

The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence.

Titus J. Galama, Adriana Lleras-Muney, and Hans van Kippersluis

NBER Working Paper No. 24225

January 2018

JEL No. I26

ABSTRACT

Education is strongly associated with better health and longer lives. However, the extent to which education causes health and longevity is widely debated. We develop a human capital framework to structure the interpretation of the empirical evidence. We then review evidence on the causal effects of education on mortality and its two most common preventable causes: smoking and obesity. We focus attention on evidence from Randomized Controlled Trials, twin studies, and quasi-experiments. There is no convincing evidence of an effect of education on obesity, and the effects on smoking are only apparent when schooling reforms affect individuals' track or their peer group, but not when they simply increase the duration of schooling. An effect of education on mortality exists in some contexts but not in others, and seems to depend on (i) gender; (ii) the labor market returns to education; (iii) the quality of education; and (iv) whether education affects the quality of individuals' peers.

Titus J. Galama
Dornsife Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA 90089
galama@usc.edu

Hans van Kippersluis
Erasmus University Rotterdam
P.O. Box 1738
3000 DR Rotterdam
The Netherlands
hvankippersluis@ese.eur.nl

Adriana Lleras-Muney
Department of Economics
9373 Bunche Hall
UCLA
Los Angeles, CA 90095
and NBER
alleras@ECON.UCLA.EDU

1 Introduction

More educated individuals live longer, healthier lives. A large literature has documented substantial associations between education and mortality, health (self-reported health, obesity, etc.) and health behaviors (smoking, excessive drinking, exercise, preventive care use, etc.). Although the strength of these relationships varies, such associations have been observed in many countries and time periods. Moreover, the gap between the educated and less educated appears to be growing (Pappas et al. 1993; Meara et al. 2008). Such disparities in health are large. For example, in the United States, at age 25, those with more than a college degree can expect to live up to seven years longer than those without a college degree (Meara et al. 2008; Hummer and Hernandez, 2013). Substantial theoretical and empirical research is devoted to understanding the causes of these disparities.

The earliest theoretical contribution was made by Michael Grossman, who developed a theory in which health is a form of capital that individuals can produce (Grossman, 1972). Grossman's seminal theory laid the foundation for a substantial body of empirical and theoretical research on the determinants of health. It was also the first work to discuss the various ways in which education might affect health and longevity. In this review, we develop a theoretical framework based on this early model but inspired by recent advances in theory, to investigate the relationship between human capital, schooling, mortality, and health behaviors. Our theory updates the early Grossman model in some important aspects. First, health, skills, health behavior, schooling, and longevity are all treated as endogenous, and they are allowed to vary as a function of initial endowments, such as genes and parental characteristics. Second, inspired by the seminal work of Heckman and others (e.g., Heckman et al. 2006; Cunha and Heckman, 2007), we distinguish between skills developed in school and time spent in school. Finally, we explicitly model the role of laws and institutions that affect schooling – this allows us to make predictions about the effects of compulsory schooling and other education policies on health outcomes. Changes in such policies have been widely exploited in the empirical literature to estimate the causal effects of education on health.

We then review recent empirical work in this area. A very large and recent literature has investigated whether the association between education and health is causal or not. Does going to school increase longevity? Does it improve health behaviors? The growing availability of large data sets in combination with a revolution in empirical methods addressing causality has produced much new evidence on this old question. We cover papers written since 2005 that investigate how education affects i) mortality, and ii) smoking and obesity, the two most important determinants of preventable death and disease in the US and most developed countries (Mokdad et al. 2004). We restrict attention to papers assessing causality using randomized controlled trials, twin differences or quasi-experiments. The vast majority of empirical studies use data from the developed world, which should be kept in mind when interpreting results.

Our findings are not uniform. We touch upon some methodological differences across studies. But

our focus is more substantive. While discrepancies in findings across studies can result from differences or flaws in the empirical methodologies applied, we conclude that there is substantial evidence of genuine heterogeneity in the estimated effects of education. We focus on understanding this heterogeneity, in light of our theoretical model and of findings in this large literature. We do not conduct a formal meta-analysis here, but our observations nevertheless allow us to draw broad conclusions from the existing empirical evidence. We find no convincing evidence of an effect of education on obesity. There appear to be effects of education on smoking and mortality, but these are only observed in some contexts and not in others. Among others, the effects seem to differ by (i) gender; (ii) the labor market returns to education; (iii) the quality of education; and (iv) whether education affects the quality of individuals' peers. We discuss these findings extensively and end with some suggestions for future research.

2 A Theory of Schooling, Skills and Health

2.1 Theoretical formulation

The theory we discuss below draws substantially on Galama and Van Kippersluis (2015a; 2015b). Consider a model in which we decompose human capital into two components: health $H(t)$ and skill $\theta(t)$, where skill includes both cognitive and non-cognitive skills (Cunha and Heckman, 2007; Conti et al. 2010).¹ Although health cannot directly be bought, it can be produced by purchasing inputs such as food and medical care. Similarly, skills can be enhanced by investments, for example by spending time in school. Like Grossman's 1972 model, individuals make optimal lifetime decisions concerning skill and health inputs. These choices determine skills and health, and thus affect labor-market outcomes and consumption choices, which in turn determine health and mortality.

Individuals derive utility from consumption, health and not being in school, and make choices in order to maximize their lifetime utility

$$\mathbb{U} = \max_{Z(t)} \left\{ \int_0^S U[\cdot] e^{-\rho t} dt + \int_S^T U[\cdot] e^{-\rho t} dt \right\}, \quad (1)$$

where $Z(t)$ is the set of choice variables (to be discussed later), $t = 0$ corresponds to the age at which individuals start school,² S denotes years of schooling, T denotes the age of death, and ρ is a subjective discount factor. Individual preferences are represented by a concave utility function $U[\cdot]$ that increases with consumption goods and services $X_C(t)$, health $H(t)$, and decreases with being in school $c_S(t)$.

¹We assume that health and skills are orthogonal traits. There are of course exceptions to this separation, e.g., dementia and other health conditions that impair skills, which we ignore here for simplicity.

²We assume here that parents make the decision about when to start school but that individuals decide when to end it. We ignore issues regarding possible conflicts between parents and children, and model children as fully rational decision makers.

Individuals have access to assets and labor income, and must use them over their lifetime to pay for the goods and services they consume. In addition, they must also spend time producing skills and improving their health. The intertemporal budget constraint for assets $A(t)$ is given by

$$\begin{aligned} \frac{\partial A}{\partial t} = & rA(t) + Y(t) - [1 - \lambda_S(t)]p_S(t) - p_\theta(t)X_\theta(t) \\ & - p_H(t)X_H(t) - p_C(t)X_C(t), \end{aligned} \quad 0 \leq t < S \quad (2)$$

$$\begin{aligned} \frac{\partial A}{\partial t} = & rA(t) + Y(t) - f^*(t)I[t - \underline{S}] - p_\theta(t)X_\theta(t) \\ & - p_H(t)X_H(t) - p_C(t)X_C(t). \end{aligned} \quad S \leq t < T \quad (3)$$

Assets $A(t)$ (equations 2 and 3) provide a return r (the rate of return on capital), increase with income $Y(t)$, and decrease with purchases of skill inputs $X_\theta(t)$, health inputs $X_H(t)$, and consumption goods $X_C(t)$, bought at prices $p_\theta(t)$, $p_H(t)$, and $p_C(t)$, respectively. During the schooling period (up to S) individuals pay a tuition fee $p_S(t)$, but receive a state (or parental) transfer that is a fraction $\lambda_S(t)$ of tuition, where both are exogenously determined. Thus, $[1 - \lambda_S(t)]p_S(t)$ is the effective net-of-subsidy tuition rate. Individuals incur a fine $f^*(t)I[t - \underline{S}]$ if they enter the labor market before the minimum school-leaving age \underline{S} , where $I[t - x]$ is the indicator function, which is 1 for $0 \leq t \leq x$ and zero otherwise.³

Income $Y(t) \equiv Y[t, S, \theta(t), H(t); \underline{S}]$ consists of the product of the wage rate $w[t, S, \theta(t); \underline{S}]$ and the time spent working $\tau_w[H(t)]$

$$Y[t, S, \theta(t), H(t); \underline{S}] = w[t, \theta(t); \underline{S}] \times \tau_w[H(t)], \quad 0 \leq t < S \quad (4)$$

$$Y[t, S, \theta(t), H(t); \underline{S}] = w[t, S, \theta(t); \underline{S}] \times \tau_w[H(t)]. \quad S \leq t < T \quad (5)$$

We further implicitly assume that skills $\theta(t)$ and the schooling duration S (largely) determine the wage rate, while health capital $H(t)$ (largely) determines the time spent working. Wages depend explicitly on the exogenous minimum school-leaving age \underline{S} . For example, labor-laws might impose a fine on those who employ individuals younger than \underline{S} so that wages are lower below that age if the law is enforced.

When an individual is still in school (equation 4), years of schooling is measured by t (i.e. the wage rate is not a function of optimal years of schooling S). After schooling has been completed (equation 5), the wage rate $w(t)$ is an (increasing) function of years of completed schooling S and the individual's level of skill $\theta(t)$.⁴ While schooling S is often used as a proxy for skills $\theta(t)$, there are reasons to believe that schooling S may capture benefits beyond skill formation. For example, the effect of schooling or college

³A useful interpretation of the fine $f^*(t)$ is as a probability of getting caught and having to pay a cost when the individual is not in school but should have been ($S < \underline{S}$). For this reason it is modelled not as a one-off cost but as a cost that operates for as long as the individual has not reached the minimum schooling age. Historically fines were the way in which compulsory schooling laws were enforced.

⁴In our formulation, we ignore possible sheepskin (or diploma) effects of completing a certain level of education.

graduation S on wages may reflect a potential signaling effect as skills, perseverance, and work ethic are difficult to observe by the employer (Arrow, 1973; Spence, 1973; Lang and Kropp, 1986; Bedard, 2001).

The time constraint during the schooling and working phases of life are given by

$$\Omega = \tau_w(t) + \tau_S(t) + \tau_H(t) + s[H(t)], \quad 0 \leq t < S \quad (6)$$

$$\Omega = \tau_w(t) + \tau_\theta(t) + \tau_H(t) + s[H(t)], \quad S \leq t < T \quad (7)$$

where

$$\tau_S(t) = \tau_\theta(t) + \tau_L(t). \quad (8)$$

The total available time Ω (in let's say a day) is divided between work $\tau_w(t)$, time spent in school $\tau_S(t)$ or devoted to skill investment $\tau_\theta(t)$, time investments in health $\tau_H(t)$, and time lost due to illness $s[H(t)]$ (assumed to be a decreasing function of health). A fixed amount of time $\tau_S(t)$ spent in school can be used for skill investments $\tau_\theta(t)$ but also unproductively $\tau_L(t)$ (see 8). The “time lost” $\tau_L(t)$ is the difference between compulsory hours spent in school and the desired hours of skill investment, $\tau_S(t) - \tau_\theta(t)$. While unproductive in terms of skill formation, we allow $\tau_L(t)$ to decrease the disutility of schooling, $c_S[t; \tau_L(t)]$. Intuitively, some children experience low returns and high costs of investing in their skill. They may prefer “staring out of the window” $\tau_L(t)$ over paying attention $\tau_\theta(t)$. This distinction is important since it implies that skill does not automatically increase with longer (compulsory) schooling duration. As we will see, these intuitive constructs allow us to better understand heterogeneity in responses to policy changes aimed at encouraging schooling.

The levels (stocks) of skills $\theta(t)$ and health $H(t)$ evolve over time. Individuals can increase them through investments (using production functions $g_\theta(t)$ and $g_H(t)$) but they also depreciate, according to the following equations

$$\frac{\partial \theta(t)}{\partial t} = g_\theta [t, X_\theta(t), \tau_\theta(t), \theta(t), H(t); \xi(t)] - d_\theta [t, \theta(t); \xi(t)], \quad (9)$$

$$\frac{\partial H(t)}{\partial t} = g_H [t, X_H(t), \tau_H(t), \theta(t), H(t); \xi(t)] - d_H [t, H(t), X_C(t); \xi(t)]. \quad (10)$$

While in school, individuals invest in skill capital $g_\theta(t)$ through outlays $X_\theta(t)$ (e.g., books) and time investments $\tau_\theta(t)$ (e.g., paying attention in class, devoting effort to study).⁵ While working, individuals learn on the job, and devote goods and services and time to learning outside of work $X_\theta(t)$, $\tau_\theta(t)$ (Becker, 1964). Individuals invest in health $g_H(t)$ (equation 10) through expenditures $X_H(t)$ and time investments $\tau_H(t)$ (e.g., medical care, flu shots, exercise).

⁵Further, while in school, individuals may be provided with healthy nutritious foods, or simply be kept of the street which may keep them out of trouble (incarceration effect; Weiss, 1995; Lochner and Moretti, 2004; Black et al. 2008; Lochner, 2011). This would suggest that skills and health could also be a function of $\tau_S(t)$ as well as $\tau_\theta(t)$.

The efficiency of the production functions $g_\theta(t)$, $g_H(t)$ is assumed to be a function of the stocks of skill $\theta(t)$ and health $H(t)$. This allows us to model self-productivity, where skills produced at one stage augment skills at later stages, and dynamic complementarity, where skills produced at one stage raise the productivity of investment at later stages (Cunha and Heckman, 2007). At a deeper level, the production functions also depend on $\xi(t)$, which are time-invariant and predetermined endowments, including genes and family background, and time-varying exogenous characteristics of the schooling and work environment (for instance one's peers, or the quality of one's teachers).

The stock of skill deteriorates at the rate $d_\theta(t)$, assumed to be a function of age t , the level of skill $\theta(t)$ and endowments $\xi(t)$. The stock of health deteriorates at the rate $d_H(t)$, assumed to be a function of age t , the level of health $H(t)$, consumption $X_C(t)$, and endowments $\xi(t)$.⁶ Consumption goods and services can be healthy (e.g., consumption of fruits / vegetables) or unhealthy (e.g., smoking, consumption of fatty / sugary foods). Healthy consumption provides utility and slows down health depreciation, $\partial d_H / \partial X_C \leq 0$ as in Case and Deaton (2005) and Galama and Van Kippersluis (2015b). Unhealthy consumption also provides utility but instead increases health depreciation, $\partial d_H / \partial X_C > 0$.

We assume individuals start life with a given level of health H_0 , assets A_0 and skills θ_0 , which may be influenced by genetic endowments and parental investments. Following the health-capital literature (Grossman, 1972) life cannot be sustained below a certain minimum health: $H(T) = H_{\min}$. Also, we assume individuals spend all of their assets during their lives, i.e. $A(T) = 0$. By contrast, following the human-capital literature, the stock of skill at the end of life can be chosen freely, i.e. $\theta(T)$ is unconstrained (e.g., Ben-Porath, 1967).

Thus individuals maximize life-time utility (1) subject to $Z(t)$, the set of control variables: $Z(t) \equiv \{S, T, X_\theta, \tau_\theta, X_H, \tau_H, X_C\}$. The Lagrangian (Seierstad and Sydsaeter, 1987; Caputo, 2005) of this problem is:

$$\begin{aligned} \mathfrak{L}(t) = & U[\cdot]e^{-\rho t} + q_\theta(t)\frac{\partial \theta}{\partial t} + q_H(t)\frac{\partial H}{\partial t} + q_A(t)\frac{\partial A}{\partial t} \\ & + \lambda_L(t)[\tau_S(t) - \tau_\theta(t) - \tau_L(t)] + \lambda_{H\min}(t)[H(t) - H_{\min}(t)], \end{aligned} \quad (11)$$

where $q_\theta(t)$, $q_H(t)$, and $q_A(t)$ are the co-state variables associated with, respectively, the dynamic equation (9) for skill capital $\theta(t)$, (10) for health $H(t)$, and (2, 3) for assets $A(t)$. These co-state variables can be interpreted as the marginal values (or shadow prices) of the relevant capital stock (see Galama and Van Kippersluis, 2015a). For example, $q_\theta(t)$ is the marginal value in terms of the remaining life-time utility (from t onwards) of an additional increment of skill capital $\theta(t)$. $\lambda_L(t)$ is the Lagrange multiplier associated with restriction (8) and $\lambda_{H\min}(t)$ is the multiplier associated with the restriction $H(t) \geq H_{\min}(t)$.

⁶An important criticism of the Grossman model, made by Dalgaard and Strulik (2014), is that health decline is arguably smaller for people in better health. Our health production function is a flexible function of health and health investment, and encompasses various specifications, including the original Grossman model as well as Dalgaard and Strulik's health deficit model.

2.2 Comparison with the literature

Our theoretical formulation builds upon and extends workhorse human-capital models. In Ben-Porath's (1967) model, individuals invest in human capital throughout life to increase their productivity, but human capital is one-dimensional (skill), a schooling period does not feature in the model, and longevity is exogenous. Card (2001) models a schooling period similar to ours, but does not distinguish between schooling (spending time in school) and skill formation, and does not include investments in health and longevity.

The seminal health-capital model of Grossman (1972) treats health as a capital stock that individuals invest in, but assumes longevity T and education are exogenously given. Ehrlich and Chuma (1990), Galama (2015) and Galama and Van Kippersluis (2015b) model both health and longevity as endogenous, but treat education as exogenously given. Becker (2007) presents a two-period model where individuals can invest in skill, and health, but has no schooling period and health investments solely affect mortality risk. A closely related model is presented in Strulik (2016), where both health and human capital are endogenously determined, and in which individuals accrue, so-called, health deficits as opposed to facing health depreciation.

As in Galama and Van Kippersluis (2015a) and Strulik (2016), we jointly model health $H(t)$, skill $\theta(t)$, optimal schooling S and optimal longevity T . Hence, we distinguish explicitly between schooling duration S and skill $\theta(t)$. Schooling S is the choice to spend a certain number of years in school, while skill is a capital stock subject to investment and depreciation. In contrast to Strulik (2016), in which the number of years spent in school is the only endogenous input into skill capital, we allow for (i) other inputs into skill capital besides schooling duration; and (ii) the possibility that the stock of skill does not increase with schooling duration if those years are spent unproductively.

While skill and schooling are separate concepts, the two are clearly connected since schooling is a period of life that is characterized by a low opportunity cost of time, encouraging time investments in skill. This is because 1) skills are low early in life and hence are wages, and 2) schooling imposes a fixed amount of time $\tau_S(t)$ that cannot be devoted to work. We argue that both schooling and skills determine wages, as skills (at least initially) are hard to observe for employers and so schooling may serve as a signal for skills. However, only skills influence the production function of skill formation and health, not the duration of schooling per se.

Importantly, the model of section 2.1 explicitly adds to Galama and Van Kippersluis (2015a) the role of institutions and laws that regulate prices, wages, whether individuals are legally required to be in school and until what age, and the penalties and enforcement associated with such laws. This allows us to derive theoretical predictions about the institutional reforms that are often exploited as quasi-experiments in empirical research, an important focus of the current review. However, as is true for most models that investigate this question, ours does not incorporate general equilibrium effects.

2.3 Optimal schooling, consumption, and longevity

Schooling: Assuming that the (dis)utility of schooling is additively separable from consumption and health, $U[\cdot] = U[X_C(t), H(t)] - c_S(t, \tau_L(t))I[t - S]$, and making some further simplifying assumptions detailed in the Appendix, we obtain a condition for the optimal years of schooling S ,

$$\begin{aligned} & \left[Y(S_+) - Y(S_-) \right] + q_A(S)^{-1} c_S[S, \tau_L(S)] e^{-\rho S} + [1 - \lambda_S(S_-)] p_S(S_-) \\ & + p_\theta(S) [X_\theta(S_-) - X_\theta(S_+)] = \int_S^T \frac{\partial Y(t)}{\partial S} dt + f^*(S) I(S - \underline{S}) + q_{\theta/a}(S) [g_\theta(S_-) - g_\theta(S_+)], \quad (12) \end{aligned}$$

where $q_{\theta/a}(t) \equiv q_\theta(t)/q_A(t)$, S_- indicates the limit in which S is approached from below, and S_+ when approached from above, and we have replaced S_- and S_+ with S for functions that are continuous in S . This condition states that individuals will join the labor market at the age S , the age at which the (net) benefits of work exceed the (net) benefits of staying in school. The left-hand side (LHS) represents the benefits of entering the labor market, while the right-hand side (RHS) represents the benefits of staying in school. The benefits of entering the labor market consist of (i) additional labor income $Y(S_+) - Y(S_-)$; individuals have more time to work, since they no longer have to spend a fixed amount of time in school $\tau_S(t)$;⁷ (ii) they no longer suffer disutility from schooling $c_S[S, \tau_L(S)]$;⁸ (iii) they do not have to pay (subsidized) tuition $[1 - \lambda_S(S_-)] p_S(S_-)$; and (iv) they incur fewer monetary costs related to skill formation $p_\theta(S) [X_\theta(S_-) - X_\theta(S_+)]$.⁹

The benefits of staying in school are (i) increased *future* earnings, (ii) not incurring a fine $f^*(S)$ before the minimum school-leaving age \underline{S} , and (iii) the value of additional skill investment while in school $q_{\theta/a}(S) [g_\theta(S_-) - g_\theta(S_+)]$.

Investment: The first order conditions for investment in skill and health are given by

$$q_{\theta/a}(t) = \pi_\theta(t), \quad (13)$$

and

$$q_{h/a}(t) = \pi_H(t), \quad (14)$$

where $q_{\theta/a}(t)$ is the relative marginal value of skill, $q_{\theta/a}(t) \equiv q_\theta(t)/q_A(t)$, $q_{h/a}(t)$ is the relative marginal value of health, $q_{h/a}(t) \equiv q_H(t)/q_A(t)$, $\pi_\theta(t)$ is the marginal cost of skill investment, and $\pi_H(t)$ is the

⁷Potentially the wage rate changes too right after dropping out of school, if the minimum school-leaving age coincides with labor laws.

⁸Of course, many people enjoy being in school and for them this would represent a cost rather than a benefit of working.

⁹Theory predicts skill investment drops after graduation as the opportunity cost of time investment is lower during schooling since individuals have to spend a fixed amount of time $\tau_S(t)$ in school. If there are complementarities between goods / services $X_\theta(t)$ and time inputs $\tau_\theta(t)$ then investment goods will also be lower.

marginal cost of health investment (see 23, 24, 26 and 27 for explicit expressions). The marginal cost of skill investment $\pi_\theta(t)$ and the marginal cost of health investment $\pi_H(t)$ increase in the level of investment in skill and health (due to diminishing returns to scale in investment), increase in prices $p_\theta(t)$, $p_H(t)$, and increase in the wage rate $w(t)$ (opportunity cost of time inputs, see Galama and Van Kippersluis [2015a,2015b] for details). Leaving prices and wages aside, a higher relative marginal value of health $q_{h/a}(t)$ or skill $q_{\theta/a}(t)$ implies higher investment in health or skill (see 13 and 14).

Consumption: The first-order condition for consumption is given by

$$\frac{1}{q_A(t)} \frac{\partial U}{\partial X_C} e^{-\rho t} = p_C(t) + q_{h/a}(t) \frac{\partial d_H}{\partial X_C}. \quad (15)$$

Consider an unhealthy consumption good (e.g., cigarettes) that increases the health deterioration rate $\partial d_H / \partial X_C > 0$. The marginal benefits (LHS of 15) consist of the discounted marginal utility of the consumption good, while the marginal costs (RHS) consists of the monetary cost $p_C(t)$ and the health cost $q_{h/a}(t)(\partial d_H / \partial X_C)$. The health cost is the product of the relative marginal value of health, $q_{h/a}(t)$, and the amount of health lost due to unhealthy consumption $\partial d_H / \partial X_C$.

Longevity: Optimal longevity is determined by the point where there are no longer net benefits of staying alive (this is the point where the Lagrangian is zero; see 28 in the Appendix),

$$\frac{\mathfrak{L}(T)}{q_A(T)} = \frac{1}{q_A(T)} U[X_C(T), H(T)] e^{-\rho T} + q_{h/a}(T) \frac{\partial H}{\partial t} + \frac{\partial A}{\partial t} = 0. \quad (16)$$

The relative marginal value of skill $q_{\theta/a}(t)$ approaches zero near the end of life as individuals can choose the terminal level of skill optimally. They then choose its level such that small increments in skill no longer have value at the end of life: $q_{\theta/a}(T) = 0$.¹⁰ Thus, the second term in (11) vanishes. Health declines near the end of life as it approaches the minimum health level H_{\min} , and assets eventually decline as declining health reduces earnings and increases medical expenditure, and because terminal assets are constrained to zero. With declining health and declining assets, the second and third terms in (16) compete with the additional utility provided by adding an increment of longevity (first term in 16). The optimal point of death T occurs when the utility of consumption $X_C(T)$ and health $H(T)$ no longer outweighs the increasing costs of maintaining health.

¹⁰This is not true for wealth $A(T)$ and health $H(T)$ as individuals cannot choose their terminal levels optimally. Hence, the transversality conditions for the co-state equations are: $q_\theta(T) = 0$, while $q_H(T) \geq 0$ and $q_A(T) \geq 0$.

2.4 Implications for empirical analyses

2.4.1 Supply side reforms and compliers

Equation (12) provides a useful reference to understand the effect of model parameters on the optimal schooling decision, and to understand heterogeneity in schooling choices. In particular, the expression suggests there are various ways to encourage additional schooling. First, the government or parents can fund schooling $\lambda_S(t)p_S(t)$ (e.g., financial aid, conditional on being in school). Second, labor-laws might stipulate that it is illegal to employ individuals younger than age \underline{S} so that earnings $Y(S)$ are effectively zero below that age if the law is enforced. Third, in addition to the schooling subsidy $\lambda_S(t)p_S(t)$, the government may subsidize skill investment $p_\theta(t)X_\theta(t)$ (e.g., books, access to libraries, additional classes, uniforms, computers). Fourth, the government may set a minimum to hours of instruction or term length, operating through $\tau_S(t)$. Last, governments often set a minimum school-leaving age \underline{S} , and individuals may incur a fine $f^*(S)$ if they drop out of school before this age.

All supply side reforms work to lower the costs or to increase the benefits of schooling (equation 12), and can often be considered as plausibly exogenous for a given individual. The most popular supply side reform is undoubtedly changes in minimum school-leaving ages. In this section we discuss its effects.

Equation (12) highlights that increasing the minimum school-leaving age \underline{S} has two effects: it increases the period over which individuals are exposed to a possible fine $f^*(S)$ (second term on the RHS of 12), and potentially raises the legal age at which one can earn wages, and hence income $Y(S_+) - Y(S_-)$ (first term on the LHS of 12). These effects alter the net benefits for each individual of staying in school longer. Taking the derivative of equation (12) with respect to \underline{S} , and assuming that individuals value skill $q_{\theta/a}(t)$ more than health $q_{h/a}(t)$ early in life, then (see 32 in the Appendix) the comparative dynamic effect of the minimum school-leaving age on schooling can be summarized as:

$$\frac{dS}{d\underline{S}} = f \left[\underbrace{\frac{\partial q_A(S)}{\partial \underline{S}}}_{+/-}; \underbrace{\frac{\partial q_{\theta/a}(S)}{\partial \underline{S}}}_{+}; \underbrace{\frac{\partial \theta(S)}{\partial \underline{S}}}_{+}; \underbrace{\frac{\partial H(S)}{\partial \underline{S}}}_{+}; \underbrace{\frac{\partial T}{\partial \underline{S}}}_{+}; \underbrace{\beta_{\underline{S}}}_{+/-} \right], \quad (17)$$

where + indicates a positive effect, - indicates a negative effect, +/- denotes an ambiguous effect, and $\beta_{\underline{S}}$ is defined in (32). The terms on the RHS represent various effects of raising the minimum school-leaving age on the optimal schooling decision: wealth effects (1st term),¹¹ increased relative marginal value of skill (2nd term), a higher stock of skill (3rd term), better health (4th term), a longer life (5th term), and effects operating through labor laws (6th term).¹²

¹¹The sign of this term depends on whether time in school provides utility or disutility, $c_S(t)$.

¹²The very last term represents the possibility that wages depend explicitly on the exogenous minimum school-leaving age \underline{S} : labor-laws might impose a fine on those who employ individuals younger than \underline{S} so that wages are lower below that age if the law is

Expression (17) captures a characteristic of the comparative dynamic analyses that many effects operate through the marginal values of wealth, skill and health. In what follows, it is useful to adopt the usual assumption of diminishing returns to wealth, skill and health. This implies that the marginal value of wealth $q_A(t)$ is smaller and the relative marginal values of skill $q_{\theta/a}(t)$ and of health $q_{h/a}(t)$ are larger for those with greater wealth. Thus, if raising the minimum schooling age affects lifetime wealth, then these effects are operational.

For some individuals, raising the minimum school-leaving age \underline{S} does not affect their schooling decision S , since they would have chosen to complete far more or far less schooling anyway. Consider an individual who starts life endowed with skill and ability to learn, excels in school and is in an environment where her skill formation is enabled and valued. She is efficient in absorbing and internalizing the information provided in class, has the necessary materials at her disposal, the quality of the teaching and the school and home environment are conducive to skill formation, and she potentially enjoys being in school (utility rather than disutility $c_S(t)$). As a result of these many factors, her skills $\theta(t)$ grow: the process of skill-formation is “productive” (raises skill) and “efficient” (uses the inputs time and goods / services effectively). Further, she is healthy and expects to live a long life T , the institutional environment (few barriers, such as discrimination, corruption, crime, etc.) and the labor market (high wages, high employment) enable and value the use of her skills (high returns to skill, $\int_S^T (\partial Y / \partial S) dt$). She would not be influenced by the minimum school-leaving age \underline{S} since she would optimally choose a far greater number of years of schooling $S > \underline{S}$. Raising the minimum schooling age \underline{S} would not have an effect on her choice S . Effectively, since there is no change in schooling, and thereby in wealth, all terms in equation (17) are zero (no effects on the marginal values of wealth, skill and health).¹³

Another way of understanding this result is by noting that condition (12) is not the usual equilibrium equation, but a switching equation. The optimal level of schooling S occurs when the net marginal benefits of staying in school equal, for the first time, the net marginal costs. In this case, the net marginal benefits of staying in school outweigh the net marginal costs for all ages before and much past the minimum school leaving age \underline{S} . Increasing the benefits of schooling for ages well before the optimal schooling age S for this individual does not affect her choice since marginal benefits already exceeded marginal costs for those ages.

The second type of individual who is unlikely to alter her schooling decision is one who is endowed with few skills and less ability to learn to begin with, which renders investments in skill rather unproductive. Further, she strongly dislikes going to school, and in order to decrease the disutility of schooling, she chooses to spend time unproductively $\tau_L(t)$ rather than making the effort to acquire skills $\tau_\theta(t)$. Before the change in the minimum school-leaving age, she already dropped out well before the legal age. Increasing the net

enforced. This would represent a discontinuity in the wage rate at \underline{S} . The effect would operate only on those for whom the optimal schooling age S and the minimum school-leaving age \underline{S} coincide, i.e. it affects those individuals who are potential compliers.

¹³If the completion of schooling by the compliers (those at the margin) increases competition in the labor-force then “economy” wide (general equilibrium) effects may reduce the returns to her schooling, resulting in a negative wealth effect.

marginal benefits of staying in school at ages well after the age she dropped optimally out of school S does not affect her choice (net marginal costs already exceeded net marginal benefits well before the old minimum age).¹⁴

Now consider the marginal individual. She would prefer to enter the labor market to make more money, but at the same time does not want to incur a fine for dropping out early and recognizes that investing in skill increases her future earnings. She optimally decides to drop out of school exactly at \underline{S} , the minimum school leaving age. Now, if the government raises the minimum school leaving age, \underline{S} , this individual has to re-optimize. For this individual, the net benefits of staying in school longer were negative under the old regime. Hence, the fine has to be set such that the individual will comply with the reform, and chooses to stay in school until the new school-leaving age. This group of individuals at the margin forms the group of ‘compliers’.

Equation (17) illustrates the various factors entering the decision whether or not to comply. First, the costs of dropping out at the previous optimal S (i.e. before the new \underline{S}) increase since she would incur a fine and forego a potential subsidy. This provides incentives to comply and increase schooling to \underline{S} , and is reinforced by the potential wealth effects associated with an extra year in school if additional schooling increases skills and earnings (terms 1 to 3 on the RHS). Second, she might invest more in skill because of the lower opportunity cost of time as she needs to spend a given amount of time $\tau_S(t)$ in school (term 2). If, for the marginal individual, the institutions and labor-market conditions are such that the additional schooling increases life-time skills, and skills lead to higher earnings, then the resulting greater investment in health, healthier behavior and longer lives may provide additional benefits of staying in school longer (terms 4 and 5). Finally, labor laws may stipulate that employment is not gainful before the new minimum school-leaving age (term 6). Thus, the complier increases (by definition) her level of schooling, but only if additional schooling increases skills and earnings do we expect to observe improvements in health, health behaviors and longevity. The magnitude of these health effects among compliers depends on institutions and on economic and social conditions that may vary by cohort, gender, level of a country’s development, etc., as we discuss below. Also note that it is important that individuals correctly perceive these benefits, and are not credit-constrained, in order for them to change their decisions. The theoretical model assumes perfect information and perfect credit markets, but in practice both assumptions may not hold for certain subgroups of the population. We discuss these possible explanations in section 4.

¹⁴But she is not unaffected. When the minimum school-leaving age is increased, and the individual does not increase her schooling duration, she now faces an even longer duration over which she pays a fine and / or cannot earn wages. This negatively affects her lifetime wealth, and hence her skill and health production (terms 1-4 on the RHS of 17). The model also allows for situations in which this group of non-compliers minimally increase or decrease their schooling duration, with ambiguous effects on lifetime wealth, skill, and health.

2.4.2 Effects of compulsory schooling on consumption and mortality

Consumption The effect of a change in the minimum-school leaving age \underline{S} on (unhealthy) consumption can be summarized by:

$$\frac{\partial X_C(t)}{\partial \underline{S}} = f \left[\underbrace{\beta_{1,C} \frac{\partial q_A(t)}{\partial \underline{S}}}_{+}; \underbrace{\beta_{2,C} \frac{\partial q_{h/a}(t)}{\partial \underline{S}}}_{-}; \underbrace{\beta_{3,C} \frac{\partial H(t)}{\partial \underline{S}}}_{+/-} \right] \quad (18)$$

where $\beta_{1,C}$ to $\beta_{3,C}$ are defined in (33).

The effect of the minimum school-leaving age \underline{S} on unhealthy consumption is ambiguous. The derivative can be decomposed into two main terms.¹⁵ The first term represents a wealth effect, and is positive among compliers: increased schooling leads to an increase in wealth enabling more unhealthy consumption ($\partial q_A(S)/\partial \underline{S} < 0$ and $\beta_{1,C} < 0$).

Yet, an increase in schooling, and the associated wealth effect, also leads to a higher marginal value of health relative to wealth $q_{h/a}(t)$ (second term on the RHS). This is quite intuitive (see also Hall and Jones, 2007): due to diminishing marginal utility of consumption, richer people eventually start caring less about consumption and more about other goods, such as health, since health extends life (adding additional periods of utility rather than marginal improvements in utility from higher consumption). A higher relative marginal value of health increases the health cost of unhealthy consumption [$q_{h/a}(t)\partial d/\partial X_C$], and therefore reduces the demand for unhealthy consumption. This health cost effect competes with the wealth effect, and the net effect of the minimum-school leaving age \underline{S} on unhealthy consumption $X_C(t)$ is ambiguous.

We can say a bit more on the competition between the wealth effect and the health cost effect. The health cost increases in the severity of its impact on health, $\partial d_H/\partial X_C$ (the degree of “unhealthiness” of the consumption good). This suggests that for moderately unhealthy goods the direct wealth effect dominates, while for severely unhealthy goods the indirect wealth effect dominates (Van Kippersluis and Galama, 2014; Galama and Van Kippersluis 2015b).

In sum, a rise in the minimum school-leaving age may impact unhealthy behaviors such as smoking and poor diet through wealth effects (schooling leads to higher lifetime earnings enabling more consumption) and health-cost effects (schooling increases the relative marginal value of health through better job prospects and higher lifetime wealth, thereby reducing demand for unhealthy consumption). Because the health cost increases in the degree of unhealthiness of the good, we expect the wealth effect to dominate among moderately unhealthy goods, and the health cost effect to dominate among severely unhealthy goods. Note that the results depend on the assumption that individuals are perfectly informed about the precise health

¹⁵The third term on the RHS of (18) shows that an increase in the minimum school-leaving age may impact health at later ages, which, depending on whether unhealthy consumption and health are complements or substitutes in utility and in the rate of depreciation, may reinforce either the wealth or health-cost effect (see 18). This term is therefore difficult to sign.

consequences of consumption goods $\partial d/\partial X_C$ – if people are unaware of the health effects of unhealthy behavior, the term capturing its effect in the optimization would vanish and the wealth effect would dominate, in which case more schooling would lead to more unhealthy behavior. This could, for example, explain why higher educated individuals were *more* likely to smoke before the 1964 Surgeon General’s report warned about its effects.

Longevity The condition for the effect of the minimum school-leaving age \underline{S} on optimal length of life T can be written as (see 37 and 40 in the Appendix)

$$\frac{\partial T}{\partial \underline{S}} = f \left[\underbrace{\beta_{1,T} \frac{\partial q_A(T)}{\partial \underline{S}}}_{+/-}; \underbrace{\beta_{2,T} \frac{\partial \theta}{\partial \underline{S}}}_{+}; \underbrace{\beta_{3,T} \frac{\partial q_{h/a}(T)}{\partial \underline{S}}}_{+} \right] \quad (19)$$

where $\beta_{1,T}$ to $\beta_{3,T}$ are defined in (40).

Conditions (19) and (40) illustrate the various ways through which a rise in the minimum school-leaving age \underline{S} may affect longevity. The first term on the RHS is the wealth effect, among compliers, if extra schooling boosts lifetime wealth, i.e. if schooling enhances skills that are valued in the labor market and because of signalling. Individuals with more resources can afford to devote resources to increasing life expectancy (first term on the RHS, $\beta_{1,T} < 0$ and $\partial q_A(T)/\partial \underline{S} < 0$).¹⁶ However, there also exist scenarios in which wealthier individuals would choose to spend more resources on unhealthy consumption, and live *shorter* lives as a result. Hence the ambiguous sign. The second term on the RHS of (19 and 40) shows that, to the extent that an increase in the minimum school-leaving age boosts skills among compliers, these additional skills improve skill and health production, and increase earnings, all of which enable life extension through greater lifetime resources and better health (2nd term on the RHS, $\beta_{2,T} > 0$ and $\partial \theta/\partial \underline{S} > 0$). Finally, through diminishing returns to wealth and skill, and/or because of improved health knowledge, a higher minimum school-leaving age may lead individuals to value health relatively more compared with wealth, invest more in health (see 14), and live longer as a result (third term on the RHS, $\beta_{3,T} \partial q_{h/a}(T)/\partial \underline{S}|_{T,S} > 0$).

In sum, a rise in the minimum school-leaving age may impact longevity through wealth effects, improved skill formation, and a higher relative marginal value of health, improving health behavior. The magnitude of these effects, and in the case of the wealth effect also its sign, depends on the relative importance of terms in equation (19). Studying these terms more thoroughly, using equation (40) in the Appendix, reveals that the effect of a raise of the minimum school-leaving age on longevity among compliers will be stronger when (i) the labor market returns to schooling are high, (ii) the quality of education and the motivation of students

¹⁶For the never-takers who drop out of school before the minimum school-leaving age, the wealth effect is negative. For them this may lead to poorer health and lower life expectancy.

is such that additional schooling boosts skill formation, and (iii) the non-monetary returns to this additional skill are large (e.g., skills improving skill formation and skills improving health production). We will discuss whether the empirical findings are consistent with these predictions in sections 4.1 and 4.2. Obviously, the magnitude and importance of these terms varies across time periods and settings, and our theory suggests there will be cases where a rise in the minimum school-leaving age does not improve skills, increase wealth, or improve health outcomes (see section 4.3). Finally, there may be other channels through which minimum school-leaving laws affect health outcomes that are not captured in our model. We discuss those in sections 4.4 and 4.5.

3 Empirical evidence

In the following review of the literature we focus on the effects of measures of skill and schooling on mortality, smoking and obesity. Mortality is an unambiguous and precise measure of health that captures circumstances throughout the lifetime. For the living, a large number of health measures are available but there is no unique health indicator, aside from self-reported health status. Because self-reported health status is a subjective measure, we instead focus on the 1st and 2nd leading causes of preventable disease and death in the United States: smoking and obesity (Mokdad et al. 2004). To compare across studies, we focus on whether or not an individual smokes currently (a binary indicator), and whether their Body-Mass Index (BMI) exceeds the threshold for obesity ($BMI \geq 30$). Both obesity and smoking are unambiguous measures of bad health.¹⁷

We organize the sections by methodology. We first discuss Randomized Controlled Trials (RCTs), arguably the gold standard for establishing causal effects (Imbens 2010). Unfortunately, there are only a few RCTs for childhood education, and the samples are small and not representative. For this reason, we rely mostly on quasi-experimental methods, which seek to replicate experimental conditions (sometimes also referred to as natural experiments). We separately discuss (i) twin studies, where within-twin-pair differences in education are related to within-twin-pair differences in health outcomes, implicitly accounting for all genetic and family characteristics shared by twin pairs; and (ii) quasi-experiments, where instrumental variables (IV) or Regression Discontinuity Designs (RDD) are used to estimate a treatment effect of education on health outcomes.

There are alternative approaches to assess causality. In the absence of random assignment, one can use observational data to infer causal effects based on a number of econometric methods. The control-function

¹⁷Although many studies report effects of education on BMI or overweight (BMI between 25 and 30) these results are difficult to interpret. Changes in BMI within normal ranges (18.5-25) are not strongly associated with poor health and mortality. Similarly, overweight individuals appear to have lower mortality than individuals in the normal BMI range (e.g., Flegal et al. 2013), also making this an ambiguous measure of health. Obesity, on the other hand, is clearly associated with increased morbidity and mortality (ibid).

approach to solving the omitted variable bias problem allows researchers to control for essential observable and unobservable variables, often using data from a specific cohort (e.g., Conti et al. 2010; Savelyev, 2014; Bijwaard et al. 2015). A limitation of the control-function approach is that it relies on generalizations of the conditional independence assumption. We do not include papers that follow this approach in our review since they vary widely in terms of methodology and are therefore hard to classify. Finally, the literature on Ordinary Least Squares (OLS) estimations is large and diverse in terms of included control variables, which makes it challenging to summarize. We refer the reader to Grossman (2015) for a review of methods that do not rely on experimental or quasi-experimental randomization.

We include studies that (i) were published after 2005; (ii) focus on the causal effect of education on all-cause mortality,¹⁸ current smoking or obesity; and (iii) used an RCT, twin difference design or quasi-experiment to assess the causal effect of education on mortality, current smoking or obesity.¹⁹

3.1 RCTs

Randomized experiments are the preferred method to establish treatment effects. When a random subset of individuals is treated, there are no average differences in the characteristics of the treated and the controls. Therefore, the controls can provide a counterfactual outcome. For practical and ethical reasons, there are no RCTs that give individuals “one more year” of formal schooling. There is, however, useful indirect evidence based on interventions that provide incentives to individuals to attend school.

In the United States, the most studied randomized interventions of this type are the Perry Preschool Program (implemented in the 1960s) and the Abecedarian (ABC) Program (implemented in the 1970s), both of which offered Early Childhood Education (ECE). Perry provided access to preschool education to a random subset of children between the ages of three and five. It offered intensive cognitive and language skill activities. The ABC program treated children between ages zero and five, and a subset up to age eight. In addition to cognitive stimulation, children received nutritional support and other health services. Because the Perry and ABC interventions occurred decades ago, it is now possible to investigate how they affected formal years of schooling, IQ, wages and health in adulthood. But it is too early to use these to investigate mortality.

Table 1 summarizes the results of RCT interventions. Males, randomly assigned to Perry, were less

¹⁸Palme and Simeonova (2015) and Leuven et al. (2016) focus on the effect of education on cancer risk and mortality.

¹⁹The exact search command in Google Scholar we used was “*instrumental variable*” OR “*regression discontinuity*” OR “*natural experiment*” OR “*exogenous variation*”, AND “*causal effect of education on mortality*” for the quasi-experiment studies, where mortality was replaced by obesity or smoking for the other outcomes. This delivered 64 hits for mortality, 30 for obesity, and 24 for smoking. For the twin studies we used “*twin difference*” OR “*twin fixed effects*” OR “*co-twin*”, AND “*causal effect of education on mortality*”, where again mortality (5 hits) was replaced by obesity (1 hit) and smoking (1 hit) for the other outcomes. We manually went through all of these hits, and the references therein, and applied our selection criteria to arrive at the selection of papers reviewed here. We have done our utmost to identify all relevant papers but recognize we may have missed a few in the process.

likely to smoke at age 27 (Heckman et al. 2013) and less likely to be a daily smoker at age 40 (Conti et al. 2016). This was not true for females. ABC recipients were less likely to be obese, though these effects are not statistically significant (Campbell et al. 2014, Conti et al. 2016). The likely mechanisms, by which these two ECE interventions improved health, are interesting. Neither Perry nor ABC had statistically significant impacts on IQ, and Perry did not affect formal measures of education (years of school). But the interventions affected non-cognitive skills, such as externalizing behaviors, and also increased achievement scores, such as grades, which reflect motivation in addition to skill and knowledge. In adulthood, treated individuals had higher employment, higher earnings, lower participation in crime and higher rates of marriage.

Although these findings provide strong evidence of a causal effect of ECE, they have several limitations. First, the Perry and ABC studies are each based on small samples of about 100 observations. Second, these studies concentrate on very disadvantaged children from poor families. In addition, Perry preschool targeted only African Americans within an IQ range of 70 to 85 – many would be considered cognitively impaired today – so it is unclear to what extent results apply to less disadvantaged and higher IQ individuals. ABC participants were not exclusively African Americans, and they had higher IQs than those in Perry, but were still drawn from poor populations with low baseline IQs. Lastly, the ABC intervention included a large health component, so effects on health and health behaviors do not necessarily stem from education alone.

We also report the results of one more RCT. Jensen and Lleras-Muney (2014) investigate the effects of an intervention in the Dominican Republic. The intervention informed male students at the end of middle school (around age 14) of the wage increases associated with attending high school (the randomization was done at the school level). Treated boys were less likely to smoke by age 18 (and delayed the onset of alcohol consumption). This study also investigated mechanisms. Treated boys were more likely to stay in school, worked less and earned less pocket money while in school, and were less likely to interact with smokers. The intervention did not affect participants' knowledge of the harms of smoking, nor did it affect their discount rates or their attitudes towards risk. This study also has several limitations. The children were also disadvantaged, drawn from poor neighborhoods with low high school attendance rates, females were not included, and all measures, except for schooling, are based on self reports.

Despite their limitations, some interesting conclusions can be drawn from these interventions. First, interventions that provide some form of schooling appear to improve some health behaviors. Because the ABC intervention included a health component that could directly explain the obesity results, the most conservative conclusion is that education, early in life or in high school, appears to lower smoking among disadvantaged males (more evidence is needed for obesity). Second, results differ for males and females, though it is not entirely clear why this is so. Third, in all cases, earnings, social skills and connections, appear to have been improved by the intervention, suggesting they could be mediators. Lastly, the Perry preschool program affected some important non-cognitive skills and motivation, despite having no effect on IQ or formal years of schooling, in line with predictions of the model (better use of school time, e.g., because of

better quality teachers, would result in greater skill without increases in schooling duration). It appears that important non-cognitive skills were formed by participation in these programs. These non-cognitive skills provided important benefits such as greater earnings, greater social connections and more stable marriages later in life. These improvements may explain better health behaviors in later life.

3.2 Twin studies

An alternative approach to RCTs is to mimic the results of experimental variation by finding a setting where almost all determinants of health are identical, but there is some variation in education that is close to random. Since important determinants of health can be traced to family inputs early in life, a substantial literature has pursued a within-twin estimation approach. The intuition behind this approach is that twins face very similar conditions – they share food, parents, neighbours, age, genetic traits, and have the same number of siblings. Further, identical twins share the same gender and have identical genetic endowments. Under the assumption that differences in education between twins are due to random factors, differences in outcomes across twins can be thought of as mostly resulting from differences in their education.

Table 2 reports the findings of studies that use twins to estimate the effect of education on mortality and smoking. The data for these studies come mostly from twin registries in various developed countries and contain a large number of observations (at least a thousand). The two U.S. studies rely on the Midlife U.S. Survey, and have smaller samples (of about 350 and 650). We report results for identical twins.

Twin studies usually start by reporting the effect of education using an OLS regression. These estimates compare the health of those with higher and lower levels of schooling. Thus the differences in outcomes reflect differences in education within and across twins. All OLS estimates in Table 2 find that for both men and women, education is associated with lower mortality and reduced smoking. The twin-fixed effect estimates only use within twin variation for identification, controlling for shared genetic and family characteristics. We discuss the twin studies in more detail below.

Four studies report effects of education on mortality (Table 2, panel A). A Swedish study (Lundborg et al. 2016) finds statistically significant twin FE estimates that are about the same size as OLS estimates. The results suggest that education lowers mortality very substantially for both men and women – a year of education is estimated to reduce mortality by 4 to 5 percent – and also that the bias in OLS estimates is low.

Three studies use Danish data. Behrman et al. (2011) finds statistically insignificant twin FE effects on mortality, that are also of “the wrong sign”. Van den Berg et al. (2012) report within-twin FE estimates of education on mortality that are negative and statistically significant for men, and these are similar to OLS estimates. Finally, Madsen et al. (2010) uses a larger sample of Danish twins and measure duration until death (rather than whether death has occurred by 2003 as in Behrman et al.). This study finds mixed results. The within twin FE estimates are not statistically significant, but they are larger than OLS in magnitude for the oldest male cohort, and smaller and close to zero for the younger male cohort. In all three Danish studies

the effects for women are insignificant and within pair estimates are always smaller than OLS estimates (or have the wrong sign).

The three studies for smoking (Table 2, panel B) find a negative, but very small and statistically insignificant, effect of education on smoking, once twin-fixed effects are included. However these studies have very small sample sizes. Finally, there are no studies that report effects on obesity.²⁰

There are some well-known methodological issues with twin approaches, which have been extensively discussed elsewhere (e.g., Bound and Solon, 1999). First, twin studies have important power limitations. The variation in education within twins is small, because the share of twins with identical education levels is very high (more than 40 percent in all studies). And, when they exist, the observed differences in education are often due to measurement error.

A second issue in twin studies is whether they can control for birth weight and for health conditions since conception. Even though identical twins share identical genetic material, they have differential access to nutrition in the womb and one twin is typically born larger and heavier than his / her counterpart. Because nutrition is different in utero, traits that are affected by nutrition, such as IQ or heart function, could also differ across twins. Moreover, parents might make compensating or reinforcing investments in response to these initial differences (Aizer and Cunha, 2012). Therefore, even after controlling for birth weight, there could be unobserved differences in inputs or environments within identical twins resulting in higher schooling and better health. The fundamental difficulty is that it is not clear in these studies why one twin has more schooling than the other, i.e., the source of variation is unclear.

A final question regarding twin studies is their external validity. Even though twins come from selected parents, they are on average much closer to the general population in their characteristics and outcomes than the participants of RCTs mentioned above. However, they have unique experiences relative to children from singleton births with siblings – for instance their in utero environment is shared, and they are typically born early and with low birth weights. Growing up, twins have an identical sibling, also a unique experience. Thus, twin studies cannot estimate the average treatment effect (or ATE) of one more year of schooling.

Where does this leave us? The study with the largest sample of identical twins is the Swedish study by Lundborg et al. (2016). This study also includes birth weight and height – two proxies for nutrition in the womb and in childhood, and has excellent mortality measures. It also uses register data from administrative sources, which typically contain less measurement error in education. Thus, the concerns raised above are likely less severe in the Lundborg et al. (2016) analysis. This study finds large effects of education on mortality for both women and men. The results of the three Danish studies are consistent with these findings for men, but inconclusive for women and generally statistically insignificant (except in Van den Berg et al.

²⁰Within twin pairs, some studies find large declines of education on overweight for men (Webbink et al. 2010) but not for women (Webbink et al. 2010, Amin et al. 2013). Lundborg (2013) finds large but statistically insignificant effects of education on BMI.

2012). Thus, some evidence of effects of education on mortality exists for men, subject to the important caveat of all twin studies, that within-twin differences in education are not necessarily random. There is some hint, that these effects vary by birth cohort and that they might apply to females as well, though potentially effects are smaller for females. For both males and females, the evidence suggests no effects of education on smoking, but these findings are less conclusive since they come from small samples.

3.3 Quasi-experimental evidence

Rather than investigating populations with similar characteristics, as in twin studies, one can seek to identify circumstances where differences in education in a certain population did not result from individual and family choices, but are due to external factors outside of the control of the individual and family. These are often referred to as quasi-experiments, since they attempt to mimic the ideal experiment where an outside force causes some individuals to obtain more schooling and others not.

Many studies have investigated the effect of compulsory schooling legislation (CSL hereafter) as a quasi-experiment. The intuition is that when the law is changed, typically to require individuals to stay more years in school, the affected cohort is very similar to the unaffected cohort, at least within a given state or geographic unit and close to the cutoff date specified by the legislation. Many countries around the world have implemented compulsory schooling legislation. If the legislation is shown to be binding – that is, it indeed forced individuals to attend school for longer, then one can compare individuals forced to go to school for X years with those forced to go to school for $X + \Delta X$ years, and assess whether those that are forced to attend school longer live longer lives, have better health and improved health behaviors.

Quasi-experimental studies typically start by reporting the OLS association with education and then report the 2SLS estimates of education, using minimum school-leaving ages as instrumental variables. As discussed in section 2.4.1, the estimated effect of increasing the minimum school-leaving age should be interpreted as treatment effects among the group of compliers, often referred to as a local average treatment effect (LATE, see Imbens and Angrist, 1994). If different individuals have different returns to school, then this LATE estimate will not coincide with the average treatment effect that OLS seeks to estimate. Therefore, apart from selection bias, another reason why OLS estimates may differ from estimates derived from quasi-experiments (supply side reforms) is that the affected population is different.

The type of individual that is induced to go to school longer by supply-side reforms could also differ across quasi-experiments. Therefore 2SLS estimates across settings could differ as well. These LATE estimates do inform policy regarding the effects of compulsory schooling in a given setting, but not necessarily in other settings, and are not necessarily informative about the effects of other policies that would also raise educational attainment.²¹

²¹This has led to a substantial amount of criticism of quasi-experiments because LATE estimates are often uninformative about other settings and other policies (Heckman and Urzua 2010; Deaton 2010).

Table 3 documents the findings. It documents substantial divergence across studies, with some studies finding large causal effects while others find no effects. We review the mortality studies in detail, after which we more briefly comment on the smoking and obesity studies.

3.3.1 Mortality

The first paper to investigate the causal effect of education on mortality, using compulsory schooling laws as instruments for education, was Lleras-Muney (2005)'s study of white cohorts born in the U.S. between 1900 and 1925. It found large effects of education on adult mortality, which was measured by tracking cohort sizes across successive censuses. The instrumental variables (IV hereafter) estimates were negative and large, though not statistically different from their OLS counterparts. No substantive differences across gender were found.

Several studies have investigated these results in the U.S. using the same approach. These studies have pointed out many limitations of the original data, methodology and findings. Mazumder (2008) documented that the results are not robust to including state-specific trends. Black et al. (2015) use more precise aggregate mortality data from vital statistics and document that almost all of the variation in mortality is explained by cohort and state fixed effects. Finally, a recent study by Fletcher (2015) shows that despite large samples and precise individual measures of age at death, the effects of education are not precisely estimated, though they appear to be large.²²

The fundamental limitation of the U.S.-based studies is that CSLs in the U.S. had small effects on the average education of the population (e.g., in Lleras-Muney's study one more year of compulsory schooling resulted in 0.05 years of additional schooling on average, or equivalently, only 1 in 20 individuals obtained one more year of schooling). Small effects are difficult to separate from the large secular improvements in education and mortality that occurred over the 20th century. This results in weak instruments, particularly when a large number of fixed effects (cohort, state) and trends (state-specific) are included. Additionally, only a small, and selected, number of people were treated (effectively forced to go to school longer). In the U.S., these are typically individuals at the low end of the education distribution that likely come from particular backgrounds. Even with sufficient power, the CSL approach cannot estimate the average treatment effect in the population, but rather the LATE.

Other papers have therefore investigated settings where the reforms had larger impacts on education – a larger first stage improves the statistical properties of the estimators and makes them more likely to be representative of average population effects. The country with the largest effects of compulsory schooling reforms on education is Great Britain, where years of education increased by 0.45 (0.35) years when the school-leaving ages were increased from 14 to 15 (15 to 16) in 1947 (1972). Yet, in an excellent study of

²²Mazumder (2008) uses similar data as Lleras-Muney (2005), and Black et al. (2015) argue that once fixed effects are added there is no variation left to estimate the additional impact of CSL. Therefore we do not report these findings in the table.

the effects of these reforms, Clark and Royer (2013) find no decrease in mortality.

Clark and Royer's results are based on large representative aggregate mortality rates from vital statistics, and are estimated using a regression discontinuity approach, which compares individuals born right before and right after the cut-off birth dates specified by the law. This approach is therefore less susceptible to the issue of cohort trends that IV studies face, since the RDD compares individuals very close to a well-specified month of birth cut-off. A limitation of this study is that, due to data limitations, it cannot estimate the OLS effect of education on mortality, so one cannot fully assess the extent of the bias for this population.

A recent re-examination of the British 1972 CSL reform by Davies et al. (2016), finds that it resulted in statistically significant declines in mortality. Methodological issues may explain the discrepancy in the estimates between Clark and Royer (2003) and Davies et al. (2016). Davies et al. use UK Biobank data that only includes people that volunteered to participate. Importantly, they only observe a small number of deaths between 2006 and 2014 (only 191 out of 22,138 die), whereas Clark and Royer use vital statistics. Not surprisingly given the very low prevalence of mortality, results in Davies et al. (2016) appear to be sensitive to the choice of bandwidth and the degree of the polynomials fitted to account for trends. The main issue is that the cohorts affected by the 1972 reform are still relatively young and have low mortality rates. So, ultimately, Clark and Royer's results based on the 1947 experiment are most informative, and these are precisely estimated zeroes.

Many other papers use CSLs in other contexts. Some suffer from weak instruments, e.g., Albouy and Lequien (2009) who investigate compulsory schooling reforms in France. But in most other studies, the first stage is large and the instruments do not appear to be weak.²³ These studies find a range of effects.

Three studies look at Sweden. Lager and Torssander (2012) and Meghir et al. (2017) investigated Swedish reforms that affected cohorts born between 1940 and 1960 and increased their education by 0.25-0.30 years. They both find very small and statistically insignificant declines in mortality. Fischer et al. (2013) investigated reforms that affected earlier cohorts in Sweden (born in 1924-1931) and had a substantially larger effect on educational attainment (about twice as large as the the more recent study by Meghir et al). They do find larger and statistically significant (at the 10% level) declines in mortality.

Van Kippersluis et al. (2011) use a Dutch reform in the early 20th Century (1928) that raised the compulsory years of schooling from 6 to 7, which for many implied having to attend at least one year of secondary school. Mortality is observed between ages 81 and 87, which is a limitation since the sample is heavily selected by mortality before age 81. The upside is that the mortality rate in this cohort is rather high (around 50% for the pivotal cohort born in 1917), providing power to detect an effect, should one exist, in this population. Using administrative data in a two-sample two-stage least squares (TS2SLS) setting, the study finds a 2.5 percentage point reduction in the mortality rate for an additional year of schooling, an effect

²³Though of course, as discussed above, the strength of the instrument is a function of the covariates that are included and in particular the number of aggregate controls for cohort, age, location and trends.

that is statistically significant among men at the 1% level. The Swedish and Dutch findings are consistent with the twin findings: effects seem larger in magnitude for older cohorts.

It is not clear whether the differences in findings across studies are due to methodology or reflect true heterogeneity in the effects of schooling. Gathmann et al. (2015) conduct a useful systematic investigation of heterogeneity by pooling data from 19 European countries and exploiting CSL reforms at various times during the 20th century. Because the same methods are applied to all countries, the difference in the findings across countries are not methodological. The downside of the exercise is that it is difficult to account correctly for all the institutional details across countries. And the samples in this study are not as large as those used in some of the individual-country studies. They find significant effects of education on mortality for men, but not for women. Among men, effects appear sizable in some countries, such as Belgium and the Netherlands, but there are no effects in other countries, such as Spain or Italy. This paper also finds that older cohorts tend to have larger effects, and the effects are also larger for poorer countries or those with initially higher mortality baselines. But they are not larger for countries that start with lower levels of compulsory schooling.

One study uses a different quasi-experiment. Buckles et al. (2016) uses the Vietnam war to generate experimental variation in college attendance. The main difficulty with this approach is that those who were drafted ended up serving, so that one must separate the effect of going to college from the effect of serving in Vietnam, both of which were affected by drafting procedures. The authors use variation in the risk of induction (the risk of being called to serve) during the War to generate two instruments that independently predict college attendance and veteran status—both of which are endogenous. This is the only study that investigates the causal effect of college and it finds that college education reduces the mortality rate in middle-age by 2.6 percentage points after accounting for both the endogeneity of education and veteran status. Interestingly it also does not find much bias in the OLS estimates.

In sum, there appears to be evidence that education leads to lower mortality for men, but this is only true for certain times and places. These effects appear to be larger for men born at the turn of the century – when GDP per capita was substantially lower and mortality was higher – and smaller more recently. Findings for women are substantially less robust, and statistically insignificant or small when significant.

3.3.2 Smoking and obesity

The results for smoking and obesity also vary across studies, as shown in Table 3, Panels B and C. A few issues are worth mentioning from the outset. We review only studies that focused on whether an individual is a current smoker, and whether an individual is currently obese. There are undoubtedly important differences between such studies and those that focus on the initiation and cessation of smoking, or BMI and overweight, respectively. Smoking and obesity information come from survey data, rather than from vital statistics or censuses. This results in several problems. First, these outcomes are typically (though not always) self

reported, and have known reporting biases,²⁴ though not much is known about whether education levels predict misreporting. Secondly, some survey data are not representative of the population. Finally, some of these studies have much smaller sample sizes than the vital statistics and censuses used in the mortality studies – this poses a problem for studies that rely on aggregate reforms such as CSLs, particularly if one wants to flexibly control for trends.

We start with studies that focus on smoking. British schooling reforms did not significantly affect smoking prevalence: Clark and Royer (2013) and Davies et al. (2016) both find that laws raising the minimum school-leaving age did not significantly affect smoking prevalence. Their point estimates are significantly smaller than the corresponding OLS estimates, and are relatively precisely estimated zeroes. Braakmann (2011), exploiting a slightly different feature of schooling laws in England, where children born in February are 3 percent more likely to obtain a qualification than those born in January due to the timing of exams, also finds no statistically significant effect on smoking prevalence. Here, the 2SLS estimates are of similar magnitude as the OLS results, but imprecisely estimated. Finally, although James (2015) uses a slightly less convincing design — education effects are identified by checking whether deviations from long-term trends of education for specific cohorts are associated with similar deviations in smoking — he reaches the same conclusion: educational expansions in Great Britain have not led to reduced smoking prevalence. Interestingly, Silles (2015) reports that education lowers smoking for males in Northern Ireland, a much poorer country.

In Germany, a group of authors (Reinhold and Jürges, 2010; Jürges, Reinhold and Salm, 2011; Kemptner, Reinhold and Jürges, 2011) used various sources of exogenous variation that increased years of education (the abolition of fees in academic track schools; academic track school construction that differed across German provinces; and compulsory-schooling laws, respectively) in conjunction with the sizable German Microcensus (between 70,000 and 120,000 observations). Interestingly, their papers reach different conclusions. Reinhold and Jürges (2010) and Kemptner et al. (2011) show, in line with the British evidence, no statistically significant effect of an extra year of secondary schooling on smoking behavior. Jürges et al. (2011), however, investigate the construction of new academic track schools, which, apart from increasing years of schooling, also changed the composition of students within the different school tracks. The induced changes, in both the quantity and quality of schooling, led to a 40 and 21 percent decrease in smoking prevalence for men and women, respectively, although the authors acknowledge that the results for men are not robust.

Etilé and Jones (2011) use a French expansion of secondary schooling in a Difference-in-Difference (DiD) design, using those with the highest level of education as a control group. They find sizable, 6-8%, effects of an extra year of education on current smoking. Heckman et al. (2016) compare college graduates

²⁴Smoking is typically under-reported, reporting biases for height and weight are more complex and differ by gender. See for example Gorber (2007).

with high-school graduates and use a structural equation framework with largely plausible exclusion restrictions. They also find sizable effects on smoking prevalence.²⁵

In sum, very convincing regression discontinuity designs in Great Britain and Germany suggest that another year in secondary school does not affect smoking prevalence. A similar conclusion is drawn by Li and Powdthavee (2015) for Australia, albeit using a smaller sample ($N < 10,000$).²⁶ However, it seems that exposure to a completely different type of schooling (e.g., completing high school as in Kenkel et al. 2006; college versus high-school as in Grimard and Parent, 2007; De Walque, 2007; Heckman et al. 2016, or academic track versus regular track as in Jürges et al. 2011) can lead to sizable reductions in smoking prevalence. This suggests that the schooling track and the associated peer group may be more important in smoking decisions than spending a certain number of years in secondary school. This interpretation would be consistent with the RCT results reported above: schooling affects smoking if it changes one's peers but not otherwise. However, the IV strategies used for these comparisons typically involve stronger assumptions compared with the RDD approaches that study a raise in the compulsory years of schooling, and the IV assumptions cannot be directly assessed for their validity.

Next we consider studies of the effects of schooling on obesity. Only a few studies met our criteria, in part because we limited attention to obesity, rather than including BMI or overweight as outcomes. The OLS results, without exception, indicate a negative correlation between education and obesity, of the order of a two percentage point reduction in the probability to be obese for every year of schooling. These results hold even though the studies cover substantially different birth cohorts and time periods.

Evidence for a causal protective effect of education on obesity is, however, at best very weak. In part because of the relatively small sample sizes, the standard errors of the 2SLS estimates are often large, rendering the effects insignificant. Only Kemptner et al. (2011) and James (2015) find a protective effect of an extra year of schooling in Germany and the UK, respectively. However, James' effect sizes seem implausibly large (68% reduction in the probability to be obese for an extra year of schooling), and the effects in Kemptner et al. are only present for men and not robust to including a higher-order polynomial in state-specific trends. Further, Jürges et al. (2011) obtain a statistically significant result (at the 1% level) that an additional year of education increases, not decreases, obesity. In all other cases, the point estimates are insignificant, close to zero (e.g., Kenkel et al. 2006; Brunello et al. 2013), and in some cases positive though not significantly different from zero, (e.g., Reinhold and Jürges, 2010; Li and Powdthavee, 2015). Hence, there seems to be no convincing evidence to date that education causally reduces the likelihood of being obese.

²⁵Park and Kang (2008) also use a structural modelling approach, combined with exclusion restrictions. They obtain sizable point estimates, but their small sample size results in very imprecisely estimated coefficients.

²⁶In a recent working paper, Huang (2016) presents evidence that an extra year of education in China led to a 5 percent reduction in smoking prevalence, although this estimate is only marginally significant at 10%.

4 Understanding heterogeneity and mechanisms

Why do the quasi-experimental estimates differ so much from the OLS estimates and from each other? Some differences can be due to methodology. OLS estimates indicate a strong association between education and health outcomes, which declines somewhat but remains strong even with an extensive set of control variables. In some cases, the point estimates obtained through 2SLS using educational reforms as IV are larger than the corresponding OLS estimate, which is counterintuitive if we believe the selection bias of OLS estimates is positive. This apparent contradiction could be explained by larger returns for the compliers of reforms than for the average individual (see the discussion about LATE in section 3.3). Alternatively, it could reflect a publication bias or weak instruments. In two cases, the 2SLS estimates of education on mortality estimate precise zeroes (Clark and Royer, 2013; Meghir et al. 2017). In most cases, in line with the OLS results, the 2SLS estimates suggest positive effects of education on health outcomes, but have wide confidence intervals that cannot rule out the OLS estimates.

At a more substantial level, what the findings suggest is that there is substantial true heterogeneity in these effects across time, space and population. Our reading of the evidence is that there is an effect of education on mortality, but not for all populations. Evidence for women is somewhat more limited, but suggests there is no, or a much smaller, effect on mortality for females than for males. Effects appear to be larger for earlier schooling reforms and smaller for later ones. The effects of education on smoking are frequently statistically insignificant, but not in RCTs and not in a few IV studies. There seems to be no convincing evidence that education causally reduces the likelihood of obesity.

How, in light of the theory, might we interpret the finding that education reduces mortality or smoking sometimes? As section 2.4.2 illustrated, in our theory education can affect unhealthy consumption and mortality through two prominent channels. First, additional schooling may raise life-time earnings and wealth ($\partial q_A(t)/\partial A(t) \leq 0$). This wealth (or permanent income) effect raises the relative marginal value of health ($\partial q_{h/a}(t)/\partial A(t) \geq 0$) and thereby the health benefit of health investment and the health cost of unhealthy consumption (see section 2.4.2), both improving healthy behavior. A second important channel through which schooling may affect health behavior and longevity is through skill formation improving the efficiency of health production $g_H(\cdot)$ (see 10, e.g., health knowledge, access to high-quality medical care, connections, etc.). We discuss each of these pathways, and some others that do not directly follow from the theory, in turn.

4.1 The role of income and access to resources

One first reason why education may lead to better health is that it increases lifetime earnings (in our model, higher earnings $Y(t)$, smaller marginal value of wealth $q_A(t)$), which in turn can be used to purchase health inputs. The monetary returns to education in the labor market vary. In the U.S. they were very high in the

beginning of the 20th century, plummeted to their lowest level in the early 1970s and have been rising rapidly since then (Goldin and Katz 2009). Our model suggests the effect of education on health should follow a similar pattern, if the permanent income pathway to health is of relative importance. Indeed, our review of the literature found larger effects on mortality for earlier schooling reforms. In line with this reasoning, Cutler et al. (2015) find that the health returns to education are larger for cohorts graduating in bad economic conditions, partly because those without education suffer large employment and wage declines if they enter the labor market during bad times (Oreopolous et al. 2012).

Some papers investigate directly whether increases in education due to CSL raised incomes. For instance, Devereux and Hart (2010) find that the British CSL reforms of 1947 (studied also by Clark and Royer, 2013) did not result in an increase in wages for women but did increase wages by 3-4 percent for men. Other papers document small or no returns to education on earnings using CSLs. Meghir and Palme (2005) and Fischer et al. (2016) find rather small income increases for men in Sweden (2 percent), and no effects for women, as a result of the 1948 CSL reforms. Perhaps not surprisingly, the corresponding effects on mortality in Sweden are small or not significantly different from zero. The effects of schooling on income are substantially lower in Great Britain and Sweden than those found in the U.S. for older cohorts. Clay et al. (2016) find wage returns for white males born between 1885-1912 to be between 6.5 and 8 percent, perhaps explaining why Lleras-Muney (2005) and Fletcher (2015) cannot rule out relatively large effects on mortality.

These results suggest one possible explanation for the observed heterogeneity in effects: perhaps the returns to education in the labor market are small for some cohorts and some populations. Low income returns to education in Great Britain, for example, are consistent with a high fraction of individuals dropping out of school exactly at the minimum school-leaving age – this is indeed what the model predicts individuals would do if perceived returns were low but enforcement penalties for non-compliance are high. This suggests that, while settings where a lot of people are affected by CSL are likely to estimate average treatment effects, they *could* also precisely be the settings where returns are expected to be small.²⁷ It also suggests that the effects of education on health are small for cohorts for whom the monetary returns to an additional year of education are small because of institutional and labor-market conditions (e.g., high barriers, such as discrimination, corruption, crime, low labor-market returns such as low wages, high unemployment, low quality of schools, etc.). In contrast, it could be that in the U.S. returns were high and enforcement penalties low. Thus, the small group of compliers were individuals who were credit constrained and had higher returns.

Another reason education might not have large health returns in Great Britain and other European countries is that it is not associated with other (non-wage) benefits in the labor market. In most European

²⁷Of course, a high fraction of individuals at the margin could also be consistent with other explanations, such as, high opportunity costs of schooling, or a large fraction of parents that do not appreciate the returns to schooling for their kids.

countries, health and disability insurance is publicly provided, making income differences potentially less important. In the U.S., e.g., the college educated have better health insurance, but in most European countries health care access is universal.

Related to the discussion about the LATE, the discrepancy across studies could also stem from the fact that different populations are affected by the reforms. Many studies find that the monetary returns to education, estimated using CSL reforms, is much larger for low socioeconomic status (SES) men. This is true in Sweden (Meghir and Palme, 2005) and in the U.S. (Clay et al. 2016). It also appears to be the case in the UK. A recent paper by Barcellos et al. (2017) finds that while the average effect of the UK 1972 reform on BMI is zero – consistent with Clark and Royer (2013), the effects of the reform are larger at the bottom of the BMI distribution, where it resulted in reductions in obesity. Davies et al. (2016) also report that the same 1972 UK reform increased incomes at the bottom of the income distribution but not at the top. In the U.S., Lleras-Muney and Shretzer (2015) find that reforms that limited work in adolescence had greater wage returns for immigrant children than for natives, and immigrants tend to be of lower SES.

The discussion so far assumes that resources, such as income and health insurance, are the main mechanisms whereby education causes health. The theory and the empirical literature suggest that this is however not the only, or even most important, channel. As already discussed, RCTs find effects on incomes but also document that education interventions are associated with improvements in non-cognitive skills and in peers. These could easily affect health. Further, there are many health behaviors that are not immediately determined by income, such as wearing a seat belt. This suggests that factors other than resources could be at play. In what follows, we consider alternative explanations for the differences between studies.

4.2 The role of quality of education, skills and the difference between time in school and skills.

Effects of education on health could also differ because of variation in the quality of schools across time, place and population. When CSL reforms are implemented, and when they affect a lot of individuals, it is possible that they lower the quality of instruction and therefore the skills associated with time in school. Evidence from large expansions of education in Italy suggests this can occur. Bianchi (2015) investigates an Italian reform in 1961 that dramatically increased university enrollment in science fields. But wages of affected individuals did not increase because congestion (higher pupil/teacher ratios and overall lower per pupil resources) and peer effects lowered overall learning rates (measured by grades). Many scholars have pointed out that the returns to education (on any outcome) must depend on the quality of education and the skills it imparts (e.g., Bold et al. 2017). But measuring school quality (separate from the characteristics of students that attend school) is extremely difficult.²⁸

²⁸Nevertheless wage returns to schooling are much larger once quality is accounted for (Hanushek and Zhang, 2009).

A related question is whether more or different education leads to different occupations, which depends on the type of skills that are acquired in school – some occupations are much more beneficial to health than others. Secondary schooling increased access to white-collar jobs at the beginning of the 20th century, but not so much at the end. These jobs were substantially safer than were jobs in agriculture and manufacturing. This could also explain why older cohorts saw larger returns from staying in school.

A related issue is whether we are measuring the effect of what happens in school or the effects of what happens outside of school. When children left school in the past, they worked in agriculture and in manufacturing – occupations that were detrimental to children’s health and growth. Or they roamed the streets. As compulsory schooling expanded, so did health-related nutrition, vaccination and health programs provided in school. As in the ABC program, schools and school policies evolved to guarantee and improve the health of its students in an attempt to better their learning capacities. It is possible, that school leads to better health in some periods because it kept kids in safe and healthy environments relative to the alternative.

It is also possible, that in some contexts, compulsory schooling extended time in school without increasing the skills individuals benefit from. For example, Pischke and Von Wachter (2008) study the effect of compulsory schooling changes in Germany in the post-WWI period and find they had no impact on wages. They hypothesize this is because “the basic skills most relevant for the labor market are learned earlier in Germany than in other countries”. This is consistent with the distinction in our model between skills and schooling. Schooling is in essence time and if time is not used productively then additional schooling may not increase skills.

This provides a segue into a related, but more disturbing, possibility that compulsory schooling is in fact a bad experiment to assess the causal effect of education because these laws force individuals, who are not interested in staying in school longer, to do exactly that. In the absence of market failures, such as financial constraints, and if individuals are rational, informed and forward looking, as in our theory, the variation in schooling in the population is driven by voluntary attendance decisions – that is, individuals want to attend school because its benefits exceeds its costs. In terms of equation (8), when there is no disutility of schooling and/or skill investment is productive, individuals spend all available time in school on skill production $\tau_S(t) = \tau_\theta(t)$. Hence, when empirical researchers compare individuals with different levels of schooling in an OLS setting, implicitly they are comparing individuals with different “effective” schooling durations *and* skills. In contrast, most IV studies use compulsory schooling reforms, which is an increase in \underline{S} . The complier population consists of individuals who would have liked to drop out of school in the absence of the reform. It is not clear whether the production of skill is equally productive for this group, compared to those who voluntarily remain in school. They may invest less (or not at all) in their skill during this additional year than do others: $\tau_\theta(t) < \tau_S(t)$. Hence, in compulsory schooling settings, schooling duration and skill capital are potentially weakly connected, and the skill capital gained may be lower in school than in other settings (e.g., they might be better off learning on the job).

The broader question is whether compulsory schooling solves a market failure: why are individuals not attending school to begin with? If they are not attending because they are credit constrained, because they are poorly informed about its returns, or because they have different objectives than their parents or legal guardians (e.g., children want to go to school but their parents don't want them to), then CSL could be beneficial. There could also be externalities that rationalize increasing education beyond what appears to be optimal for the individual (societal benefits). But in the absence of such failures, economic theory predicts that CSL would have negative returns.

4.3 The role of period effects: disease, technology, and information

Only two studies investigate sources of heterogeneity beyond gender. Gathmann et al. (2015) report that earlier reforms appear to have had larger effects than did later reforms, but not necessarily at lower levels of schooling. Earlier cohorts had larger mortality rates – this suggests that for more recent cohorts, effects might not materialize or become statistically detectable until older ages. This might explain why studies that investigate pre-WWII cohorts are more likely to find effects. Also, effects appear larger in poorer countries. Glied and Lleras-Muney (2008) suggest another reason why the effects of education may differ across time periods: they find that CSLs lowered mortality in the U.S., but these effects were much larger for diseases for which there was more medical innovation.

Scholars have documented that the health returns to education measured by OLS are also quite heterogeneous, particularly if one investigates these effects over long spans of time and across different populations. For example, the association between education and BMI varies with the level of development. Among poor countries, education is associated with higher weight, but the relationship becomes negative as countries get richer and fatter (Cutler and Lleras-Muney 2014). Thus the effect of education depends on the stage of the “nutrition transition”: in countries where nutrition levels are low and infectious disease is prevalent, being overweight is an advantage, so the more educated and those with more resources will be overweight. But as nutrition levels improve and infectious diseases disappear, being overweight becomes costly as it leads to chronic diseases, such as cardiovascular disease. Those with knowledge and resources avoid being overweight. Indeed, a recent meta analysis by Smith et al. (2015) finds that the OLS effect of education on chronic disease is, on average, negative, but it varies systematically from 0 to a negative number depending on the stage of the epidemiological transition. Our theory suggests an alternative (but not necessarily competing) hypothesis, operating solely through wealth, rather than knowledge or chronic disease. As countries are poor, due to diminishing returns to wealth, improvements in wealth raise the demand for unhealthy consumption more than the health costs decreases it. At higher levels of wealth (in more developed countries) the opposite happens: gains in wealth raise the demand for unhealthy consumption by less than the health costs decreases it.

Similarly, the effect of education on smoking has evolved over time. In the U.S. in the 1950s, more

educated individuals were more likely to smoke, but this reversed as knowledge of the harms of smoking disseminated. What these examples suggest is that wealth, and the availability of information, play an important role in determining whether individuals undertake healthy behaviors.

4.4 Multiple inputs into health

Clearly, smoking is determined by several factors, including knowledge but also income, wealth, cigarette taxes and prices, prices of complementary (or substitute) goods, peers and possibly many other factors, including genetic risk. A different set of factors might affect (over)eating and exercise and therefore obesity. Thus, at a particular point in time, one could observe educated individuals undertaking one but not the other behavior because different factors are at play.

Indeed, the bias in OLS estimates of the effect of schooling seems to differ across outcomes, even within studies. Almost all studies considered here, find statistically insignificant effects of education on smoking, and the IV estimates are generally smaller than their OLS counterparts. By contrast, many studies investigating mortality find IV effects that are larger than (or statistically identical to) OLS effects.

Several papers report IV estimates for more than one outcome, using the same sample and identification strategy. The pattern that emerges from these studies is somewhat puzzling, but consistent with the idea that different outcomes have different inputs. Consider the results by Kempter et al. (2011). They find that for men OLS estimates of the effect of education on smoking are upwards biased, whereas OLS estimates for obesity are downwards biased. Jürges et al. (2011) find that for women the OLS estimate of the effect of education on smoking is downward biased but obesity estimates are upward biased.

This suggests that the relevant omitted variables are different for different outcomes (peers may matter for smoking but maybe less for exercise), or that the same omitted variable has different effects on different outcomes (e.g., income may increase smoking but also increase exercise and thus lower obesity). Thus, simple stories that suggest that a single factor, such as IQ, generates upwards bias in the effect of education is too simplistic.²⁹ Advances in the genetics and biology of health behaviors may improve understanding of why certain, otherwise identical, individuals engage more or less in specific unhealthy behaviors.

4.5 Gender, culture and peers

Across the studies we consider, many find substantial differences across gender, with the evidence typically weaker for women. Why this is the case is not clear. One reason could be that more educated women are more likely to delay births, have fewer children, and are more likely to use modern hormone-based contraceptives. These behaviors are thought to increase the chance of cancers of the reproductive system,

²⁹Cutler and Glaeser (2005) also point out that the correlation between healthy behaviors within individuals is small, so a single factor is unlikely to explain differences across individuals.

because these cancers are a function of lifetime exposure to hormones.³⁰

Pregnancy, which only women experience, is related to important changes in weight, health-seeking behaviors, and labor force participation. In many developed countries pregnant women and women with small children make more use of the health care system, regardless of their level of education – this may somewhat attenuate differences between more and less educated women. For instance, all women today are strongly discouraged from smoking during pregnancy. Moreover, pregnancy causes persistent decreases in hourly earnings and labor supply (e.g., Lundborg et al. 2017), potentially rendering the relationship between education and lifetime earnings, and in turn health, weaker among women versus men.

Another possibility is that the returns to education for women operated through the marriage market, rather than through their own labor income, since labor-force participation of women was low until recently. Finally, it is worth noting that some behaviors, such as smoking or drinking, were considered taboo for females at the turn of the 20th century – but as women entered the labor force they became acceptable, and in some cases, were adopted by the most successful women as symbols of independence (Amos and Haglund, 2000). Thus, for more recent cohorts of women, it is possible that more successful women undertake more unhealthy behaviors, despite their health cost. This discussion suggests that for women it is key to incorporate the marriage and fertility effects of education, in addition to other behaviors that are partially culturally determined, to understand the net effect of education on women’s health.

More broadly, economic theories largely ignore social ties and norms as determinants of health. Many empirical papers suggest that peers are a key determinant of some behaviors, such as eating, smoking and drinking (e.g., Cawley and Ruhm, 2012). And, as we already discussed, the evidence from RCTs and IV studies suggests that one reason why schooling may affect smoking is that, in some instances, it alters the peer group. But research rarely considers the effects of isolation and networks on health directly, independently of these behaviors. Many empirical studies find strong associations between social and community ties and mortality (starting with Berkman and Syme 1979, see Kawachi et al. 2008; Holt-Lunstad et al. 2010 for reviews). Education, and the type of education individuals receive, affect the size and quality of their social connections and their integration into the communities in which they live. There are important differences across gender in how ties are formed and in the type of ties they have. Women also appear to be differently affected by social connections. For example, the effect of marriage on mortality is very different by gender: marriage lowers mortality more among men than among women, and widowhood increases mortality more for men than for women (Smith and Christakis, 2008).

³⁰Indeed Glied and Lleras-Muney (2008) and Meghir et al. (2012) document that CSLs increase mortality rates from reproductive cancers.

5 Some concluding thoughts and directions for future research

In this review, we have attempted to provide some insight into the findings of a large literature seeking to understand the effect of education on health outcomes. While there are some discrepancies across studies that appear to be driven by methodological differences, there appears to be substantial true heterogeneity in the effect of education on health outcomes. This heterogeneity varies depending on the outcome studied, the country and the time period. We have attempted to understand this heterogeneity by investigating the theoretical determinants of health. We have also attempted to gain insight into this heterogeneity by hypothesizing *ex-post* about the factors that might explain observed differences across empirical studies.

Our conclusion, based on reviewing twin studies and IV studies, is that education lowers mortality among men, but only for some populations and time periods. However, we caution that the results based on IV are often very imprecise so it remains possible that differences in the findings across studies are spurious. A second conclusion is that education does not appear to affect smoking in a causal sense, except for disadvantaged populations, or for those whose peer group changes with the amount or type of schooling they obtain. Third, the evidence on the effect of education on obesity is weak. Finally, across all studies, methods and outcomes, there are substantial differences between men and women, with men's outcomes depending more strongly on education than women's.

Next, we summarize a few broader conclusions. There is an important difference between time spent in school, which is an input, and the actual skills that are developed during that time. As our theory highlights, it may very well be that additional schooling does not lead to additional skills, i.e. the complier affected by a particular reform does not learn much from schooling (but more from, e.g., hands-on work on the job), or that the skills learned are not useful to the complier given her labor-market prospects. Research should seek to move towards a better understanding of observed heterogeneity in the effects of education by evaluating the benefits of reforms to those affected.

Related, the quasi-experiments used to estimate the effects of education on health outcomes almost exclusively derive from changes in the minimum school-leaving age. Whereas these reforms often provide convincing, random-like variation in attended years of schooling for cohorts born (in some cases) only a few months apart, they only affect a specific group of compliers at a particular margin – those for whom the minimum school-leaving age happened to coincide with the end of their educational career. The theory suggests there are various other policy instruments encouraging skill formation that are plausibly exogenous to the individual, for example prices p_S and p_θ and subsidies λ_S , or the term length τ_S . It would be informative to see how these other type of reforms – with different types of compliers, and at different margins – affect later-life health outcomes.³¹

³¹Fischer et al. (2016), for example, show that increases in the total time spent in school, as a result of increases in the term length, affect later-life income much more than do comparable increases in time spent in school due to raising the minimum school-leaving age.

Future work should also link short, medium and long term outcomes. Most studies of mortality do not observe important determinants of mortality, such as health behaviors or income across the life cycle, and few studies include measures of what occurred in school in terms of nutrition, and cognitive and non-cognitive skill development. Studies must move beyond asking whether or not “there is an effect”, to tracing the pathways by which particular reforms impact individuals in a comprehensive manner. This deeper investigation, linking short and long term outcomes, can shed light on why there is, what we believe to be, true heterogeneity across studies in the effects of education on health.

Studies could also benefit from investigating the returns to the skills that are learned in school – these returns depend greatly on economic and social circumstances and institutions. Greater attention should be paid to the fact that different health behaviors or health outcomes have different determinants: the simple notion that education has the same (beneficial or detrimental) effect across outcomes does not seem to be supported by the data.

The current availability of measures in, and sample sizes of, existing data sources are probably too limited to undertake these kinds of detailed studies, particularly if one wants to use quasi-experiments for identification. Thus, a possible future direction for research would be to combine results from studies using large scale administrative databases with wide population coverage but a small number of outcomes, with analyses using data with smaller sample sizes but that are richer in their survey measures.

In general, heterogeneity is investigated without much theoretical guidance. Even the simple theory we presented here makes rich predictions as to how and why effects of education may differ. A greater integration of theoretical and experimental work would be useful.³² In particular, a promising avenue seems to be the exploration, guided by theory, of “interactions” between determinants of education and health. For instance, education might make more of a difference when there is information about the dangers of smoking, but less so when there are smoking bans.

Both our theory and our review of the empirical evidence fall short in some specific areas. One important area is the modeling and documenting of the role of peer effects and social factors in the determination of health gaps by education. Our model operates at the individual level, and important factors, such as the influence of spouses and friends are not incorporated. We also do not model, and did not discuss, genetic factors. A new, important, and exciting area of research investigates the role of genetic risk, and more specifically the interaction between genes and the environment. Genetic make-up is treated typically as a potential omitted variable, and is implicitly taken into account when using within-twins designs or carefully conducted IV studies. But this approach assumes genetic factors are additive and do

³²Many recent authors have moved in this direction. For instance, Todd and Wolpin (2006) estimate a structural model using experimental variation from Progreso, an experiment in Mexico. Work in other fields is also increasingly using this combined approach (e.g., Heckman et al. 2013; Finkelstein et al. 2015). Another approach that tries to combine both is the work by Chetty (2009) on the use of sufficient statistics.

not interact with the environment. This view of the effect of genetic endowments has proven to be overly simplistic: it is increasingly appreciated that complex interactions exist between genes and environment (Turkheimer, 2000; Heckman, 2007). Lastly, many common health behaviors, such as smoking and eating, have important addictive characteristics – our theory and analysis of the empirical evidence is based on a set of strong assumptions that include perfect information and rational decision making. Greater incorporation of bounded rationality into health decision making, and a greater understanding of how education affects imperfect decision making would be useful.

We also reach some methodological conclusions. A key difficulty with quasi-experiments is that they rely on variation over time or across cohorts. Health and mortality vary substantially with age. As Davies et al. (2016) explain: “... *people affected by the reform, are one year younger than our control group, those unaffected by the reform. (...) Many of the outcomes we investigated increase linearly or log-linearly over time. This means it is difficult to determine if any differences we observed are due to an additional year of aging or the reform...*” Similarly, there have been large secular improvements in health resulting in much lower mortality and much better health at any age among recent cohorts. It is very difficult to identify the effects of reforms separately from age effects and from secular trends across cohorts, particularly if one wants to account for these trends flexibly, in a non-parametric fashion. This suggests that exploiting cross-sectional variation, i.e. comparing individuals born around the same time but who obtain different levels of schooling through a randomized or natural experiment, could be very informative since it alleviates the need to control for secular trends over cohorts and time.

Finally, we note that in compiling our tables we found it difficult to compare across studies – the information empirical researchers report varies greatly across studies. Not all studies report the number of observations, or basic summary statistics, such as the the mean and standard deviation of the dependent variable or of education. Some quasi-experimental studies do not report OLS results, and some papers do not report results separately by gender. The choice of the dependent variable and the functional form used across studies also varies widely. Studies do not systematically report whether results are sensitive to these choices or compare their estimates to existing estimates in the literature. Neither do studies report how results vary when one includes covariates that were included in other existing studies. This lack of uniformity makes it very challenging to summarize and compare findings, particularly the magnitudes of their effects. Future work should try to systematically report basic information, and include results of models that are identical to those previously estimated, for greater transparency and so that comparisons can be made more easily.

Bibliography

Aizer, A., & Cunha, F. (2012). The production of human capital: Endowments, investments and fertility. National Bureau of Economic Research Working Paper No. w18429.

Albouy, V. & Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics* 28 (2009) 155-168.

Almond, D., Currie, J. & Duque, V. (2017). Childhood Circumstances and Adult Outcomes: Act II (No. w23017). National Bureau of Economic Research.

Amin, V., Behrman, J. R. & Spector, T. D. (2013). Does more schooling improve health outcomes and health related behavior? Evidence from U.K. twins. *Economics of Education Review*. 35 (2013) 134-148.

Amos, A., & Haglund, M. (2000). From social taboo to “torch of freedom”: the marketing of cigarettes to women. *Tobacco Control*, 9(1), 3-8.

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, 2(3), pp. 193-216.

Barcellos, S.H., Carvalho, L.S. & Turley, P. (2017). Distributional effects of education on health. Mimeo.

Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research.

Becker, G. S. (2007). Health as human capital: synthesis and extensions. *Oxford Economic Papers*, 59(3), 379-410.

Bedard, K. (2001). Human capital versus signaling models: university access and high school dropouts. *Journal of Political Economy*, 109(4), 749-775.

Behrman, J. R., Kohler, H. P., Jensen, V. M., Pedersen, D., Petersen, I., Bingley, P. & Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography*, (2011) 48:1347-1375.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political*

Economy, 75(4, Part 1), 352-365.

Berkman, L. F., & Syme, S. L. (1979). Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *American Journal of Epidemiology*, 109(2), 186-204.

Bianchi, N. (2016). The Indirect Effects of Educational Expansions: Evidence from a Large Enrollment Increase in STEM Majors. Working paper.

Bijwaard, G. E., Van Kippersluis, H. & Veenman, J. (2015). Education and health: the role of cognitive ability. *Journal of Health Economics*, 42, 29-43.

Black, D. A., Hsu, Y. C. & Taylor, L. J. (2015). The effect of early-life education on later-life mortality. *Journal of Health Economics*, 44, 1-9.

Black, S. E., Devereux, P. J. & Salvanes, K. G. (2008). Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *The Economic Journal*, 118(530), 1025-1054.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J. & Wane, W. (2017). Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa. *Journal of Economic Perspectives* 31(4): 185-204.

Bound, J. & Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review*, 18(2), 169-182.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M. & Wozniak, A. (2016). The effect of college education on mortality. *Journal of Health Economics*, 50 (2016), 99-144.

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

Brunello, G., Fabbri, D. & Fort, M. (2013). The causal effect of education on body mass: evidence from Europe. *Journal of Labor Economics*, vol. 31. No. 1. 195-223.

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. & Pan, Y. (2014). Early

childhood investments substantially boost adult health. *Science*, vol. 343. 1478-1485.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127-1160.

Case, A., & Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging* (pp. 185-212). University of Chicago Press.

Cawley, J., & Ruhm, C. J. (2012). The Economics of Risky Health Behaviors. In: *Handbook of Health Economics*, Mark V. Pauly, Thomas G. McGuire, Pedro Pita Barros (eds.); Chapter 3, p. 95-199.

Chetty, R. (2009). Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. *Annual Review of Economics* 1:451-488.

Conti, G., Heckman, J.J & Urzua, S. (2010). The Education-Health Gradient. *American Economic Review Papers and Proceedings* 100: 234-238.

Conti, G., Heckman, J. J. & Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal*, 126, F28-F65.

Contoyannis, P., Jones, A. M. & Rice, N. (2004), The dynamics of health in the British Household Panel Survey, *Journal of Applied Econometrics*, 19(4): pp. 473-503.

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

Clay, K., Lingwall, J. & Stephens Jr, M. (2016). Laws, Educational Outcomes, and Returns to Schooling: Evidence from the Full Count 1940 Census (No. w22855). National Bureau of Economic Research.

Cunha, F. & Heckman, J. (2007). The Technology of Skill Formation. *The American Economic Review*, 97(2), 31.

Cutler, D. M., & Glaeser, E. (2005). What Explains Differences in Smoking, Drinking, and Other Health-Related Behaviors?. *The American Economic Review*, 95(2), 238-242.

Cutler, D.M. & Lleras-Muney, A. (2010), Understanding Differences in Health behaviors by Education, *Journal of Health Economics*, 29(1): pp. 1-28.

Cutler, D.M., Lleras-Muney, A. & Vogl, T. (2011), Socioeconomic Status and Health: Dimensions and Mechanisms. In: the *Oxford Handbook of Health Economics*.

Cutler, D. M., Huang, W., & Lleras-Muney, A. (2015). When does education matter? The protective effect of education for cohorts graduating in bad times. *Social Science & Medicine*, 127, 63-73.

Dalgaard, C. J., & Strulik, H. (2014). Optimal aging and death: understanding the Preston curve. *Journal of the European Economic Association*, 12(3), 672-701.

Davies, N. M., Dickson, M., Smith, G. D., Van den Berg, G. & Windmeijer, F. (2016) The causal effects of education on health, mortality, cognition, well-being, and income in the UK Biobank. Preprint.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424-455.

Devereux, P. J., & Hart, R. A. (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, 120(549), 1345-1364.

Ehrlich, I., & Chuma, H. (1990). A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political economy*, 98(4), 761-782.

Etilé, F. & Jones, A. M. (2011). Schooling and smoking among the baby boomers – An evaluation of the impact of educational expansion in France. *Journal of Health Economics* 30 (2011) 811-831.

Finkelstein, A., Hendren, N. & Luttmer, E.F.P. (2015). The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment. NBER Working Paper 21308.

Fischer, M., Karlsson, M. & Nilsson, T. (2013). Effects of compulsory schooling on mortality: Evidence from Sweden. *International Journal of Environmental Research and Public Health*, 2013, 10 3596-3618.

Flegal, K.M., Kit, B.K., Orpana, H. & Graubard, B.I. (2013) Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis.

Journal of the American Medical Association 309:71-82.

Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science & Medicine*, 127 (2015), 101-107.

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

Galama, T.J. (2015). A Contribution to Health-Capital Theory. CESR Working Paper #2015-004.

Galama, T.J. & Van Kippersluis, H. (2015a). A Theory of Education and Health. CESR Working Paper #2015-001.

Galama, T.J. & Van Kippersluis, H. (2015b). A Theory of Socioeconomic Disparities in Health Over the Life Cycle CESR Working Paper #2015-016.

Glied, S. & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography*, 45(3), 741-761.

Goldin, C. D. & Katz, L. F. (2009). *The race between education and technology*. Harvard University Press.

Gorber, S. C., Tremblay, M., Moher, D. & Gorber, B. (2007) A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews*, 8: 307–326. doi:10.1111/j.1467-789X.2007.00347.

Grimard, F. & Parent, D. (2007). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics* 26 (2007) 896-926.

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

Grossman, M. (2015). The Relationship between Health and Schooling: What's New? *Nordic Journal of Health Economics*, 3(1), pp-7-17.

Hall, R. E. & Jones, C. I. (2007). The value of life and the rise in health spending. *The Quarterly Journal of*

Economics, 122(1), 39-72.

Holt-Lunstad, J., Smith, T. B. & Layton, J. B. (2010). Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine*, 7(7), e1000316.

Fujiwara, T. & Kawachi, I. (2009). Is education causally related to better health? A twin fixed-effects study in the USA. *International Journal of Epidemiology*, 2009 38:1310-1322.

Hanushek, E. A. & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, 3(2), 107-143.

Heckman, J. J., Stixrud, J. & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411-482.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(33), 13250-13255.

Heckman J. J., Humphries, J. E. & Veramendi, G. (2016). Returns to education: The causal effects of education on earnings, health, and smoking. National Bureau of Economic Research Working Paper No. 22291.

Heckman J., Pinto, R. & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6): 2052-2086.

Heckman, J. J., & Urzua, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, 156(1), 27-37.

Holt-Lunstad, J., Smith, T.B. & Layton, J.B. (2010) Social Relationships and Mortality Risk: A Meta-analytic Review. *PLoS Medicine* 7(7): e1000316. <https://doi.org/10.1371/journal.pmed.1000316>

Huang, W. (2016). Understanding the effects of education on health: evidence from China. Mimeo.

Hummer, R. A., & Hernandez, E. M. (2013). The effect of educational attainment on adult mortality in the United States. *Population Bulletin*, 68(1), 1.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, 48(2), 399-423.

James, J. (2015). Health and education expansion. *Economics of Education Review* 49 (2015) 193-215.

Jensen, R. & Lleras-Muney, A. (2012). Does staying in school (and not working) prevent teen smoking and drinking? *Journal of Health Economics*, 31 (2012) 644-657.

Jürges, H., Reinhold, S. & Salm, M. (2011). Does schooling affect health behavior? Evidence from educational expansion in Western Germany. *Economics of Education Review*, 30 (2011) 862-872.

Fischer, M., Karlsson, M., Nilsson, T., & Schwarz, N. (2016). The Sooner the Better? Compulsory Schooling Reforms in Sweden. IZA Working Paper No. 10430.

Kawachi, I., Subramanian, S. V., & Kim, D. (2008). Social capital and health. *Social capital and health*, 1-26.

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

Kenkel, D., Lillard, D. & Mathios, A. (2006). The roles of high school completion and GED receipt in smoking and obesity. *Journal of Labor Economics*, vol. 24, no. 3.

Khang, Y.H., Lynch, J.W., Yang, S., Harper, S., Yun, S.C., Jung-Choi, K., et al. (2009), The contribution of material, psychosocial, and behavioral factors in explaining educational and occupational mortality inequalities in a nationally representative sample of South Koreans: relative and absolute perspectives, *Social Science & Medicine*, 68(5): pp. 858-866.

Lager, A. C. J. & Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences* 2012 109 (22) 8461-8466.

Lang, K. & Kropp, D. (1986). Human capital versus sorting: the effects of compulsory attendance laws.

The Quarterly Journal of Economics, 101(3), pp. 609-624.

Lemieux, T. (2006). The “Mincer equation” thirty years after schooling, experience, and earnings. In Jacob Mincer A Pioneer of Modern Labor Economics (pp. 127-145). Springer US.

Leuven, E., Plug, E. & Rønning, M. (2016). Education and cancer risk. Labour Economics, 43, 106-121.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. The Review of Economic Studies, 72 1 189-221.

Lleras-Muney, A. & Shertzer, A. (2015). Did the Americanization movement succeed? An evaluation of the effect of english-only and compulsory schooling laws on immigrants. American Economic Journal: Economic Policy, 7(3), 258-290.

Li, J. & Powdthavee, N. (2015). Does more education lead to better health habits? Evidence from the school reforms in Australia. Social Science & Medicine, 127 (2015) 83-91.

Lochner, L. & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. The American Economic Review, 94(1), 155-189.

Lochner, L. (2011). Nonproduction Benefits of Education: Crime, Health, and Good Citizenship. Handbook of the Economics of Education, 4, 183.

Lundborg, P. (2013). The health returns to schooling—what can we learn from twins? Journal of Population Economics, (2013) 26:673-701.

Lundborg, P., Lyttkens, C. H. & Nystedt, P. (2016). The effect of schooling on mortality: New evidence from 50,000 Swedish Twins. Demography, (2016) 53:1135-1168.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. American Economic Review, 107(6), 1611-1637.

Mackenbach, J.P., Huisman, M., Andersen, O., Bopp, M., Borgan, J.K., Borrell, C. et al. (2004), Inequalities in lung cancer mortality by the educational level in 10 European populations, European Journal of Cancer, 40(1): pp. 126-135.

Madsen, M., Andersen, A. M. N., Christensen, K., Andersen, P. K. & Osler, M. (2010). Does educational status impact adult mortality in Denmark? A twin approach. *American Journal of Epidemiology*, vol. 172, no. 2. 225-234.

Mazumder, B. (2008). Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws. *Economic Perspectives*, 32(2), 2.

Meara, E. R., Richards, S. & Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27(2), 350-360.

Meghir, C. & Palme, M. (2005). Educational reform, ability, and family background. *The American Economic Review*, 95(1), 414-424.

Meghir, C., Palme, M. & Simeonova, E. (2017). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics*, forthcoming.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

Mokdad, A. H., Marks, J. S., Stroup, D. F. & Gerberding, J.L. (2004), Actual causes of death in the United States, *Journal of the American Medical Association*, 291(10): pp. 1238-45.

Oreopoulos, P., Von Wachter, T., & Heisz, A. (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics*, 4(1), 1-29.

Palme, M. & Simeonova, E. (2015). Does women's education affect breast cancer risk and survival? Evidence from a population based social experiment in education. *Journal of Health Economics*, 42, 115-124.

Park, C. & Kang, C. (2008). Does education induce health lifestyle? *Journal of Health Economics*, 27 (2008) 1516-1531.

Reinhold, S., Jürges, H. (2010). Secondary school fees and the causal effect of schooling on health behavior. *Health Economics*, 19: 994-1001 (2010).

Pappas, G., Queen, S., Hadden, W. & Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *New England Journal of Medicine*, 329(2), 103-109.

Savelyev, Peter A. (2014), *Conscientiousness, Education, and Longevity of High-Ability Individuals*, Unpublished manuscript, Vanderbilt University, Department of Economics.

Silles, M. (2015). The causal effect of schooling on smoking behavior. *Economics of Education Review*, 48 (2015) 102-116.

Smith, W. C., Anderson, E., Salinas, D., Horvatek, R., & Baker, D. P. (2015). A meta-analysis of education effects on chronic disease: the causal dynamics of the Population Education Transition Curve. *Social Science & Medicine*, 127, 29-40.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annual Review of Sociology*, 34:1, 405-429.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Strulik, H. (2016). *The Return to Education in Terms of Wealth and Health*. Cege Discussion Paper 293.

Todd, P. E., & Wolpin, K. I. (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *The American Economic Review*, 1384-1417.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9(5), 160-164.

Van den Berg, G., Janys, L. & Christensen, K. (2012). *The causal effect of education on mortality*. Mimeo.

Van Kippersluis, H., O'Donnell, O. & Van Doorslaer, E. (2011). Long-run returns to education: Does schooling lead to an extended old age? *The Journal of Human Resources* 46 4 695-721.

Van Kippersluis, H. & Galama, T. J. (2014). Wealth and health behavior: Testing the concept of a health cost. *European Economic Review*, 72, 197-220.

Pischke, J. S. & Von Wachter, T. (2008). Zero returns to compulsory schooling in Germany: Evidence and interpretation. *The Review of Economics and Statistics*, 90(3), 592-598.

de Walque, D. (2007). Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics*, 26 (2007) 877-895.

Webbink, D., Martin, N. G. & Visscher, P. M. (2010). Does education reduce the probability of being overweight?. *Journal of Health Economics*, 29(1), 29-38.

Weiss, A. (1995). Human capital vs. signalling explanations of wages. *The Journal of Economic Perspectives*, 9(4), pp. 133-154.

Further Reading (selected papers)

Almond, D., Currie, J. & Duque, V. (2017). Childhood Circumstances and Adult Outcomes: Act II (No. w23017). National Bureau of Economic Research.

Cawley, J., & Ruhm, C. J. (2012). The Economics of Risky Health Behaviors. In: *Handbook of Health Economics*, Mark V. Pauly, Thomas G. McGuire, Pedro Pita Barros (eds.); Chapter 3, p. 95-199.

Cutler, D. M., & Glaeser, E. (2005). What Explains Differences in Smoking, Drinking, and Other Health-Related Behaviors? *The American Economic Review*, 95(2), 238-242.

Cutler, D.M. & Lleras-Muney, A. (2014), Education and Health: Insights from International Comparisons. In: *Encyclopedia of Health Economics*, Anthony J. Culyer (ed.), Vol 1. San Diego: Elsevier; 2014. pp. 232-245.

Galama, T.J. & Van Kippersluis, H. (2015). A Theory of Education and Health. CESR Working Paper #2015-001.

Grossman, M. (2015). The Relationship between Health and Schooling: What's New? *Nordic Journal of Health Economics*, 3(1), pp-7-17.

Heckman J. J., Humphries, J. E. & Veramendi, G. (2016). Returns to education: The causal effects of

education on earnings, health, and smoking. National Bureau of Economic Research Working Paper No. 22291.

Lochner, L. (2011). Nonproduction Benefits of Education: Crime, Health, and Good Citizenship, in E. Hanushek, S. Machin, and L. Woessmann (eds.), *Handbook of the Economics of Education*, Vol. 4, Ch. 2, Amsterdam: Elsevier Science, 2011.

Tables

Table 1: Effect of Randomized Education Interventions on Smoking and Obesity

Paper	Data & Sample	Outcomes (overall means)	Program	Treatment Effect (M=Males, F=Females, P=Pooled)
A. Smoking				
Conti et al. (2016)	Perry Preschool Program Follow-up at age 40 Male children age 3-5 Birth cohorts early 1960s <i>N</i> = 66	Not a daily smoker at age 40 (0.56)	Perry Preschool Program	M ^a 0.194** [0.01] ^b [Table 4, Column 6]
Heckman et al. (2013)	Perry Preschool Program Follow-up at age 27 Male children age 3-5 Birth cohorts early 1960s <i>N</i> = 51	Tobacco use at age 27 (0.48)	Perry Preschool Program	M ^a -0.119* [0.09] ^b [Table 1, Column 1]
Jensen and Lleras-Muney (2012)	Dominican Republic Survey data (2001-2005) Male students around age 14 Birth cohort 1986-1987 <i>N</i> = 2,011	Currently smoking around age 18 (0.05)	Providing information on the returns to education to Dominican youths	M -0.039** (0.018) [Table 4, Column 4]
B. Obesity				
Conti et al. (2016)	Carolina Abecedarian Project Male children age 0-5 Birth cohorts 1972-1977 <i>N</i> = 26	Obesity at age 35 (0.56)	Carolina Abecedarian Project	M ^a -0.069 [0.23] [Table 5, Column 6]
Campbell et al. (2014)	Carolina Abecedarian Project Children age 0-5 Birth cohorts 1972-1977 <i>N</i> = 26/40 ^c	Obesity at age 35 (0.56/0.73)	Carolina Abecedarian Project	M/F -0.069/-0.061 [0.233]/[0.790] [Table 1/2, Column 3]

^aFemale estimates are not significant and not reported in the paper

^bSquare brackets are p-values, parentheses are standard errors

^cMen/Women, respectively

Table 2: Health Effects of Education from Twin Studies

Paper	Data & Sample ^a	Variables (mean)	Estimates (M=Males, F=Females, P=Pooled) ^b	
			OLS	Fixed Effects
A. Mortality				
Lundborg et al. (2016)	Swedish Twin Registry	$H =$ Mortality by 2007, hazard ratio	M/F	M/F
	Birth cohorts 1886-1958 $N = 8,602/10,084^d$	$(0.60/0.65)^c$ $E =$ Years of education (9.74/9.59)	0.963***/0.951*** (0.006)/(0.007) [Table 3, Column 3]	0.948***/0.962*** (0.017)/(0.019) [Table 3, Column 4]
Van den Berg et al. (2012) ^e	Danish Twin Registry	$H =$ Mortality (0.62)	M/F	M/F
	Birth cohorts 1888-1897 $N = 2,839/2,856$	$E =$ Eligible for reform that expanded voluntary education (0.5/0.5)	-0.194**/-0.063 (0.086)/(0.080) [Table 11/12, Column 1]	-0.183**/-0.054 ^f (0.088)/(0.087) [Table 11/12, Column 3]
Behrman et al. (2011)	Danish Twin Registry	$H =$ Mortality by 2003 (0.16)	M/F	M/F
	Birth cohorts 1921-1950 $N = 3,234/2,060$	$E =$ Years of education (10.8)	Cohort 1921-1935 -0.011***/-0.010*** (0.005)/(0.004) Cohort 1936-1950 -0.006**/-0.003 (0.002)/(0.002) [Table 6, Column 2]	Cohort 1921-1935 0.005/0.013 (0.008)/(0.008) Cohort 1936-1950 -0.001/-0.001 (0.004)/(0.004) [Table 6, Column 4]

^aSample size is for monozygotic twins, if available

^bReport for monozygotic twins, if available

^cMeans for Male/Female, respectively

^dMen/Women, respectively

^eWorking paper

^fCorrelated frailty model, where unobservables within twins are assumed correlated

Madsen et al. (2010)	Danish Twin Registry Statistics Denmark Census of Death Registry Birth cohorts 1921-1950 $N = 5,260$	$H =$ Mortality 1980-2008 (Hazard ratio) $E =$ Less than 7 years of education (0.60)	M/F Cohort 1921-1935 1.17*/1.41*** (0.10)/(0.11) Cohort 1936-1950 1.36**/1.51*** (0.15)/(0.18) [Table 3, Column 4]	M/F Cohort 1921-1935 1.32/0.92 (0.25)/(0.29) Cohort 1936-1950 1.05/0.94 (0.31)/(0.36) [Table 3, Column 7]
B. Smoking				
Amin et al. (2013)	TwinsUK Database Women, birth cohorts 1924-1974 $N = 1,482$	$H =$ Currently smoking (0.12) $E =$ Years of education (13.44)	F -0.021**** ^a (0.004) [Table 4, Column 2]	F -0.003 (0.014) [Table 4, Column 6]
Lundborg (2013)	Midlife in the U.S. Survey Birth cohorts 1921-1970 $N = 664$	$H =$ Currently smoking (0.213) $E =$ Years of education (13.68)	P -0.046*** (0.007) [Table 3, Column 1]	P -0.014 (0.012) [Table 3, Column 2]
Fujiwara and Kawachi (2009)	Midlife in the U.S. Survey Birth cohorts 1920-1970 $N = 168/183$	$H =$ Currently smoking (odds ratio) $E =$ Years of education	M/F 0.75***/0.75*** (0.05)/(0.05) [Table 3, Column 1/3]	M/F 0.60/1.00 (0.34)/(0.24) [Table 4, Column 1/3]

^aOwn education instrumented by twin's report

Table 3: Health Effects of Education from Quasi-Experiments

Paper	Data & Sample	Variables (mean, full sample if available)	Identification, first stage ^a	Estimates (M=Males, F=Females, P=Pooled)	
				OLS	2SLS
A. Mortality					
Meghir et al. (2017)	Swedish Education Register	H = Mortality, hazard ratio	CLS 7 or 8 to 9 ^b	P	M/F
	National Cause of Death Register	in 1973-2015 (0.200)	in 1948	0.929***	0.994/1.007
	Birth cohorts 1940-1957	E = Years of education (11.4)	First stage: 0.313/0.177	(0.001)	((0.018)/(0.018)
	N = 812,719/749,702		(0.023)/(0.031)	[Table 2, Column 1]	[Table 4, Column 1]
Buckles et al. (2016)	U.S. Vital Statistics Mortality	H = Mortality rate per 1,000,	Vietnam War	M	M
	U.S. Census and NHIS ^c	1981-2007 (138.58)	Draft Avoidance	-22.92***	-25.86***
	White men, Cohorts 1942-53	E = Years of college (1.99)	First stage: 11.41 (2.40)	(3.60)	(4.66)
	G = 600 (birth state-cohort)			[Table 3, Column 1]	[Table 3, Column 5]
	N = 1,994,459				
Davies et al. (2016) ^d	UK Biobank	H = Mortality rate 2006-2014,	CSL ^e 15 to 16 (1972)	M/F	M/F
	National Death Statistics	risk difference (0.9%)	F-stat: 629/689	0.850/1.501***	-0.044***/-0.016*
	Birth cohorts 1935-1971	E = Leaving school after age 15		(0.461)/(0.124)	(0.010)/(0.010)
	N = 9,699/12,439 ^f			[Table S6/S7, Column 8]	[Table S6/S7, Column 2]
Fletcher (2015)	US Diet & Health Study	H = 10-year Mortality rate (0.18)	Various CSL by state	P	P
	1995-96, Low Education	E = Self-reported education	F-stat: 16.36	-0.019***	-0.069
	Birth cohorts 1925-1945	categories (12 years)		(0.002)	(0.078)
	N = 127,550			[Table 3, Column 8]	[Table 3, Column 9]

^aThe F-statistic is reported in cases of multiple instruments or unreported first-stage

^bExtension of compulsory schooling from 7 or 8 depending on municipality to 9 nationally

^cNational Health Interview Survey

^dWorking Paper

^eCompulsory Schooling Laws

^fMen/Women, respectively

Gathmann et al. (2015)	European Social Survey SHARE ^a Human Mortality Database Birth cohorts 1880-1986 <i>N</i> = 21,979/27,237	<i>H</i> = 20-year mortality (odds ratio) <i>E</i> = Years of education	Various CSL by country (1903-1976) First stage: 0.50/0.54 (0.22)/(0.18)	M/F 0.971***/0.986 (0.01)/(0.01) [Table 2/3, Column 2]	M/F 0.983**/0.991 (0.007)/(0.01) [Table 5, Column 2/7]
Clark and Royer (2013)	UK National Death Statistics Health Survey for England General Household Survey Census in 2001 Birth cohorts 1926-40, 1950-65	<i>H</i> = Mortality 1970-2007 (log odds of death) <i>E</i> = Years of education (15.11/16.55) ^b	CSL 14 to 15 in 1947 First stage: 0.45 (0.04) CSL 15 to 16 in 1972 First stage: 0.35 (0.06)		P 0.01/-0.11 ^{b,c} (0.01)/(0.73) [Table 3, Column 1]
Fischer et al. (2013)	Sweden National School Authority Census in 1935 Swedish Death Index Birth cohorts 1924-1931 <i>G</i> = 400 (Region-cohort-gender) <i>N</i> = 731,791	<i>H</i> = 10-year mortality rate in region (0.007) <i>E</i> = Share of schools adopting CSL in region (0.4)	Differential CSL (1936) adoption		M/F -0.010*/0.002 ^d (0.005)/(0.006) [Table 5, Column 1]
Lager and Torssander (2012)	Swedish Census of 2007 Birth cohorts 1943-1955 <i>N</i> = 639,473/608,394	<i>H</i> = Mortality after 40 (hazard ratio) <i>E</i> = Having been through the reform 1949-1962	CSL 8 to 9, differential roll-out (1949-1962)		M/F 0.98/0.97* (0.02)/(0.02) [Table 2, Column 1/2]
Van Kippersluis et al. (2011)	Statistics Netherlands Survey and administrative data Men, Birth cohorts 1912-1922 <i>N</i> = 66,891	<i>H</i> = Mortality rate 81-87 <i>E</i> = Years of education	CSL 6 to 7 in 1928 First stage: 1.039 (0.05)	M -0.011*** (0.002) [Table 1, Column 3]	M -0.026*** (0.004) [Table 4, Column 2]

^aSurvey of Health, Ageing and Retirement in Europe

^b1947/1972 reform, respectively

^cCalculated by dividing reduced-form estimate by the first stage estimate.

^dReduced form estimate

Albouy and Lequien (2009)	French EDP ^a	$H =$ Survival at 50	CSL 14 to 16, Berthoin	P	P
	Cohorts 1950-55 for Berthoin $N = 47,337$	for the Berthoin reform	First stage: 0.28 (0.06)	0.049***/0.028*** ^b	0.13/-0.24 ^b
	Cohorts 1920-25 for Zay	$H =$ Survival at 80	CSL 13 to 14, Zay	(0.0035)/(0.0024)	(0.11)/(0.32)
	$N = 35,828$	for the Zay reform	First stage: 0.11 (0.06)	[Table 2, Column 2/1]	[Table 6, Column 3/2]
Glied and Lleras-Muney (2008)	U.S. Census, 1980 and 1990	$H =$ 4-year mortality rate	Various CSL by state		M/F
	Mortality Detail Files	(0.0067)	1915-1940		-0.0013*/-0.0002
	White US-born cohorts 1901-25 $N = 119,975$	$E =$ Years of education	First stage: 0.054/0.041 (0.012)/(0.012)		(0.0007)/(0.0007) [Table 4, Column 3]
Lleras-Muney (2005)	U.S. Census 1960, 1970, 1980	$H =$ 10-year mortality rate	Various CSL by state	P	P
	NHEFS ^c	1960-70 & 1970-80 (0.11)	1915-1939	-0.017**	-0.061*
	Birth cohorts 1901-1925 $G = 4,792$ (cohort-gender-state) $N = 814,806$	$E =$ Years of education (10.67)	First stage: 0.05 (0.008)	(0.004) [Table 3, Column 1]	(0.025) [Table 4, Column 4]
B. Smoking					
Davies et al. (2016) ^d	UK Biobank	$H =$ Currently smoking	CSL 15 to 16 (1972)	M/F	M/F
	Office of National Statistics	(11.8%)	F-stat: 625/702	-0.171***/-0.116***	0.012/0.025
	Birth cohorts 1935-1971 $N = 9,681/12,405$	$E =$ Leaving school after 15		(0.012)/(0.012) [Table S6/S7, Column 8]	(0.023)/(0.022) [Table S6/S7, Column 2]
Heckman et al. (2016) ^d	National Longitudinal Survey	$H =$ Currently smoking	Unemployment rate,	M	M
	of Youth 1979	$E =$ college graduate	college presence,	-0.209***	-0.172***
	Men, Birth cohort 1979 $N = 2,242$	(vs high school dropout)	college tuition	(0.032) [Table 3, Column 1]	(0.043) [Table 3, Column 7]

^aEchantillon Démographique Permanent^bEstimates for the Berthoin/Zay reform, respectively.^cNational Health and Nutrition Examination Survey Epidemiologic Follow-up Study^dWorking Paper

Huang (2016) ^a	Chinese Family Panel Studies Health and Nutrition Survey Birth cohorts 1955-1993 <i>N</i> = 104,634	<i>H</i> = Currently smoking (0.26) <i>E</i> = Years of education (8.86)	CSL to 9 in 1986, differential adoption by province F-stat: 25.78	P -0.004*** (0.0004) [Table 5, Column 3]	P -0.0134* (0.007) [Table 5, Column 3]
James (2015)	Health Survey of England Birth cohorts 1962-1980 <i>N</i> = 27,927	<i>H</i> = Currently smoking <i>E</i> = Age left school	Expansion of post- compulsory education during 1980s F-stat: 19.02	P -0.069*** (0.002) [Table 5, Column 2]	P 0.001 (0.034) [Table 5, Column 2]
Li and Powdthavee (2015)	Household, Income and Labour Dynamics in Australia (HILDA) Birth cohorts 1939-1972 <i>N</i> = 9,099	<i>H</i> = Currently smoking (0.204) <i>E</i> = Years of education (12.171)	CSL age 14 to 15 differential adoption by state in 1970s F-stat: 71,601	P -0.033*** (0.000) [Table 3, Column 1]	P -0.006 (0.009) [Table 3, Column 1]
Silles (2015)	General Household Survey for Great Britain, 1978-2004 Continuous Household Survey for Northern Ireland, 1983-2004 Birth cohorts 1923-1981 <i>N</i> = GB: 79,271/90,666 NI: 15,298/19,629	<i>H</i> = Currently smoking (0.32/0.32) <i>E</i> = Years of education (11.57/11.45)	CSL 1940s and 1970s schooling age from 14 to 16 GB: F-stat: 61.22/109.95 NI: F-stat: 21.91/23.39	M/F Great Britain -0.034***/-0.039*** (0.001)/(0.001) [Table 5, Column 1]	M/F Northern Ireland 0.012/-0.023 (0.023)/(0.017) [Table 5, Column 1] -0.090**/0.041 (0.002)/(0.001) [Table 6, Column 1]
Clark and Royer (2013)	UK Office of National Statistics Health Survey for England General Household Survey Cohorts 1926-40 & 1950-65 <i>N</i> = 49,421 / 47,177 ^b	<i>H</i> = Currently smoking (0.23/0.30) ^b <i>E</i> = Years of education (15.11/16.55)	CSL 14 to 15 in 1947 First stage: 0.45 (0.04) CSL 15 to 16 in 1972 First stage: 0.35 (0.06)	P -0.06***/-0.140*** ^b (0.009)/(0.008) [Table 5A, Column 2]	P -0.022/-0.001 ^b (0.018)/(0.038) [Table 5A, Column 4]

^aWorking paper^b1947/1972 reform, respectively

Braakmann (2011)	British Labour Force Survey Health Survey in England Birth cohorts 1957-1970 $N = 15,822$	$H =$ Currently smoking (0.3) $E =$ Having any degree (CSE/O-level or above) (0.84)	CSL and birth-month First stage: 0.021 (0.006)	P -0.176*** (0.011) [Table 8, Column 1]	P -0.207 (0.375) [Table 8, Column 1]
Etilé and Jones (2011)	EPCV and ES in France ^a Birth cohorts 1945-1965 $N = 18,785/20,335$	$H =$ Currently smoking (0.31) $E =$ Years of education (9.68)	Post-war expansion of education in France F-stat: 502/283	M/F -0.010***/-0.003*** (0.001)/(0.001) [Table 1, Column 7/3]	M/F -0.027***/-0.019*** (0.003)/(0.004) [Table 1, Column 8/4]
Jürges et al. (2011)	German Micro-census for 1999 and 2003 Birth cohorts 1940-1980 $N = 71,388/71,353$	$H =$ Currently smoking (0.38/0.29) $E =$ Years of education (10.15/10.02)	Differential expansion of education in Germany 1960s-1970s First stage: 41.2/51.8 (9.95)/(9.49)	M/F -0.045***/-0.037*** (0.002)/(0.002) [Table 2, Column 1]	M/F -0.150***/-0.061* (0.056)/(0.036) [Table 3, Column 1]
Kemptner et al. (2011)	German Micro-census in 1989-2003 Birth cohorts 1930-1960 $N = 121,318/124,314$	$H =$ Currently smoking (0.36/0.24) $E =$ Years of education (9.42/9.17)	Differential roll-out of CSL 8 to 9 from 1949-1969 First stage: 0.616/0.654 (0.03)/(0.02)	M/F -0.032***/-0.021*** (0.001)/(0.002) [Table 4, Column 1]	M/F -0.005/-0.001 (0.010)/(0.009) [Table 5, Column 2]
Reinhold and Jürges (2010)	German Microcensus for 1999 and 2003 Birth cohorts 1934-1982 $N = 85,698/86,405$	$H =$ Currently smoking (0.37/0.28) $E =$ Years of education (10.16/10.00)	Abolition of secondary school fees in 1950s First stage: -0.11/-0.10 F-stat: 13.91/14.69	M/F -0.043**/-0.036** (0.001)/(0.001) [Table 2, Column 1]	M/F 0.016/-0.027 (0.068)/(0.069) [Table 3, Column 1]
Park and Kang (2008)	Labor & Income Panel Korean Men High school cohorts 1965-85 $N = 1,611$	$H =$ Refrains from smoking (0.304) $E =$ Years of education (12.479)	Expansion of secondary education and birth order 1970s F-stat: 6.65	M 0.076** (0.017) [Table 2, Column 2]	M 0.088 (0.169) [Table 4, Column 2]

^aEPCV: Permanent Survey on the Conditions of Living of French Households 1996-2003; ES: Health Surveys 1992 and 2003

Grimard and Parent (2007)	U.S. CPS ^a Tobacco Supplements Birth cohorts 1935-1974 <i>N</i> = 227,027	<i>H</i> = Smoking regularly <i>E</i> = Years of education	Vietnam war draft avoidance (dummy for male cohorts, 1945-1950) First stage: 0.161 (0.045)	M -0.043*** (0.001) [Table 2, Column 2]	M -0.0826* (0.0429) [Table 2, Column 2]
De Walque (2007)	U.S. National Health Interview Survey Birth cohorts 1937-1956 <i>N</i> = 73,952	<i>H</i> = Currently smoking <i>E</i> = Years of education past high school	Vietnam War draft avoidance (induction risk) First stage: 1.779 (0.23)	M -0.040*** (0.004) [Table 1, Column 5]	M -0.038* (0.020) [Table 2, Column 5]
Kenkel et al. (2006)	NLSY 1979 Birth cohort 1979 <i>N</i> = 3,205/3,286	<i>H</i> = Currently smoking (0.27/0.26) <i>E</i> = High school graduate (0.72/0.76) ^b	State education policies F-stat: 35.84/19.22	M/F -0.226***/-0.194*** (0.029)/(0.030) [Table 3, Column 2/6]	M/F -0.229***/-0.102 (0.088)/(0.124) [Table 3, Column 3/7]
C. Obesity					
James (2015)	Health Survey of England (HSE) Birth cohorts 1962-1980 <i>N</i> = 25,888	<i>H</i> = Obese (0.147) <i>E</i> = Age left school	Expansion of post- compulsory education during 1980s F-stat 16.50	P -0.020*** (0.002) [Table 3, Column 3]	P -0.100*** (0.030) [Table 3, Column 3]
Li and Powdthavee (2015)	Household, Income and Labour Dynamics in Australia (HILDA) Birth cohorts 1939-1972 <i>N</i> = 8,873	<i>H</i> = Obese (0.279) <i>E</i> = Years of education (12.171)	CSL age 14 to 15 in 1970s differential adoption by state F-stat: 1.5x10 ⁷	P -0.020*** (0.001) [Table 5, Column 4]	P 0.025* (0.014) [Table 5, Column 4]
Brunello et al. (2013)	ECHP ^c SHARE ^d ELSA ^e Birth cohorts 1927-1970 <i>N</i> = 9,013/11,872	<i>H</i> = Obese (0.14/0.13) <i>E</i> = Years of education (12.2/11.1)	Various CSL by country (1949-1972) First stage: 0.42/0.57 (0.059)/(0.065)	M/F -0.006**/-0.012*** (0.001)/(0.001) [Table 5, Column 7/3]	M/F -0.002/-0.014 (0.018)/(0.016) [Table 5, Column 8/4]

^aCurrent Population Survey^bSeparate dummies for high school graduates and GED. We focus on high school graduates.^cEuropean Community Household Panel^dSurvey for Health, Ageing and Retirement in Europe^eEnglish Longitudinal Study of Ageing

Jürges et al. (2011)	German Micro-census for 1991 and 2003 Birth cohorts 1940-1980 $N = 61,892/60,139$	$H = \text{Obese } (0.12/0.09)$ $E = \text{Years of education } (10.15/10.02)$	Differential expansion of education in Germany First stage: 35.2/52.73 (9.80)/(9.79)	M/F -0.018***/-0.019*** (0.001)/(0.001) [Table 2, Column 5]	M/F 0.133***/0.012 (0.051)/(0.024) [Table 3, Column 5]
Kemptner et al. (2011)	German Micro-census in 1989-2003 Birth cohorts 1934-1960 $N = 48,640/49,225$	$H = \text{Obese } (0.16/0.13)$ $E = \text{Years of education } (9.42/9.20)$	CSL 8 to 9 1949-1969 First stage: 0.595/0.663 (0.04)/(0.03)	M/F -0.018***/-0.020*** (0.001)/(0.001) [Table 4, Column 1]	M/F -0.030***/-0.004 (0.014)/(0.010) [Table 5, Column 2]
Reinhold and Jürges (2010)	German Microcensus years 1999, 2003 Birth cohorts 1934-1982 $N = 74,638/73,233$	$H = \text{Obese } (0.10/0.12)$ $E = \text{Years of education } (10.16/10.00)$	Abolition of secondary school fees in 1950s First stage: -0.11/-0.010 F-stat: 11.67/12.61	M/F -0.018**/-0.019** (0.001)/(0.001) [Table 2, Column 5]	M/F 0.030/0.120 (0.059)/(0.070) [Table 3, Column 5]
Kenkel et al. (2006)	NLSY 1979 Birth cohort 1979 $N = 3,248/3,274$	$H = \text{Obese } (0.25/0.27)$ $E = \text{High school graduate } (0.72)^a$	State education policies F-stat: 36.87/21.33	M/F -0.014/-0.039 (0.024)/(0.028) [Table 6, Column 1/5]	M/F -0.008/-0.021 (0.082)/(0.139) [Table 6, Column 3/7]

^aSeparate dummies for high school graduates and GED. We focus on high school graduates.

A Derivations

A.1 Optimality conditions

Optimal schooling age S : The condition for the optimal length of schooling S follows from the dynamic envelope theorem (e.g., Caputo 2005, p. 293):

$$\mathcal{L}(S) \equiv \frac{\partial}{\partial S} \int_0^T \mathfrak{S}(t) dt = \mathfrak{S}(S_-) - \mathfrak{S}(S_+) + \int_0^T \frac{\partial \mathfrak{S}(t)}{\partial S} dt = 0, \quad (20)$$

where S_- indicates the limit in which S is approached from below, and S_+ when approached from above. Noting that state and co-state functions are continuous in S , and $\lambda_L(S)[\tau_S(S) - \tau_\theta(S) - \tau_L(S)] = 0$, we obtain

$$\begin{aligned} & Y(S_+) - Y(S_-) + q_A(S)^{-1} c_S[S_-, \tau_L(S_-)] e^{-\rho S} + p_S(S_-) \\ &= \lambda_S(S_-) p_S(S_-) + f^*(S) I(S - \underline{S}) + \int_S^T \frac{\partial Y(t)}{\partial S} dt \\ &+ q_{\theta/a}(S) [g_\theta(S_-) - g_\theta(S_+)] + q_{h/a}(S) [g_H(S_-) - g_H(S_+)] \\ &+ [p_\theta(S_+) X_\theta(S_+) - p_\theta(S_-) X_\theta(S_-)] + p_H(S) [X_H(S_+) - X_H(S_-)], \end{aligned} \quad (21)$$

where we have replaced the limits S_- and S_+ with S for functions that are continuous in S . The LHS of (21) represents the benefits of entering the labor market consisting of gains in labor income $Y(S_+) - Y(S_-)$ (e.g., no longer does time need to be devoted to schooling $\tau_S(t)$ and / or wages may be higher after graduation than while in school), the monetary value of no longer suffering disutility from being in school $c_S(t)0$, and not having to pay tuition $p_S(t)$. The RHS represents the benefit of staying in school, consisting of the schooling subsidy $\lambda_S(S_-) p_S(S_-)$ (first term), not having to pay a fine if younger than the minimum school-leaving age \underline{S} (second term), higher lifetime earnings from additional schooling (third term), and the value of higher levels of skill investment and of health investment while in school due to the possibly lower opportunity cost of time resulting from lower wages (e.g., if laws constrain labor before the school-leaving age) and from less time that can be devoted to work during schooling years (fourth and fifth term). The sixth term reflects the possibility that the cost of skill investment $p_\theta(t) X_\theta(t)$ is subsidized when in school ($t < S$), providing another benefit of schooling. Finally, if time substitutes for goods and services $X_H(t)$ in the production of health investment, then the final term on the RHS represents the benefit of schooling in terms of reduced expenditures on health investment if the opportunity cost of time is lower during schooling.

Condition (21) weighs multiple costs and benefits of schooling that factor into the decision of the optimal age at which to leave school and enter the labor market. Some of these costs and benefits are arguably not as important as others for most individuals. For example, school-age individuals are generally in good health, resulting in a relatively low marginal value of health $q_{h/a}(t)$ and few medical expenditures $p_H(t) X_H(t)$

(e.g., Galama and Van Kippersluis, 2015b). Therefore, it seems safe to assume that terms involving health production are small.

Optimal skill investment: The first order condition for skill investment is given by

$$q_{\theta/a}(t) = \pi_{\theta}(t), \quad (22)$$

where $\pi_{\theta}(t)$ is the marginal cost of skill investment

$$\pi_{\theta}(t) = \frac{p_{\theta}(t)}{\partial g_{\theta}/\partial X_{\theta}} = \frac{w(t)}{\partial g_{\theta}/\partial \tau_{\theta}}, \quad (23)$$

and $q_{\theta/a}(t)$ is the relative marginal value of skill, which evolves according to

$$\frac{\partial q_{\theta/a}}{\partial t} = q_{\theta/a}(t) \left[r + \frac{\partial d_{\theta}}{\partial \theta} - \frac{\partial g_{\theta}}{\partial \theta} \right] - q_{h/a}(t) \frac{\partial g_H}{\partial \theta} - \frac{\partial Y}{\partial \theta}. \quad (24)$$

Optimal health investment: The first-order condition for health investment is given by

$$q_{h/a}(t) = \pi_H(t), \quad (25)$$

where $\pi_H(t)$ is the marginal cost of health investment

$$\pi_H(t) = \frac{p_H(t)}{\partial g_H/\partial X_H} = \frac{w(t)}{\partial g_H/\partial \tau_H}, \quad (26)$$

and $q_{h/a}(t)$ is the relative marginal value of health, which evolves according to

$$\frac{\partial q_{h/a}}{\partial t} = q_{h/a}(t) \left[r + \frac{\partial d_H}{\partial H} - \frac{\partial g_H}{\partial H} \right] - q_{\theta/a}(t) \frac{\partial g_{\theta}}{\partial H} - \frac{1}{q_A(0)} \frac{\partial U}{\partial H} e^{(\rho-r)t} - \frac{\partial Y}{\partial H}. \quad (27)$$

Optimal consumption The condition for optimal consumption is obtained by taking the derivate of the Lagrange function (11) with respect to X_C . This leads to equation (15).

Optimal length of life T : The condition for the optimal length of life T follows from the dynamic envelope theorem (e.g., Caputo 2005, p. 293):

$$\frac{\partial}{\partial T} \int_0^T \mathfrak{S}(t) dt = \mathfrak{S}(T) = 0. \quad (28)$$

This leads to equation (16).

A.2 Comparative dynamic analyses

Optimal length of schooling S : Using condition (20), we can explore variations in model parameters Z_0 on the optimal schooling decision. Note that $\mathcal{L}(S) = 0$ implies that

$$\left. \frac{\partial \mathcal{L}(S)}{\partial Z_0} \right|_{T,S} \partial Z_0 + \left. \frac{\partial \mathcal{L}(S)}{\partial S} \right|_{T,Z_0} \partial S + \left. \frac{\partial \mathcal{L}(S)}{\partial T} \right|_{Z_0,S} \partial T = 0 \quad (29)$$

This can be rewritten as

$$\frac{\partial S}{\partial Z_0} = \frac{-\left. \frac{\partial \mathcal{L}(S)}{\partial Z_0} \right|_{T,S} - \left. \frac{\partial \mathcal{L}(S)}{\partial T} \right|_{Z_0,S} \frac{\partial T}{\partial Z_0}}{\left. \frac{\partial \mathcal{L}(S)}{\partial S} \right|_{T,Z_0}}. \quad (30)$$

We can further develop (30) into

$$\begin{aligned} & \left[-\left. \frac{\partial \mathcal{L}(S)}{\partial S} \right|_{Z_0,T} \right] \frac{dS}{dZ_0} = \left. \frac{\partial \mathcal{L}(S)}{\partial Z_0} \right|_{T,S} - \left. \frac{\partial \mathcal{L}(S)}{\partial T} \right|_{Z_0,S} \frac{\partial T}{\partial Z_0} \\ & = \left\{ \frac{1}{q_A(S)^2} c_S[S, \tau_L(S)] e^{-\rho S} \right\} \times \left. \frac{\partial q_A(S)}{\partial Z_0} \right|_{T,S} \\ & + [g_\theta(S_-) - g_\theta(S_+)] \times \left. \frac{\partial q_{\theta/a}(S)}{\partial Z_0} \right|_{T,S} \\ & + [g_H(S_-) - g_H(S_+)] \times \left. \frac{\partial q_{h/a}(S)}{\partial Z_0} \right|_{T,S} \\ & - \left[\frac{\partial w(S_+)}{\partial \theta} \tau_w(S_+) - \frac{\partial w(S_-)}{\partial \theta} \tau_w(S_-) \right] \times \left. \frac{\partial \theta(S)}{\partial Z_0} \right|_{T,S} \\ & + \left\{ q_{\theta/a}(S) \left[\frac{\partial g_\theta(S_-)}{\partial \theta} - \frac{\partial g_\theta(S_+)}{\partial \theta} \right] + q_{h/a}(S) \left[\frac{\partial g_H(S_-)}{\partial \theta} - \frac{\partial g_H(S_+)}{\partial \theta} \right] \right\} \times \left. \frac{\partial \theta(S)}{\partial Z_0} \right|_{T,S} \\ & - [w(S_+) - w(S_-)] \left[-\frac{\partial s(S)}{\partial H} \right] \times \left. \frac{\partial H(S)}{\partial Z_0} \right|_{T,S} \\ & + \left\{ q_{\theta/a}(S) \left[\frac{\partial g_\theta(S_-)}{\partial H} - \frac{\partial g_\theta(S_+)}{\partial H} \right] + q_{h/a}(S) \left[\frac{\partial g_H(S_-)}{\partial H} - \frac{\partial g_H(S_+)}{\partial H} \right] \right\} \times \left. \frac{\partial H(S)}{\partial Z_0} \right|_{T,S} \\ & - \frac{\partial Y(T)}{\partial S} \times \frac{\partial T}{\partial Z_0} \\ & + [X_\theta(S_+) - X_\theta(S_-)] \times \frac{\partial p_\theta(S)}{\partial Z_0} + [X_H(S_+) - X_H(S_-)] \times \frac{\partial p_H(S)}{\partial Z_0} \\ & - \left[\frac{\partial w(S_+)}{\partial \underline{S}} \tau_w(S_+) - \frac{\partial w(S_-)}{\partial \underline{S}} \tau_w(S_-) \right] \times \frac{\partial \underline{S}}{\partial Z_0} \\ & + p_S(S) \times \frac{\partial \lambda_S(S)}{\partial Z_0} - (1 - \lambda_S(S)) \times \frac{\partial p_S(S)}{\partial Z_0} + \frac{\partial f^*(S)}{\partial Z_0} I[S - \underline{S}], \end{aligned} \quad (31)$$

where the last three rows represent the direct effect of specific variations with respect to prices $p_\theta(S)$, $p_H(S)$, $p_S(S)$ and institutional characteristics \underline{S} , $\lambda_S(S)$, and $f^*(S)$. By contrast, the other terms in the

expression apply more generally to all types of variations δZ_0 . These reflect indirect effects on the marginal value of wealth $q_A(S)$, the relative marginal value of skill $q_{\theta/a}$ and of health $q_{h/a}(S)$, and the stocks of skill $\theta(S)$ and of health $H(S)$.

Further, by substituting $Z_0 = \underline{S}$, and assuming that individuals value skill $q_{\theta/a}(t)$ more than health $q_{h/a}(t)$ early in life, we obtain the comparative dynamic effect of the minimum school-leaving age on schooling:

$$\begin{aligned}
& \left[-\frac{\partial \mathcal{L}(S)}{\partial S} \Big|_{\underline{S}, T} \right] \frac{dS}{d\underline{S}} \\
\approx & \left\{ \frac{1}{q_A(S)^2} c_S[S, \tau_L(S)] e^{-\rho S} \right\} \times \frac{\partial q_A(S)}{\partial \underline{S}} \Big|_{T, S} \\
+ & [g_\theta(S_-) - g_\theta(S_+)] \times \frac{\partial q_{\theta/a}(S)}{\partial \underline{S}} \Big|_{T, S} \\
- & \left\{ \left[\frac{\partial w(S_+)}{\partial \theta} \tau_w(S_+) - \frac{\partial w(S_-)}{\partial \theta} \tau_w(S_-) \right] - q_{\theta/a}(S) \left[\frac{\partial g_\theta(S_-)}{\partial \theta} - \frac{\partial g_\theta(S_+)}{\partial \theta} \right] \right\} \times \frac{\partial \theta(S)}{\partial \underline{S}} \Big|_{T, S} \\
- & \left\{ [w(S_+) - w(S_-)] \left[-\frac{\partial s(S)}{\partial H} \right] - q_{\theta/a}(S) \left[\frac{\partial g_\theta(S_-)}{\partial H} - \frac{\partial g_\theta(S_+)}{\partial H} \right] \right\} \times \frac{\partial H(S)}{\partial \underline{S}} \Big|_{T, S} \\
+ & \frac{\partial Y(T)}{\partial S} \times \frac{\partial T}{\partial \underline{S}} \\
- & \left[\frac{\partial w(S_+)}{\partial \underline{S}} \tau_w(S_+) - \frac{\partial w(S_-)}{\partial \underline{S}} \tau_w(S_-) \right], \tag{32}
\end{aligned}$$

where the first term on the LHS $-\partial \mathcal{L}(S)/\partial S|_{\underline{S}, T}$ is positive under the assumption of diminishing returns to schooling. Note that $\partial T/\partial \underline{S}$ is also a function of $\partial q_A(S)/\partial \underline{S}$, $\partial q_{\theta/a}(S)/\partial \underline{S}$, etc. However, when fully substituting $\partial T/\partial \underline{S}$ using the information in (37) and (38), the expression becomes too cumbersome to work with. We therefore choose to informally discuss the full effect using (17) in the main text.

Consumption The comparative dynamic effect of the minimum school-leaving age \underline{S} on the optimal consumption path $X_C(t)$ is given by

$$\begin{aligned}
& - \left[\frac{1}{q_A(t)} \frac{\partial^2 U}{\partial X_C^2} e^{-\rho t} - q_{h/a}(t) \frac{\partial^2 d_H}{\partial X_C^2} \right] \frac{\partial X_C(t)}{\partial \underline{S}} \\
= & - \left[\frac{1}{q_A(t)^2} \frac{\partial U}{\partial X_C} e^{-\rho t} \right] \times \frac{\partial q_A(t)}{\partial \underline{S}} \\
- & \left[\frac{\partial d_H}{\partial X_C} \right] \times \frac{\partial q_{h/a}(t)}{\partial \underline{S}} \\
+ & \left[\frac{1}{q_A(t)} \frac{\partial^2 U}{\partial X_C \partial H} - q_{h/a}(t) \frac{\partial^2 d_H}{\partial H \partial X_C} \right] \times \frac{\partial H(t)}{\partial \underline{S}} \tag{33}
\end{aligned}$$

where the term on the LHS of equation (33) is positive under the plausible assumptions of diminishing marginal utility of consumption and constant or increasing returns to scale in the health cost of unhealthy

consumption (see Van Kippersluis and Galama, 2014). The term $\beta_{1,C}$ in (18) is given by the coefficient in front of $\partial q_A(t)/\partial \underline{S}$ and the term $\beta_{2,C}$ is given by the coefficient in front of $\partial q_{h/a}(t)/\partial \underline{S}$, respectively. The sign of both $\partial^2 U/\partial X_C \partial H$ and $\partial^2 d_H/\partial H \partial X_C$ are unknown and hence $\beta_{3,C}$ is hard to sign.

Optimal length of life T : Varying an initial condition, end condition, or model parameter Z_0 in equation (28), we have

$$\frac{\partial \mathfrak{S}(T)}{\partial Z_0} \Big|_{S,T} \partial Z_0 + \frac{\partial \mathfrak{S}(T)}{\partial S} \Big|_{T,Z_0} \partial S + \frac{\partial \mathfrak{S}(T)}{\partial T} \Big|_{S,Z_0} \partial T = 0. \quad (34)$$

From (34) we have

$$\frac{\partial T}{\partial Z_0} = \frac{-\frac{\partial \mathfrak{S}(T)}{\partial Z_0} \Big|_{S,T} - \frac{\partial \mathfrak{S}(T)}{\partial S} \Big|_{T,Z_0} \frac{\partial S}{\partial Z_0}}{\frac{\partial \mathfrak{S}(T)}{\partial T} \Big|_{S,Z_0}}. \quad (35)$$

Consistent with diminishing returns to life extension (Ehrlich and Chuma, 1990), we assume

$$\frac{\partial \mathfrak{S}(T)}{\partial T} \Big|_{S,Z_0} < 0, \quad (36)$$

(see Galama and Van Kippersluis 2015a, 2015b), in which case we can identify the sign of the variation in life expectancy from

$$\text{sign} \left(\frac{\partial T}{\partial Z_0} \right) = \text{sign} \left(\frac{\partial \mathfrak{S}(T)}{\partial Z_0} \Big|_{S,T} + \frac{\partial \mathfrak{S}(T)}{\partial S} \Big|_{T,Z_0} \frac{\partial S}{\partial Z_0} \right). \quad (37)$$

Taking the first derivative of the optimality condition $\mathfrak{S}(T) = 0$ (see 11 and 28) with respect to an initial condition, end condition, or model parameter Z_0 , and holding length of life T and schooling S fixed, we obtain

$$\begin{aligned} \frac{\partial \mathfrak{S}(T)}{\partial Z_0} \Big|_{S,T} &= -\frac{\partial q_\theta(t)}{\partial t} \Big|_{S,T,t=T} \frac{\partial \theta(T)}{\partial Z_0} \Big|_{S,T} + \frac{\partial q_A(T)}{\partial Z_0} \Big|_{S,T} \frac{\partial A(t)}{\partial t} \Big|_{S,T,t=T} \\ &+ \frac{\partial q_H(T)}{\partial Z_0} \Big|_{S,T} \frac{\partial H(t)}{\partial t} \Big|_{S,T,t=T}, \end{aligned} \quad (38)$$

where $\partial \mathfrak{S}(T)/\partial \theta = -\partial q_\theta(t)/\partial t|_{t=T}$, $\partial A(T)/\partial Z_0 = \partial H(T)/\partial Z_0 = 0$ since $A(T)$ and $H(T)$ are fixed, and $\partial q_\theta(T)/\partial Z_0|_T = 0$ since $q_\theta(T) = 0$, regardless of Z_0 . Further, $\partial \mathfrak{S}(T)/\partial \xi(t)|_{S,T} = 0$ for any control function $\xi(t)$, since these are the necessary first-order conditions.³³

The condition for the effect of the minimum school-leaving age \underline{S} on optimal length of life T can be written as (see 37)

$$\text{sign} \left(\frac{\partial T}{\partial \underline{S}} \right) = \text{sign} \left(\frac{\partial \mathfrak{S}(T)}{\partial \underline{S}} \Big|_{S,T} + \frac{\partial \mathfrak{S}(T)}{\partial S} \Big|_{T,\underline{S}} \frac{\partial S}{\partial \underline{S}} \right). \quad (39)$$

³³Note that we distinguish in notation between $\partial f(t)/\partial t|_{t=T}$, which represents the derivative with respect to time t at time $t = T$, and $\partial f(t)/\partial T|_{t=T}$, which represents variation with respect to the parameter T at time $t = T$.

Hence we can infer the sign of $\partial T/\partial \underline{S}$ by studying

$$\begin{aligned}
& \frac{\partial \mathfrak{S}(T)}{\partial \underline{S}} \Big|_{S,T} + \frac{\partial \mathfrak{S}(T)}{\partial S} \Big|_{T,\underline{S}} \frac{\partial S}{\partial \underline{S}} \\
= & - \frac{\partial q_\theta(t)}{\partial t} \Big|_{t=T} \frac{\partial \theta(T)}{\partial \underline{S}} \Big|_{T,S} + \frac{\partial A(t)}{\partial t} \Big|_{t=T} \frac{\partial q_A(T)}{\partial \underline{S}} \Big|_{T,S} \\
+ & \frac{\partial H(t)}{\partial t} \Big|_{t=T} \frac{\partial q_H(T)}{\partial \underline{S}} \Big|_{T,S} + q_A(T) \frac{\partial Y(T)}{\partial S} \Big|_{T,\underline{S}} \frac{\partial S}{\partial \underline{S}} \\
= & \left[\frac{\partial A(t)}{\partial t} \Big|_{t=T} + q_H(T) \frac{\partial H(t)}{\partial t} \Big|_{t=T} \right] \times \frac{\partial q_A(T)}{\partial \underline{S}} \Big|_{T,S} \\
- & \frac{\partial q_\theta(t)}{\partial t} \Big|_{t=T} \times \frac{\partial \theta(T)}{\partial \underline{S}} \Big|_{T,S} \\
+ & q_A(T) \frac{\partial H(t)}{\partial t} \Big|_{t=T} \times \frac{\partial q_{h/a}(T)}{\partial \underline{S}} \Big|_{T,S} \\
+ & q_A(T) \frac{\partial Y(T)}{\partial S} \Big|_{T,S} \times \frac{\partial S}{\partial \underline{S}} \tag{40}
\end{aligned}$$

where $\beta_{1,T}$ in (19) is given by the coefficient front of $\partial q_A(T)/\partial \underline{S}$, $\beta_{2,T}$ is given by the coefficient in front of $\partial \theta/\partial \underline{S}$, etc. Since health $H(t)$ declines by definition near the end of life (approaching H_{\min} from above) and assets $A(t)$ do too (since individuals tend to build wealth early in life and spent it later in life), the first term is positive if increasing the minimum schooling age \underline{S} increases schooling S and schooling in turn generates wealth, extending life (wealth effect). The second term is positive since the marginal value of skills declines near the end of life (skill effect). The third term is positive (marginal value of health effect) since $\partial H(t)/\partial t|_{t=T} < 0$ and $\partial q_{h/a}(T)/\partial \underline{S}|_{T,S} < 0$ (see Galama and Van Kippersluis, 2015b). The last term is negligible, since individuals tend to reduce working hours in old-age (a retirement phase), and earnings at the point of death are zero $Y(T) = 0$.