Discussion Paper No. 836

THE CAUSAL EFFECT OF EDUCATION ON HEALTH: WHAT IS THE ROLE OF HEALTH BEHAVIORS?

Giorgio Brunello
Margherita Fort
Nicole Schneeweis
Rudolf Winter-Ebmer

March 2012

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

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

Giorgio Brunello (University of Padua, IZA and CESifo)

Margherita Fort (University of Bologna, IZA and CHILD)

Nicole Schneeweis (University of Linz)

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

March 6, 2012

Abstract

We study the contribution of health-related behaviors to the health 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. We use an empirical approach that addresses the endogeneity of education and behaviors in the health production function. Focusing on self-reported poor health, we find that education has a protective effect for European males and females aged 50+. We also find that the mediating effects of health behaviors - measured by smoking, drinking, exercising and the body mass index - account in the short-run for 17% to 31% and in the long-run for 23% to 45% of the entire effect of education on health, depending on gender.

Keywords: Health, education, health behaviors, Europe.

JEL Codes: J1, I12, I21

*We would like to thank David Card, Lance Lochner and the participants to seminars in Berkeley, Bologna, Bressanone, Catanzaro, UC Davis, Firenze, Hangzhou, Helsinki, Linz, Munich, Nurnberg, Padova, Regensburg and Wurzburg for comments and suggestions on an earlier version of the paper. 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. The usual disclaimer applies.

1 Introduction

The relationship between education and health - the "education gradient" - is widely studied. There is abundant evidence that a gradient exists (Cutler and Lleras-Muney (2010)). Yet less is known as to why education might be related to health. In this paper we explore the contribution of health-related behaviors (shortly, behaviors) - which we measure with smoking, drinking, exercising and the body mass index[1] - to the education gradient. To do so, we decompose the gradient into two parts: a) the part mediated by health behaviors; b) a residual, which includes for instance stress reduction, better decision making, better information collection, healthier employment and better neighborhoods (Lochner (2011)).[2]

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 the health status. 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.

[1] Conditional on exercising, drinking and smoking, the body mass index essentially captures the effects of poor diets and low intake of fruit and vegetables, two key behaviors affecting health (see Cawley and Ruhm, 2011). We use the body mass rather than these behaviors because of lack of suitable data.

[2] The residual also includes the contribution of unmeasured behaviors.

Our identification strategy is based on aggregation, differencing and selection on observables (ADS) and allows us to estimate average education effects for an individual randomly picked from the population. Using a cross-country dataset, in which we observe a rich set of parental and early life information, this strategy combines selection on observables and fixed effects assumptions to estimate the parameters of both a dynamic health equation, which depends on education and lagged health behaviors, and a static health equation, where health depends only on education. The effect of education on health in the second equation is the education gradient (shortly, the gradient), i.e. the total effect of education on health that results from both mediated and residual effects of education. The identification strategy we propose presents significant improvements with respect to a strategy simply based on selection on observables since it removes the effect of all unobserved factors that are on average common within clearly defined cells.

To corroborate the internal validity of our identification approach, we contrast these estimates of the health education gradient with those obtained with a completely different methodology, instrumental variables (IV) estimation, where the key exogenous variation is provided by the changes in compulsory schooling laws across countries and birth cohorts. While the IV strategy generates causal estimates that are internally valid for individuals affected by mandatory schooling laws (*compliers*), it cannot be used for the decomposition of the education gradient because of the lack of valid and relevant instruments for behaviors.

We apply this approach to a multi-country dataset, 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. These 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 modeled following the US Health and Retirement Study.

Focusing on self-reported poor health as health outcome, we find that education has a protective effect, both for males and females, although the effects for females are typically somewhat higher. When evaluated at the sample mean of the dependent variable, one additional year of education reduces self-reported poor health by about 7% for females and 3% for males. These effects are smaller than those found by others. Our explanation is that we use a sample of older individuals (50+) than typically done in the literature, and that the protective role of education on health declines with age. Our qualitative findings are robust to the choice of the identification strategy.

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, however, remains unexplained. Potential candidates accounting for this part include direct effects of education on health as well as indirect effects operating through unobserved health behaviors, wealth and cognitive abilities.

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 empirical 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,

see for instance Adams (2002), Mazumder (2008) and Oreopoulos (2007) for the US, Arendt (2008) for Denmark, Kempter et al. (2011) for German males and Silles (2009) and Oreopoulos (2007) for the UK. Others find small or no effects. While Clark and Royer (2010) 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), Juerges et al. (2009) and Powdthavee (2010) for the UK and Kempter 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.[3] The contribution of behaviors, which include smoking, drinking and eating calorie-intensive food, has been examined in the economic and sociological literature, starting with the contribution by Ross and Wu (1995).[4] 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 which includes both. In the study closest to this paper, Contoyannis and Jones (2004) partly address this concern by explicitly modeling the optimal choice of health behaviors. They jointly estimate a health equation - where health depends 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) be-

[3]Conti et al. (2010) argue that non-cognitive skills may be an important factor as well.
[4]See the reviews by Feinstein et al. (2006) and Cawley and Ruhm (2011).

haviors to the education gradient varies between 23% to 73%, depending on whether behaviors are treated as exogenous or endogenous.

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 limited set of countries (Denmark, France, Germany, 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.[5]

We contribute to this literature in several directions. First, we distinguish explicitly between the short-run and long-run mediating effects of health behaviors. While the former only includes the effects of behaviors in the immediate past, the latter takes into account the contribution of the entire history of behaviors. This qualification is empirically relevant, as we show in Section 6. Furthermore, our study is the first to cover a substantial number of European countries (13), using a multi-country dataset which includes also Southern European countries, which have not been studied before. We are also the first to offer an identification strategy - aggregation, gender differencing and selection on observables (ADS) - which addresses the endogeneity of both education and health behaviors in the health production function. The assessment of the health education gradient obtained relying on this approach proves to be broadly very similar to the estimate obtained relying on a a more conventional -and widely accepted- IV strategy, which exploits the exogenous variation in education across countries and cohorts induced by changes in mandatory schooling. We interpret this as supportive evidence for the causal interpretation of our main results.

3 Contribution of Health Behaviors to the Education Gradient

In the empirical literature (Ross and Wu (1995) and Cutler et al. (2008)) the contribution of health behaviors to the education gradient (HEG) is evaluated by adding the

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

vector of either 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 a constant and v is the error term. We assume stationarity in the parameters ($c_t = c$; $\alpha_{t-1} = \alpha$; $\beta_t = \beta$). Behaviors themselves depend on education. The education gradient $\alpha \frac{\partial B_{t-1}}{\partial E_i} + \beta$ can be decomposed into: a) the effect operating via health behaviors lagged once B_{t-1}, or $\alpha \frac{\partial B_{t-1}}{\partial E_i}$; b) the residual effect β. As reviewed by Lochner (2011), channels through which education may improve health without affecting behaviors include stress reduction, better decision making, healthier and safer employment, healthier neighborhoods and peers. The ratio between the effect operating via health behaviors and the overall effect measures the relative contribution of health behaviors in the immediate past to the education gradient.

To illustrate with an example, assume that the instantaneous utility function is given by $U(C_{it}, B_{it}, \eta_{it}) - h(E)H_{it}$, where η is a vector of unobservables affecting preferences, and let ρ be the discount factor and p_t the price of the bundle of goods not affecting health.[6] As shown in the Appendix, the maximization of the inter-temporal utility function subject to the health production function (1) and the budget constraint yields the vector of optimal behaviors $B_{it} = B(E_i, p_t, \rho, X_{it}, \eta_{it})$, where X is a vector of exogenous covariates. Ignoring for the time being the price vector p, the discount factor and the vector X, a linear approximation of behaviors is

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

[6]The price of the bundle of goods affecting health, which include risky health behaviors B, is normalized to one. The utility function does not take into consideration habit formation mechanisms. An extension which deals with the effects of these mechanisms on behaviors is left to future research.

Substituting (2) into (1) yields

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

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

3.1 The History of Behaviors

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. Using the instantaneous utility function introduced above and ignoring again the price vector p, the discount factor and the vector X, a linear approximation of optimal behaviors is given by equation (2), which combined with (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 only for 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

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

Since equation (6) can be written as equation (4) and we retain the same instantaneous utility function, the linear approximation of optimal health behaviors in equation (2) is unchanged.[7] Using this approximation into (6), we obtain

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

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 $B_{it} = \sigma_0 + \sigma_1 E_i + \eta_{it}$ into (7) yields the "reduced form" 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 education gradient HEG.

[7] We ignore again prices, the vector X and the discount factor.

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 behaviours, 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 $t-T$.[8]

3.2 Estimating Short- and Long-Run Mediating Effects

One of the aims of this paper is to provide estimates of $SRME$ and $LRME$. Our empirical strategy is based on the estimation of the parameters of the dynamic health equation (6) and the "reduced form" health equation (8). Using these estimates, we can compute

$$\widehat{\pi\sigma_1} = \widehat{\chi}_1(1-\widehat{\phi}) - \widehat{\nu} \tag{11}$$

and

$$\widehat{LRME} = \frac{\widehat{\chi}_1(1-\widehat{\phi}) - \widehat{\nu}}{\widehat{\chi}_1(1-\widehat{\phi})} \tag{12}$$

[8]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 also change signs.

$$\widehat{SRME} = (1 - \widehat{\phi})\widehat{LRME} \tag{13}$$

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. For that, we would need to estimate also equations such as (2) for each available behavior. We leave this development for future research.

4 Empirical Strategy

We start this section with the following assumptions on the error terms e in the dynamic health equation (6) and η in the behavior equation (2):

$$e_{it} = f_i + \xi_{it}; \quad \eta_{it} = g_i + \zeta_{it} \tag{14}$$

$$(\xi_{it}, \zeta_{it}) \perp (\xi_{ik}, \zeta_{ik}) \text{ for all } j \neq k \qquad | \qquad f_i, g_i, X, E_i, B_{it} \tag{15}$$

$$E[\xi_{it}|f_i, g_i, X, E_i, B_{it}] = 0 \quad E[\zeta_{it}|f_i, g_i, X, E_i, B_{it}] = 0$$

$$E[f_i|E_i] \neq 0 \quad E[f_i|B_{it-1}] \neq 0 \quad E[f_i|H_{it-1}] \neq 0 \tag{16}$$

$$E[g_i|E_i] \neq 0 \tag{17}$$

Each error term is composed of a time invariant individual effect (either f or g in equation (14)) and a strictly exogenous transitory serially uncorrelated effect (either ξ or ζ). Conditional on f, g, E and B and the vector X of exogenous variables, transitory effects are uncorrelated (Eq.(15)) and have zero mean (Eq.(16)). Education, behaviors and lagged health are correlated to the errors e and η only via their correlations with the individual effects f and g, i.e. we do not assume that these effects are independent of E_i, B_{it}, H_{it-1}.[9]

[9]In our set-up, individual effects could be regarded as random without loss of generality given that we are interested in partial regression coefficients holding these individual effects constant. We regard $(H_{it}, E_i, B_{it}, H_{it-1}, X, f_i, g_i)$ as a random sample from an artificial multivariate population with joint distribution $p(H_{it}, E_i, B_{it}, H_{it-1}, X, f_i, g_i) = p(H_{it}|E_i, B_{it}, H_{it-1}, X, f_i, g_i)p(E_i, B_{it}, H_{it-1}, X, f_i, g_i)$ and focus on the conditional distribution of H_{it}.

Since optimal education depends on the unobservables that affect preferences (η) and health production (e) - see the illustrative example in the Appendix - ordinary least squares estimates of the health production function fail to uncover causal relationships. A similar problem affects the OLS estimates of the "reduced form" health equation, because health depends both on education and on the sequence of shocks affecting preferences and health production (see equation (8)). An important drawback of the empirical studies investigating the mediating effect of health behaviors on the education gradient is that they fail to consider the endogeneity of education and behaviors (Lochner (2011)). In this paper, we address these problems in an attempt to give a causal interpretation both to the gradient and to the mediating role of behaviors.

In the past few years, several papers have estimated the causal effect of education on health using the exogenous variation in educational attainment generated by compulsory schooling laws. This instrumental variables (IV) approach can be used to estimate the "reduced form" health equation (8). In principle, the same approach can also be applied to the dynamic health production function (6), provided that we can find additional credible sources of exogenous variation which affect risky health behaviors without influencing individual health (conditional on behaviors). This is a very difficult task with the data at hand.[10] Therefore, we turn to an identification strategy that combines aggregation, fixed effects and selection on observables to estimate both the dynamic health production function and the "reduced form" health equation.[11] For the latter equation, we compare the results obtained following this approach to those obtained with a more conventional IV approach, using changes in compulsory education as instrument for education and provide additional evidence supporting the causal interpretation of our main results. We illustrate these two approaches in turn.

[10] Using instruments such as the price of alcohol or cigarettes does not work in our setup because these variables influence all cohorts in one country alike.

[11] Card and Rothstein (2007) use a similar method to investigate ethnic segregation in US schools and its impact on the black-white test score gap.

4.1 Aggregation, Differencing and Selection on Observables

Estimating the dynamic health equation (6) by OLS gives biased estimates if education and health behaviors are correlated with unobservable characteristics affecting health. To illustrate the endogeneity problem, consider unobserved ability. While this variable is omitted from the regression, it might influence current health and it is likely to be correlated with our variables of interest, education and lagged behaviors. We address this problem by combining aggregation, differencing and selection on observables. In short, we aggregate our data into cells defined by gender, cohort, country and time period (two waves of data). 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 and capture residual gender-specific unobservables with observable controls, which include a rich set of parental and early life conditions.

More in detail, consider the following empirical version of the dynamic health production function

$$H_{icgbt} = \alpha_{g0} + \alpha_{g1} B_{icgb(t-1)} + \alpha_{g2} E_{icgb} + \alpha_{g3} X_{icgb} + \alpha_{g4} H_{icgb(t-1)} + \widetilde{f}_{icgb} + \xi_{icgbt} \quad (18)$$

where i denotes the individual, c the country, g gender (M: males; F: females), b the birth cohort, t calendar time, X a vector of control variables and we allow each explanatory variable to have a gender-specific effect on health. The error term in equation (19) can be decomposed as follows

$$\widetilde{f}_{icgb} + \xi_{icgbt} = f_{cgb} + \xi_{cgbt} + \epsilon_{icgbt} \quad (19)$$

where $f_{cgb} + \xi_{cgbt}$ represent a common error component for individuals of the same country c, gender g, birth cohort b and time t and ϵ_{icgbt} is an individual-specific idiosyncratic error component for which we assume

$$E[\epsilon_{icgbt} | c, g, b, t] = 0 \quad (20)$$

We aggregate individual data into cells identified by country, gender, birth cohort and time and obtain the aggregated health equation (21), where $\overline{H}_{cgbt}$ denotes $E[H|c,g,b,t]$ and similarly for the other key regressors E, B_{t-1}, X, H_{t-1}

$$\overline{H}_{cgbt} = \alpha_{g0} + \alpha_{g1}\overline{B}_{cgb(t-1)} + \alpha_{g2}\overline{E}_{cgb} + \alpha_{g3}\overline{X}_{cgb} + \alpha_{g4}\overline{H}_{cgb(t-1)} + f_{cgb} + \xi_{cgbt} \qquad (21)$$

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

$$\Delta\overline{H}_{cbt} = \alpha_0 + \alpha_{M1}\Delta\overline{B}_{cb(t-1)} + \alpha_1\overline{B}^F_{cb(t-1)} + \alpha_{M2}\Delta\overline{E}_{cb} + \alpha_2\overline{E}^F_{cb} + \alpha_{M3}\Delta\overline{X}_{cb} + \alpha_3\overline{X}^F_{cb} +$$

$$+\alpha_{M4}\Delta\overline{H}_{cbt-1} + \alpha_4\overline{H}^F_{cbt-1} + \Delta f_{cb} + \Delta\xi_{cbt} \qquad (22)$$

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 for males and females respectively. Similarly, the gender gap in the "returns" to education is given by coefficient α_2.[12]

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.[13] Even after eliminating common unobservables, however, one may argue that the residual error component Δf_{cb} could still be correlated with education and lagged health behaviors. This could happen, for instance, if health conditions and parental background during childhood differ systematically by gender or if labor market discrimination affects individual income and access to health care, conditional on educational attainment. In this case, Δf_{cb} would differ from zero and may drive the endogeneity

[12]To avoid confusion, we stress that $\Delta\xi_{cbt}$ is the difference between ξ_{cbFt} and ξ_{cbMt}, not the difference between ξ_{cbFt} (ξ_{cbMt}) and ξ_{cbFt-1} (ξ_{cbMt-1}), i.e. we are taking differences between genders in a given calendar time, not differences within gender over time.

[13]See Zweifel and Breyer (1997).

of E, B_{t-1} in the health equation. To remove this potential correlation, we model this residual as

$$\Delta f_{cb} = \psi_b + \psi_c + \psi_{M1}\Delta\overline{Z}_{cb} + \psi_1\overline{Z}^F_{cb} + \psi_{M2}\Delta\overline{Y}_{cb} + \psi_2\overline{Y}^F_{cb} + \kappa_{cb} \tag{23}$$

where $\psi_s = \psi_{Fs} - \psi_{Ms}$, with $s = 1, 2$, ψ_b includes cohort dummies and country-specific linear or quadratic trends in birth cohorts, ψ_c is a vector of country dummies, Z a vector of observables, which includes a rich set of parental background characteristics and health conditions during childhood[14] and Y is real income. Our identifying assumption is that, conditional on these variables which capture gender-specific genetic and environmental effects, the error term κ_{cb} is orthogonal to levels and changes in health behaviors and educational attainment.[15] For the sake of brevity, we call this method ADS (aggregation cum differencing cum selection on observables).

Our identifying assumptions would only be violated if there are residual omitted factors which affect gender differences in health, education and behaviors after having conditioned for country and cohort effects as well as for gender differences in early life conditions. We can hardly think about good examples of such factors.

Going back to the illustrative example at the beginning of this section, suppose that the key unobservable in (21) 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.[16] 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.

[14]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.

[15]The inclusion of cohort and country dummies in (23) implies that cohort and country effects may differ by gender.

[16]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. Other than this the conditional distribution is left unrestricted and the inference is conditional on this effect.

Conditional on our identifying assumptions, equation (22) 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).

4.2 The IV approach

As an alternative approach, we estimate the "reduced form" health equation (8) by instrumental variables, using the number of years of compulsory education YC as instrument for individual schooling years E. This strategy is widely considered as credible and has been used extensively in the literature (see Lochner, 2011 for a review). As in Brunello et al. (2009), Brunello et al. (forthcoming) and Fort et al. (2011), we apply this strategy to a multi-country setup and exploit the fact that school reforms have occurred at different points in time in several countries.

For each country and reform included in our sample, we construct pre-treatment and post-treatment samples by identifying for each reform the pivotal birth cohort, defined as the first cohort potentially affected by the change in mandatory years of schooling. 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 YC "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 the within country exogenous variation in the instrument to identify the causal effects of schooling on health.

We include in equation (8) 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.[17] On the other hand, country fixed effects control for national differences both in reporting styles and in institutions affecting health.

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

[17] "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)

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

5 Data

The estimation of the "reduced form " and the dynamic health equation requires data on health outcomes, risky health behaviors, 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 requirements. SHARE is a longitudinal dataset on the 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, Switzerland, Denmark, Spain, France, Germany, Italy, Greece, The Netherlands and Sweden.[18] ELSA has similar characteristics and covers England.[19] 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.

The measure of health used in this paper is self-reported poor health ($SRPH$), 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.[20] 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.

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

[19]For England, we use waves 2 (2004/5) and 3 (2006/7).

[20]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.

In contrast, the percentage of individuals in good health with similar diseases was 28, 44 and 33 percent, respectively. Moreover, the latter group experienced only 1.10 chronic diseases.[21]

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.[22] However, we also present in the robustness section of this paper estimates based on the number of chronic diseases.

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.

The ADS strategy used in this paper requires that there is gender variation in educational attainment and self-reported health. Figure 1 documents that such variation exists: using our cell aggregation approach, the figure shows gender differences by cell in education and self reported poor health. The two variables are negatively correlated (slope coefficient of the weighted regression: -0.027; standard error: 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 refrains from vigorous activity and the body mass index (BMI). Conditional on the first three measures, the body mass index captures the effects of poor diets and low intake

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

[22] While self-perceived health is a comprehensive health measure, specific chronic diseases capture only part of individual health. Previous studies have shown that self-perceived health and future mortality are strongly correlated - see Heiss, 2011 for the US and Boppetal, 2012 for Switzerland.

of fruit and vegetables, two key behaviors affecting health.[23] Table 1 reports the country averages of the health outcome $SRPH$, years of education and annual income (thousand euro at 2005 prices, PPP) in 2006/07 as well as the means of the four health behaviors (in 2004/05), separately by gender.[24]

There is important cross-country and cross-gender variation, both in the outcome 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.[25]

As discussed above, we use the ADS approach to estimate the dynamic health equation (6) and both the ADS and the IV approach for the "reduced form" 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[26], and retain those cells that include at least two observations. We use data from 12 countries, all of which have participated in at least two waves in the surveys.

We implement the IV approach by selecting 7 countries where the individuals in our sample experienced at least one compulsory school reform: Austria, the Czech

[23]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.

[24]The Table does not include the Czech Republic because the ADS strategy cannot be applied to this country, where a single wave of data is available.

[25]Table B.1 in the Appendix reports the country averages of the parental background variables included in the vector Z. The table shows that the between genders variations in parental background and childhood characteristics are small: we interpret this as suggestive evidence that parental background characteristics are substantially removed by gender differencing, since - within country and cohort - they are largely common between males and females, on average.

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

Republic, Denmark, England, France, Italy and the Netherlands.[27] A short description of the compulsory school reforms used in this paper can be found in Appendix C. Since the "reduced form" equation (8) is static, we can use individuals who participated in at least one of the two first waves. When available, we measure the key variables (health, education) using the information provided by individuals during their second interview. For those who did not participate to the second wave, we use the first wave.

6 Results

This section describes the results of our empirical analysis and is organized as follows: first, the baseline estimates of the ADS model for the "reduced form" and the dynamic health equation are presented. Next, we discuss in Section 6.2 the reduced form estimates based upon the IV approach. Finally, the ADS estimates of the "reduced form" health equation are compared to the IV results (Section 6.3). Section 6.4 concludes with several robustness checks.

6.1 Baseline Estimates of the Reduced Form and Dynamic Health Equations

As reviewed in Section 2, most of the earlier contributions to this literature fail to consider the endogeneity of education and health behaviors in the health regressions. For the sake of comparison with this literature, we start the illustration of our empirical findings with estimates of the "reduced form" and the dynamic health equations based on micro data. We use a linear probability model, treat education and behaviors as exogenous, and regress self-reported poor health on years of education and a vector of variables, which varies according to whether we consider the "reduced form" or the dynamic health equation and whether we include parental and early life controls or not.

[27] We include the Czech Republic in the IV estimates of the reduced form model because these estimates only use a single wave of the data, which is available also for this country. 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.

For each regression, we pool males and females but allow for the full set of interactions of each explanatory variable with a gender dummy. Preliminary testing suggests that we cannot reject the null hypothesis that cohort, country, time and early life effects do not vary by gender.[28] We therefore report only the estimates with micro data in which country, cohort, time and early life effects do not vary by gender.

Table (2) is organized in two columns, one for the "reduced form" and the other for the dynamic health equation, which includes health behaviors lagged once, the first lag of health and current income. In the "reduced form" equation the marginal effect of one additional year of schooling on poor health is equal to -0.017 for females and to -0.012 for males, a relatively small effect when compared to the existing literature for Europe, which points to an effect in the range -0.026 to -0.081 (Lochner (2011), Table 6). This difference can be explained, at least in part, if the education gradient declines with age, given that our sample consists of individuals aged 50+ and the samples used in the literature typically include also younger individuals. Coefficients of parental and early life conditions, including poor health at age 10, are statistically significant and point in the expected direction: poor health conditions at 10 or 15 as well as poor parental environments at early ages increase self-perceived poor health at age 50+. Importantly, the inclusion of these variables reduces the gradient by 15 to 20 percent with respect to a more parsimonious specification without early life controls (not shown in the table), which suggests that these variables capture at least in part the positive correlation between educational attainment and unobserved individual effects such as ability and initial health.

Turning to the dynamic health equation, we find that our measures of risky health behaviors have statistically significant coefficients, with predictable correlations: smoking, refraining from vigorous activity and poor diet leading to higher BMI are positively related to self-perceived poor health. Somewhat unexpectedly, however, drinking alcohol almost every day is negatively related to self-reported poor health, both for males

[28]The joint hypothesis 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: 0.42), background variables (p-value: 0.263), trends in cohorts (p-value: 0.112) and we never reject the null at conventional significance levels.

and females. Annual real income is negatively associated to perceived poor health. Finally, the lagged dependent variable has a coefficient close to 0.5 (but statistically distinct from 1), which suggests that self-reported health is persistent over time.

Adding health behaviors, income and lagged health reduces the coefficient of education from -0.017 to -0.006 for females, and from -0.012 to -0.005 for males. Assuming that the returns to education for the sample of countries under study is equal to 0.07[29], the estimated mediating effect of behaviors lagged once is 16.8% for females and 9.7% for males - see Table 3. In the long run, when we include the effect of earlier health behaviors, the mediating effect almost doubles to 32.3% for females and 18.9% for males, suggesting that considering only their first lag may substantially underestimate the contribution of health behaviors to the education gradient. Our estimated long-run effects are smaller than those found by Cutler et al. (2008), who use a different approach but conclude that measured health behaviors account for over 40% of the education gradient (on mortality) in a sample of non-elderly Americans.[30]

Although the inclusion of parental and early life controls in our regression is likely to attenuate the correlation between education, health behaviors and unobservables, there is no guarantee that this correlation will disappear entirely. In order to identify the causal impact of education on health and behaviors, we apply the ADS procedure discussed in Section 4.1, which combines aggregation and gender differentiation with selection on observables. The specification tests carried out on the micro data suggest that cohort, country and early life effects do not differ significantly by gender. As a consequence, when we take gender differences of cell data, these common effects are removed together with common unobservables.

Our results for the ADS model are shown on the right-hand side of Table 2, both for the "reduced form" and for the dynamic health equation. When we consider the former, we find that the overall effect of education on poor health is negative and larger

[29]See for instance the estimates in Brunello et al. (2009). Adding income to equation (6) implies that $LRME$ is equal to $\frac{\pi\sigma_1}{(\pi\sigma_1+\nu+q\rho\overline{Y})}$, where q is the coefficient of income in the dynamic health equation, ρ is the estimated return to education and $\overline{Y}$ is average income.

[30]These authors estimate a static health equation, which includes income and occupation among the explanatory variables, and use the following measures of health behaviors: current smoker, ever smoker, number of cigarettes per day, obesity, regular exercise and use of seat belts always.

in absolute values for females (−0.026) than for males (−0.010). 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. Turning to the dynamic health equation, we find that the effect of education conditional on behaviors is much smaller (−0.015 for females and −0.003 for males) than in the "reduced form". While the precision of the estimates of the effects of behaviors declines in the cell data with respect to the micro data, we cannot reject the null hypothesis that these effects are jointly statistically significant. Finally, income effects are insignificant and the persistence of self-reported poor health over time is substantially reduced with respect to the estimates based on micro data.

Aggregation and differentiation increases the absolute value of the overall education gradient for females from 0.017 to 0.026 but has limited effects on the gradient for males, which marginally declines in absolute values from 0.012 to 0.010. The short-run and the long-run mediating effects of health behaviors are also affected. As shown in Table 3, when compared to the estimates based on micro data, the long-run mediating effect for males declines in absolute value (from 0.007 to 0.004) but increases as a share of the gradient (from 18.9 to 44.5%). The opposite happens for females, for whom this effect increases in absolute value from 0.005 to 0.006 but declines as share of the gradient (from 32.3% to 22.8%).

In sum, when we explicitly take into account the endogeneity of education and health behaviors, we find that the long-run mediating effect of the latter ranges between 23 (for females) and 45% (for males) of the total education gradient. While the effect of education on behaviors accounts for an important share of the gradient, 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.

6.2 IV Estimates of the Reduced Form Health Equation

In this section, we present the estimates of the "reduced form" health equation obtained using instrumental variables. We instrument education with the number of years of compulsory education, which varies across countries and cohorts because of compulsory schooling reforms. For each country, we construct a sample of treated individuals, who have experienced a change in compulsory education, and a control sample, with no change in compulsory schooling. Since our data include only individuals aged 50+, we need to focus on school reforms which took place between the 1940s and the 1960s, and to restrict our attention to a sub-sample of 7 countries affected by these reforms. Table 4 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 (see Appendix C for a short description of the education reforms used in this paper).

In order to ensure that individuals spent their schooling in their host country, we restrict our sample to individuals, who participated in the first or second wave of SHARE (second or third wave in ELSA), and were born in the country or migrated there before age 5. Additionally, 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. In particular, we estimate two specifications, one with a linear trend and one with a quadratic trend.

Since the key identifying assumption that changes in average education within country can be fully attributed to the reforms is more plausible when the window around the pivotal cohort is small, we estimate our model using two alternative samples, one including individuals who were born up to 10 years before and after the reforms and another where the relevant window is +7,-7. The two samples consist of 15,960 and 12,294 individuals respectively. Table 5 shows summary statistics by country for the larger sample.

Table 6 presents our estimates of the health-education gradient for both males and females. We report OLS, 2SLS, ITT (Intention-To-Treat, i.e. the effect of compulsory schooling on health), first stage and IV-Probit estimates for both samples, using two alternative specifications for the country-specific trends (linear or quadratic). The OLS estimates of the gradient are $-0.024/-0.025$ for females and -0.017 for males. The estimated magnitude of the gradient increases with instrumental variables: we find that one additional year of schooling decreases the probability of poor health by $4-8.5$ percentage points for females and by $5-6$ percentage points for males, depending on the selected window. IV-Probit estimations yield very similar results. The IV strategy works well: our first stage regressions show that the instrument is relevant and not weak (F-Statistics between around 13 and 42) and that one additional year of compulsory schooling increases actual schooling by a quarter to a third of a year, broadly in line with previous findings in the literature using similar identification strategies. We interpret the IV estimates as Local Average Treatment Effects, 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 distribution of education.

6.3 IV and ADS Results Compared

Next, we compare the education gradients estimated with the IV and the ADS approaches (Table 7). For the IV approach, we report the estimates with the linear trend specification and the larger window (± 10). We find that education reduces self-perceived poor health by 4 and 4.8 percentage points for females and males respectively. The ADS approach yields smaller estimates - in absolute values - for females (2.6 percentage points) and especially for males (1 percentage point).

Since the two approaches are based on a different set of countries and cohorts, we re-estimate the ADS model for the same sample used in the IV approach. The results are shown in the last column of Table 7. The magnitudes of the ADS estimates on this new sample increase somewhat, to 2.8 percentage points for females and to 2 for males, but remain smaller in absolute value than the IV estimates. To explain this difference,

we notice that, while IV estimates are Local Average Treatment Effects, i.e. the causal effects of education on health for the individuals affected by the compulsory schooling reforms, the estimates obtained with the ADS approach pertain to a randomly drawn individual from the entire sample. If the protective effect of education on health is more pronounced for persons with lower education, this would explain the higher magnitudes obtained with the IV approach.

6.4 Robustness Checks

In this section, we focus on the ADS approach and consider 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.[31] 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. However, when we decompose the gradient into the effect mediated by behaviors and the residual effect, we find that the $LRME$ in this sub-sample is smaller than in the full sample, and is equal to 8.5% and 11.1% of the gradient for females and males respectively.[32]

Furthermore, we notice that the older cohorts in our data are strongly selected by mortality patterns.[33] To control for this, we add to the regressions the level and the gender difference of life expectancy at birth, which vary by country, gender and birth cohort. Since these data are not available for Greece[34], we are forced to omit

[31] We do not restrict our attention to individuals interviewed in both waves to build the pseudo-panel.

[32] 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.

[33] Age in our sample ranges from 50 to 86.

[34] 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.

this country from the sample. As displayed by the last two columns in the table, life expectancy is never statistically significant in the "reduced form" 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 closer to those found in the literature. Since survivors aged 70 to 86 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 have a higher 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. 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 "reduced form" and the dynamic health equation, and the IV estimates of the "reduced form". 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 directions 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 have studied 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. We have proposed a strategy to estimate and decompose the education gradient which takes into account both the endogeneity of educational attainment as well as the endogenous choice of health behaviors. This strategy combine aggregation, gender differencing and selection on a broad set of observables.

Our results show that one additional year of schooling reduces self-reported poor health by about 7% for females and by about 3% for males.[35] These effects are smaller than those found in comparable empirical literature. We have argued that this discrepancy can be due to the fact that our sample consists of senior individuals aged 50+ and that the gradient declines with age.

Health behaviors - measured by smoking, drinking, exercising and the body mass index - contribute to this gradient. We find 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 in a significant way - especially for males - if we only consider behaviors in the immediate past, as usually done in the empirical

[35]The elasticities are evaluated at the sample mean value of self-reported health (0.367 for females and 0.316 for males).

literature. Using a completely different strategy - instrumental variables estimation - we find corroborating results for the education gradient.

Since the gradient is key to understanding inequality in health and life expectancy and is also used to assess 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, 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 about subsidizing schooling. 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

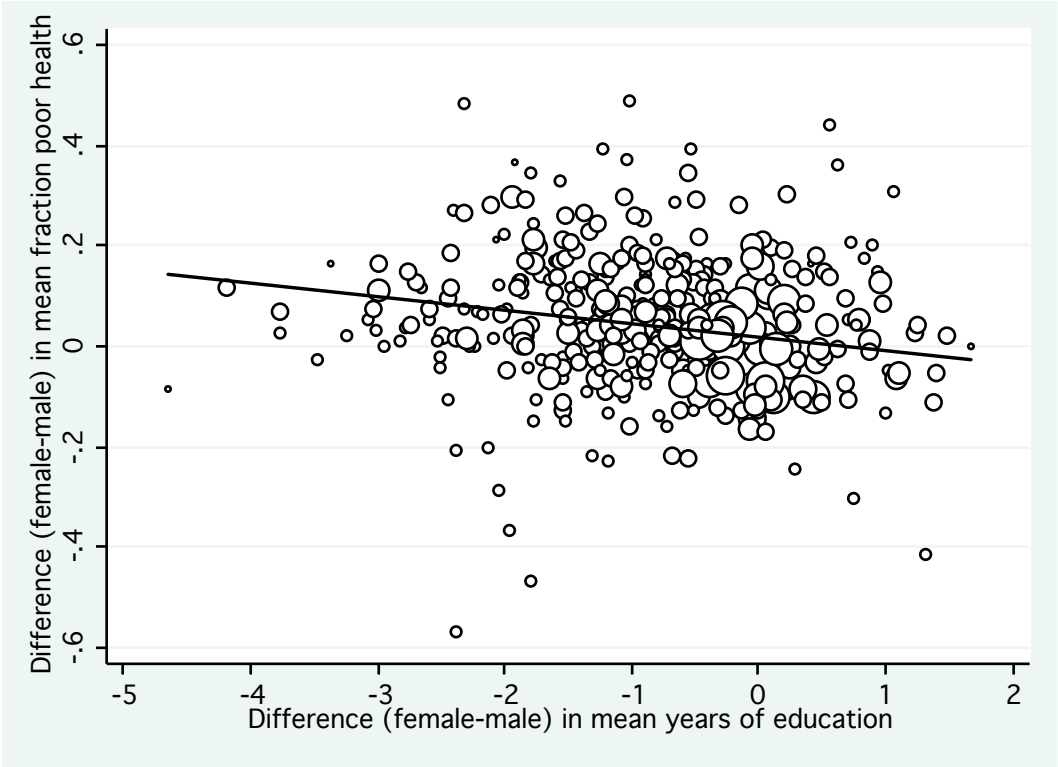

Figure 1: Gender differences in education and self-perceived poor health at time t. Aggregated data by gender, cohort and country. Circle areas are proportional to weights based on the number of individuals used for aggregation $(N_M^{-1} + N_F^{-1})^{-1}$.

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

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.

Table 2: Baseline Results - Micro and ADS Model

	Micro-estimates		ADS-model	
	Reduced form	Dynamic HE	Reduced form	Dynamic HE
Females				
education	-0.017	-0.006	-0.026	-0.015
	(0.001)***	(0.001)***	(0.005)***	(0.005)***
self-rep poor health$_{t-1}$		0.479		0.246
		(0.012)***		(0.046)***
drinking$_{t-1}$		-0.025		-0.013
		(0.012)**		(0.053)
smoking$_{t-1}$		0.052		-0.034
		(0.012)***		(0.056)
No vigorous		0.032		0.040
exercise$_{t-1}$		(0.009)***		(0.042)
BMI$_{t-1}$		0.007		0.003
		(0.001)***		(0.004)
income$_t$		-0.000		-0.002
		(0.000)**		(0.001)
Males				
education	-0.012	-0.005	-0.010	-0.003
	(0.001)***	(0.001)***	(0.005)*	(0.005)
self-rep poor health$_{t-1}$		0.486		0.308
		(0.014)***		(0.046)***
drinking$_{t-1}$		-0.041		-0.062
		(0.010)***		(0.038)
smoking$_{t-1}$		0.030		0.043
		(0.011)***		(0.042)
No vigorous		0.049		0.089
exercise$_{t-1}$		(0.009)***		(0.041)**
BMI$_{t-1}$		0.006		0.011
		(0.001)***		(0.005)**
income$_t$		-0.000		-0.001
		(0.000)**		(0.001)
Early life				
few books in HH	0.043	0.022	0.053	0.040
	(0.009)***	(0.008)***	(0.035)	(0.033)
serious diseases at 15	0.017	0.004	0.028	0.004
	(0.008)**	(0.007)	(0.036)	(0.035)
poor health at 10	0.117	0.062	0.158	0.135
	(0.014)***	(0.012)***	(0.052)***	(0.049)***
hospital at 10	0.032	0.025	0.004	0.042
	(0.016)**	(0.014)*	(0.063)	(0.061)
Principal components				
parents drunk or had	0.036	0.018	0.011	0.025
mental problems at 10	(0.009)***	(0.008)**	(0.039)	(0.038)
parental absence at 10	0.011	0.007	-0.008	-0.009
	(0.011)	(0.009)	(0.039)	(0.037)
poor housing at 10	0.016	0.013	0.023	0.014
	(0.004)***	(0.004)***	(0.017)	(0.016)
Observations	15,201	15,201	736	734

NOTES: Country and cohort fixed effects as well as country-specific quadratic trends in cohorts are included in the first two columns. ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 3: Decomposition - Micro and ADS Model

	Females		Males	
	Micro-model	ADS-model	Micro-model	ADS-model
Health-Education Gradient (HEG)	-0.017	-0.026	-0.012	-0.010
- behaviors (short-term)	-0.003	-0.004	-0.004	-0.003
- behaviors (long-term)	-0.005	-0.006	-0.007	-0.004
- residual (direct effect)	-0.012	-0.020	-0.010	-0.006
Mediating effect as fraction of HEG				
- SRME (short-term)	0.168	0.172	0.097	0.308
- LRME (long-term)	0.323	0.228	0.189	0.445

NOTES: Computations based on the estimates reported in Table 2.

Table 4: Compulsory schooling reforms in Europe

Country	Reform	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 5: Summary Statistics IV - Sample 10

Country	Self-rep poor health	Education	Compulsory Edu	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

Table 6: Health-Education Gradient - IV approach

	Sample 10		Sample 7	
	lin-trend	qu-trend	lin-trend	qu-trend
Females				
OLS	-0.024	-0.024	-0.025	-0.025
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
2SLS	-0.040	-0.064	-0.041	-0.085
	(0.024)*	(0.034)*	(0.035)	(0.032)***
ITT	-0.014	-0.017	-0.011	-0.023
	(0.008)*	(0.008)**	(0.009)	(0.008)***
First Stage	0.344	0.253	0.263	0.271
	(0.053)***	(0.058)***	(0.053)***	(0.058)***
IV-Probit	-0.042	-0.057	-0.041	-0.073
	(0.022)*	(0.025)**	(0.032)	(0.017)***
F-Stat (First Stage)	41.93	18.95	24.89	21.66
Observations	8,602	8,602	6,631	6,631
Males				
OLS	-0.017	-0.017	-0.017	-0.017
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
2SLS	-0.048	-0.054	-0.062	-0.064
	(0.029)*	(0.029)*	(0.029)**	(0.034)*
ITT	-0.016	-0.018	-0.020	-0.020
	(0.009)*	(0.008)**	(0.008)**	(0.010)**
First Stage	0.323	0.318	0.313	0.298
	(0.076)***	(0.078)***	(0.079)***	(0.082)***
IV-Probit	-0.047	-0.051	-0.056	-0.057
	(0.024)**	(0.022)**	(0.019)***	(0.022)***
F-Stat (First Stage)	17.87	16.62	15.66	13.07
Observations	7,358	7,358	5,663	5,663

NOTES: ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 7: Health-Education Gradient - IV and ADS compared

		ADS-model	
	IV-estimate	All countries	IV-sample
Females	-0.040	-0.026	-0.028
	(0.024)*	(0.005)***	(0.007)***
Males	-0.048	-0.010	-0.020
	(0.029)*	(0.005)*	(0.008)**

NOTES: ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

Table 8: Robustness - ADS Model

	ADS year pseudo-panel		ADS w/o ENG		ADS l-exp, w/o GRC	
	Red form	Dynamic HE	Red form	Dynamic HE	Red form	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)***
self-rep 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)
self-rep 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)**
serious 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 or 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: ***, ** and * indicate statistical significance at the 1-percent, 5-percent and 10-percent level.

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

	ADS-Model		IV (Sample 10, lin-trend)
	Reduced form	Dynamic HE	Reduced form
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		0.279	
exercise$_{t-1}$		(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		0.220	
exercise$_{t-1}$		(0.198)	
BMI$_{t-1}$		0.041	
		(0.016)*	
income$_t$		-0.004	
		(0.005)	
Early life			
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: ***, ** 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 (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).

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 (forthcoming), 'The causal effect of education on the body mass: Evidence from Europe', *Journal of Labor Economics* .

Brunello, Giorgio, Guglielmo Weber and Christoph Weiss (2012), Books are forever: early life opportunities, education and lifetime earnings 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.

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 (2010), The effect of education on adult health and mortality: Evidence from Britain, Working Paper 16013, National Bureau of Economic Research (NBER).

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.

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.

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 6015, Institute for the Study of Labor (IZA, Bonn).

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.

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

Juerges, Hendrik, Eberhard Kruk and Steffen Reinhold (2009), The effect of compulsory schooling on health: Evidence from biomarkers, Discussion Paper 183–2009, Mannheim Institute for the Economcis of Aging (MEA).

Kempter, 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.

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 (forthcoming), 'Heterogeneity in health responses and anchoring vignettes', *Empirical Economics* .

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.

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).

Zweifel, Peter and Friedrich Breyer (1997), *Health Economics*, Oxford University Press, Oxford.

Appendix A 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

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

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

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

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]

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

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

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

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

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

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

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

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

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

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

[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

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

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

Appendix B 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 B.1.

Table B.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

Appendix C Educational 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. (forthcoming) and Brunello et al. (2009) 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.