The Causal Effect of Education on Health: What is the Role of Health Behaviors?

The Causal Effect of Education on Health: What is the Role of Health Behaviors?*

Giorgio Brunello (University of Padua, CESifo, IZA and ROA)

Margherita Fort (University of Bologna, CESifo, IZA)

Nicole Schneeweis (University of Linz and IZA)

Rudolf Winter-Ebmer (University of Linz, CEPR, IZA and IHS)

July 21, 2014

Abstract

We investigate the causal effect of education on health and the part of it which is attributable to health behaviors by distinguishing between short-run and long-run mediating effects: while in the former only behaviors in the immediate past are taken into account, in the latter we consider the entire history of behaviors. We use two identification strategies: instrumental variables based on compulsory schooling reforms and a combined aggregation, differencing and selection on observables technique to address the endogeneity of both education and behaviors in the health production function. Using panel data for European countries we find that education has a protective effect for European males and females aged 50+. We find that the mediating effects of health behaviors - measured by smoking, drinking, exercising and the body mass index - account in the short run for around a quarter and in the long run for around a third of the entire effect of education on health.

Keywords: SHARE, health, education, health behaviors

JEL Codes: I1, I12, I21

*We thank David Card, Lance Lochner and the participants at seminars in Bologna, Bressanone, Catanzaro, Chicago, Firenze, Hangzhou, Helsinki, Linz, Munich, Nurnberg, Padova, Regensburg, Rotterdam and Wurzburg for comments and suggestions. We acknowledge the financial support of Fondazione Cariparo, MIUR- FIRB 2008 project RBFR089QQC-003-J31J10000060001 and the Austrian Science Funds ("The Austrian Center for Labor Economics and the Analysis of the Welfare State"). The SHARE data collection has been primarily funded by the European Commission through the 5th, 6th and 7th framework programme as well as from the U.S. National Institute on Aging and other national Funds. Fort was member of CHILD during the early stages of this project. The usual disclaimer applies.

1 Introduction

The relationship between education and health - the "health-education gradient" - is widely studied. There is abundant evidence that a gradient exists (Cutler and Lleras-Muney, 2010). Yet less has been done to understand why education might be related to health. A potential channel is that education may improve decision making abilities, which may lead to better health decisions and to a more efficient use of health inputs (Lochner, 2011). In addition, education can reduce stress and generate healthier behaviors. Better educated individuals are also more likely to have healthier jobs, live in healthier neighbourhoods and interact with healthier peers and friends. Education may also lead to better health outcomes because it raises income levels.

In this paper, we estimate the causal impact of education on health using a multi-country set-up. We explore the contribution of health-related behaviors (shortly, behaviors) - which we measure with smoking, drinking, exercising and the body mass index - to the education gradient. To do so, we decompose the gradient into two parts: a) the part mediated by health behaviors, and b) a residual, which includes for instance stress reduction, better decision making, better information collection, healthier employment and better neighborhoods (Lochner, 2011)

We are not the first to investigate the mediating role of health behaviors. Our contribution is two-fold: first, we distinguish between short-run and long-run mediating effects. Typically, the empirical literature considers only the former and focuses either on current behaviors or on behaviors in the immediate past, thereby ignoring the contribution of the history of behaviors. By ignoring this history, short-run mediating effects are likely to underestimate the overall mediating effect of behaviors whenever there is some persistence in health status. Our empirical approach combines the estimates of a static health equation - where health depends only on education - and a dynamic health equation, that relates current health both to education and to the entire history of health behaviors through past health.

Second, as recently pointed out by Lochner (2011), a problem with the existing empirical literature is that most contributions fail to address the endogeneity of education and behaviors in health regressions and therefore ignore that there are possibly many confounding factors which influence both education and behaviors on the one hand, and health outcomes on the other hand. While some studies have dealt with endogenous education, our approach is novel because we address the endogeneity of both education and behaviors in the health production function, and therefore can give a causal interpretation to our estimates.

In this paper, we combine two identification strategies. We first estimate a static health equation using an instrumental variables (IV) approach, and exploiting the

exogenous variation provided by the changes in compulsory schooling laws which occurred in several European countries between the 1940s and the 1960s. While this strategy allows us to estimate the total effect of education on health, it does not help us in estimating the mediating effects of behaviors because we do not have credible instruments for health behaviors. We therefore propose an alternative identification strategy, which combines aggregation, differencing and selection on observables (ADS), to estimate the parameters of both a static health equation - as in the IV approach - and a dynamic health equation. By combining the estimates of these two equations, we are able to evaluate the mediating effects of health behaviors both in the short and in the long run.

We use a multi-country data-set, which includes 13 European countries (Austria, Belgium, Czech Republic, Denmark, England, France, Germany, Greece, Italy, the Netherlands, Spain, Sweden and Switzerland) and provides information on education, health and health behaviors for a sample of males and females aged 50+. By focusing on older individuals, we consider the long-term effects of education on health. The data are drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE) and from the English Longitudinal Study of Ageing (ELSA). Both surveys are constructed following the US Health and Retirement Study.

Focusing on self-reported (poor) health, we present two sets of estimates of the gradient: the IV estimates, which apply to individuals whose education is affected by mandatory schooling reforms (compliers), and the estimates based on the ADS strategy, which apply to the average individual in the sample. Both estimates show that education has a protective effect for males and females, although the effects for females are typically larger in magnitude.

Our IV results show that one additional year of schooling reduces self-reported poor health by 4 to 6.4 percentage points for females and by 4.8 to 5.4 percentage points for males. Compared to the recent empirical literature for Europe, which uses compulsory school reforms to estimate the gradient, these estimates are larger in magnitude than the 0.5 percentage points estimated by Clark and Royer (2013) and smaller than the 8.4 percentage points found by Powdthavee (2010) for the UK. When we apply the ADS strategy to the IV sample and restrict our sample to potential compliers by excluding those with college education, we obtain estimates of the gradient that are reasonably close to the IV estimates, especially for females.

We show that health behaviors - measured by smoking, drinking, exercising and the body mass index - contribute to explaining the gradient. The size of this contribution is larger when we consider the entire history of behaviors rather than only behaviors in the immediate past. In the former case, we find that the effects of education on smoking, drinking, exercising and eating a proper diet account for 23% to 45% of

the entire effect of education on health, depending on gender. In the latter case, the mediating effects are about 17% for females and 31% for males. The largest part of the gradient remains, however, unexplained. Potential candidates accounting for this part include both the direct effects of education on health operating through knowledge and skills and the indirect effects operating through differences in wealth and the socio-economic environment as well as other unobserved health behaviors.

The paper is organized as follows: Section 2 is a brief review of the relevant literature. The theoretical model is presented in section 3 and our empirical strategy is discussed in section 4. Section 5 describes the data. The results are discussed in section 6. Conclusions follow.

2 Review of the Literature

As recently reviewed by Lochner (2011), empirical research on the causal effect of education on health has produced mixed results. This literature typically focuses on single countries and identifies the effect of education on health with the exogenous variation generated by mandatory schooling laws. Most of these studies consider self-reported health as well as other outcomes. Some find that education improves health and reduces mortality, see for instance Adams (2002) and Mazumder (2008) for the US, Arendt (2008) for Denmark, Kemptner et al. (2011) for German males, Van Kippersluis et al. (2011) for the Netherlands and Silles (2009) and Powdthavee (2010) for the UK. Others find small or no effects. While Clark and Royer (2013) and Oreopoulos (2007) find very small effects for Britain, ambiguous or no effects are obtained by Albouy and Lequien (2009) for France, Arendt (2008) for Denmark, Braakmann (2011) and Juerges et al. (2013) for the UK (with some positive effects for females) and Kemptner et al. (2011) for German females. Overall, the existing literature is inconclusive.

There are many possible channels through which education may improve health. Lochner (2011) lists the following: stress reduction, better decision making and information gathering, higher likelihood of having health insurance, healthier employment, better neighborhoods and peers and healthier behaviors. Conti et al. (2010) argue that non-cognitive skills are an important factor as well.

Some authors have also investigated the causal impact of education on health-related behaviors, such as smoking, drinking, exercising, eating healthy food and the BMI. On the one hand, Clark and Royer (2013), Arendt (2005) and Braakmann (2011) find no evidence of a causal link between education and health behaviors. On the other hand, Kemptner et al. (2011) present evidence of significant protective effects of education on BMI but not on smoking. In addition, Brunello et al. (2013) use

the exogenous variation provided the compulsory schooling laws in nine European countries and find that education has a protective effect on the BMI of European females. Additional research investigating the relationship between education and the BMI includes Spasojevic (2003) for Sweden and Grabner (2008) for the US. Both studies find that education has a statistically significant causal (protective) effect on body weight.

While the adverse effects of smoking on health are well-known in the medical literature, the effects of alcohol consumption are more complex. A meta-analysis on the relationship between alcohol dosage and total mortality shows a J-shaped relationship, with lowest mortality found for low levels of alcohol intake as compared to abstinence or high levels of drinking (Di Castelnuovo et al., 2006). Physical inactivity is also strongly related to health, as inactivity was found to cause nine percent of premature mortality worldwide in 2008 (I-Min et al., 2012). Furthermore, overweight and obesity are at the root of many chronic diseases, such as diabetes, coronary heart disease, gallstones or hypertension (Field et al., 2001; Must et al., 1999).

The contribution of behaviors, such as smoking, drinking, eating calorie-intensive food and refraining from exercising, has been examined in the economic and sociological literature, starting with the contribution by Ross and Wu (1995).[1] These authors use US data, regress measures of health on income, social resources and behaviors and treat both, behaviors and education, as exogenous. They find that behaviors explain less than 10% of the education gradient. Cutler et al. (2008) discuss possible mechanisms underlying the education gradient. Using data from the National Health Interview Survey (NHIS) in the US, they find that behaviors account for over 40% of the effect of education on mortality in their sample of non-elderly Americans.

A problem with these studies is that they fail to consider the endogeneity of education and behaviors in a health equation including both. In the study most closely related to our paper, Contoyannis and Jones (2004) partly address this concern by explicitly modeling the optimal choice of health behaviors. They jointly estimate a health equation - with health depending on education and behaviors - and separate behavior equations - where behaviors depend on education - by Full Information Maximum Likelihood (FIML), treating education as exogenous. Using Canadian data, they show that the contribution of lagged (7 years earlier) behaviors to the education gradient varies between 23% to 73%, depending on whether behaviors are treated as exogenous or endogenous.[2]

We summarize the existing evidence as follows: first, the available empirical evidence on the causal effect of education on health is mixed and covers a rather lim-

[1] See the reviews by Feinstein et al. (2006) and Cawley and Ruhm (2011).

[2] Tubeuf et al. (2012) find that health behaviors account for 25% of health inequalities.

ited set of countries (Denmark, France, Germany, the Netherlands, the UK and the US); second, the estimated contribution of behaviors to the education gradient varies substantially across the few available studies, depending on model specification and identification strategy.[3]

We contribute to this literature by providing a framework to distinguish between the short-run and long-run mediating effects of health behaviors, and a method to estimate these effects on a sample of twelve European countries. While the short-run only includes the effects of behaviors in the immediate past, the long-run takes the contribution of the entire history of behaviors into account. This distinction is empirically relevant, as we show in Section 6.

We are also the first to combine a conventional – and widely accepted – IV-strategy[4] with a more flexible identification approach based on aggregation, gender differencing and selection on observables (ADS). Using this new approach, we address the endogeneity of education and health behaviors in the health production function.

3 Health Behaviors and the Education Gradient

In the empirical literature (Cutler et al., 2008; Ross and Wu, 1995) the contribution of health behaviors to the education gradient (HEG) is evaluated by adding the vector either of current behaviors (B) - which include smoking, the use of alcohol or drugs, unprotected sex, excessive calorie intake and poor exercise - or of behaviors in the immediate past (first lag) to a regression of (poor) health status (H) on education (E) and other covariates. The lag is often justified with the view that the impact of health behaviors on health requires time. Consider the following empirical model

$$H_{it} = c_t + \alpha_{t-1} B_{i,t-1} + \beta_t E_i + \nu_{it} \tag{1}$$

where i is the individual, t is time, c is the intercept and v is the error term. We assume stationarity in the parameters ($c_t = c$; $\alpha_{t-1} = \alpha$; $\beta_t = \beta$) and the following linear approximation of the relationship between behaviors B and education E[5]

$$B_{it} = \sigma_0 + \sigma_1 E_i + \eta_{it} \tag{2}$$

[3]See also Stowasser et al. (2011) for a discussion of causality issues in the relationship between socio-economic status and health.

[4]We estimate the causal effect of education on health using a multi-country data set including several European countries. This multi-country set-up allows us to exploit both the within-country and between-cohorts variation and the between-countries variation in mandatory years of schooling.

[5]See the Appendix for an illustrative model of optimal education and health behaviors.

Substituting (2) into (1) yields the following static health equation

$$H_{it} = (c + \alpha\sigma_0) + (\alpha\sigma_1 + \beta)E_i + \alpha\eta_{it} + \nu_{it} \tag{3}$$

In this simple model, the education gradient HEG is given by $(\alpha\sigma_1 + \beta)$ and the mediating effect of behaviors in the immediate past to the gradient is $\frac{\alpha\sigma_1}{(\alpha\sigma_1+\beta)}$.

By focusing on behaviors in the immediate past, specification (1) assumes that, conditional on B_{it-1}, earlier behaviors do not contribute to current health. To illustrate the implications of this assumption, let the "true" health production function be given by

$$H_{it} = k_0 + k_1 B_{it-1} + k_2 B_{it-2} + ... + k_T B_{it-T} + \theta E_i + \varepsilon_{it} \tag{4}$$

where we assume again stationarity in the coefficients. This function is more general than (1) because current health depends both on behaviors lagged once and on all previous lags from $(t-2)$ to the initial period T.

Combining equation (2) and (4) yields

$$H_{it} = [k_0 + \sigma_0(k_2 + ... + k_T)] + k_1 B_{it-1} + [\sigma_1(k_2 + ... + k_T) + \theta] E_i + \upsilon_{it} \tag{5}$$

where $\upsilon_{it} = \varepsilon_{it} + \sum_{s=2}^{T} k_s \eta_{it-s}$.

When the health production function depends on the entire sequence of risky health behaviors, from period 1 to T, the contribution of behaviors in the immediate past to the education gradient is $\frac{\sigma_1 k_1}{[\sigma_1(k_1+k_2+...+k_T)+\theta]}$, where the denominator includes both the effect of education on health conditional on behaviors θ and the mediating effects of behaviors. This contribution differs from the contribution of the entire sequence of health behaviors from lag 1 to T, which is given instead by $\frac{\sigma_1(k_1+k_2+...+k_T)}{[\sigma_1(k_1+k_2+...+k_T)+\theta]}$. If the parameters k_i are positive, ignoring the contribution of higher lags leads to an underestimation of the overall mediating effect of risky health behaviors.

When the available data do not include information on behaviors from lag $t-2$ to lag T, as it happens in our case, an alternative approach is to adopt a dynamic health equation (see for instance Park and Kang (2008))

$$H_{it} = d + \pi B_{it-1} + \nu E_i + \phi H_{it-1} + e_{it} \tag{6}$$

which requires data for the periods t and $t-1$. Under the additional assumptions that $H_{t-T} = 0$, $\phi < 1$ and $T \to \infty$, equation (6) is equivalent to equation (4) when the following restrictions on the parameters hold

$$k_1 = \pi;\ k_2 = \pi\phi;\ k_s = \pi\phi^{s-1}, \forall s = 3, \ldots, T;\ \theta = \frac{\nu}{1-\phi};\ k_0 = \frac{d}{1-\phi}; \varepsilon_{it} = \frac{e_{it}}{1-\phi}$$

We repeatedly substitute lagged health in (6) to obtain health as a function of education and the lags of behaviors from $t-1$ to $t-T$. We then substitute $B_{it-2}...B_{it-T}$ using (2) to obtain

$$H_{it} = \frac{d + \phi\pi\sigma_0}{1-\phi} + \pi B_{it-1} + \left[\frac{\nu + \phi\sigma_1\pi}{1-\phi}\right] E_i + \overline{e}_{it} \tag{7}$$

for $T \to \infty$, where $\overline{e}_{it} = \sum_{k=0}^{T-1} \phi^k \varepsilon_{it-k} + \pi \sum_{k=1}^{T-1} \phi^k \eta_{it-k-1}$. Furthermore, placing $B_{it-1} = \sigma_0 + \sigma_1 E_i + \eta_{it-1}$ into (7) yields the static health equation

$$H_{it} = \chi_o + \chi_1 E_i + \widetilde{e}_{it} \tag{8}$$

where $\chi_o = \frac{\pi\sigma_0 + d}{1-\phi}$, $\widetilde{e}_{it} = \sum_{k=0}^{T-1} \phi^k (\varepsilon_{it-k} + \eta_{it-k-1})$ and $\chi_1 = \frac{\pi\sigma_1 + \nu}{1-\phi}$ is the health-education gradient HEG.

The relative contribution of health behaviors in the immediate past B_{it-1} to the education gradient (short-run mediating effect, $SRME$) is

$$SRME = \frac{(1-\phi)\pi\sigma_1}{(\pi\sigma_1 + \nu)} \tag{9}$$

The overall relative contribution of health behaviors (or long-run mediating effect, $LRME$) to the education gradient adds to the contribution of health behaviors in the immediate past the contribution of previous behaviors, from $t-2$ to $t-T$, and is equal to

$$LRME = \frac{\pi\sigma_1}{(\pi\sigma_1 + \nu)} \tag{10}$$

This implies that $SRME = (1-\phi)LRME$. Under these assumptions, for any $\phi > 0$, $SRME$ under-estimates $LRME$, and the degree of under-estimation is larger the higher is ϕ (persistence of health status over time). Therefore, if we only estimate $SRME$, we may find a small contribution of health behaviors to the overall education gradient not because health behaviors have a small mediating effect but because we

have ignored the contributions of health behaviors from period $t-2$ to $t-T$.[6] An important channel through which education influences health is income. Incorporating income into the dynamic health equation $H_{it} = d + \pi B_{it-1} + qY_{it} + \nu E_i + \phi H_{it-1} + e_{it}$ and assuming that $Y_{it} = mE_i$, the long-run mediating effect is $\pi\sigma_1/(\pi\sigma_1 + \widetilde{\nu})$ where $\widetilde{\nu} = v + qm$.

4 Empirical Strategy

The estimates of the static health equation (8) and the dynamic health equation (6) can be used to compute $\widehat{\pi\sigma_1} = \widehat{\chi}_1(1-\widehat{\phi}) - \widehat{\nu}$ and obtain estimates of the short and long-run mediating effects

$$\widehat{LRME} = \frac{\widehat{\chi}_1(1-\widehat{\phi}) - \widehat{\nu}}{\widehat{\chi}_1(1-\widehat{\phi})} \tag{11}$$

$$\widehat{SRME} = (1-\widehat{\phi})\widehat{LRME} \tag{12}$$

This strategy has the advantage that it only requires the estimation of two equations and the drawback that we cannot separately identify the mediating effect of each single health behavior.[7] Adding income to equation (6) implies that $LRME$ and $SRME$ are equal to

$$\widehat{\widetilde{LRME}} = \frac{\widehat{\chi}_1(1-\widehat{\phi}) - \widehat{\nu} - \widehat{q}\widehat{m}}{\widehat{\chi}_1(1-\widehat{\phi})} \tag{13}$$

$$\widehat{\widetilde{SRME}} = (1-\widehat{\phi})\widehat{\widetilde{LRME}} \tag{14}$$

4.1 Endogeneity of education and health behaviors

Education, health behaviors in the immediate past and lagged health (the history of behaviors) are not exogenous in the dynamic health equation and very likely correlated with unobservable individual characteristics affecting health. Consider the error terms (e) in the dynamic health equation (6) and (η) in the behavior equation (2). Since optimal education depends on the unobservables that affect preferences (η) and health production (e) – see the illustrative model in the Appendix – OLS fails to uncover causal relationships. A similar problem affects the OLS estimates of the static health

[6] If the overall education gradient HEG is negative, sufficient conditions for the indicator $LRME$ ($SRME$) to fall within the range $[0,1]$ are $\pi\sigma_1 \leq 0$ and $\nu \leq 0$. If HEG is positive, these conditions change signs.

[7] For this purpose, we would need to estimate equation (2) for each single health behavior. We leave this development for future research.

equation (8), because health depends both on education and on the sequence of shocks affecting preferences and health production.

An important drawback of the empirical studies investigating the mediating effect of health behaviors on the education gradient is that they fail to simultaneously consider the endogeneity of education and behaviors (Lochner, 2011). In this paper, we address endogeneity in order to give a causal interpretation to the gradient and to the mediating role of behaviors. For this purpose, we use two identification approaches, which are illustrated in turn below.

4.2 The IV approach

We estimate the static health equation (8) by instrumental variables, using the number of years of compulsory education YC as instrument for individual years of schooling E. This strategy is widely considered as credible and has been used extensively in the literature. As in Brunello et al. (2009), Brunello et al. (2013) and Fort et al. (2011), we apply this strategy to a multi-country setup and exploit the fact that compulsory school reforms have occurred at different points in time during the 1940s-1960s in several European countries, affecting adjacent cohorts differently.[8]

For each country and reform included in our sample, we construct pre-treatment and post-treatment samples. We identify for each country the pivotal birth cohort, i.e. the first cohort potentially affected by the change in mandatory years of schooling, for each country. We include in the pre- and post-treatment samples all individuals born either before, at the same time or after the pivotal cohort. By construction, the number of years of compulsory education "jumps" with the pivotal cohort and remains at the new level in the post-treatment sample. The timing and intensity of these jumps varies across countries, and we use both the within and between country exogenous variation in the instrument to identify the causal effects of schooling on health.

In our estimations, we control for country fixed effects, cohort fixed effects and country-specific linear or quadratic trends in birth cohorts. These trends account for country-specific improvements in health that are independent of educational attainment.[9] Country fixed effects control for national differences, including differences in institutions affecting health or in reporting styles. Notice that the older cohorts in our data are healthier than average, having survived until relatively old age. Since the comparison of positively selected pre-treatment individuals with younger post-

[8] Brunello et al. (2013) address the cross-country heterogeneity of the first stage and IV effects in a similar sample of European countries and show that the estimates obtained by using all available countries and the sub-sample of countries that can be pooled according to standard statistical tests are qualitatively similar. We therefore disregard the issue of heterogeneity in this paper.

[9] "Failure to account for secular improvements in health may incorrectly attribute those changes to school reforms, biasing estimates toward finding health benefits of schooling." (Lochner (2011), p.41)

treatment samples is likely to result in a downward bias in the estimates, we control for this selection process by including cohort fixed effects.

In principle, the same IV approach could also be applied to the estimation of the dynamic health production function (6), provided that we can find additional credible sources of exogenous variation for health behaviors. This is a very difficult task with the data at hand. For instance, using instruments such as the price of alcohol or cigarettes does not work in our setup because these variables – being only time-dependent – influence all cohorts in one country alike. In the absence of credible instruments, we follow an approach introduced by Card and Rothstein (2007) and turn to a different identification strategy that combines aggregation, fixed effects and selection on observables to estimate both the static and the dynamic health production function.[10]

4.3 Aggregation, Differencing and Selection on Observables

We aggregate our data into cells defined by gender, cohort and country.[11] By doing so, we average out individual unobserved idiosyncracies. We difference data by gender to eliminate all those unobservables which are shared by males and females in each cell (country by cohort) and capture residual gender-specific unobservables with observable controls, including a rich set of parental and early life conditions.

Consider the following empirical version of the dynamic health production function

$$H_{icgb} = \alpha_{g0} + \alpha_{g1} B_{icgb}^{t-1} + \alpha_{g2} E_{icgb} + \alpha_{g3} X_{icgb} + \alpha_{g4} H_{icgb}^{t-1} + \epsilon_{icgb} \tag{15}$$

where i denotes the individual, c the country, g gender (m: males; f: females), b the birth cohort and X is a vector of control variables. Importantly, we allow each explanatory variable, including education, to have a gender-specific effect on health. Thus, we do not impose the unrealistic restriction that health production is equal for males and females.

The error term in equation (15) can be decomposed as follows

$$\epsilon_{icgb} = \mu_{cgb} + \nu_{icgb} \tag{16}$$

[10] Card and Rothstein (2007) investigate ethnic segregation in US schools and its impact on the black-white test score gap.

[11] Since the dynamic health equation relates current health to behaviors and health in the previous period, we use two waves of data and aggregate also by time period. To avoid confusion, we suppress the time dimension.

where μ_{cgb} represent a common error component for individuals of the same country c, gender g and birth cohort b and ν_{icgb} is an individual-specific error component for which we assume

$$E[\nu_{icgb}|c,g,b] = 0 \tag{17}$$

We aggregate individual data into cells identified by country, gender and birth cohort and obtain the aggregated health equation (18), where $\overline{H}_{cgb}$ denotes $E[H|c,g,b]$ and the same applies for the other regressors

$$\overline{H}_{cgb} = \alpha_{g0} + \alpha_{g1}\overline{B}_{cgb}^{t-1} + \alpha_{g2}\overline{E}_{cgb} + \alpha_{g3}\overline{X}_{cgb} + \alpha_{g4}\overline{H}_{cgb}^{t-1} + \mu_{cgb} \tag{18}$$

Furthermore, we take gender differences for each cell (Δ =females - males) and define $\alpha_s = \alpha_{Fs} - \alpha_{Ms}$, with $s = 0, .., 4$. We obtain

$$\Delta\overline{H}_{cb} = \alpha_0 + \alpha_{m1}\Delta\overline{B}_{cb}^{t-1} + \alpha_1\overline{B}_{cb}^{t-1,f} + \alpha_{m2}\Delta\overline{E}_{cb} + \alpha_2\overline{E}_{cb}^{f} + \alpha_{m3}\Delta\overline{X}_{cb} + \alpha_3\overline{X}_{cb}^{f} +$$

$$+\alpha_{m4}\Delta\overline{H}_{cb}^{t-1} + \alpha_4\overline{H}_{cb}^{t-1,f} + \Delta\mu_{cb} \tag{19}$$

where the superscript f refers to females. In this specification, α_{m1} and $\alpha_1+\alpha_{m1}$ are the effects of health behaviors lagged once on health for males and females, respectively. Similarly, the gender gap in the "returns" to education is given by coefficient α_2.

Differencing by gender eliminates all unobserved factors that are common to males and females for a given country c and birth cohort b, including genetic and environmental effects, income components, medical inputs and the organization of health care. Even after eliminating common unobservables, however, one may argue that the residual error component $\Delta\mu_{cb}$ could still be correlated with education and lagged health behaviors. This could happen, for instance, if health conditions and parental background during childhood are excluded from vector X in (15) and differ systematically by gender or if unaccounted labor market discrimination by gender correlates with income, education, behaviors and health.

We add additional structure to our empirical specification by modeling the residual $\Delta\mu_{cb}$ as

$$\Delta\mu_{cb} = \psi_b + \psi_c + \psi_{m1}\Delta\overline{Z}_{cb} + \psi_1\overline{Z}_{cb}^{f} + \psi_{m2}\Delta\overline{Y}_{cb} + \psi_2\overline{Y}_{cb}^{f} + \kappa_{cb} \tag{20}$$

where $\psi_s = \psi_{fs} - \psi_{ms}$, with $s = 1, 2$, ψ_b is a vector of cohort effects and country-specific linear or quadratic trends in birth cohorts, ψ_c a vector of country effects, Z a vector of observable characteristics, which includes a rich set of parental background

characteristics and health conditions during childhood[12] and Y is real income. By including income, we control for the monetary effects of labor market discrimination. By adding trends in cohorts, cohort and country fixed effects in the gender difference equation, we allow for the possibility that these effects vary by gender.

Consider for instance trends in childbearing. These trends may have gender-specific effects on health outcomes (eg. breast cancer). Since childbearing trends are likely to be correlated with education and health behaviors, omitting them from (19) may generate biased estimates. By including cohort dummies as well as country specific trends in birth cohorts in (20), we remove this threat. In addition, suppose that the key unobservable in (18) is latent time invariant average ability. The ADS method assumes that part of this latent factor is common across genders and can be differenced out. The residual gender-specific component is captured by cohort and country dummies as well as by gender differences in parental background during childhood and initial health conditions.

Our identifying assumption is that, conditional on these variables - which capture gender-specific childhood and environmental effects - the error term κ_{cb} is orthogonal to health behaviors and educational attainment. For the sake of brevity, we call this method ADS (aggregation cum differencing cum selection on observables). With respect to the standard fixed effect model we assume that the conditional distribution of the individual fixed effect given $(E_i; B_{it}; H_{it-1}; X)$ is common between genders rather than over time for a given individual. Other than this, the conditional distribution is left unrestricted and the inference is conditional on this effect. Notice that we cannot apply the standard fixed effect approach here because education is time-invariant. Conditional on our identifying assumptions, equation (19) is estimated by weighted least squares, using as weight $\left(\frac{1}{N_M} + \frac{1}{N_F}\right)^{-1}$, where N_M and N_F are the number of males and females in each cell (see Card and Rothstein, 2007).

In our data, both lagged health behaviors and lagged health, which captures all previous health behaviors, are observed two years prior to the measurement of current health. Since our sample consists of individuals aged 50+, these behaviours are measured way after the end of education. Yet, there might be a concern that the omission of behaviors early in life - and before school is completed - affects our ADS estimates. While we do not have measures of early behaviors, we indirectly control for them by including in the ADS regressions a rich set of early life conditions, which

[12]There is a growing literature on the impact of childhood health on adult economic outcomes (Banks et al. (2011), Smith (2009) and Brunello et al. (2012)). The vector Z includes: childhood poor health, hospitalization during childhood, presence of serious diseases, had at most 10 books at home at age 10, mother and father in the house at age 10, mother or father died during childhood, number of rooms in the house at age 10, had hot water in the house at age 10, parents drunk or had mental problems at 10, had serious diseases at age 15, born in the country.

affect these behaviors. By taking gender differences, we also eliminate all common unobserved factors for a given country and cohort of birth, including those relating to early behaviors.

5 Data

In principle, we would like to estimate the impact of the history of past health behaviors (drinking, smoking, etc.) on current health, as in eq. (4). This would require, however, fairly long longitudinal data with information on these behaviors, that are typically not available in most European countries. A more practical alternative is to estimate a dynamic health equation - (eq. (6) in the paper) - which relates current health to education, behaviors in the immediate past and lagged health, which captures all previous health behaviors. By estimating (6) and by adding a few restrictions to the set of parameters, we can recover the health production function (eq. (4)) by repeatedly substituting lagged health in eq. (6). The advantage of this approach is that we only need to estimate equations (6) and (8) to identify the short-run and long-run mediating effects of health behaviours. By using information on current health, lagged health and behaviors in the immediate past, these equations have much less stringent data requirements than eq. (4). We also need information on education, parental background and early socio-economic and health conditions.

The Survey of Health, Ageing and Retirement in Europe (SHARE), the English Longitudinal Study of Ageing (ELSA) and their retrospective interviews satisfy these data requirements. SHARE is a longitudinal dataset on health, socio-economic status and social relations of European individuals aged 50+, and consists of two waves - 2004/5 and 2006/7 - plus a retrospective wave in 2008/9 (SHARELIFE), covering several European countries - Austria, Belgium, the Czech Republic, Denmark, France, Germany, Greece, Italy, The Netherlands, Spain, Sweden and Switzerland.[13] ELSA has similar characteristics and covers England. For England, we use waves 2 (2004/5) and 3 (2006/7). Since education is typically accumulated in one's teens or twenties, by focusing on individuals aged 50+ we are considering the long-run effects of education on health. Moreover, we are using some family-background information which is available before major schooling decisions have been taken in order to control for the parental influence on schooling. Early life conditions are available from the SHARELIFE module which asks individuals a number of questions concerning their childhood at (approximately) age 10.

[13]The Czech Republic, Poland, Israel and Ireland joined in the second wave.

The measure of health used in this paper is self-reported poor health ($SRPH$), which is based on a question whether the individual considers her health as poor, good, very good or excellent. To attenuate the risks of over- or under-reporting, we recode this variable as a dummy equal to 1 if the individual considers her health as fair or poor and to 0 if she considers it as good, very good or excellent. This is a subjective and comprehensive measure of health, which is conventionally used in the applied literature (Lochner, 2011). One may object that self-reported information is likely to be dominated by noise and may fail to capture differences in more objective measures of health.[14] This is not the case here: among the individuals in the sample who reported poor health, 46% were diagnosed with hypertension, 69% with cardiovascular diseases and 79% suffered some long-term illness. On average, they had 2.44 chronic diseases certified by doctors. In contrast, the percentage of individuals in good health with similar diseases was 28, 44 and 33%, respectively. Moreover, the latter group experienced only 1.10 chronic diseases.[15]

While our data contain information on chronic diseases, which can be argued to be more objective than self-reported health, we have chosen to focus on the latter in order to be able to compare our results with the bulk of estimates in the relevant literature. Moreover, self-perceived health has the advantage of being the most comprehensive measure of health.

Previous studies have shown that self-perceived health and future mortality are strongly correlated (Bopp et al., 2012; Heiss, 2011). We present estimates based on the number of chronic diseases in the robustness section of this paper.

We measure educational attainment with years of education. The second wave of SHARE provides information on the number of years spent in full time education. In the first wave, however, participants were only asked about their educational qualifications. Thus, for the individuals participating only in the first wave, we calculate their years of schooling using country-specific conversion tables. In ELSA, years of education are computed as the difference between the age when full-time education was completed and the age when education was started.

We implement the IV approach by focusing on the seven countries where the individuals in our sample experienced at least one compulsory school reform: Austria,

[14]For an early discussion about the importance of measurement error in self-reported health see Bound (1991) and Butler et al. (1987) as well as Baker et al. (2004). These authors were primarily concerned with the impact of measurement error in equations determining the impact of health on retirement and other labor market outcomes. Justification bias, i.e. non-working persons over-reporting specific conditions, is an obvious problem there.

[15]Peracchi and Rossetti (2012) use anchoring vignettes with SHARE and find that gender differences in self-reported health are somewhat reduced. As these vignettes are asked only in some countries and not in the general SHARE survey, we refrain from extending our analysis to these vignette comparisons.

the Czech Republic, Denmark, England, France, Italy and the Netherlands.[16] In each country, we use all individuals who participated in the first or second wave of SHARE (second or third wave in ELSA).[17] To ensure that individuals spent their schooling in their host country, we restrict our sample to those who were born in the country or migrated there before age 5. Table 1 shows the selected countries, the years and the content of the reforms as well as the pivotal cohorts, i.e. the first cohorts potentially affected by the reforms. A short description of the compulsory school reforms used in this paper can be found in section 9.3.

For each country, we construct a sample of treated and control individuals. Since the key identifying assumption that changes in average education within counties can be fully attributed to the reforms is more plausible when the window around the pivotal cohort is relatively small, we estimate our model using individuals who were born up to 10 years before and after the reforms. This IV-sample consist of 15,960 individuals. Table 2 shows summary statistics of key variables by country.

To implement the ADS strategy, we use a sample of twelve countries with at least two data waves (the Czech Republic is excluded because this country participated only in the second wave of SHARE), aggregate individual data by cohort and country and difference the resulting cell data by gender. This strategy requires that there is gender variation in the variables of interest. Figure 1 plots gender differences in poor health and education and documents that such variation exists. The figure also shows that these differences are negatively correlated: the slope coefficient of the weighted regression is -0.027, with a standard error of 0.006.

We have four measures of risky health behaviors: whether the individual is currently smoking, whether he or she drinks alcohol almost every day, whether he or she engages in vigorous physical activity, such as sports, heavy housework or a job that involves physical labor and the body mass index.[18] Whether BMI should be considered as health outcome or as an health behavior is controversial. In our paper, we would like to use calorie intake as health behavior, but this is not available. In its place, we use BMI, which, conditional on the health behaviors we can measure, captures the effects of poor diet and low intake of fruit and vegetables, two key behaviors affecting health (Cawley and Ruhm, 2011).

[16] We exclude Germany and Sweden because school reforms in these countries were implemented at the regional level and our information on the region where the individuals completed their education is not accurate.

[17] When available, we measure the key variables (health, education) using the information provided by the respondents during their second interview. When this is not possible, the first interview is used.

[18] Smoking, drinking alcohol, exercising and diet are among the seven listed factors that affect individual health by the World Health Organization - the remaining three being low fruit and vegetable intake, illicit drugs and unsafe sex.

Table 3 shows country by gender averages of self-reported health, years of education, age and annual income (in thousand Euro at 2005 prices, PPP) in 2006/07, and averages of smoking, drinking, exercising and the BMI in 2004/05 for the ADS-sample. We notice the presence of important cross-country and cross-gender variation, both in health and in health behaviors. As expected, both income and years of education are higher among males aged 50+ than among females of the same age group. The percentage of females reporting poor health is higher than that of males (32 versus 27 percent). Females are less likely to smoke and drink than males. They have a slightly lower body mass index (26.7 versus 27.1) and tend to exercise vigorously less often than males.[19] Figure 2 plots gender differences in health behaviors by birth cohort. We detect a positive trend in the relative drinking behavior of females, and a negative trend in the percent overweight (BMI$\geq$ 25).

As discussed above, we use the ADS approach to estimate the dynamic health equation (6) and the ADS and the IV approach for the static health equation (8). The estimation of the dynamic health equation requires information on the current and the previous period. The two waves of SHARE and ELSA used in this paper include individuals who appear in both waves and individuals who are interviewed only in a single wave. We compute cell averages at time t and $t-1$ by using all individuals rather than only the longitudinal subsample. Each cell is defined by gender, country, wave and semester of birth. We use semesters rather than years to increase the number of available cells in the estimation.[20]

6 Results

This section describes the results of our empirical analysis. In section 6.1, we present the IV estimates of the education gradient for the static health equation and compare them with those obtained with the ADS strategy. In section 6.2, we show the ADS estimates of the dynamic health equation and decompose the total effect of education on health into the mediating effect of health behaviors and the residual effect. We also distinguish between short and long-run mediating effects. Section 6.3 concludes the presentation of results with several robustness checks.

[19]Table A1 in the Appendix reports the country by gender averages of the parental background variables included in the vector Z. The table shows that the gender variation in parental background and childhood characteristics is small. We interpret this as evidence that parental background characteristics are substantially removed by gender differencing.

[20]Since we do not have information on the month of birth for England, we aggregate by year of birth for this country.

6.1 The Health-Education Gradient

We estimate the education gradient in the static health equation by instrumental variables, using as instrument for endogenous education the number of years of compulsory education, which varies across countries and cohorts because of compulsory schooling reforms. We control for country fixed effects, cohort fixed effects as well as for some individual characteristics (whether the individual is foreign-born, whether there was a proxy respondent for the interview and indicators for the interview year). We capture smooth trends in education and health by adding country-specific polynomials in cohorts. The sample for the IV approach consists of at most 10 birth cohorts before and after the pivotal cohort in each country.

Table 4 presents our estimates by gender with two alternative specifications of the country-specific trends (linear or quadratic). In each case, we also report OLS, ITT (Intention-To-Treat, i.e. the effect of compulsory schooling on health), first stage (the effect of the instrument on the endogenous variable) and IV-Probit estimates. The numbers in the table are coefficients/marginal effects, and the estimated standard errors are clustered by country and cohort.[21]

The OLS estimates of the gradient are -2.4 percentage points for females and -1.7 percentage points for males. The estimated magnitude of the gradient increases when we instrument individual years of education with compulsory schooling. We find that one additional year of schooling decreases the probability of poor health by 4 to 6.4 percentage points for females and by 4.8 to 5.4 percentage points for males. The IV-Probit estimates are very similar to the linear IV estimates and more precise. Compared to the recent empirical literature for Europe, which uses the exogenous variation generated by compulsory school reforms to estimate the gradient, our findings are larger in absolute value than the 0.5 percentage points estimated by Clark and Royer (2013) and smaller than the 7 to 8.4 percentage points found by Powdthavee (2010) (see Lochner (2011), Table 6).[22]

Our first stage regressions show that the instrument is relevant and not weak – the F-Statistics are between 16.62 and 41.93 – and that one additional year of compulsory schooling increases actual schooling by a quarter to a third of a year, broadly in line

[21] Clustering by country, with or without using the wild bootstrap procedure suggested by Cameron et al. (2008), yield standard errors similar to those reported in the paper. Pischke and von Wachter (2008) also cluster by state and cohort, as we do in this paper.

[22] The analysis by Clark and Royer (2013) for England and Wales differs from this study in many ways. One explanation for the smaller estimated effects in their study might be that they consider individuals aged 45-69 years old. Our individuals are significantly older (age 72 on average). The causal effect of education on health might be stronger later in life, especially when mortality is the outcome variable.

with previous findings in the literature using similar reforms in European countries.[23] Figure 3 shows the first stage graphically, by allocating cohorts before and after the pivotal cohorts associated to each school reform (cohorts 0). While there is a general upward trend in years of schooling over time, the increase in compulsory schooling experienced by pivotal and younger cohorts definitely shifts education upwards. We interpret the IV estimates as local average treatment effects (LATE), i.e. the effects of schooling on health for the individuals affected by the reforms. These individuals typically belong to the lower portion of the education distribution.

We also estimate a static health equation with the ADS strategy. For each regression, we pool male and female cells and include the full set of interactions of each explanatory variable with a gender dummy. We start with a general specification which allows for the possibility that cohort, country and early life effects vary by gender. Preliminary testing, however, suggests that we cannot reject a more parsimonious specification which omits these effects.[24] We therefore report only the results using the latter specification hereafter. Table 5 shows the ADS estimates of the static health equation (columns (2),(3) and (4)) and compares the results to the IV estimates (column (1)).

While the ADS estimates pertain to a randomly drawn individual from the entire sample, the IV estimates measure the causal effects of education on health for the individuals affected by the compulsory schooling reforms. To compare ADS with IV, we report ADS estimates based on different samples: the full sample of twelve countries (column (2)), the sub-sample of the seven countries for which we have IV estimates (column (3)) and the sub-sample which excludes individuals with college education (column (4)). We believe that the comparability of IV and ADS estimates is highest in the last column because college graduates are typically not affected by compulsory schooling reforms. When we consider the largest sample, the ADS estimates show that one additional year of schooling reduces the prevalence of poor health by 2.6 percentage points for women and by 1 percentage point for men. When we reduce the sample to the same countries and cohorts used for our IV regressions, the magnitudes of the ADS estimates increase in absolute value. Finally, when we exclude highly educated individuals, the estimated marginal effects become closer to the 2SLS estimates shown in Table 4, especially for women.

[23] Our first stage estimates are broadly similar to those reported in previous studies based on European data (Brunello et al., 2013, 2009; Fort et al., 2011).

[24] The joint hypothesis that cohort, country, trends and early life effects do not vary by gender is not rejected at the 5 percent level of confidence (p-value: 0.094). We tested separately also the null that the following effects are common between genders: cohort effects (p-value: 0.894), country effects (p-value: 0.420), early life conditions (p-value: 0.263), trends in cohorts (p-value: 0.112). We never reject the null at conventional significance levels.

6.2 The Mediating Effects of Health Behaviors

In this section we present the results obtained by applying the ADS procedure to estimate the dynamic health equation in the sample of 12 European countries and evaluate the mediating effects of health behaviors. Table 6 presents the estimates of the static (column 1) and the dynamic health equation (column 2). Although we estimate gender differenced equations, we report separate estimates for males and females. This is possible because we allow the coefficients of our covariates, with the exception of early life conditions, to vary by gender. As already mentioned, our preliminary specification tests suggest that cohort, country and early life effects do not differ significantly by gender. Therefore, in our empirical specification, we omit cohort and country dummies and include only the gender differences in early life conditions.

The estimates of the static health equation show that the gradient is negative and larger in absolute value for females than for males. As already shown in Table 5, we estimate that an additional year of schooling reduces poor health by 2.6 percentage points for females and by 1 percentage point for males. Parental and early life variables are jointly statistically significant (p-value: 0.009), mainly because of the gender differences in poor health at age 10.[25] The estimates also suggest that few books in the house when age 10 and poor health during childhood increase self-reported poor health at age 50+.

Turning to the dynamic health equation, we find that our measures of health behaviors attract statistically significant coefficients, with predictable correlations: smoking, refraining from vigorous activity and poor diet leading to higher BMI are positively correlated to self-perceived poor health. Somewhat unexpectedly, however, drinking alcohol almost every day is negatively correlated to self-reported poor health, both for males and females. While the precision of the effects of behaviors is not high, we cannot reject the null hypothesis that these effects are jointly statistically significant. We also find that annual real income is negatively associated to perceived poor health, and the lagged dependent variable has a coefficient close to 0.3 (statistically distinct from 1), indicating the presence of some persistence in self-reported health over time and that the short-run mediating effect of health behaviors is close to 70% of the long-run effect. Finally, adding health behaviors, income and lagged health to the static health equation reduces the coefficient of education from -0.026 to -0.015 for females, and from -0.010 to -0.003 for males.

[25] Since we have many early life variables, we use principal component analysis to summarize some of the available information with the following three variables: poor housing at age 10, parental absence at age 10 and parents drunk/had mental problems at age 10. See the Appendix for further details.

In Table 7, we show our calculations of the short and long-run mediating effects. We use our estimates of the health education gradient ($\widehat{\chi}_1$), which equals -0.026 for females and -0.010 for males, our estimates of health persistence ($\widehat{\phi}$) and the direct effects of education $\widehat{\nu}$ and income $\widehat{m}$ on health in the dynamic health equation to calculate $\widehat{LRME}$ and $\widehat{SRME}$. In doing so, we assume that the income return to education is 0.07.[26] Our calculations based on equation (13) and equation (12) give a short term mediating effect of health behaviors equal to 17.2% for females and to 30.8% for males. In the long run, when we include the effect of earlier health behaviors, the estimated mediating effect increases to 22.8% for females and to 44.5% for males. This suggests that using only the first lag of behaviors - as is often done in the empirical literature - is likely to underestimate the contribution of health behaviors to the education gradient.

In the case of males, our estimated long-run effects are similar to those found by Cutler et al. (2008), who use a different approach and conclude that measured health behaviors account for over 40% of the education gradient (on mortality) in a sample of non-elderly Americans. In the case of females, we find that health behaviors contribute less to the gradient in the long run. While the effect of education on behaviors accounts for an important share of the gradient, especially for males, much remains to be explained, either by the role played by unmeasured behaviors or by effects that do not involve behaviors, such as better decision making, stress reduction and more health-conscious peers.

6.3 Robustness Checks

In this section, we focus on the ADS approach and show several robustness checks. We start by collapsing data by gender, country and year rather than semester of birth. By doing so, we reduce the sample size by almost a half. As shown in the first two columns of Table 8, the effect of education on health is virtually unaffected for females but declines for males. Next, we omit England to take into account that English data are drawn from a different (although quite similar) survey and can only be collapsed by year of birth. The next two columns of Table 8 show that the education gradient changes only marginally.[27]

Furthermore, we notice that the older cohorts in our data - age in our sample ranges from 50 to 86 - are strongly selected by mortality patterns. To control for this, we add to the regressions the level and the gender difference of life expectancy at birth, which varies by country, gender and birth cohort. Since these data are not available

[26] See for instance the estimates in Brunello et al. (2009).

[27] We have also estimated our equations on two sub-samples of countries, based on their proximity to the Mediterranean Sea, but cannot reject the hypothesis that the estimated coefficients are not statistically different.

for Greece[28], we are forced to omit this country from the sample. As displayed by the last two columns in the table, life expectancy is never statistically significant in the static health equation, and only marginally significant (at the 10% level of confidence) in the dynamic health equation. We conclude that adding this variable does little to our empirical estimates.

We also run our estimates for the sub-sample of individuals aged 50 to 69 and find that one additional year of schooling reduces self-reported poor health by 22.4% for females and by 11.5% for males. These percentages are significantly higher than those estimated for the full sample (-8.1% for females and -3.7% for males). Since survivors aged 70 to 86 in our sample might be better educated and might experience a stronger protective role of education on health than the average individual in the same age group - i.e. they might face a larger education gradient - it is unlikely that the decline of the gradient with age is driven by selection effects.

One may think of several factors affecting changes in the education gradient by age group. On the one hand, the gradient could decline among older individuals because cognitive abilities decline with age. On the other hand, the effect of behaviors on health accumulates over time, which should increase the gradient with age. At the same time, one may speculate that differences by education increase with age because the older care more about their health. While these factors go in different directions, our empirical results suggest that their balance is tilted in favor of the first.

Finally, we consider an alternative and more objective measure of health outcome, the number of chronic diseases.[29] While this number is reported by interviewed individuals, it is conditional on screening, i.e. each condition must have been detected by a doctor. Table 9 presents both, the ADS estimates of the static and the dynamic health equation, and the IV estimates of the static equation. Using the ADS method, we find evidence of a negative and statistically significant gradient for females (-0.057) and of a positive, small and imprecisely estimated gradient for males (0.012). The direc-

[28]We use data on life expectancy at birth from the Human Mortality & Human Life-Table Databases. The databases are provided by the Max Planck Institute for Demographic Research (www.demogr.mpg.de). The data are missing for some cohorts and for Greece. We use period measures of life expectancy at birth since cohort measures are not available for all the cohorts considered in the study.

[29]The respondents were asked whether a doctor has ever told them they had any of the following conditions: a heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure, high blood pressure or hypertension, high blood cholesterol, a stroke or cerebral vascular disease, diabetes or high blood sugar, chronic lung disease such as chronic bronchitis or emphysema, asthma, arthritis, including osteoarthritis or rheumatism, osteoporosis, cancer or malignant tumor, including leukaemia or lymphoma, but excluding minor skin cancers, stomach or duodenal ulcer, peptic ulcer, parkinson disease, cataracts, hip fracture or femoral fracture or other fractures, Alzheimer's disease, dementia, organic brain syndrome, senility or any other serious memory impairment, benign tumor (fibroma, polypus, angioma) or other unspecified conditions.

tions of these effects are confirmed but their magnitudes in absolute values are larger (-0.157 for females and 0.080 for males) when we apply the IV method. Defining $P(D)$ as the probability of reporting a condition, this probability is the product of the probability of undergoing screening $P(S)$ and the probability of having a disease conditional on screening, $P(D|S)$. We speculate that in the case of males the positive effect of education on the number of diseases may be driven by the fact that better educated males choose more intensive screening.

Turning to the decomposition of the gradient into the mediating effect of behaviors and the residual effect, we find that SRME and LRME for females are equal to 16.5 and 28.1 percent respectively, not far from the effects estimated for self-reported poor health. In the case of males, the estimated parameters do not meet the conditions for both SRME and LRME to be well defined within the range $[0, 1]$.

7 Conclusions

In this paper we estimate the causal effect of education on health in a sample of seven European countries, using the exogenous variation generated by compulsory school reforms. We also study the contribution of health behaviors to the education gradient by distinguishing between short-run and long-run mediating effects: while in the former only behaviors in the immediate past are taken into account, in the latter we consider the entire history of behaviors. In the absence of credible instruments for health behaviors, we propose a strategy to estimate and decompose the education gradient which takes into account both the endogeneity of educational attainment and the endogenous choice of health behaviors. We call this approach ADS because it combines aggregation (A), gender differencing (D) and selection on observables (S).

Our IV estimates show that one additional year of schooling reduces self-reported poor health by 4 to 6.4 percentage points for females and by 4.8 to 5.4 percentage points for males. Using a larger sample, our ADS estimates produce smaller effects but a larger gap between females (2.6 percent) and males (1.0 percent). One reason for the somewhat higher returns for females might originate from the fact that females in our sample are less educated than males, and that marginal returns might decline with education. Moreover, it might be that females take health-related information – coming with additional education – more seriously than males. While they may not change their health-related behaviours to a larger extent (see the decomposition results in Table 7), they may visit a doctor more often. Indeed, when we look at the number of chronic diseases which have been diagnosed by a doctor (Table 9), the gender difference is even stronger.

Compared to the recent empirical literature for Europe, which also uses the exogenous variation generated by compulsory school reforms to estimate the gradient, our estimates are larger in magnitude than the -0.5 percentage points estimated by Clark and Royer (2013) and smaller than the -7.0 to -8.4 percentage points found by Powdthavee (2010). We show that health behaviors - measured by smoking, drinking, exercising and the body mass index - contribute to the education gradient. Our estimates suggest that the long-run mediating effect of behaviors accounts for 23% to 45% of the entire effect of education on health, depending on gender. This contribution is reduced to 17% for females and to 31% for males, if we only consider behaviors in the immediate past, as usually done in the empirical literature.

Since the gradient is key to understanding inequalities in health and life expectancy and is also used to assess the overall returns to education (Lochner, 2011), it is important to understand the mechanisms governing it. Many of the discussed health behaviors are individual consumption decisions and changes thereof come at personal costs, e.g. abstaining from smoking or drinking good wine. Increases in health achieved by such costly changes in behavior have, thus, to be distinguished from changes resulting from the free benefits of education, such as lower stress or better decision making. This distinction is relevant for political decisions on school subsidies. If individuals are aware of the health-fostering effects of schooling and these are private, then there is no room for public policy. If individuals are unaware of these benefits, the case for public policy is stronger if the health benefits of schooling are primarily free rather than being based on the costly health behavior decisions of individuals (Lochner, 2011).

8 Figures and Tables

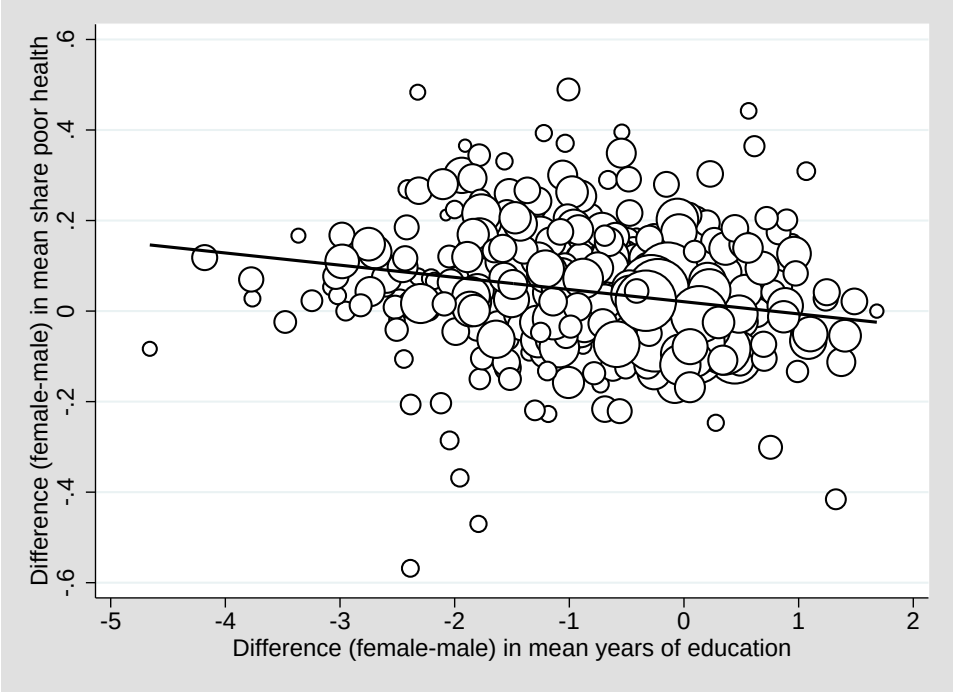

Figure 1: Gender differences in education and self-perceived poor health. Aggregated data by gender, cohort and country. Circle areas are proportional to weights based on the number of individuals used for aggregation $(N_M^{-1} + N_F^{-1})^{-1}$.

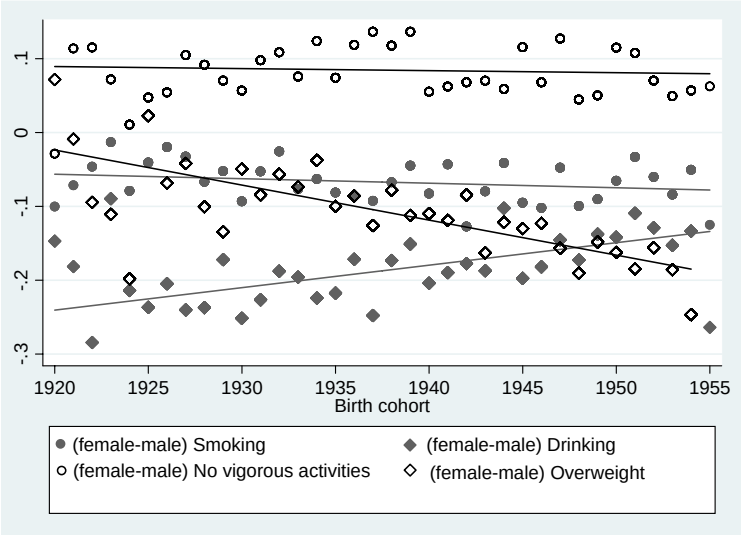

Figure 2: Gender differences by birth cohorts (differences in fractions of currently smoking, drinking alcohol almost every day, no vigorous activities and overweight).

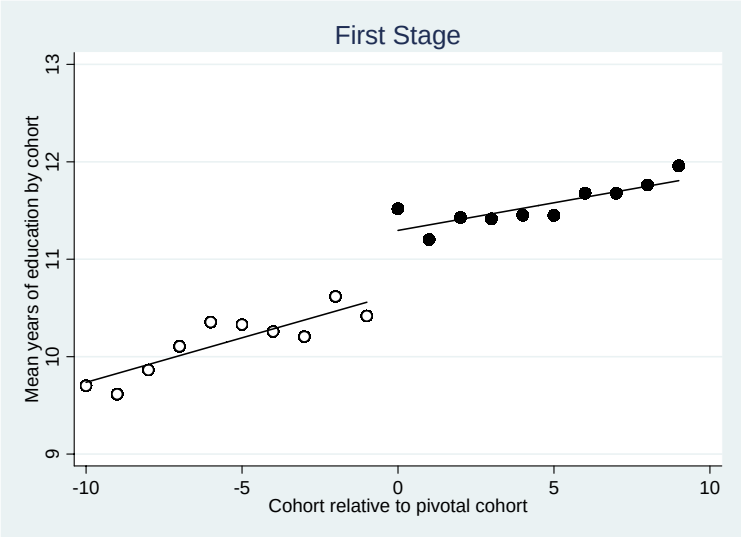

Figure 3: Mean years of education before and after various reforms. 0 on the x-axis is the first cohort affected by the increase in compulsory schooling in each country. In the Czech Republic and the Netherlands the first reform is shown in the graph. The picture does not qualitatively change if other reforms for these countries are included in the graph.

Table 1: Compulsory schooling reforms in Europe

Country	Reform	Changes in Years of Compulsory Education	Pivotal Cohort
Austria	1962/66	8 to 9	1951
Czech Republic	1948	8 to 9	1934
	1953	9 to 8	1939
	1960	8 to 9	1947
Denmark	1958	4 to 7	1947
England	1947	9 to 10	1933
France	1959/67	8 to 10	1953
Italy	1963	5 to 8	1949
Netherlands	1942	7 to 8	1929
	1947	8 to 7	1933
	1950	7 to 9	1936

Table 2: Descriptive Statistics for the IV-sample (Window: plus/minus 10 years around the pivotal cohort)

Country	Self-rep poor health	Education	Comp. Education	Age	Obs
Austria	0.233	11.363	8.237	58.971	782
Czech Republic	0.418	12.026	8.535	63.304	2,452
Denmark	0.208	11.802	5.642	59.194	1,898
England	0.373	10.713	9.585	72.355	4,672
France	0.331	11.324	8.275	63.668	2,223
Italy	0.337	8.822	6.032	59.631	2,093
Netherlands	0.338	10.613	8.263	69.95	1,840
All	0.339	10.901	8.088	65.588	15,960

NOTES: The sample consists of all individuals who participated in either the first wave of SHARE/second wave of ELSA in 2004/05 or in the second wave of SHARE/third wave of ELSA in 2006/07.

Table 3: Descriptive statistics for the ADS-sample by country and gender (M: males, F: females)

Country	Self-rep poor health		Education		Income		Age		Obs	
	M	F	M	F	M	F	M	F	M	F
Austria	0.27	0.31	11.04	9.47	18.74	10.74	65.14	66.18	260	364
Belgium	0.24	0.29	12.36	11.55	16.09	10.82	65.24	65.59	905	1044
Denmark	0.21	0.26	11.25	10.98	16.34	13.02	64.57	65.68	385	399
England	0.28	0.29	11.26	11.20	20.67	14.25	67.50	67.35	1673	2050
France	0.32	0.38	12.17	11.29	23.53	14.04	65.36	66.35	486	638
Germany	0.29	0.35	13.58	12.23	24.50	8.57	65.23	63.69	310	342
Greece	0.19	0.25	9.49	8.16	14.95	6.90	65.10	64.78	717	801
Italy	0.38	0.50	8.08	7.11	13.07	6.55	66.42	65.16	602	722
Netherlands	0.26	0.29	11.88	11.23	22.92	11.29	65.33	64.66	526	599
Spain	0.39	0.52	7.99	7.50	13.65	5.52	67.30	66.44	364	458
Sweden	0.22	0.26	11.42	11.61	16.81	13.00	65.94	65.38	512	615
Switzerland	0.12	0.18	12.25	10.68	29.89	14.10	66.01	64.85	197	232
All	0.27	0.32	11.02	10.37	18.66	11.17	66.03	65.86	6937	8264

Country	Smoking_{-1}		Drinking_{-1}		$\text{No vigorous exercise}_{-1}$		BMI_{-1}	
	M	F	M	F	M	F	M	F
Austria	0.21	0.05	0.17	0.17	0.64	0.73	27.46	26.94
Belgium	0.37	0.20	0.20	0.12	0.61	0.75	26.95	26.06
Denmark	0.37	0.20	0.31	0.28	0.48	0.52	26.49	25.57
England	0.22	0.14	0.13	0.12	0.75	0.81	27.81	28.15
France	0.52	0.24	0.19	0.09	0.59	0.73	26.57	25.74
Germany	0.26	0.11	0.21	0.14	0.44	0.43	26.83	26.04
Greece	0.18	0.03	0.36	0.20	0.60	0.67	27.11	26.73
Italy	0.60	0.29	0.25	0.14	0.65	0.74	27.11	26.56
Netherlands	0.38	0.28	0.24	0.24	0.52	0.54	26.26	26.17
Spain	0.45	0.11	0.29	0.10	0.63	0.74	27.62	27.98
Sweden	0.10	0.03	0.12	0.20	0.48	0.60	26.55	25.53
Switzerland	0.34	0.19	0.24	0.19	0.48	0.57	25.78	24.76
All	0.32	0.16	0.21	0.15	0.61	0.70	27.07	26.72

NOTES: The upper panel refers to the second wave of SHARE/third wave of ELSA in 2006/07 and the lower panel refers to the first wave in SHARE/second wave in ELSA in 2004/05. The Czech Republic is excluded because only one wave is available for this country. Descriptives statistics are based on individual level data.

Table 4: Health-Education Gradient - IV approach

	Females		**Males**	
	lin-trend	qu-trend	lin-trend	qu-trend
OLS	-0.024	-0.024	-0.017	-0.017
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
2SLS	-0.040	-0.064	-0.048	-0.054
	(0.024)*	(0.034)*	(0.029)*	(0.029)*
ITT	-0.014	-0.017	-0.016	-0.018
	(0.008)*	(0.008)**	(0.009)*	(0.008)**
First Stage	0.344	0.253	0.323	0.318
	(0.053)***	(0.058)***	(0.076)***	(0.078)***
IV-Probit	-0.042	-0.057	-0.047	-0.051
	(0.022)*	(0.025)**	(0.024)**	(0.022)**
F-Statistics (First Stage)	41.93	18.95	17.87	16.62
Observations	8,602	8,602	7,358	7,358

NOTES: Each coefficient/marginal effect represents a separate regression. Estimations are based on the IV-sample and include an indicator for foreign born individuals (who migrated before age 5), an indicator for interviews which have partly or fully been given by proxy respondents, interview-year dummies, country-fixed effects, cohort-fixed effects and country-specific trends in birth cohorts. The trends are linear and quadratic as indicated above. Standard errors are clustered at the country-cohort-level. ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 5: Health-Education Gradient - IV and ADS compared

	IV approach	ADS approach		
	IV-sample	ADS-sample	IV-sample	IV-sample w/o college educated
Females	-0.040	-0.026	-0.028	-0.042
	(0.024)*	(0.005)***	(0.007)***	(0.013)***
Males	-0.048	-0.010	-0.020	-0.020
	(0.029)*	(0.005)*	(0.008)**	(0.008)*

NOTES: Column (1) shows the baseline results of the IV approach (compare Table 4), column (2) gives the baseline estimates of the ADS approach using the ADS-sample (all 12 countries, compare Table 6), column (3) gives ADS-results for the sample of all countries and cohorts that are used in the IV approach and in column (4) the ADS approach is applied to the IV-sample but further excludes individuals who have college education. Standard errors are clustered at the country-cohort-level. ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 6: Baseline Results - ADS Model

	Static HE	Dynamic HE
Females		
education	-0.026	-0.015
	(0.005)***	(0.005)***
self-rep poor health$_{t-1}$		0.246
		(0.046)***
drinking$_{t-1}$		-0.013
		(0.053)
smoking$_{t-1}$		-0.034
		(0.056)
no vigorous exercise$_{t-1}$		0.040
		(0.042)
BMI$_{t-1}$		0.003
		(0.004)
income$_t$		-0.002
		(0.001)
Males		
education	-0.010	-0.003
	(0.005)*	(0.005)
self-rep poor health$_{t-1}$		0.308
		(0.046)***
drinking$_{t-1}$		-0.062
		(0.038)
smoking$_{t-1}$		0.043
		(0.042)
no vigorous exercise$_{t-1}$		0.089
		(0.041)**
BMI$_{t-1}$		0.011
		(0.005)**
income$_t$		-0.001
		(0.001)
Early life conditions		
few books in the household at 10	0.053	0.040
	(0.035)	(0.033)
serious diseases at 15	0.028	0.004
	(0.036)	(0.035)
poor health at 10	0.158	0.135
	(0.052)***	(0.049)***
hospital at 10	0.004	0.042
	(0.063)	(0.061)
Principal components		
parents drunk/had mental problems at 10	0.011	0.025
	(0.039)	(0.038)
parental absence at 10	-0.008	-0.009
	(0.039)	(0.037)
poor housing at 10	0.023	0.014
	(0.017)	(0.016)
Observations	736	734

NOTES: Each column represents a separate weighted OLS regression (coefficients on education, health-behaviors and income were allowed to differ for females and males) based on the ADS-sample. Data has been aggregated by country, birth cohort/semester and gender. Column (1) gives an estimate of the static health equation (8) and column (2) shows the dynamic health equation (6). Weights are inversely related to the number of observations used for the aggregation, $((1/N_M + (1/N_F))^{-1}$, where N_M and N_F are the number of males and females in each cell. ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 7: Decomposition of the Health-Education Gradient

	Females	**Males**
Health-Education Gradient (HEG)	-0.026	-0.010
- behaviors (short-term)	-0.004	-0.003
- behaviors (long-term)	-0.006	-0.004
- residual (direct effect)	-0.020	-0.006
Mediating effect as fraction of HEG		
- SRME (short-term)	0.172	0.308
- LRME (long-term)	0.228	0.445

NOTES: Calculations are based on the estimates reported in Table 6 using the static and the dynamic health equation (eq. (6) and (8)). The SRME and LRME are calculated using equations (14) and (13).

Table 8: Robustness - ADS approach

	ADS yearly pseudo-panel		ADS w/o ENG		ADS life-exp, w/o GRC	
	Static HE	Dynamic HE	Static HE	Dynamic HE	Static HE	Dynamic HE
Females						
education	-0.025	-0.011	-0.023	-0.016	-0.03	-0.018
	(0.006)***	(0.007)	(0.005)***	(0.006)***	(0.006)***	(0.006)***
sr poor health$_{t-1}$		0.307		0.240		0.252
		(0.063)***		(0.046)***		(0.052)***
drinking$_{t-1}$		0.017		-0.017		-0.031
		(0.069)		(0.052)		(0.056)
smoking$_{t-1}$		-0.080		-0.043		-0.031
		(0.076)		(0.056)		(0.063)
no vigorous		-0.016		0.021		0.036
exercise$_{t-1}$		(0.057)		(0.044)		(0.045)
BMI$_{t-1}$		0.001		0.000		0.002
		(0.005)		(0.005)		(0.004)
income$_t$		-0.001		-0.003		-0.003
		(0.002)		(0.002)*		(0.002)*
Males						
education	-0.006	0.004	-0.008	-0.004	-0.010	-0.004
	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)*	(0.006)
sr poor health$_{t-1}$		0.301		0.319		0.295
		(0.060)***		(0.046)***		(0.051)***
drinking$_{t-1}$		-0.011		0.078		-0.067
		(0.051)		(0.038)**		(0.042)
smoking$_{t-1}$		0.001		-0.038		0.038
		(0.056)		(0.042)		(0.049)
no vigorous		0.076		0.090		0.077
exercise$_{t-1}$		(0.054)		(0.043)**		(0.044)*
BMI$_{t-1}$		0.005		0.014		0.011
		(0.007)		(0.006)**		(0.006)**
income$_t$		-0.002		-0.001		-0.001
		(0.001)		(0.001)		(0.001)
Early life						
few books in HH	0.024	-0.006	0.050	0.051	0.085	0.076
	(0.048)	(0.047)	(0.035)	(0.034)	(0.038)**	(0.036)**
diseases at 15	0.110	0.070	0.021	0.007	0.021	-0.006
	(0.051)**	(0.050)	(0.037)	(0.035)	(0.038)	(0.037)
poor health at 10	0.185	0.170	0.137	0.109	0.164	0.146
	(0.073)**	(0.070)**	(0.053)***	(0.050)**	(0.053)***	(0.051)***
hospital at 10	-0.078	-0.028	0.060	0.097	-0.009	0.016
	(0.093)	(0.091)	(0.065)	(0.062)	(0.065)	(0.062)
Principal components						
parents drunk/had	-0.015	0.010	0.029	0.043	-0.009	0.011
mental problems at 10	(0.054)	(0.053)	(0.041)	(0.039)	(0.041)	(0.040)
parental absence at 10	0.047	0.029	-0.022	-0.016	0.009	0.005
	(0.056)	(0.054)	(0.040)	(0.038)	(0.041)	(0.039)
poor housing at 10	0.039	0.029	0.022	0.010	0.014	0.004
	(0.023)*	(0.022)	(0.017)	(0.016)	(0.018)	(0.018)
Life-expectancy						
females					0.007	0.009
					(0.005)	(0.005)*
males					0.005	0.007
					(0.003)	(0.004)*
Observations	389	387	701	701	640	638

NOTES: Each column represents a separate weighted OLS regression similar to those presented in Table 6. In the first two columns the aggregation is based on country, birth cohort and gender (not semester of birth), the second two columns show estimations without England and the third two columns give estimates when cohort-level life-expectancy is included in the regressions (Greece is excluded due to missing life-expectancy data). ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 9: Number of chronic diseases - ADS and IV approach

	ADS-approach		IV-approach (lin-trend)
	Static HE	Dynamic HE	Static HE
Females			
education	-0.057	-0.024	-0.157
	(0.015)***	(0.016)	(0.091)*
# chronic diseases$_{t-1}$		0.413	
		(0.044)***	
drinking$_{t-1}$		-0.044	
		(0.161)	
smoking$_{t-1}$		0.007	
		(0.178)	
no vigorous exercise$_{t-1}$		0.279	
		(0.131)***	
BMI$_{t-1}$		0.012	
		(0.305)	
income$_t$		-0.002	
		(0.004)	
Males			
education	0.012	-0.006	0.080
	(0.017)	(0.016)	(0.066)
# chronic diseases$_{t-1}$		0.337	
		(0.046)***	
drinking$_{t-1}$		-0.089	
		(0.116)	
smoking$_{t-1}$		0.045	
		(0.147)	
no vigorous exercise$_{t-1}$		0.220	
		(0.198)	
BMI$_{t-1}$		0.041	
		(0.016)*	
income$_t$		-0.004	
		(0.005)	
Early life conditions			
few books in HH	-0.135	-0.133	
	(0.110)	(0.102)	
serious diseases at 15	0.067	0.084	
	(0.114)	(0.106)	
poor health at 10	0.084	-0.004	
	(0.164)	(0.151)	
hospital at 10	0.081	0.112	
	(0.200)	(0.186)	
Principal components			
parents drunk or had	0.149	0.124	
mental problems at 10	(0.124)	(0.117)	
parental absence at 10	-0.128	-0.112	
	(0.123)	(0.114)	
poor housing at 10	0.069	0.037	
	(0.054)	(0.050)	
Observations	736	734	8,602 females, 7,358 males

NOTES: The first two columns show estimates of the static and the dynamic health equation using the ADS approach for the number of chronic diseases as health outcome (similar to the estimations reported in Table 6). The last column shows the IV-regressions for the number of chronic diseases as health outcome (similar to the estimations reported in Table 4). ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

References

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

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

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

Arendt, Jacob Nielsen (2008), 'In sickness and in health - till education do us part: Education effects on hospitalization', *Economics of Education Review* **27**(2), 161–172.

Baker, Michael, Mark Stabile and Catherine Deri (2004), 'What do self-reported, objective, measures of health measure?', *Journal of Human Resources* **39(4)**, 1067–1093.

Banks, James, Zoe Oldfield and James P. Smith (2011), Childhood health and differences in late-life helath outcomes between England and the United States, Working Paper 17096, National Bureau of Economic Research (NBER).

Bopp, Matthias, Julia Braun, Felix Gutzwiller and David Faeh (2012), 'Health risk or resource? gradual and independent association between self-rated health and mortality persists over 30 years', *Public Library of Science ONE* **7(2)**.

Bound, John (1991), 'Self-reported versus objective measures of health in retirement models', *Journal of Human Resources* **26(1)**, 106–138.

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

Brunello, Giorgio, Daniele Fabbri and Margherita Fort (2013), 'The causal effect of education on the body mass: Evidence from Europe', *Journal of Labor Economics* **31**(1), 195–223.

Brunello, Giorgio, Guglielmo Weber and Christoph Weiss (2012), Books are forever: early life opportunities, education and lifetimeearnings in europe, Discussion Paper 6386, Institute for the Study of Labor (IZA, Bonn).

Brunello, Giorgio, Margherita Fort and Guglielmo Weber (2009), 'Changes in compulsory schooling, education and the distribution of wages in Europe', *Economic Journal* **119**(March), 516–539.

Butler, J. S., Richard V. Burkhauser, Jean M. Mitchell and Theodore P. Pincus (1987), 'Measurement error in self-reported health variables', *The Review of Economics and Statistics* **69(4)**, 644–650.

Cameron, A. Colin, B. Jonah Gelbach and L. Douglas Miller (2008), 'Bootstrap-based improvements for inference with clustered errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Card, David and Jesse Rothstein (2007), 'Racial segregation and the black-white test score gap', *Journal of Public Economics* **91**, 2158–2184.

Cawley, John and Christopher Ruhm (2011), The economics of risky health behaviors, Working Paper 17081, National Bureau of Economic Research (NBER).

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

Conti, Gabriella, James J. Heckman and Sergio Urzua (2010), 'The education-health gradient', *American Economic Review: Papers and Proceedings* **100**, 234–238.

Contoyannis, Paul and Andrew Michael Jones (2004), 'Socio-economic status, health and lifestyle', *Journal of Health Economics* **23**, 965–995.

Cutler, David M. and Adriana Lleras-Muney (2006), Education and health: Evaluating theories and evidence, Working Paper 12352, National Bureau of Economic Research.

Cutler, David M. and Adriana Lleras-Muney (2010), 'Understanding differences in health behaviour by education', *Journal of Health Economics* **29**, 1–28.

Cutler, David M., Adriana Lleras-Muney and Tom Vogl (2008), Socioeconomic status and health: Dimensions and mechanisms, Working Paper 14333, National Bureau of Economic Research.

Cutler, David M., Edward L. Glaeser and Jesse M. Shapiro (2003), 'Why have Americans become more obese?', *Journal of Economic Perspectives* **17**(3), 93–118.

Di Castelnuovo, Augusto, Simona Costanzo, Vincenzo Bagnardi, Maria Benedetta Donati, Licia Iacoveillo and de Gaetano Giovanni (2006), 'Alchohol dosing and total mortality in men and women: An updated meta-analysis of 34 prespective studies', *Archives of Internal Medicine* **166**(22), 2437–2445.

Feinstein, Leon, Ricardo Sabates, Tashweka Anderson, Annik Sorhaindo and Cathie Hammond (2006), *Measuring the Effects of Education on Health and Civic Engagement: Proceedings of the Copenhagen Symposium*, Paris, chapter What are the effects of education on health?, pp. 171–354.

Field, Alison E., Eugenie H. Coakley, Aviva Must, Jennifer L. Spadaon, Nan Laird, William H. Dietz, Eric Rimm and Graham A. Colditz (2001), 'Impact of overweight on the risk of developing common chronic diseases during a 10-year period', *Archives of Internal Medicine* **161**(13), 1581–1586.

Fort, Margherita (2006), 'Education reforms across Europe: A toolbox for empirical research'. Paper version: May 11, 2006, mimeo.

Fort, Margherita, Nicole Schneeweis and Rudolf Winter-Ebmer (2011), More schooling, more children: Compulsory schooling reforms and fertility in Europe, Discussion Paper No. 8609, CEPR.

Garrouste, Christelle (2010), *100 years of educational reforms in Europe: A contextual database*, European Commission Joint Research Center, Luxembourg: Publications Office of the European Union.

Grabner, Michael (2008), 'The causal effect of education on obesity: Evidence from compulsory schooling laws'. Unpublished manuscript, Department of Economics, University of California, Davis.

Grossman, Michael (1972), 'On the concept of health capital and the demand for health', *Journal of Political Economy* **80**, 223–255.

Heiss, Florian (2011), 'Dynamics of self-rated health and selective mortality', *Empirical Economics* **40**, 119–140.

I-Min, Lee, Eric J. Shiroma, Felipe Lobelo, Pekka Puska, Steven N. Blair and Peter T. Katzmarzyk (2012), 'Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy', *The Lancet* **380**(9838), 219–229.

Juerges, Hendrik, Eberhard Kruk and Steffen Reinhold (2013), 'The effect of compulsory schooling on health: Evidence from biomarkers', *Journal of Population Economics* pp. 645–72.

Kemptner, Daniel, Hendrik Juerges and Steffen Reinhold (2011), 'Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany', *Journal of Health Economics* **30**(2), 340–354.

Levin, Jesse and Erik J. S. Plug (1999), 'Instrumenting education and the returns to schooling in the Netherlands', *Labour Economics* **6**, 521–534.

Lochner, Lance (2011), Non-production benefits of education: Crime, health, and good citizenship, Working Paper 16722, National Bureau of Economic Research (NBER).

Mazumder, Bhashkar (2008), 'Does education improve health: A reexamination of the evidence from compulsory schooling laws', *Economic Perspectives* **33**(2), 1–15.

Must, Aviva, Jennifer Spadaon, Eugenie H. Coakley, Alison E. Field, Graham Colditz and William H. Dietz (1999), 'The disease burden associated with overweight and obesity', *The Journal of the American Medical Association* **282**(16), 1523–1529.

Oreopoulos, Phillip (2007), 'Do dropouts drop out too soon? Wealth, health, and happiness from compulsory schooling', *Journal of Public Economics* **91**(11–12), 2213–2229.

Park, Cheolsung. and Changhui Kang (2008), 'Does education induce healthy lifestyle?', *Journal of Health Economics* **27**(6), 1516–1531.

Peracchi, Franco and Claudio Rossetti (2012), 'Heterogeneity in health responses and anchoring vignettes', *Empirical Economics* **42**(2), 513–538.

Pischke, J-S. and Till von Wachter (2008), 'Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation', *Review of Economics and Statistics* **90**, 592–598.

Powdthavee, Nattavudh (2010), 'Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England', *Journal of Human Capital* **4**(2), 173–202.

Rosenzweig, Mark R. and T. Paul Schultz (1983), 'Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight', *Journal of Political Economy* **91**(5), 723–746.

Ross, Catherine E. and Chia-ling Wu (1995), 'The links between education and health', *American Sociological Review* **60**(5), 719–745.

Silles, Mary A. (2009), 'The Causal Effect of Education on Health: Evidence from the United Kingdom', *Economics of Education Review* **28**(1), 122–128.

Smith, James P. (2009), 'The Impact of Childhood Health on Adult Labor Market Outcomes', *The Review of Economics and Statistics* **91**(3), 478–489.

Spasojevic, Jasmina (2003), 'Effects of education on adult health in Sweden: Results from a natural experiment'. PhD thesis. Graduate School for Public Affairs and Administration. Metropolitan College of New York.

Stowasser, Till, Florian Heiss, Daniel McFadden and Joachim Winter (2011), "Healthy, wealthy and wise?" Revisited: An analysis of the Causal Pathways from Socio-Economic Status to Health, Working Paper 17273, National Bureau of Economic Research (NBER).

Tubeuf, Sandy, Florence Jusot and Damien Bricard (2012), 'Mediating role of education and lifestyles in the relationship betwen early-life conditions and health: Evidence from the 1958 British cohort', *Health Economics* **21**(Suppl. 1), 129–150.

Van Kippersluis, Hans, Owen O'Donnell and Eddy van Doorslaer (2011), 'Long run returns to education: Does schooling lead to an extended old age?', *Journal of Human Resources* **46**(4), 695–721.

9 Appendix

9.1 An Illustrative Model

Following Grossman (1972), Rosenzweig and Schultz (1983) and Contoyannis and Jones (2004), assume that individuals have preference orderings over their own poor health H and two bundles of goods, C and B, where only the latter affects health. The vector B includes risky health behaviors or habits - such as smoking, the use of alcohol or drugs, unprotected sex, excessive calorie intake and poor exercise - which increase the utility from consumption but damage health.[1] In this illustrative example, we assume - as in Cutler et al. (2003) - that instantaneous utility U is concave in C and B but linear in H. We also assume that the marginal utility of (poor) health declines as individual education E increases, reflecting the view that better educated individuals have access to higher income and can therefore extract higher utility from better health and a longer life.[2] The intertemporal utility function for individual i is given by

$$\Omega_i = \sum_{k=0}^{T} \rho^k [U_{it+k}(C_{it+k}, B_{it+k}, \eta_{it+k}) - h(E_i)H_{it+k}] \tag{1}$$

where ρ is the discount factor, η is a vector of unobservable influences on U, $h(E)$ is increasing in E and the expression within brackets is the instantaneous utility function.

We posit that the stock of individual poor health H is positively affected by behaviors B and negatively affected by individual education E. Using a linear specification and assuming stationarity in the parameters, the health production function for individual i at time t is given by

$$H_{it} = \alpha B_{it} + \beta E_i + e_{it} \tag{2}$$

where e is a vector of unobservable influences on H and $\beta < 0$.

[1] See the discussion in Feinstein et al. (2006)

[2] As argued by Cutler and Lleras-Muney (2006), the higher weight placed on health by the better educated could reflect the higher value of the future: "...if education provides individuals with a better future along several dimensions - people may be more likely to invest in protecting that future"(p.15)

Rational individuals maximize (A.1) with respect to consumption and behaviors, subject to the health production function and to the budget constraint, defined by[3]

$$p_t C_{it} + B_{it} = Y_{it}(E_i, X_{it}) \tag{3}$$

where Y is income, which varies with education and a vector of observable controls X, p is the vector of consumption prices for goods C and the prices of B are normalized to 1. Assuming that an internal solution exists, the necessary conditions for a maximum are

$$U_{it}^C - \lambda p_t = 0 \tag{4}$$

$$U_{it}^B + \rho\alpha h(E_i) - \lambda = 0 \tag{5}$$

where λ is the Lagrange multiplier and the superscripts are for partial derivatives. By totally differentiating (A.4) and (A.5) and using (A.2) we obtain that

$$\frac{\partial B_{it}}{\partial E_i} = \frac{-\rho\alpha p_t \frac{\partial h(E_i)}{\partial E_i}}{\Delta} \tag{6}$$

where Δ is the determinant of the bordered Hessian, which is positive if the second order conditions for a maximum hold. It follows that higher education reduces optimal risky behaviors if $\frac{\partial h(E_i)}{\partial E_i} > 0$.

Equations (A.3), (A.4) and (A.5) yield optimal health behaviors

$$B_{it} = B(E_i, p_t, \rho, X_{it}, \eta_{it}) \tag{7}$$

Using (A.2), (A.7) and a similar expression for consumption C in (A.1) yields the indirect utility function

$$\Gamma_{it} = \Gamma(E_i, p_t, \rho, X_{it}, \eta_{it}, e_{it}) \tag{8}$$

[3]Rosenzweig and Schultz (1983), and Contoyannis and Jones (2004), use a similar formulation.

Letting $\Upsilon(E_i, Q_{it})$ be the cost of investing in education, where Q are cost shifters, the condition

$$\Gamma_{it}^{E} = \Upsilon_{it}^{E} \tag{9}$$

defines optimal education, which depends both on health production shocks e and on preference shocks η.

9.2 Synthetic Indicators for Parental Background

We have built synthetic indicators of parental background by extracting the first principal component from several groups of variables, in order to reduce the dimensionality of the vector of controls. Since most indicators are discrete we use the polychoric or polyserial correlation matrix instead of the usual correlation matrix as the starting point of the principal component analysis. The polychoric correlation matrix is a maximum likelihood estimate of the correlation between ordinal variables which uses the assumption that ordinal variables are observed indicators of latent and normally distributed variables. The polyserial correlation matrix is defined in a similar manner when one of the indicators is ordinal and the others are continuous. We list below the synthetic indicators, the observed variables used for each indicator and the interpretation we propose, based on the sign of the scoring coefficients. The scoring coefficients are the same across males and females (otherwise, we argue, results would not be comparable and we could not proceed with the aggregation-differentiation strategy).

Poor Housing at 10 based on the number of rooms in the house at age 10 and facilities in the house (hot water) at age 10. The extracted first principal component decreases as the number of rooms in the house (where the individual lived at age 10) increases and if there was no hot water: we interpret this indicator as *poor housing conditions at age 10*;

Parents drunk or had mental problems at 10 based on binary indicators of whether parents drunk or had mental problems when the individual was aged 10. Since the extracted principal component increases if parents drunk or had mental problems, we interpret it as *poor parental background at age 10*;

Parental absence at 10 based on three binary indicators: whether the mother died early, whether the father died early and whether the mother and the father where present when the individual was aged 10. The extracted principal component increases if any parent died early and decreases when parents were present at age 10. We interpret this indicator as *poor care at young age*.

Descriptive statistics on the background variables used to build the synthetic indicators and the additional background variables used in the baseline specification are reported in Table 1.

Table 1: Descriptive statistics, baseline estimation sample (micro data), males (M) and females (F).

Country	Serious dis. at 15		Poor Health at 10		Hospital at 10		Few books at 10		No hot water at 10		Rooms at 10	
	M	F	M	F	M	F	M	F	M	F	M	F
Austria	0.33	0.32	0.13	0.13	0.11	0.10	0.42	0.48	0.37	0.37	3.3	3.1
Belgium	0.27	0.28	0.06	0.09	0.04	0.05	0.49	0.46	0.30	0.33	5.1	5.2
Denmark	0.25	0.25	0.08	0.08	0.09	0.09	0.23	0.24	0.13	0.14	4.4	4.3
England	0.36	0.31	0.10	0.13	0.11	0.11	0.30	0.24	0.04	0.21	2.9	3.0
France	0.29	0.28	0.10	0.13	0.04	0.04	0.47	0.48	0.24	0.26	4.3	4.0
Germany	0.30	0.33	0.13	0.12	0.09	0.08	0.32	0.31	0.10	0.10	3.9	4.0
Greece	0.21	0.17	0.00	0.00	0.00	0.01	0.64	0.64	0.38	0.33	2.7	2.8
Italy	0.16	0.21	0.05	0.08	0.02	0.03	0.79	0.75	0.47	0.45	3.1	2.9
Netherlands	0.23	0.22	0.11	0.11	0.08	0.08	0.35	0.30	0.05	0.04	4.7	4.6
Spain	0.14	0.17	0.09	0.11	0.02	0.02	0.66	0.65	0.46	0.44	3.6	3.5
Sweden	0.24	0.24	0.06	0.08	0.09	0.08	0.20	0.18	0.14	0.13	3.7	3.6
Switzerland	0.30	0.32	0.06	0.14	0.07	0.07	0.28	0.31	0.03	0.05	4.8	4.9
All	0.27	0.26	0.08	0.08	0.07	0.07	0.43	0.41	0.21	0.21	3.7	3.7

Country	Parents drunk at 10		Parents ment. prob. at 10		Moth/Fath present at 10		Mother died early		Father died early	
	M	F	M	F	M	F	M	F	M	F
Austria	0.09	0.09	0.02	0.02	0.80	0.71	0.0	0.0	0.0	0.0
Belgium	0.09	0.09	0.01	0.03	0.92	0.92	0.0	0.0	0.0	0.0
Denmark	0.07	0.09	0.08	0.09	0.89	0.90	0.0	0.0	0.0	0.0
England	0.05	0.06	0.05	0.06	0.89	0.89	0.01	0.0	0.01	0.01
France	0.10	0.10	0.01	0.01	0.90	0.86	0.01	0.01	0.01	0.01
Germany	0.07	0.08	0.04	0.05	0.79	0.84	0.0	0.0	0.0	0.0
Greece	0.05	0.05	0.00	0.00	0.97	0.97	0.0	0.0	0.0	0.0
Italy	0.10	0.11	0.01	0.00	0.92	0.93	0.0	0.0	0.0	0.0
Netherlands	0.02	0.05	0.02	0.03	0.92	0.92	0.0	0.0	0.0	0.0
Spain	0.08	0.07	0.01	0.01	0.87	0.88	0.0	0.0	0.0	0.0
Sweden	0.07	0.08	0.02	0.02	0.87	0.88	0.0	0.0	0.0	0.0
Switzerland	0.09	0.09	0.03	0.03	0.91	0.94	0.0	0.0	0.0	0.0
All	0.07	0.08	0.03	0.03	0.90	0.90	0.0	0.0	0.01	0.0

9.3 Education Reforms in Europe

In this section, we briefly describe the compulsory schooling reforms we are using in this study. Our choice of reforms differs somewhat from Brunello et al. (2009) and Brunello et al. (2013) because the individuals in our data are aged 50 or older at the time of the interviews in 2004/2006. Therefore, we need to focus only on relatively early reforms. For further details on educational reforms in Europe see Fort (2006).

Austria A federal act was passed in 1962 that increased compulsory schooling from 8 to 9 years. The law came into effect on September 1, 1966. Pupils who were 14 years old (or younger) at that time had to attend school for an additional year. Since compulsory education starts at the age of 6 and the cut-off date for school-entry is September 1, (mostly) individuals born between September and December 1951 were the first ones affected by the reform. Thus, the pivotal cohort is 1951.

Czech Republic In the 20^{th} century, compulsory education was reformed several times. In 1948, compulsory schooling was increased from 8 to 9 years (age 6 to 15). It was reduced to 8 in 1953 and increased to 9 again in 1960. Two further changes took place in 1979 and 1990. We consider the first three reforms for our analysis. The pivotal cohorts are 1934 (for the first reform), 1939 (for the second) and 1947 for the reform in 1960. See Garrouste (2010) for more information on compulsory schooling reforms in the Czech Republic.

Denmark Compulsory education was increased in 1958 by 3 years, from 4 to 7. In 1971, compulsory schooling was further increased by 2 years, from 7 to 9. Education started at age 7, thus pupils who were 11 years old (or younger) in 1958 were potentially affected by the first reform, i.e. children born in 1947 and after. Since our data only cover individuals 50+ in 2004/2006, we only consider the first reform for this study.

England Two major compulsory schooling reforms were implemented in the UK in 1947 and 1973. The first reform increased the minimum school leaving age from 14 to 15, the second reform from 15 to 16. Since the school-entry age is 5 in the UK, compulsory schooling was increased from 9 to 10 years in 1947 and from 10 to 11 years in 1973. Pupils who were 14 years old (or younger) in 1947 were affected by the first reform, i.e. cohorts born in 1933 and after. Due to the sampling frame of ELSA (individuals 50+), we only consider the first reform in this study.

France Two education reforms were implemented in France. Compulsory schooling was increased from 7 to 8 years (age 13 to 14) in 1936 and from 8 to 10 years (age 14 to 16) in 1959. After a long transition period, the second reform came into effect

in 1967. The first reform affected pupils born 1923 (and after) and the second reform pupils born 1953 (and after).

Italy In 1963, junior high school became mandatory in Italy and compulsory years of schooling increased by 3 years (from 5 to 8 years). The first cohort potentially affected by this reform is the cohort born in 1949.

Netherlands The Netherlands experienced many changes in compulsory education in the last century. In this paper, we consider three education reforms: in 1942, in 1947 and in 1950 (Levin and Plug (1999)). With the first reform compulsory schooling was increased from 7 to 8 years, with the second reform it drop back to 7 years and with the last reform it increased again by 2 years, from 7 to 9. Accordingly, we choose the cohorts born in 1929, 1933 and 1936 as pivotal cohorts.