the differences in the returns to those characteristics.

Figure 4 shows that employment, exercising, and parental education are the only characteristics, which have a different prevalence between the lower and higher educated and are, at the same time, significantly associated with the sub-group differences in health. The impact of these variables, however, is not the same along the health distribution. Sub-group differences in employment patterns, for example, explain a part of the health gradient starting from the 35th percentile of the distribution but their explanatory power disappears at the top decile. This relates to the fact that education predefines the likelihood of employment and the types of jobs individuals occupy. Higher educated people tend to hold occupations and perform jobs, which foresee better work conditions, lower physical workloads, more active thinking, and offer more social protection and stability (Ross and Wu, 1996; Lynch, 2003). Such jobs are also typically better paid and help individuals to maintain their health for longer periods (Leopold, 2018). However, if health deteriorates to very low levels, people cannot keep working as before regardless of the level of their education, and the sub-group differences in employment patterns disappear (as we observe at the top of the health distribution).

A further look at Figure 4 reveals that differences in the frequency of exercising also explain a substantial portion of the health gradient between the lower and higher educated. Furthermore, the contribution of this factor increases along the distribution mirroring the increase in the gradient itself. If the lower educated were exercising a similar number of times per week as the higher educated do, the health gradient between them would be at least 14.7 percent smaller at the 45th percentile of the distribution (whether the contribution of this factor becomes significant) and 17 percent smaller at the 95th percentile. At the very top of the distribution, however, the most important factor explaining the sub-group differences in health is parental background. According to our calculations, between 62 and 79 percent of the education-related health gap can be attributed to this factor between the 80th and 95th percentiles. The effect becomes insignificant and disappears at the top 5th percentile of the distribution comprising the sickest individuals in the sample.

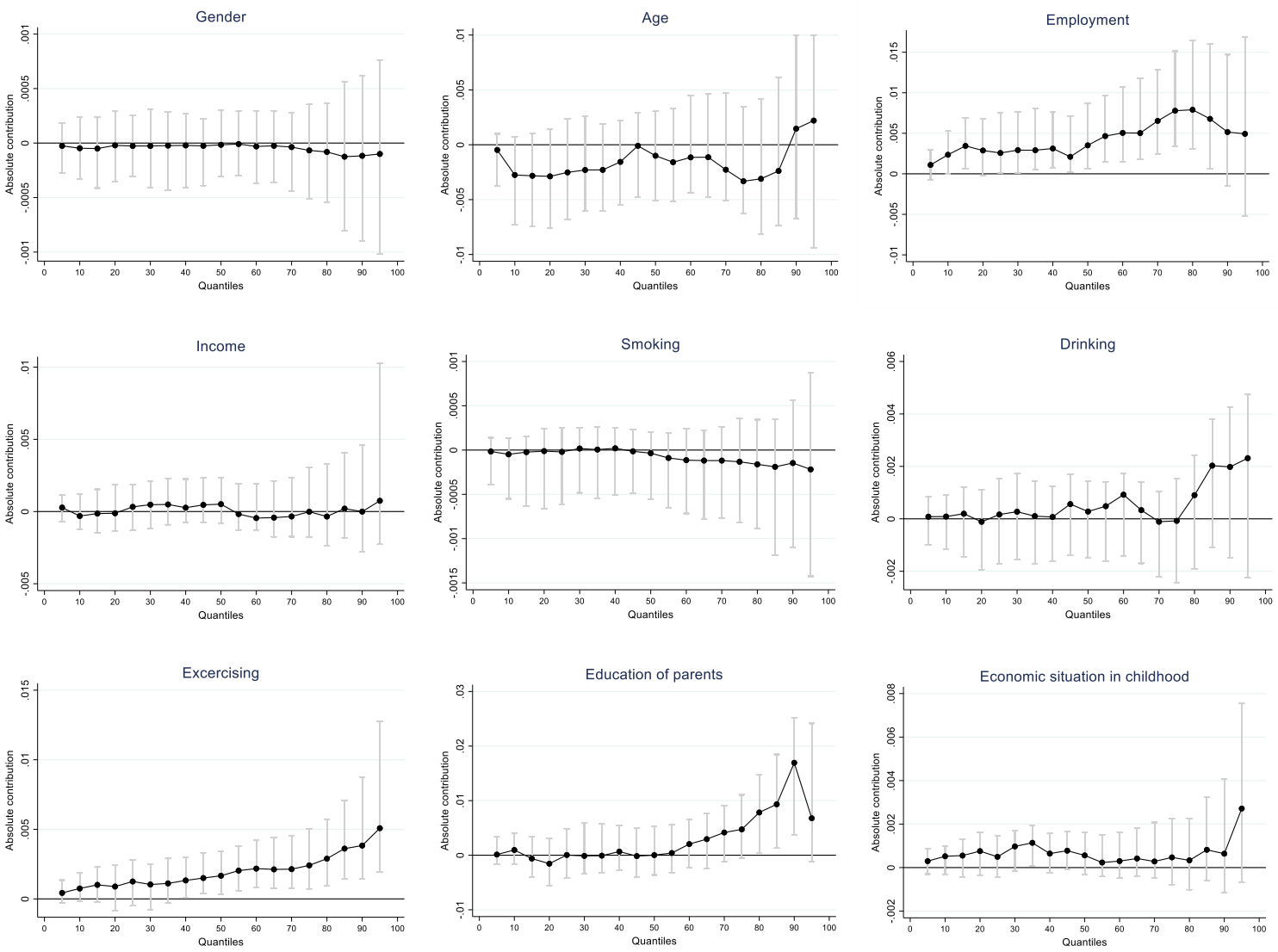

Figure 4. Contributions of the differences in individual characteristics to the education-related gradient in health

Note: All estimates are weighted using individual sample weights. Vertical lines plot 95% confidence intervals derived from 500 bootstrapped replications.

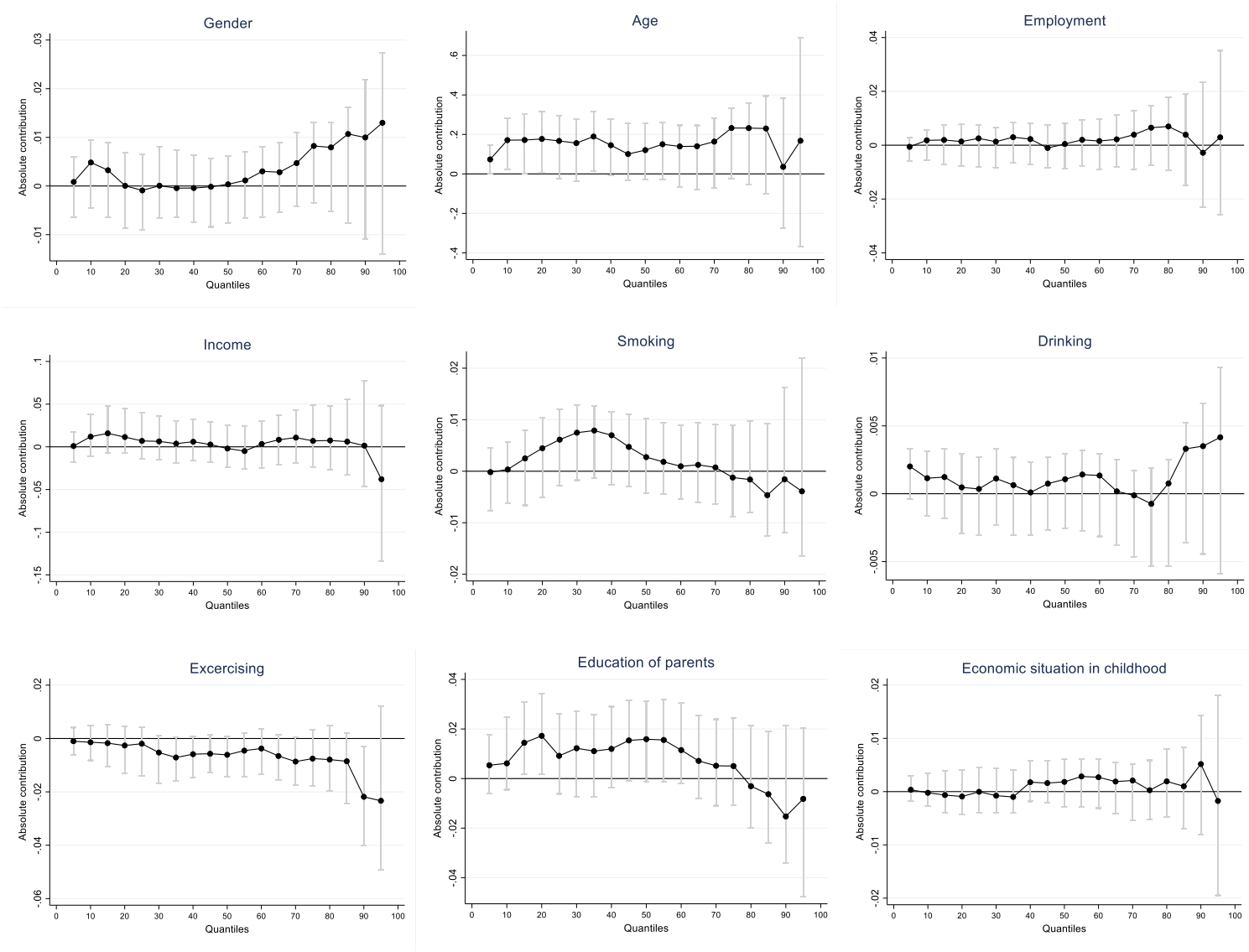

Figure 5. Contributions of the differences in returns to individual characteristics to the education-related gradient in health

Note: All estimates are weighted using individual sample weights. Vertical lines plot 95% confidence intervals derived from 500 bootstrapped replications.

The detailed decomposition of the structural component shows that a portion of the education-related gradient in the lower part of the health distribution can be explained by the sub-group differences in the returns to age and parental background. The contribution associated with age comes in line with the cumulative disadvantage hypothesis, which emphasizes that health disparities between education groups increase with age (Ross and Wu, 1996; Mirowsky and Ross, 2008, Leopold, 2018). Educated people, for example, are more likely to follow a healthy diet, exercise, do regular health checks, etc. bringing benefits to health, which tend to accumulate over the life course. As a result, the level of health of someone who is higher educated might be better than the level of health of someone who is lower educated albeit both individuals are of the same age.

Figure 5 also shows that the impact of education on health is reinforced by parental education bringing additional health benefits to the group of higher educated individuals, who come from the families with higher educated parents. The effect, however, is significant only between the 15th and 20th percentiles of the distribution comprising relatively healthy individuals. In contrast, the sub-group differences in returns to exercising are in favor of the lower educated and the estimates are significant only at the top of the distribution contributing to a decrease in the health gradient there. Among healthier individuals located at the bottom of the distribution, the same frequency of exercising brings the same health benefits to both education sub-groups. This evidence points once again that exercising is especially beneficial for the lower educated in the sickest group, helping them to restrain further health deterioration.

5. Conclusions

This paper contributes to the existing literature in two ways. First, we go beyond the mean and estimate differences in the level of health between the lower and higher educated at different points of the health distribution. Second, using distributional decomposition techniques, we investigate the factors explaining those differences, which helps us to uncover the mechanisms standing behind the education-health relationship. In order to overcome the endogeneity of education – prior to classifying individuals into two groups by the level of their education – we instrument it with a set of educational reforms, which changed the duration of compulsory schooling in different countries at different points in time.

In line with the previous literature, our findings suggest that higher educated individuals, on average, enjoy better health than those who are lower educated. We show, however, that the health gradient is not constant along the distribution: it is relatively small if we compare the healthiest 10 percent of the lower and higher educated, but multiplies by the factor of three if we compare the sickest 10 percent. These results provide two main messages: (1) education is especially important for precluding high levels of health impairments, and (2) the relationship between education and health is non-linear which might serve as an explanation of why previous studies estimating the average effect of education on health produced inconclusive results.

Our findings suggest that most of the health gradient between the higher and lower educated can be explained by the sub-group differences in individual characteristics rather than by the differences in returns to them. In particular, we find that higher educated people are more likely to keep their employment and to do physical exercising than the lower educated, which is associated with a substantial portion of the gradient in the lower part of the distribution. Differences in exercising also explain 14-17% of the health gradient in the upper part of the distribution but the largest portion of the gradient at the top is linked to the differences in parental background.

Apart from the sub-group differences in characteristics, we also find that the effect of education on health varies depending on age, exercising, and parental background revealing further mechanisms standing behind the gradient. In line with the cumulative advantage hypothesis, our results suggest that higher educated people are better managers of their health, which helps them to maintain better health throughout the life-course as compared to the lower educated. In a similar way, the protective impact of education on health increases for those who grew up in the families where at least one parent was highly educated.

From the policy perspective our results suggests that increasing the level of education will not only have a positive impact on health in the current generation (through, for example, exercising healthier behaviors or improving the quality of employment) but also will help to narrow the education gradient in health in future generations. In addition, promoting the importance of exercising and making it more accessible for lower educated individuals (who typically have low income and limited access to sports facilities) might also help decreasing the education related gradient in health.

References

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

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

Bago d'Uva, T., O'Donnell, O., & van Doorslaer, E. (2008). Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans. *International Journal of Epidemiology*, 37, 1375–1383.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., & van Doorslaer, E. (2011). Education-related inequality in healthcare with heterogeneous reporting of health. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 174(3), 639–664.

Betti, G. & Verma, V. K. (1998), *Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory*, Working Paper 22, Dipartimento di Metodi Quantitativi, Università di Siena.

Blinder, A.S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8, 436-455.

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

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

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

Braga, M., Checchi, D., & Meschi, E. (2013). Educational policies in a long-run perspective. *Economic Policy*, 28(73), 45–100.

Brunello, G., Fort, M. & Weber, G. (2009). Changes in compulsory schooling, education and the distribution of wages in Europe. *The Economic Journal*, 119, 516-539.

Brunello, G., Fabbri, D., & Fort, M. (2013). The causal effect of education on body mass: Evidence from Europe. *Journal of Labor Economics*, 31 (1), 195-223.

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

Brunello, G., Weber, G., & Weiss, C. (2017). Books are forever: Early life conditions, education and lifetime earnings in Europe. *The Economic Journal*, 127, 271-296.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92 (5), 1308–1334.
Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstances. *Journal of Health Economics*, 24, 365-389.

Chernozhukov, V., Fernandez-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6), 2205-2268.

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

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *American Economic Review: Papers & Proceedings*, *100*, 234–238.

Currie, J. (2009). Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47, 87-122.

Cutler, D. M., & Lleras-Muney, A. (2008). Education and health: Evaluating theories and evidence. In J. House, R. Schoeni, G. Kaplan & H. Pollack (Eds.) *Making Americans Healthier: Social and Economic Policy as Health Policy* (pp. 29–60). New York: Russell Sage Foundation.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5), 1001–1044.

Firpo, S., Fortin, N.M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3), 953-973.

Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). Is education always reducing fertility? Evidence from compulsory schooling reforms. *The Economic Journal*, 126 (595), 1823-1855.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in Economics. In D. Card & and O. Ashenfelter (Eds.) *Handbook of Labor Economics*, Volume 4A (pp. 1–102). North Holland Elsevier.

Galama, T.J. & van Kippersluis, H. (2018). A theory of socio-economic disparities in health over the life cycle. *The Economic Journal*, 129, 338-374.

Gathmann, C., Jürges, H., & Reinhold, S. (2015). Compulsory schooling reforms, education and mortality in twenties century Europe. *Social Science & Medicine*, 127, 74-82.

Grossman, M. (1972). On the concept of health capital and demand for health. *Journal of Political Economy*, 80 (2), 223-255.

Groot, W., & van den Brink, H.M. (2007). The health effects of education. *Economics of Education Review*, 26, 186-200.

Hummer, R. A., & Lariscy, J. T. (2011). Educational attainment and adult mortality. In R. G. Rogers & E. M. Crimmins (Eds.) *International handbook of adult mortality* (pp. 241–261). Springer.

Jürges, H. (2007). True health vs. response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16, 163–178.

Jürges, H. (2010). Health inequalities by education, income and wealth: a comparison of 11 European countries and the US. *Applied Economics Letters*, 17, 87–91.

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

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

Kunst, A. E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., Mielck, A., Cardano, M., Dalstra, J. A. A., Geurts, J. J. M., Helmert, U., Lennartsson, C., Ramm, J., Spadea, T., Stronegger, W. J., & Mackenback, J. P. (2005). Trends in socio-economic inequalities in self-assessed health in 10 European countries. *International Journal of Epidemiology*, 34(2), 295–305.

Leopold, L. (2018). Education and physical health trajectories in later life: A comparative study. *Demography*, 55, 901-927.

Lundborg, P., Nordin, M., & Rooth, D.O. (2018). The intergenerational transmission of human capital: the role of skills and health. *Journal of Population Economics*, 31, 1035-1065.

Lynch, S.M. (2003). Cohort and life-course patterns in the relationship between education and health: A hierarchical approach. *Demography*, 40 (2), 309-331.

Malter. F., & A. Börsch-Supan (Eds.) (2015). *SHARE Wave 5: Innovations & Methodology*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Malter. F., & A. Börsch-Supan (Eds.) (2017). *SHARE Wave 6: Panel innovations and collecting Dried Blood Spots*. Munich: MEA, Max Planck Institute for Social Law and Social Policy.

Mazumder, B. (2008). Does education improve health? A reexamination of the evidence from compulsory schooling laws. *Economic Perspectives*, 32 (2), 2-16.

Mirowsky, J. & Ross, C.E. (2008). Education and self-rated health: Cumulative advantage and its rising importance. *Research on Aging*, 30, 93-122.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14, 693-709.

O'Donnell, O. (2009). Measuring health inequalities in Europe. *Eurohealth*, 15 (3), 10–14.

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

Pi Alperin, M. N. (2016). A multidimensional approach to measure health. *Economics Bulletin*, *36*(3), 1553–1568.

Prince, M., Reischies, F., Beekman, A., Fuhrer, R., Jonker, C., Kivela, S. et al. (1999). Development of the EURO–D scale – a European Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*, 174 (4), 330-338.

Reiss, P.C. (2016). Just how sensitive are instrumental variable estimates? *Foundations and trends in accounting*, 10 (2-4), 204-237.

Ross, C. E. & Wu, C.-L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior*, 37 (1), 104-120.

Salonen, L. & Pöyliö, H. (2017). *Historical dataset of major educational reforms in Europe in 1950-1990*. Working Papers on Social and Economic Issues 15/2017. Turku Center for Welfare Research.

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

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

Staiger, D. & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 557-586.

Van Doorslaer, E., Wagstaff, A., van der Burg, H., Christiansen, T., De Graeve, D. et al. (2000). Equity in the delivery of health care in Europe and the US. *Journal of Health Economics*, 19, 553-583.

Appendix A: Construction of the synthetic health indicators

We measure health status of each individual using a multidimensional approach proposed by Pi Alperin (2016) based on fuzzy set theory. This approach allows aggregating various health limitations at the individual level in one synthetic health indicator, h_i, as follows:

$$h_i = \sum_{j=1}^{M} h_{ij} \cdot w_j \,/ \sum_{j=1}^{M} w_j \,, \tag{A.1}$$

where h_{ij} is defined as the health status of individual i (i=1,…,N) in the j-th health item (j = 1,…, M) and w_j stands for the weight attached to this health item. By the way it is constructed, the synthetic health indicator, h_i, can take any value between 0 (absolutely healthy) and 1 (absolutely sick) being, thus, quasi-continuous in nature.

In order to determine the relative contribution of each health item j to the synthetic indicator, three different weighting schemes are tested in this paper. The first, and the simplest one, is the "equal weight" scheme:

$$w_j^{ew} = 1/M. \tag{A.2}$$

With this weighting scheme, in Equation (A.1), h_i is the average value of h_{ij}.

The second weighting scheme is proposed by Cerioli and Zani (1990) who consider the weight as being an inverse function of the deprivation level of the item in the entire population:

$$w_j^{cz} = 1/\log(\bar{h}_j), \tag{A.3}$$

where $\bar{h}_j$ is the sample mean of the j-th health item.

The third weighting scheme, used in the main part of the paper (Sub-sections 4.1-4.3), is the one proposed by Betti and Verma (1998). This weighting scheme accounts for the relative frequency of items (as Cerioli and Zani, 1990) and the correlation of items in order to limit the

influence of highly correlated health limitations on the synthetic index. In particular, the authors propose to derive weight w_j^{bv} as a product of two components:

$$w_j^{bv} = w_j^a . w_j^b, \tag{A.4}$$

with w_j^a accounting for the prevalence of a given health item in the population and w_j^b accounting for the correlation of items. In particular:

$$w_j^a = \left(\sum_{i=1}^{N} \left(h_{ij} - \bar{h}_j\right)^2\right) / \left(\bar{h}_j . N\right)^{1/2} \tag{A.5}$$

and

$$w_j^b = \left(1 + \sum_{j=1}^{M} p_{j,j'} I\left(p_{j,j'} < p_H\right)\right)^{-1} . \left(\sum_{j=1}^{M} p_{j,j'} I\left(p_{j,j'} \geq p_H\right)\right)^{-1} \tag{A.6}$$

where $p_{j,j'}$ is the correlation coefficient between items j and j' and I(.) is the indicator function taking value 1 if the expression in the brackets holds and 0 otherwise. p_H is a predetermined correlation cut-off, which separates low correlated items from highly correlated ones. Betti and Verma (1998) suggest setting this threshold at the point that reflects the largest gap between the ordered set of correlation values, the suggestion that we also follow in the paper.[15]

Additionally, in order to build the synthetic health indicator, the health items can be aggregated as a unique set of items or can be grouped in a smaller number of K dimensions. For example, K = 2 represents the mental and the physical dimensions of health.

In general, item weights are normalized so that they sum up to one. However, when health items are grouped in dimensions, the item weights are normalized within each dimension, so that each dimension has equal weight, and this independently of the number of items each dimension contains. Thus, when using the Betti and Verma (1998) weighting scheme and grouping health items in dimensions, only the correlations between health items included in each dimension are considered.

[15] This weighs structure does not satisfy the group decomposition property. Therefore, although being item specific, the values of the weights are the same for each individual in the sample.

Appendix B: Description of health items used for the construction of health indicators

Table B.1. Depression

Depression scale Euro-d*	Assigned value	Frequency (%)
Not depressed (0 dimension)	0	21.83
Between 1 and 11 dimensions	$1 - (12 - X_i)/12$	78.15
Completely depressed (12 dimensions)	1	0.02

* Depression, pessimism, suicidal thoughts, guilty, sleep, interest, irritability, appetite, tiredness, concentration, enjoyment, tearfulness.

Table B.2. Memory

Memory		Assigned value	Frequency (%)
Four questions have been asked regarding date, day of the week, month, and year	Knows all	0	87.92
	Knows 3 of 4	0.25	9.34
	Knows 2 of 4	0.50	1.67
	Knows 1 of 4	0.75	0.59
	None of them	1	0.47

Table B.3. Cognition

Capacity to memorize words		Assigned value	Frequency (%)
How many words do you recall?*	More than 15 words	0	4.70
	More than 1 and less than 16	$(16 - X_i)/14$	91.32
	Only 1	1	3.98

* This number is the addition between the first trial and the delayed trial

Table B.4. Long-term illnesses

Long-term illnesses		Assigned value	Frequency (%)
Do you have any long-term health problems, illness, disability or infirmity?*	No	0	36.12
	One	0.75	30.34
	More than one	1	33.54

***** Doctor ever told you had: high blood pressure or hypertension; diabetes or high blood sugar; high blood cholesterol, a stroke or cerebral vascular disease; chronic lung disease such as chronic bronchitis or emphysema; cancer or malignant tumour; stomach or duodenal ulcer, peptic ulcer; Parkinson disease; cataracts; hip fracture; other fractures; Alzheimer's disease, dementia, organic brain syndrome, senility or many other serous memory impairment; other affective of emotional disorders, including anxiety, nervous or psychiatric problems; rheumatoid arthritis; osteoarthritis, or other rheumatism; chronic kidney disease; a heart attack including myocardial infraction or coronary thrombosis or any other heart problem including congestive heart failure.

Table B.5. Limitation activities 1

Health and daily activities		Assigned value	Frequency (%)
Because of a health problem, do you have difficulty doing any of the following daily activities?*	No	0	88.13
	Somewhat	$1 - (6 - X_i)/6$	10.92
	Yes	1	0.95

* Dressing, bathing or showering, eating, cutting up the food, walking across a room, getting in or out of bed

Table B.6. Limitation activities 2

Health and instrumental activities		Assigned value	Frequency (%)
Because of a health problem, do you have difficulty doing any of the following instrumental activities?*	No	0	90.41
	Somewhat	$1 - (5 - X_i)/5$	8.4
	Yes	0	1.19

* Telephone calls, taking medications, managing money, shopping for groceries, preparing a hot meal

Table B.7. Limitation activities 3

Health and general activities		Assigned value	Frequency (%)
Because of a health problem, do you have difficulty doing any of the following activities?*	No	0	69.52
	Somewhat	$1 - (4 - X_i)/4$	28.43
	Yes	1	2.05

* Walking 100 meters, walking across a room, climbing several flights of stairs, climbing one flight of stair

Table B.8. Eyesight

Eyesight distance and reading*	Assigned value	Frequency (%)
Both are E or VG	0	38.63
Any other combination	$1 - (6 - X_i)/4$	59.27
Both are P	1	2.10

*E: excellent; VG: very good; G: good; F: fair; P: poor

Table B.9. Hearing

Hearing		Assigned value	Frequency (%)
Is your hearing*	Excellent or Very good	0	39.90
	Good	0.20	39.31
	Fair	0.50	16.83
	Poor	1	3.95

*With or without a hearing aid

Appendix C: Sensitivity checks for various weighting schemes and health items aggregation

The construction of composite health indicators relies on the selection of the weighting scheme and the procedure used for aggregation of multiple items in synthetic health scores. In the main part of the paper, we use the weights proposed by Betti and Verma (1998), which we constructed within each country separately. We also do not distinguish between health items related to physical and mental dimensions of health while aggregating them in a global health indicator (for more details, see Appendix A).

In order to test the sensitivity of our findings to this choice, we re-estimated education gradient in global health using twelve alternative combinations of three components: different weighting schemes applied to each health item, different ways of defining the sample within which the weights are derived, and different ways of aggregating health items in synthetic global health scores. More precisely, we use three weighting schemes, i.e. the equal weights (EW), the weight scheme proposed by Cerioli and Zani in 1990 (CZ), and the weight scheme proposed by Betti and Verma in 1998 (BV). For health items aggregation, we either treated all health items as a unique set of items or grouped them in two dimensions related to physical and mental health before deriving scores for global health. Finally, we computed those combinations either within each country separately (country-specific weights) or for all countries pooled together (cross-country weights). The values of the weights from these twelve combinations applicable to various health items are presented in Tables C.1 and C.2.

Tables C.1 and C.2 show that the cross-country weights are quite similar to country-specific weights as soon as they are derived using the same weighting schemes. It holds for both ways of items aggregation – when health items are treated as a unique set of items (Table C.1) and when they are grouped in two different dimensions of health (physical and mental health) prior to their aggregation in a global health indicator (Table C.2). The weights, however, differ substantially depending on the weighting scheme used. These differences, nonetheless, do not affect the main conclusions of the paper. Independently of the weighting scheme chosen, we find that the health gap between the lower and higher educated not only persists but also increases along the distribution (see Tables C.3 - C.6).

Table C.1. Weights applicable to various health items (where health items are treated as a unique set of items during their aggregation in global health scores)

	Cross-country weights			Country-specific weights*		
	Weight BV	Weight CZ	Weight EW	Weight BV	Weight CZ	Weight EW
Long-term illness	0.0752	0.0333	0.1111	0.0719	0.0331	0.1111
limitation activities 1	0.1392	0.1772	0.1111	0.1376	0.1764	0.1111
Limitation activities 2	0.1498	0.1839	0.1111	0.1498	0.1839	0.1111
Limitation activities 3	0.0696	0.1165	0.1111	0.0773	0.1169	0.1111
Eyesight	0.0736	0.0729	0.1111	0.0762	0.0747	0.1111
Hearing	0.0968	0.0927	0.1111	0.0979	0.0929	0.1111
Depression	0.0693	0.0941	0.1111	0.0670	0.0931	0.1111
Orientation	0.2842	0.1855	0.1111	0.2819	0.1849	0.1111
Cognition	0.0424	0.0439	0.1111	0.0404	0.0442	0.1111
Total	1	1	1	1	1	1

* These columns show the average weight across all countries.

Table C.2. Weights applicable to various health items (where health items are grouped in two health dimensions prior to their aggregation in global health scores)

	Cross-country weights			Country-specific weights*		
	Weight BV	Weight CZ	Weight EW	Weight BV	Weight CZ	Weight EW
Long-term illness	0.0553	0.0246	0.0833	0.0541	0.0244	0.0833
limitation activities 1	0.1195	0.1309	0.0833	0.1183	0.1302	0.0833
Limitation activities 2	0.1352	0.1359	0.0833	0.1345	0.1357	0.0833
Limitation activities 3	0.0619	0.0861	0.0833	0.0627	0.0862	0.0833
Eyesight	0.0555	0.0539	0.0833	0.0566	0.0550	0.0833
Hearing	0.0727	0.0685	0.0833	0.0738	0.0686	0.0833
Sub-Total	0.50	0.50	00.50	0.50	0.50	0.50
Depression	0.1039	0.1454	0.1667	0.1029	0.1446	0.1667
Orientation	0.3384	0.2867	0.1667	0.3409	0.2868	0.1667
Cognition	0.0576	0.0679	0.1667	0.0562	0.0686	0.1667
Sub-Total	0.50	0.50	0.50	0.50	0.50	0.50
Total	1	1	1	1	1	1

* These columns show the average weight across all countries.

Table C.3. Education gradient in health based on different weighting schemes computed within each country separately

	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
	EW, country-specific weights					
Low	0.066***	0.119***	0.184***	0.254***	0.324***	0.193***
High	0.052***	0.094***	0.156***	0.218***	0.280***	0.164***
Difference	0.014***	0.024***	0.028***	0.036***	0.044***	0.030***
	CZ, country-specific weights					
Low	0.033***	0.060***	0.094***	0.138***	0.192***	0.108***
High	0.029***	0.051***	0.081***	0.118***	0.166***	0.095***
Difference	0.004	0.008***	0.012***	0.019***	0.023***	0.013***
	BV, country-specific weights					
Low	0.030***	0.067***	0.111***	0.160***	0.217***	0.122***
High	0.026***	0.056***	0.096***	0.139***	0.188***	0.106***
Difference	0.004	0.011**	0.015***	0.021***	0.030***	0.015***

Note: EW stands for equal weights; CZ stands for Cerioli and Zani weights; VB stands for Betti and Verma weights.

Table C.4. Education gradient in health based on different weighting schemes computed for all countries together

Educational category	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
	EW, cross-country weights					
Low	0.066***	0.119***	0.184***	0.254***	0.324***	0.193***
High	0.052***	0.095***	0.156***	0.218***	0.280***	0.163***
Difference	0.014***	0.024***	0.028***	0.036***	0.044***	0.030***
	CZ, cross-country weights					
Low	0.034***	0.063***	0.098***	0.141***	0.197***	0.112***
High	0.028***	0.051***	0.081***	0.118***	0.167***	0.095***
Difference	0.006**	0.012***	0.017***	0.023***	0.029***	0.017***
	BV, cross-country weights					
Low	0.033***	0.072***	0.115***	0.166***	0.217***	0.125***
High	0.027***	0.056***	0.099***	0.143***	0.089***	0.108***
Difference	0.006*	0.016***	0.016***	0.023***	0.027***	0.018***

Note: EW stands for equal weights; CZ stands for Cerioli and Zani weights; VB stands for Betti and Verma weights.

Table C.5. Education gradient in health based on different weighting schemes computed within each country separately (with the aggregation of health items in physical and mental health first)

Educational category	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
	EW by dimensions, country-specific weights					
Low	0.083***	0.132***	0.190***	0.257***	0.329***	0.201***
High	0.063***	0.107***	0.159***	0.220***	0.282***	0.169***
Difference	0.020***	0.025***	0.032***	0.037***	0.047***	0.032***
	CZ by dimensions, country-specific weights					
Low	0.040***	0.064***	0.097***	0.144***	0.201***	0.113***
High	0.036***	0.057***	0.086***	0.126***	0.174***	0.101***
Difference	0.004	0.007***	0.011***	0.018***	0.027***	0.012***
	BV by dimensions, country-specific weights					
Low	0.038***	0.065***	0.105***	0.150***	0.205***	0.117***
High	0.033***	0.055***	0.090***	0.129***	0.178***	0.101***
Difference	0.005*	0.010***	0.015***	0.020***	0.028***	0.016***

Note: EW stands for equal weights; CZ stands for Cerioli and Zani weights; VB stands for Betti and Verma weights.

Table C.6. Education gradient in health based on different weighting schemes computed for all countries together (with the aggregation of health items in physical and mental health first)

Educational category	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
	EW by dimensions, cross-country weights					
Low	0.083***	0.132***	0.191***	0.257***	0.329***	0.201***
High	0.063***	0.107***	0.159***	0.220***	0.282***	0.169***
Difference	0.020***	0.025***	0.032***	0.037***	0.047***	0.032***
	CZ by dimensions, cross-country weights					
Low	0.043***	0.068***	0.101***	0.149***	0.208***	0.117***
High	0.034***	0.056***	0.085***	0.125***	0.173***	0.100***
Difference	0.008***	0.012***	0.016***	0.024***	0.035***	0.017***
	BV by dimensions, cross-country weights					
Low	0.041***	0.069***	0.108***	0.153***	0.206***	0.120***
High	0.032***	0.052***	0.092***	0.133***	0.177***	0.103***
Difference	0.009***	0.014***	0.016***	0.021***	0.029***	0.017***

Note: EW stands for equal weights; CZ stands for Cerioli and Zani weights; VB stands for Betti and Verma weights.

Appendix D: Sensitivity check with an alternative specification of educational sub-groups

In the main part of the paper, we define educational sub-groups using the 75th percentile of the distribution of the instrumented number of years spent in education as a cut-off point. In order to check the sensitivity of our results to this definition, we repeat the analysis using the 50th percentile of the distribution as a cut-off for the definition of the educational sub-groups.

The results of this exercise summarized in Table D.1 and Figures D.1-D.2 lie relatively close to the baseline estimates presented in Sections 4. In particular, we find that the education gradient in health is not only present at the mean but also becomes evident from the 50th percentile of the distribution onwards, doubling in size towards its top. The gradient, thus, is absent at the bottom of the distribution and appears somewhat smaller when we use the 50th percentile cut-off instead of the 75th percentile, which can be explained by a larger similarity between the two educational sub-groups when a lower cut-off point is used for their definition.

The results of the decomposition also look quite similar independently of the cut-off used. Differences in employment patterns, exercising, and parental education keep playing an important role in explaining the education gradient in health in favor of the higher educated. The contributions associated with the differences in the returns to characteristics remain similar for all covariates except of smoking and exercising. The difference in the returns to exercising becomes more pronounced with the 50th percentile cut-off significantly contributing to the decrease of the gradient in the upper quantile of the distribution. In turn, smoking has a more detrimental influence on health in the lower tail of the health distribution for the lower educated than for the higher educated. A potential explanation might lie in the intensity of smoking, which we unfortunately cannot control for with our data.

Table D.1. Differences in the levels of health between the lower and higher educated at different points of the distribution of health

Educational category	Percentiles of the distribution of health					
	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile	Mean
			Health values			
Lower educated	0.028*** (0.003)	0.064*** (0.002)	0.112*** (0.002)	0.159*** (0.003)	0.219*** (0.005)	0.121*** (0.002)
Higher educated	0.030*** (0.002)	0.064*** (0.002)	0.104*** (0.002)	0.152*** (0.002)	0.203*** (0.003)	0.114*** (0.002)
	Absolute difference in health values between the lower and higher educated					
Absolute difference	-0.002 (0.003)	0.000 (0.003)	0.008** (0.003)	0.008* (0.004)	0.016* (0.007)	0.007* (0.003)
	Relative difference in health values between the lower and higher educated, in %					
Relative difference	-6.1 (11.5)	0.0 (4.9)	8.2** (3.1)	5.2* (2.6)	8.1* (3.4)	6.0* (2.5)

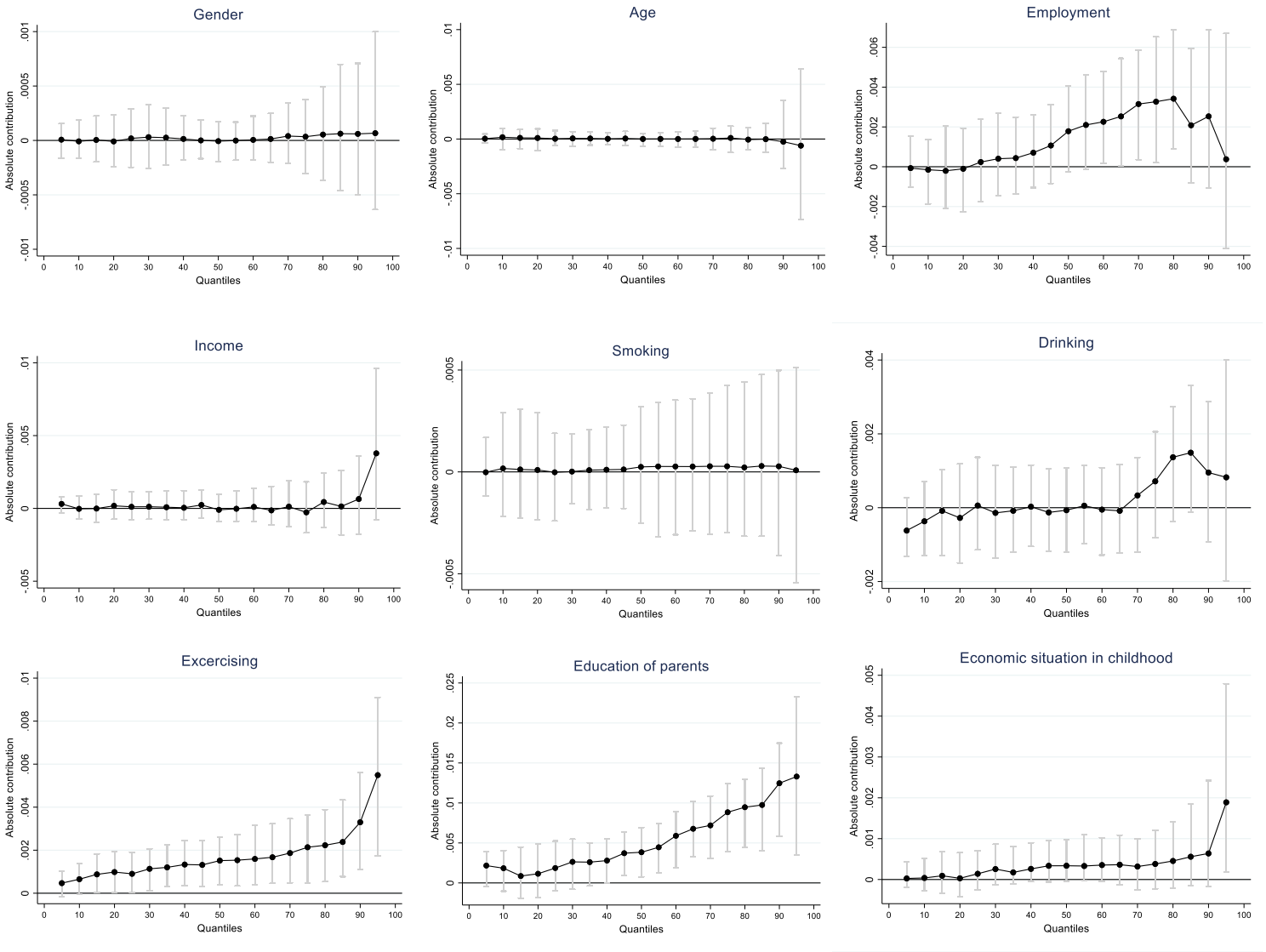

Figure D.1. Contributions of the differences in individual characteristics to the education-related gradient in health

Note: All estimates are weighted using individual sample weights. Vertical lines plot 95% confidence intervals derived from 500 bootstrapped replications.

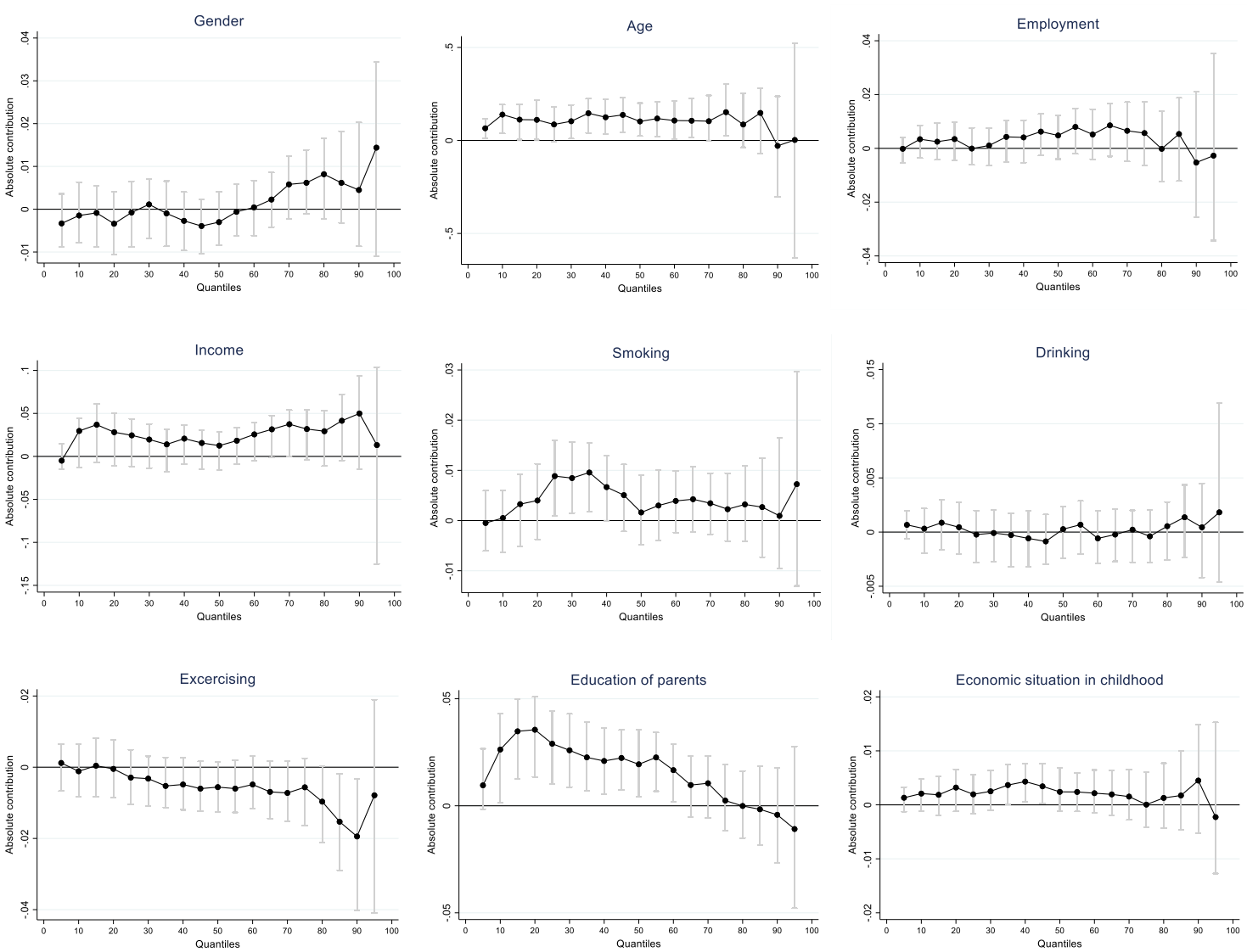

Figure D.2. Contributions of the differences in returns to individual characteristics to the education-related gradient in health

Note: All estimates are weighted using individual sample weights. Vertical lines plot 95% confidence intervals derived from 500 bootstrapped replications.