Center for
European,
Governance and
Economic Development Research

cege

Discussion Papers

Number 293 – November 2016

THE RETURN TO EDUCATION IN TERMS OF WEALTH AND HEALTH

Holger Strulik

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

ISSN: 1439-2305

The Return to Education in Terms of Wealth and Health

Holger Strulik*

October 2016

Abstract. This study presents a new view on the association between education and longevity. In contrast to the earlier literature, which focused on inefficient health behavior of the less educated, we investigate the extent to which the education gradient can be explained by fully rational and efficient behavior of all social strata. Specifically, we consider a life-cycle model in which the loss of body functionality, which eventually leads to death, can be accelerated by unhealthy behavior and delayed through health expenditure. Individuals are heterogeneous with respect to their return to education. The proposed theory rationalizes why individuals equipped with a higher return to education chose more education as well as a healthier lifestyle. When calibrated for the average male US citizen, the model motivates about 50% percent of the observable education gradient by idiosyncratic returns to education, with causality running from education to longevity. The theory also explains why compulsory schooling has comparatively small effects on longevity and why the gradient gets larger over time through improvements in medical technology.

Keywords: Health Inequality, Schooling, Aging, Longevity, Health Expenditure, Unhealthy Behavior, Smoking, Value of Life.
JEL: D91, I10, I20, J24.

* University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de. I would like to thank Daron Acemoglu, Hippolyte d'Albis, Carl-Johan Dalgaard, Gustav Feichtinger, Leonid and Natalia Gavrilov, Franz Hof, Sophia Kan, Petter Lundborg, Wolfgang Lutz, Volker Meier, Arnold Mitnitski, Alexia Prskawetz, Patrick Puhani, Fidel Perez Sebastian, and Engelbert Theurl for useful comments.

1. Introduction

Better educated individuals are, on average, healthier and live longer than less educated people. The literature refers to the strong positive association between education and health as the education gradient or just “the gradient”. According to one popular study, in the year 1990, U.S. Americans aged 25 with any college education live, on average, 5.4 years longer compared to those with only a high school education or less. By the year 2000, the gap increased to 7.0 more years for the better educated (Meara et al., 2008). A similar association between education and health has been observed in many other countries.[1]

In this paper, I propose one particular mechanism that explains the education–health gradient as the outcome of fully rational and efficient behavior of individuals facing an idiosyncratic return to education. By shutting off all other potential channels and carefully calibrating the model for an average white U.S. American male, I show that about half of the observed gradient can be motivated with causality running from education to health and longevity. Please note that I am neither arguing that health behavior is fully rational nor that there is no reverse causality. Instead, I propose a “controlled theoretical experiment” that identifies how much of the gradient can be explained by the proposed mechanism and thus how much is potentially left to be explained by other channels.

To date, the literature has focussed on productive and allocative efficiency in order to explain the gradient. The idea of productive efficiency, based on Becker’s (1965) commodity theory, postulates that less educated individuals “produce” less health from a set of given inputs of time and medical care (see e.g. Grossman, 1972). Allocative efficiency, in contrast, suggests that less educated individuals use different inputs in health production, presumably because they are less informed about their “health technology” (see e.g. Kenkel, 1991). The common theme of both ideas is that less educated people behave less efficiently. If they only had access to the health technology and the knowledge of the better educated, they would care more about their health and live longer.

Empirically, health knowledge and general attitudes seem to play only a minor role for educational differences in health behavior, whereas income (access to resources) and cognitive ability

[1] An incomplete list of the literature on the gradient includes Elo and Preston (1996), Contoyannis and Jones (2004), Case and Deaton (2005), Lleras-Muney (2005), Mackenbach et al. (2008), Conti et al. (2010), and Cutler and Lleras-Muney (2010). See Grossman (2006), Cutler and Lleras-Muney (2006), and Cutler et al., (2011) for surveys of the by now large literature. See Glied and Lleras-Muney (2008) and Cutler et al. (2010) on the rising education gradient.

each account for about 30 percent of the difference, according to one popular study (Cutler and Lleras-Muney, 2010). A recent study by Heckman et al. (2016) demonstrates substantial sorting into schooling by cognitive and non-cognitive abilities and establishes a causal link of education on smoking, physical health, and wages. Likewise, Bijwaard et al. (2015), using Dutch cohort data, argue that at least half of the education gradient can be explained by a selection effect based on cognitive ability. The studies by Contoyannis and Jones (2004), Hong et al. (2015), and Brunello et al. (2016) also document a mediating role of health behavior in the effect of education on health and longevity.

A central element of the proposed theory is an individual-specific return to education. While most of the earlier literature in labor economics assumed that latent factors, such as cognitive ability, may affect earnings but not the return to education itself, there is now increasing evidence that individuals, in fact, differ in their return to education (Card, 1999; Heckman et al., 2016). When dispersion in the return to education is explicitly taken into account, it is typically found to be large. Harmon et al. (2003), for example, estimate a standard deviation of 4 percent for a mean return of education of about 7 percent. A natural explanation for idiosyncratic differences in the return to education are cognitive and non-cognitive abilities (e.g. Heckman and Vytlacil, 2001). While other factors such as school quality the return to education as well (e.g. Card, 1991), concerning the education–health gradient, there exists particularly strong evidence that smart people live longer (Deary, 2008; Der et al., 2009; Calvin et al., 2011; Kaestner and Callison, 2011).

In a nutshell, the proposed mechanism is explained as follows. A higher return to education induces individuals to seek more education and to earn more life time labor income. Since utility derived from consumption is strictly concave at any point of time but linear in time, individuals aspire to live a long life. Simply put, individuals prefer to consume x over 2 years against consuming $2x$ over one year. Rich individuals are particularly interested in smoothing their consumption over a long life because the marginal utility from instantaneous consumption is low when consumption expenditure is high. Well educated and rich individuals are thus induced to spend a higher share of their income on health. Rational individuals balance the marginal return of improving health and the marginal cost of damaging health caused by consumption of unhealthy goods. Since the marginal return of health expenditure is declining while the marginal damage of unhealthy consumption is increasing, educated and wealthy individuals spend not only more on health but also less on unhealthy consumption than less educated ones. Beyond the education channel, however, income is predicted to contribute little to longevity in a multivariate regression. This is because labor income is based on human capital acquired from education.

Methodologically, the present paper is related to the literature on optimal health spending and longevity, to which the studies by Ehrlich and Chuma (1990) and Hall and Jones (2007) are presumably the most popular contributions. In contrast to that literature, the present paper considers education and unhealthy consumption as individual choice variables. More importantly, the present paper has the distinction of being embedded in recent research in gerontology. Existing literature has been built mostly on the health capital model by Grossman (1972). A defining feature of the health capital model assumes that health depreciation is large when the health stock is large. This means that among two persons of the same age t the one in better health, i.e. with more health capital $H(t)$, loses more health in the next period. This counterfactual assumption leads to counterfactual predictions. For example, without further amendments, the health capital model predicts eternal life (Case and Deaton, 2005; Strulik, 2015a) and when death is enforced, the model usually predicts that health investments decline in old age and near death (Wagstaff, 1986; Zweifel and Breyer, 1997; Strulik, 2015a). Health capital depreciation also implies that health shocks in early life (or in utero) have a vanishing impact on health in old age although the opposite is observed (Almond and Currie, 2011).

Most importantly, health capital is a latent variable, not observed by doctors and medical scientists, a fact that confounds any serious calibration of the model. Here, we build on the health deficit model developed by Dalgaard and Strulik (2014), which avoids the problematic features of the Grossman model, and can be calibrated in a straightforward manner using the so-called frailty index (Mitnitski et al., 2002a,b, 2005, 2006, 2007). Since the calibration provides no degrees of freedom, the model can be used to assess health issues quantitatively.[2]

The frailty index counts the proportion of the total potential deficits that an individual has, at a given age. The list of potential deficits ranges from mild deficits such as impaired vision to severe deficits such as dementia. In developed countries, the average adult accumulates 3-4% more deficits from one birthday to the next (Mitnitski et al., 2002a; Harttgen et al., 2013). Here, as in Dalgaard and Strulik (2014), we assume that health deficit accumulation can be slowed

[2] In Dalgaard and Strulik (2014) we showed that the law of health deficit accumulation has a deep gerontological foundation. As suggested by McFadden (2005) it is built upon an application of reliability theory to the functioning of the human body (see also Gavrilov and Gavrilova, 1991).

down by health expenditure. Additionally, we take into account the fact that unhealthy behavior quickens health deficit accumulation.

The present paper takes the education decision of individuals endogenously, in contrast to most of the related literature.[3] It will be shown that this treatment makes a considerable difference for the magnitude of the predicted education gradient. If, instead, education is exogenously imposed, then there is no positive impact on health beyond the point at which the time spent on education is optimal from the individual viewpoint. The model thus provides an explanation for why the health gradient for the differential education of monozygotic twins is found to be relatively small (Fujiwara and Kawachi, 2009; Lundborg, 2013). The reason is that twins are likely to be endowed with a similar return to education (cognitive ability), which trumps the length of the education period in the determination of human capital and health behavior. The model also provides an explanation for why, generally, exogenous variation in education levels can be expected to be associated with substantially smaller education gradients, as frequently found in empirical studies based on compulsory schooling laws (e.g. Lleras-Muney, 2005; Oreopoulos, 2006).

The paper is organized as follows. The next section sets up the model. The standard approach of human capital accumulation is modified to account for the empirical wage for age curve and then integrated together with the unhealthy consumption behavior into the health deficit model. Section 3 presents the analytical results. It shows that the optimal life style is governed by conditions for (i) optimal expenditure on health and unhealthy goods, (ii) optimal aging (the evolution of the expenditure profiles with age), (iii) optimal length of schooling, (iv) optimal financial management, and (v) optimal time of death. These optimality conditions are simple enough to allow for an intuitive interpretation. In Section 4, the model is calibrated for a 16 year old US American male in the year 2000. Section 5 present the results. It is shown that an increase in the return to education that motivates one more year of education results in a gain in longevity of about half a year. After corroborating the main result with a series of robustness checks the paper finishes by comparing the education gradient for voluntary and compulsory education and by showing that the education gradient widens with ongoing medical technological progress.

---

[3] A recent study in the Grossman (1972) tradition of health capital accumulation that considers endogenous education is Galama and van Kippersluis (2015). It provides a rich discussion of the potential shapes of life cycle trajectories and their comparative dynamics without considering a quantitative assessment of the explanatory power for the education gradient. Another related study by van Kippersluis and Galama (2014) investigates the effect of wealth shocks on unhealthy consumption within the Grossman framework without considering the influence of education on health behavior.

2. Model Setup

2.1. **The objective function.** Consider a young adult at the end of the compulsory schooling period. Later on, in the numerical part, this will be a 16 year old with 9 years of education. For simplicity let the initial age at this stage be normalized to zero. The – yet to be determined – date of death is denoted by T. At each age t, the person experiences utility from consumption of health-neutral goods $c(t)$ and unhealthy goods $u(t)$. We consider a minimum setup for the elaboration of the health and education nexus in which it suffices to treat health expenditure purely instrumental and not in itself utility-enhancing. Likewise, education is purely instrumental in achieving higher labor income and not in itself utility enhancing. The instantaneous utility function is assumed to be iso-elastic such that life-time utility is given by (1).

$$V = \int_0^T \mathrm{e}^{-\rho t} \left\{ v\left(\; c(t) + \beta u(t) \; \right) \right\} \; \mathrm{d}t \tag{1}$$

with $v(x) = (x^{1-\sigma} - 1)/(1-\sigma)$ for $\sigma \neq 1$ and $v(x) = \log(x)$ for $\sigma = 1$. The inverse $1/\sigma$ is the intertemporal elasticity of substitution. Consumption is measured such that x is always larger than one, implying that at each age, utility is positive, a fact that makes a longer life desirable.

The parameter β measures the pleasure from consuming the unhealthy good. If $\beta > 1$, the person likes unhealthy consumption more than health-neutral consumption. If $\beta < 1$, the person prefers health-neutral consumption and unhealthy goods are consumed only if they are less expensive. Later on, the parameter β is a useful device to pin down the price elasticity of demand for unhealthy goods. In the benchmark calibration, cigarettes will represent the unhealthy good. The feature that health-neutral and unhealthy consumption enter utility additively is harmless for the calibration since the desire for unhealthy consumption is counterbalanced by the restraint from the potential of the good to damage health. Perfect substitution between goods is a simple method in order to include the special case of complete abstention from consuming unhealthy goods.

2.2. **Budget constraint.** Let health expenditure be denoted by h. The price of health-neutral goods is normalized to unity, the price for health goods is denoted by p, and the price of unhealthy goods is denoted by q. Total expenditure is thus given by $e = c + ph + qu$. The – yet to be determined – length of the voluntary education period is denoted by s and the predetermined age of retirement is at R. From s to R, the individual receives a wage $w(t)$ per unit of human capital. Human capital $H(s,t)$ varies with age t and the time spend on education s. We allow for aggregate productivity growth such that the wage per unit of human capital grows at rate g_w, which is taken as given by the individual.[4]

For simplicity, there are no restrictions on the capital market; the individual can borrow or lend at rate r. An individual that holds capital k thus faces the budget constraint (2).

$$\dot{k}(t) = \chi w(t)H(s,t) + rk(t) - c(t) - ph(t) - qu(t), \tag{2}$$

with $k(0) = k_0$ and $k(T) = \bar{k}$. We abstain from modeling a bequest motive such that k_0, and $\bar{k}$, as well as all prices, are taken as given by the individual. In (2) χ is an indicator function, $\chi = 1$ for $t \in [s, R]$ and $\chi = 0$ otherwise. People save for consumption and for health interventions after retirement. Although aging and death are certainly stochastic phenomena, we follow the related literature (e.g. Ehrlich and Chuma, 1990; Hall and Jones, 2007) and treat, for simplicity, the problem deterministically. This means that the model neglects a precautionary savings motive and thus potentially somewhat underestimates the propensity to save for old age.[5]

2.3. **Health Deficit Accumulation.** Inspired by recent research in gerontology, we consider a physiologically founded aging process, where aging is understood as increasing loss of redundancy in the human body (Gavrilov and Gavrilova, 1991, Arking, 2006). For young people, the functional capacity of organs is about tenfold higher than needed for mere survival (Fries, 1980). Yet as people age, organs weaken, and people people become more fragile. An empirical measure of human frailty has been developed by Mitnitski and Rockwood and various coauthors in a series of articles (Mitnitski et al., 2002a,b; 2005; Rockwood and Mitnitski, 2006). They propose to compute the frailty index as the proportion of the total potential health deficits of an individual, at a given age. As suggested by aging theory, Mitnitski et al. (2002a) confirm that the frailty index number (the number of health deficits), denoted by $D(t)$ increases exponentially with age t, $D(t) = E + be^{\mu t}$. They estimate that $\mu = 0.043$ for Canadian men. This "law of health deficit accumulation" explains around 95% of the variation in the data, and its parameters are estimated with great precision. Conceptually, the rate of aging μ is given to the adult individual. From a physiological viewpoint, however, it can be explained by applying reliability theory to

[4] Fixing the retirement age shuts off reverse causality by the Ben-Porath (1967) mechanism, i.e. the potential impact of increasing longevity on desired years of education. See Hazan (2000); Strulik and Werner (2016).

[5] Strulik (2015b) investigates a stochastic version of the basic model of optimal aging by Dalgaard and Strulik (2014) and shows that the quantitative predictions are robust against the consideration of death as a stochastic event.

human functioning (Gavrilov and Gavrilova, 1991). Rockwood and Mitnitski (2007) show that elderly community-dwelling people in Australia, Sweden, and the U.S. accumulate deficits in an exponential way very similar to Canadians.

In order to utilize the findings of Mitnitski and Rockwood for the present work, I begin with differentiating the frailty law with respect to age, $\dot{D}(t) = \mu (D(t) - E)$.[6] Following Dalgaard and Strulik (2014), we assume that the factor E, which slows down exponential aging, can be increased by deliberate health expenditures. Furthermore, we assume that E can be diminished by unhealthy behavior. Specifically, I propose the following parsimonious refinement of the process of deficit accumulation:

$$\dot{D}(t) = \mu \left[\; D(t) - a - Ah(t)^{\gamma} + Bu(t)^{\omega} \; \right], \qquad 0 < \gamma < 1, \qquad \omega > 1. \tag{3}$$

The parameter a captures environmental influence on aging beyond the control of the individual, the parameters $A > 0$ and $0 < \gamma < 1$ reflect the state of health technology, and h is health investment. While A refers to the general efficiency of health expenditure in the maintenance and repair of the human body, the parameter γ specifies the degree of decreasing returns of health expenditure. Likewise, the parameter B measures the general unhealthiness of the unhealthy good and the parameter ω measures the return in terms of deficits from unhealthy consumption. It is reasonable to assume that there are increasing returns from unhealthy consumption in terms of health deficits (cf. binge drinking vs. the occasional glass of red wine). This is captured by $\omega > 1$. The parameter μ measures the biological rate of bodily deterioration when there is neither unhealthy consumption nor health expenditure. This physiological parameter is called *the force of aging*, as it drives the inherent and inevitable process of human aging.[7]

Initial health deficits are given for the young adult, $D(0) = D_0$. Furthermore, following Rockwood and Mitnitski (2006), we assume that a terminal frailty exists at which the individual expires, $D(T) = \bar{D}$. Problem (1) – (3) thus constitutes a free terminal time problem with known terminal states.

[6] In order to see that a larger autonomous component E implies less deficits at any given age, notice that the solution of the differential equation is $D(t) = (D_0 - E)\,\mathrm{e}^{\mu t} + E = D_0 \mathrm{e}^{\mu t} - E(\mathrm{e}^{\mu t} - 1)$, in which D_0 are initial health deficits.

[7] If individuals' investment in health influences E, then one may wonder if the "frailty law" should still work empirically, as E then is expected to exhibit individual-level variation. It should; but the cross-section estimate for E should be interpreted as the average level in the sample in question (see Zellner, 1969).

2.4. **Education.** The length of the schooling period s is a choice variable for individuals. Specifically, we assume that human capital of an individual of age t with s years of schooling is given by (4).

$$H(s,t) = \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t\right] - \delta \exp\left[\alpha_2 t\right], \tag{4}$$

The parameters θ and ψ capture the return to education and η is the return to experience (learning on the job). For $\psi > 0$ there are decreasing return to education. The marginal return to education is given by $\theta s^{-\psi}$. Later on, we consider variation of θ in order to investigate the impact of the return to education on education, health behavior, and health outcomes.

For $\alpha_1 = \delta = 0$ the schooling function boils down to the standard model (Bils and Klenow, 2000). The parameter α_1 controls for the impact of age on the return to education. The parameter δ controls for the feedback of age on general human capital. For sufficiently high δ and α_2, human capital starts to decline after a certain age. The parameters α_1, α_2, and δ are conceptualized as being job specific. While some cognitive skills and motor skills start deteriorating around the age of 30 or even earlier, so called *crystallized abilities*, i.e. the ability to use knowledge and experience, remain relatively stable until most of adulthood and start declining after the age of 60, or even later. As a result, some measures of psychological competence, like verbal skills and inductive reasoning start declining "only" around age 50 (Schaie, 1994). Structural change that makes occupations utilizing crystallized abilities, on average, more widespread could be described by a secular decline of α_1.[8]

To summarize, human capital varies with education and age. The return to education, captured by θ, is a shifter of the human capital function. Together, the three parameters α_1, α_2, and δ are used to estimate the empirical wage for age curve (Murphy and Welch, 1990). It turns out that α_2 is estimated to be of about the same magnitude as the force of aging μ.[9]

3. Solution

3.1. **Summary.** The problem of the individual is to maximize (1) subject to (2) – (4). In order to solve the problem conveniently, it is helpful to define a measure of aggregate consumption

[8] Structurally, (4) is isomorph to the standard Mincer equation containing a negative term for "experience squared". Conceptually, however, it is hard to imagine why human capital should start to decline when individuals have "too much experience".

[9] It may thus appear tempting to let health deficits enter the human capital equation. The direct impact of health on wages, however, is estimated to be rather small (Jaeckle and Himmler, 2010; Hokayem and Ziliak, 2014). Declining mental capabilities are insufficiently approximated by declining bodily deficits. Conceptually, avoiding health deficits in the equation has the advantage to shut down a channel of potential reverse causality running from health to education. It is thus a useful device for our controlled theoretical experiment.

$x \equiv c + \beta u$ and replace c in (1) and (2). Details of the computation are included in the Appendix. The first order conditions can be summarized by the following, nicely interpretable conditions.

$$\mathrm{e}^{-\rho t} x^{-\sigma} = \lambda_k \mathrm{e}^{-rt} \tag{5}$$

$$\lambda_k \mathrm{e}^{-rt} p = -\lambda_D \mu A \gamma h^{\gamma-1} e^{-\mu t} \tag{6}$$

$$\lambda_k \mathrm{e}^{-rt}(q-\beta) = -\lambda_D \mu B \omega u(t)^{\omega-1} \mathrm{e}^{-\mu t} \quad \text{for } \beta > q \text{ and } u = 0 \text{ otherwise.} \tag{7}$$

$$(\theta s^{-\psi} - \eta) \exp\left(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right) \frac{\mathrm{e}^{(g_w - r + \eta - \alpha_1)(R-s)} - 1}{g_w - r + \eta - \alpha_1} = \exp\left(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right) - \delta \mathrm{e}^{\alpha_2 s}, \tag{8}$$

in which λ_k is the shadow price of capital and λ_D is the shadow price of health deficits.

3.2. **Optimal Consumption Profiles.** Condition (5)–(7) determine the optimal structure of consumption. Insert (5) into (6) to obtain that $px^{-\sigma}\mathrm{e}^{-\rho t} = -\lambda_D \mu A \gamma h^{\gamma-1} e^{-\mu t}$. This condition requires that the marginal cost of health expenditure in terms of foregone utility from instantaneous consumption (the left-hand side) equals the marginal benefit of health expenditure in terms of slower health deficit accumulation (the right-hand side). Both expressions are in present value and thus expenditure is discounted by ρ while health deficits are discounted by μ. Notice that health deficits are “a bad”. The associated shadows price λ_D is thus negative, indicating that it is, ceteris paribus, better to have fewer health deficits.

When unhealthy goods are consumed, we obtain from (5) and (7) that $(\beta - q)x^{-\sigma}e^{-\rho t} = -\lambda_D \mu B \omega u(t)^{\omega-1} \mathrm{e}^{-\mu t}$. This condition requires that the marginal utility from consuming an unhealthy good instead of a health neutral good (the left-hand side) equals the marginal cost of unhealthy consumption in terms of increased speed of health deficit accumulation (the right-hand side). Again, both expressions are in present value. Combining (5), (6), and (7), we obtain

$$\frac{\mu A \gamma h^{\gamma-1}}{p} = \frac{\mu B \omega u(t)^{\omega-1}}{\beta - q}. \tag{9}$$

The left-hand side of this condition shows the marginal return in terms of health deficits when one unit of income is spent on health instead of health neutral goods. The right-hand side shows the marginal cost, in terms of health deficits, when one unit is spent on unhealthy goods instead of health neutral goods. Notice that the return to health investment declines when individuals spend more on health. Then, in order to balance the equation, the right-hand side must also decline. Given the increasing damage of unhealthy consumption ($\omega > 1$), individuals reduce their unhealthy consumption. Intuitively, this makes sense: individuals who care a lot about health and longevity are careful when consuming unhealthy goods and increase their spending to preserve their health.

Solving (9) and taking a potential corner solution into account, we obtain

$$u^{\omega-1} = \begin{cases} \frac{\gamma(\beta-q)A}{\omega p B} \cdot \frac{1}{h^{1-\gamma}} & \text{for } \beta > q \\ 0 & \text{for } \beta \leq q. \end{cases} \tag{10}$$

Consuming unhealthy goods requires that utility derived from unhealthy consumption exceeds the utility from health-neutral consumption, or that the unhealthy good is cheaper than the health-neutral good, or both. In terms of policy, it is important to note that demand for unhealthy goods is lower at higher prices and that there exists a preemptive price, $q = \beta$, which deters unhealthy consumption. If unhealthy consumption exists, then condition (10) furthermore predicts that its extent is large if the resulting health damage is low (B is low), if medical efficiency in repairing damage is large (A is large), or if the price of health goods p is low.

3.3. **Optimal Aging.** Differentiating (5) to (7) with respect to age we obtain:

$$g_x \equiv \dot{x}/x = \frac{r-\rho}{\sigma} \tag{11}$$

$$g_h \equiv \dot{h}/h = \frac{r-\mu}{1-\gamma} \tag{12}$$

$$g_u \equiv \dot{u}/u = \frac{r-\mu}{1-\omega}. \tag{13}$$

These Euler equations show how optimal expenditure evolves through life. Together, they determine optimal aging of the person since the evolution of health deficits D depends on life-cycle health behavior (h and u). Condition (11) is the familiar Euler equation for consumption, in our case, for the aggregate measure of consumption x. It has the usual textbook interpretation. The "Health-Euler" (12) suggests that one should postpone health expenditure to later periods of life in favor of financial investment if the return on investment r is relatively high. On the other hand, if the force of aging, μ is high, implying that health deficits accumulate very fast at the end of life, late-in-life health investments would then be considered a relatively ineffective way of prolonging life. In this case, one should invest more heavily early in life. The expenditure profile for health is also influenced by the curvature of the health investment function: a smaller γ implies a lower growth rate of health expenditure. Intuitively, if γ is small, diminishing returns set in rapidly, which makes it optimal to smooth health expenditure over the life cycle.

The Euler equation (13) prescribes that expenditure for unhealthy consumption should decrease with age (since $\omega > 1$). For example, intuitively, the damage caused by alcohol consumption is relatively harmless at young age when there is abundant redundancy in the body. At an advanced age, drinking frequently could be lethal and consumption is best reduced to an occasional glass of red wine. Since unhealthy consumption is negatively correlated with health expenditure, the expenditure profile is steep (in absolute value) if the interest rate is high and the rate of deficit accumulation is low.

3.4. **Optimal Schooling.** The optimal duration of education s requires that the marginal loss from postponing entry into the labor market, $\mathrm{e}^{-rs}w(s)H(s,s)$, equals the marginal gain from extending education, $\int_s^R \frac{\partial}{\partial s}\mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t$. Inserting the respective values leads to condition (8), in which g_w denotes the growth rate of the wage per unit of human capital, $w(t) = \bar{w}\exp(g_w t)$. It is assumed to be given by the rate of aggregate productivity growth and to be exogenous to the individual. The left-hand side of (8) displays the marginal gain from education and the right-hand side displays the marginal loss from postponing entry into the labor market.

Observe that the solution for (8) is independent from life span T, implying that – as long as retirement age R is smaller than T, which is henceforth assumed – there is no influence of longevity on education. This means that any causality behind the education gradient runs from increasing education to increasing health and longevity. Removing reverse causality is helpful in order to identify the size of the gradient that can be explained by increasing education.

3.5. **Optimal Financial Management and Human Wealth.** Integrating (2) and inserting initial and terminal values provides the life-time budget

$$k_0 + W(s,R)\mathrm{e}^{-rs} - \frac{x(0)}{g_x - r}\left(\mathrm{e}^{(g_x - r)T} - 1\right) - \frac{ph(0)}{g_D}\left(\mathrm{e}^{g_D t} - 1\right) - \frac{(q-\beta)u(0)}{g_\omega}\left(\mathrm{e}^{g_\omega T} - 1\right) = \bar{k}\mathrm{e}^{-rT} \tag{14}$$

with $g_D \equiv (\gamma r - \mu)/(1-\mu)$ and $g_\omega \equiv (\omega r - \mu)/(1-\omega)$. The expression $W(s,R)$ denotes human wealth, i.e. the life-time labor income acquired between leaving school and retirement, $W(s,R) = \int_s^R \mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t$. It is obtained as (15).

$$W(s,R) = \frac{\bar{w}\exp\left(\theta s^{\frac{1-\psi}{1-\psi}} - \eta s\right)}{\eta + g_w - r - \alpha_1}\cdot\left[\mathrm{e}^{(\eta+g_w-r-\alpha_1)R} - \mathrm{e}^{(\eta+g_w-r-\alpha_1)s}\right] - \frac{\bar{w}\delta\left[\mathrm{e}^{(g_w+\alpha_2-r)R} - \mathrm{e}^{(g_w+\alpha_2-r)s}\right]}{\alpha_2 + g_w - r}. \tag{15}$$

3.6. **Optimal Death.** At the individually optimal time of death two conditions have to hold. First, the accumulated health deficits must have reached the terminal value $\bar{D}$, and, second, the Lagrangian associated with Problem (1)– (4) must assume the value zero. Turning towards the first condition, integrating (3) provides $D(T)$ and thus (16).

$$\bar{D} = D(T) = D_0 e^{\mu T} - a\left(e^{\mu T} - 1\right) - \frac{\mu A h(0)^\gamma e^{\mu T}}{g_D}\left(e^{g_D T} - 1\right) + \frac{\mu B u(0)^\omega e^{\mu T}}{g_\omega}\left(e^{g_\omega T} - 1\right). \tag{16}$$

Evaluating the Lagrangian at T provides the condition (17).

$$\begin{aligned} 0 &= v(T) + \xi \bar{D}^\nu + x(T)^{-\sigma}\left\{r\bar{k} - x(T) - (q-\beta)u(T) - ph(T)\right\} \\ &\quad - \frac{x(T)^{-\sigma} p h(T)^{1-\gamma}}{\mu A \gamma}\left\{-\mu a - \mu A h(T)^\gamma + \mu B u(T)^\omega + \mu D(T)\right\} \end{aligned} \tag{17}$$

with $x(T) = x(0)e^{g_c T}$, $h(T) = h(0)e^{g_h T}$, $u(T) = u(0)e^{g_u T}$, and $v(T) = (x(T)^{1-\sigma} - 1)/(1-\sigma)$ for $\sigma \neq 1$ and $v(T) = \log(x(T)$ otherwise. Together, (8), (9), and (15) – (17) establish a system of 5 (non-linear) equations in 5 unknowns: the initial values $x(0), h(0)$, and $u(0)$, optimal education s, and the optimal age of death T. I use Matlab's fsolve routine in order to obtain the solution.

4. Calibration

In the following calibration study, we consider an average 16 year old US American male in the year 2000. The initial age is set to 16 years, corresponding to model-age zero, because individuals below roughly the age of 16 are not subject to increasing morbidity (Arking, 2006) and are presumably not well described by the law of increasing frailty. Furthermore, in many US states as well as in many countries around the world, schooling is compulsory until about 16 years of age. This means that there is not really an *individual decision* about education below this age. An implication is that a person of model age zero has spent already 9 years on education. The effect of past education is captured by the initial endowment $\bar{w}$.

In order to calibrate the model to US data, I begin with employing the Health Euler equation (7). From the data in Keehan et al. (2004), I set the growth rate of health expenditure over the life cycle g_h to 0.021. From Mitnitski et al. (2002a), I take the estimate of $\mu = 0.043$ for (Canadian) men, and I set $r = 0.06$ (e.g., Barro et al., 1995). This produces the estimate $\gamma = 1 - (r-\mu)/g_h = 0.19$, which squares well with the independent estimates obtained by Hall and Jones (2007).[10]

In the year 2000, the average life-expectancy of a 20 year old US American male was 55.6 years (death at age 75.6). From Mitnitski et al.'s (2002a) regression analysis, I infer terminal health deficits $\bar{D} = D(75.6) = 0.1005$ and initial health deficits $D(0) = D(16) = 0.0261$. In order to get an estimate of a, I

[10] As explained in Section 2, the average rate of aging within the US and Canada are similar (Rockwood and Mitnitski, 2007). Thus, using the estimate from the Canadian sample should be a good approximation. While Rockwood and Mitnitski (2007) stress the similarity of their results for US and Canadian populations, they do not report the detailed results for their US analysis.

assume that before the 20th century the impact of medical technology on *adult* mortality was virtually zero. In the year 1900, the life expectancy of a 20 year old U.S. American was 42 years (death at 62; Fries, 1980), implying that a 16 year old expected to live for 46 additional years. I set a such that a person who abstains from unhealthy consumption and has no access to life prolonging medical technology expects $T = 46$. From this value, I get the estimate $a = 0.01427$.

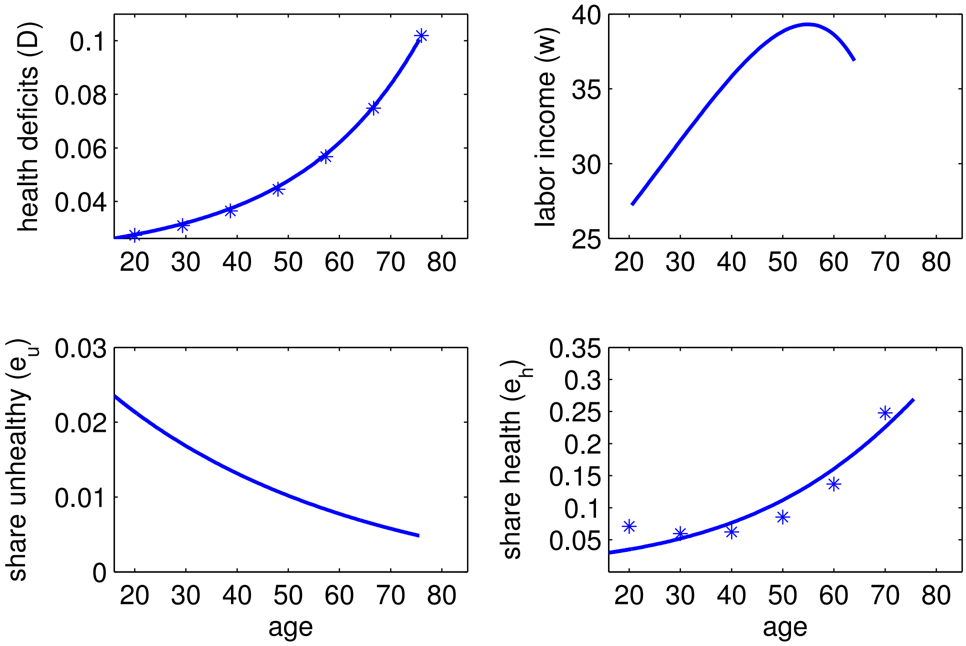
The parameters entering the equation for optimal education (8) are potentially individual-specific and varying their size is the most interesting numerical experiment. In the following, I estimate parameter values for the average US American person. I set η to 0.05 and ψ to 0.28 according to the medium scenario in Bils and Klenow (2000). Since the benchmark American attends school for 13.5 years, I adjust θ such that the marginal return to education $\theta s^{-\psi}$ equals 0.068 for $s = 13.5$, which is the return to schooling beyond the eighth year estimated by Psacharopoulos (1994) and since then applied by Hall and Jones (1999) and many others. This provides the estimate $\theta = 0.141$. I estimate α_1, α_2, and δ, by requiring that the 16 year old reference American wants to attend school for 4.5 more years such that he acquires altogether 13.5 years of schooling (which is the US average in the year 2000 according to Turner et al., 2007) and such that the wage for age curve attains its maximum at age 55 at a value of about 1.2 times the labor income at age 30, as observed by French (2005). These three constraints provide the estimate $\alpha_1 = 0.039$, $\alpha_2 = 0.046$, and $\delta = 0.135$. Remarkably, the estimate of the decay rate of human capital α_2 is close to the value of the force of aging μ (which is 0.043).

The growth rate of the wage per unit of human capital g_w, that is aggregate productivity, is set to an annual rate of one percent based on US TFP growth between 1995-2000 (Jorgenson et al., 2008). I set $R = 48 = 64 - 16$ corresponding to the average US retirement age, and adjust the initial unit wage $\bar{w}$ such that total labor income across all working ages equals \$35,320, which is the average annual pay for workers in the year 2000 (BLS, 2011). This implies $\bar{w} = 19,605$. In the benchmark run, I set $k_0 = \bar{k} = 0$.

For the benchmark run, I normalize the price of health and unhealthy goods to unity, $p = q = 1$. This is an interesting benchmark case because it eliminates the price channel through which the uneducated (and poor) may have an incentive to consume more unhealthy goods and spend less on health. I then check the robustness of results with respect to alternative prices.

Most of the available empirical literature on consumption of unhealthy goods is on cigarettes and tobacco. It thus seems reasonable to capture the characteristics of cigarette consumption in a benchmark case and then proceed with a sensitivity analysis. Preston et al. (2010) estimate that smoking takes away 2.5 years of life-expectancy of 50 year old US males. The price elasticity of demand for cigarettes is estimated to be about -0.5 (Chaloupka and Warner, 2000). On average, Americans spent \$ 319 for cigarettes in the year 2000 (BLS, 2002). Based on these observations, the remaining six parameters, A, B, β, ρ, σ and ω are calibrated jointly such that: (i) the model predicts the actual accumulation of health deficits over life (as estimated by Mitnitski and Rockwood, 2002), (ii) death occurs at the moment when $\bar{D}$ health deficits have been accumulated at an age of 75.6 years, (iii) the health share of total expenditure approximates the average age-specific expenditure shares of American adults, (iv) the average American spends on average \$ 319 per year on unhealthy goods, (v) consumption of the unhealthy good costs 2.5 years of longevity, and (vi) the price elasticity of demand for unhealthy goods is about -0.5. This leads to the estimates $A = 0.00168$, $B = 2 \cdot 10^{-7}$, $\beta = 4$, $\rho = 0.085$, $\omega = 1.4$, and $\sigma = 1.16$. The estimate of σ fits nicely with recent empirical studies suggesting that the intertemporal elasticity of substitution is around unity (e.g. Chetty, 2006).

Figure 1: Optimal Life Cycle Trajectories: Basic Run



Parameters: $a = 0.01427$, $A = 0.00168$, $\alpha_1 = 0.039$, $\alpha_2 = 0.046$, $\delta = 0.135$, $\eta = 0.05$, $\theta = 0.141$, $\bar{w} = 19,605$, $g_w = 0.01$, $\mu = 0.043$, $r = 0.06$, $\rho = 0.085$, $\sigma = 1.16$, $D_0 = 0.0261$, $\bar{D} = 0.105$, $p = q = 1$, $\beta = 4$, $B = 2 \cdot 10^{-7}$, $\omega = 1.4$, $k_0 = \bar{k} = 0$. Stars: data. Annual labor income measured in thousands of Dollars.

Figure 1 shows the implied trajectories over the life cycle of the Reference American. Stars in the health deficit panel represent the original data from Mitnitski and Rockwood (2002). The model explains the actual accumulation of health deficits quite well. The upper-right panel displays the calibrated invertedly-u-shaped trajectory of labor income across ages. The lower-left panel shows the expenditure share of unhealthy consumption $e_u \equiv qu/(c+qu+ph)$. When young, the Reference American is predicted to spend about 2.2 percent on unhealthy goods. The expenditure share is subsequently declining until death. The lower-right panel shows the health expenditure share $e_h \equiv ph/(c + qu + ph)$. Stars indicate the actual age-specific shares in the year 2000. The health expenditure data is taken from Meara et al. (2004) and the data for total expenditure is taken from BLS (2002). The model mildly underestimates the health expenditure share at young ages and overestimates it at middle ages. Altogether, however, the predicted health expenditure share fits the data reasonably well.

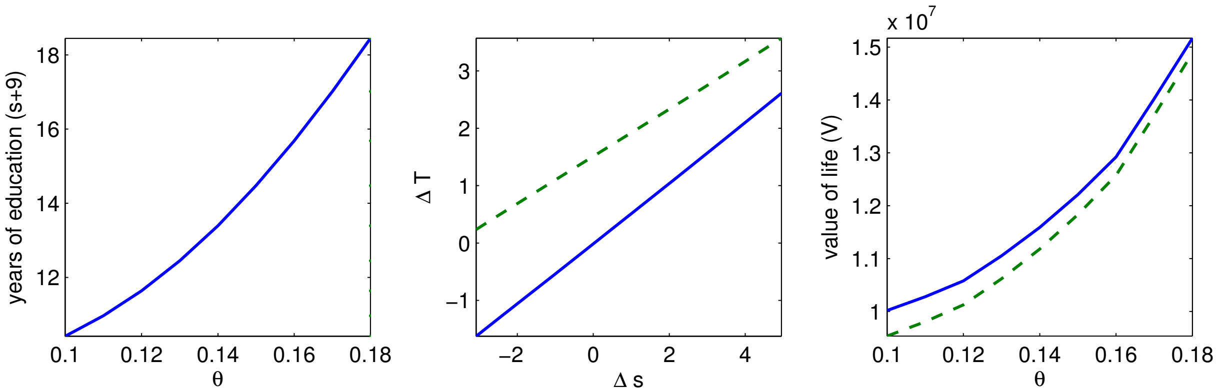
5. Results

5.1. The Return to Education in Terms of Wealth and Health, and the Value of Life. The major aim of this paper is to identify the education gradient. As such, the first experiment considers a person endowed with a higher return to education. As motivated in the Introduction, the experiment is inspired by research in labor economics which acknowledges that the "return to education is not a single parameter in the population, but rather a random variable that may vary with other characteristics of individuals" (Card, 1999). Since our experiment holds both preferences (attitudes) and time (calender year) constant, the variation of the return to education is best explained as originating from a variation in cognitive and non-cognitive abilities. Individuals who are either more or less able than the average person, acquire either more or less education (Heckman and Vytlacil, 2001; Heckman et al., 2016). Given this notion of the return to education, the results presented below fit nicely with the empirical evidence on cognitive ability and health behavior (Cutler and Lleras-Muney, 2010).

The first experiment is to increase θ such that the Reference American is motivated to acquire one more year of education. The associated optimal changes of behavior and the resulting change of longevity are shown in the first row of Table 1. Ceteris paribus, the person obtains one more year of education when θ rises from 0.141 to 0.150. This motivates the person to reduce unhealthy consumption by about 9 percent and increase health expenditure by about 5 percent, compared to the benchmark citizen. These values are calculated on the basis of average expenditure on the respective good over the life cycle. As a consequence of the behavioral changes, the better educated person lives about half a year longer. The result accords well with the empirical studies referenced in the Introduction showing that the impact of education on health outcomes is mediated through health behavior. As explained above, the model identifies the channel from education to health because the reverse causality has been shut off. A higher return to education (higher cognitive and non-cognitive skills) make education more worthwhile. Better educated persons earn higher income and thus aspire to enjoy consumption for a longer period of life by indulging less in unhealthy consumption and by spending more on health.

Figure 2 investigates the education gradient more generally. Blue (solid) lines represent the benchmark calibration. The panel on the left hand side shows the desired extra years of schooling for alternative θ. When θ varies between 0.1 and 0.18, the marginal return to education, $\theta s^{-\psi}$, evaluated at 13.5 years of education, varies between 0.048 and 0.087, and years of schooling vary between 9 years (high school drop out) and 18 years (PhD). The center panel shows the gain in longevity associated with extended education, compared to the benchmark run. When θ increases from 0.14 to 0.18 the individual obtains 5.0 more years of education and gains 2.6 more years in longevity. The effect of education on life-length is thus almost linear. Compared with the observation, cited in the introduction, that a 25 year old college graduate could expect to live 8 years longer than a high school dropout of the same age (Cutler and Lleras-Muney, 2010; Richards and Barry, 1998), we conclude that dispersion in the return to education motivates about half of the observable education gradient.

Figure 2: The Return to Education in Terms of Wealth, Health, and the Value of Life

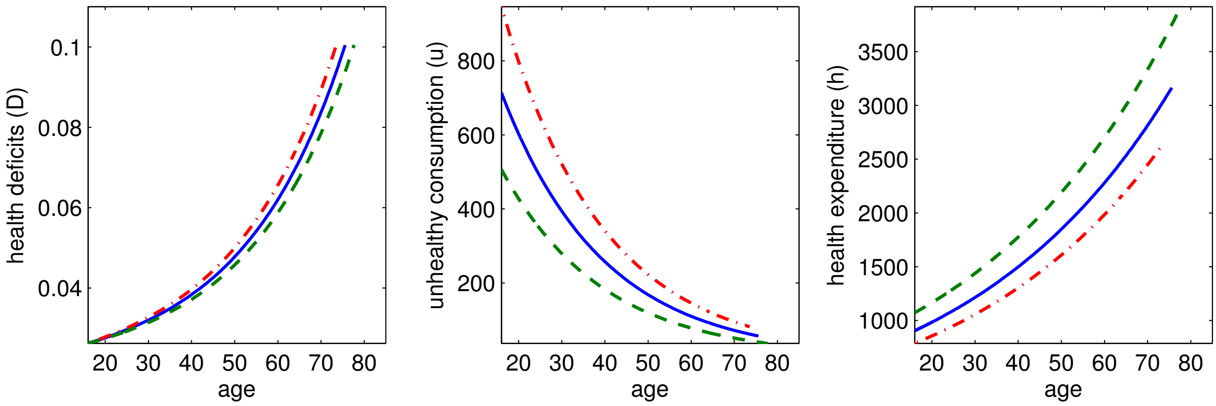


Left: Optimal length of education for alternative returns to education. Center: Extra education and subsequent gain in lifespan relative to benchmark run for alternative θ. Right: Value of life for alternative returns to education. Solid lines: Benchmark run. Dashed lines: More expensive unhealthy consumption ($q = 2$ instead of 1).

The panel on the right-hand side of Figure 2 shows the value of life, evaluated at age 25, experienced by individuals with alternative returns to education. The value of life (VOL) is the monetary expression of aggregate utility experienced during life whereby instantaneous utility is converted by the unit value of a "util", i.e. by initial marginal utility, such that the VOL at age t equals $\int_t^T \mathrm{e}^{-\rho(\tau-t)} u[x(\tau)] \mathrm{d}\tau)/u_x(t)$. A higher return to education and thus, more education, is associated with a higher value of life. Over the considered range of θ the VOL increases by 50 percent from \$ 10 million at 9 years of education to \$ 15 million at Phd-level. In terms of magnitude these VOL figures correspond nicely to the value of a statistical life of \$ 9.2 million adopted by the Department of Transportation (2014). It is somewhat higher than the estimate of Murphy and Topel (2006, Fig. 3) of about \$ 7 million at age 25. A higher return to education allows individuals to lead a more valuable life because (i) they experience higher utility at any age, and (ii) they live longer due to their healthier behavior.

Would an increasing price of unhealthy goods (cigarettes) reduce or widen the education gradient? The non-obvious answer to this question is presented by the green (dashed) lines in Figure 2. They show the outcome of the original experiment when the price of unhealthy goods is 2 (instead of 1). As a result, all social strata reduce unhealthy consumption and live longer for about 1 year (center panel). The education gradient, represented by the slope of the curve, gets somewhat flatter, indicating that the less educated benefit slightly more from the price change. In terms of VOL, however, all strata are harmed by the higher prices, and the less educated suffer the most. This is shown by the tilted downward shift of the VOL in the panel on the right-hand side. The result is intuitive since the less educated spend more on unhealthy goods. The experiment can be considered as a robustness check to the initial assumption of equal prices for all kinds of goods. The education gradient $\Delta T/\Delta s$ is almost invariant to drastic price changes.[11]

Figure 3: Education, Health, and Health Behavior



Health expenditure and unhealthy expenditure in thousands. Blue (solid) lines: basic run (from Figure 1). Green (dashed) lines: $\theta = 0.173$ (four years more education). Red (dash-dotted) lines: $\theta = 0.077$ (four years less education).

In order to more thoroughly investigate how the return to education affects health behavior and health outcomes, we turn to another experiment, shown in Figure 3. The figure shows the optimal life-cycle trajectories for the Reference American (solid lines), another person endowed with a higher return ($\theta = 0.173$, marginal return to schooling 0.083 at 13.5 years of schooling) who takes up four more years of education (dashed lines), and a third person endowed with a lower return ($\theta = 0.077$, marginal return 0.041) who obtains four years less education (dash-dotted lines). Better educated people display, at any given age, better health status, i.e. fewer health deficits, in line with the evidence provided by Harttgen et al. (2013). These health differences are explained by health behavior; the better educated people spend more on health and less on unhealthy consumption at any given age.

At the aggregate level, the result implies that the secular increase of the average return to education over the last decades (Katz and Autor, 1999) may have had a causal impact on the simultaneously observed increase in life-expectancy. This suggests a mechanism linking the long-run evolution of longevity and education that reverses the causality proposed in the macroeconomics literature (Ben-Porath; 1967, Cervellati and Sunde, 2005). The empirical fact that the return to education is greater for skilled persons and for skilled occupations (Murnane et al., 1995) may have contributed to the disproportionate increase in longevity for the well educated.

[11] Notice that this experiment does not allow for conclusions on the desirability of taxes on unhealthy goods. Such an analysis would have to take into account how tax revenue collected from unhealthy goods expenditure is redistributed.

Table 1: The Education Gradient: Alternative Mechanisms

Δs	ΔT	$\Delta u/u$	$\Delta h/h$	par. change
$+1$	$+0.51$	-9.5	$+4.7$	$\theta = 0.150$
$+1$	$+0.08$	-0.31	$+0.0$	$\psi = 0.244$
$+1$	$+2.6$	-31	$+21$	$\alpha_1 = 0.030$
$+1$	$+1.7$	-25	$+16$	$g_w = 0.0184$
$+1$	$+1.1$	$+9.5$	-28	$D_0 = 0.0237, \delta = 0.027$
$+1$	$+1.0$	$+9.2$	$+1.0$	$A = 0.0215, \ \delta = 0.027$

Δs and ΔT are measured in years, $\Delta u/u$ and $\Delta h/h$ are measured in percent.

Finally, we consider the implied estimated education gradient and the role of income. When we feed the data generated for Figure 2 into a univariate OLS regression, where the only variation comes from idiosyncratic returns to education, the estimated education gradient is 0.53 (0.53 more years of life for an additional year of education) with an R^2 of 0.999, indicating that the relationship is almost linear, a result that corresponds nicely with the empirical observation that the relationship between education and health is roughly linear after 10 years of school (Cutler and Lleras-Muney, 2010). When we consider average life-time income as an additional regressor the estimated gradient is reduced to 0.51 (and R^2 increases to 0.9999). Thus, income plays a relatively minor role for life expectancy once education is taken into account. This is, of course, an outcome "by design". It suggests that the frequent empirical finding of only a small impact of income on longevity, once education is controlled for in the regression, can be explained by the theory. Most of the observed socioeconomic gradient seems to run from the impact of the return to education (cognitive and non-cognitive skills) to education, to health, and longevity.

5.2. **Alternative Mechanisms.** Before w investigate other potential drivers of the education gradient, it is worthwhile to note that two seemingly natural candidates are already excluded by theory, namely the time preference rate, ρ, and income for given education, that is $\bar{w}$. The reason is that both parameters – while having a strong impact on life-length – leave education unaffected. To see this, reconsider equation (8) which determines optimal s and conclude that it is independent from $\bar{w}$ and ρ. The independence of the schooling decision from time-preference and the initial level of wages is not a particularity of the current approach but a standard result from the literature on optimal education (see e.g. Bils and Klenow, 2000; Card, 1999).[12]

The return to education may vary because of variation in θ (as in the benchmark experiment) or because of variation in ψ, i.e. the curvature of the return function. In order to explore the latter possibility, we set

[12] This, of course, does not mean that time preference plays no role in the education gradient in general. Generating such an impact, however, would require non-standard assumptions such as hyperbolic discounting and time-inconsistent planning. As motivated in the Introduction, such deviations from fully rational behavior are intentionally avoided in the present analysis in order to set up a clean theoretical experiment.

$\psi = 0.244$. Compared to the benchmark case, where $\psi = 0.28$, this elicits one more year of education. The elicited change of health behavior and longevity, however, is relatively small. T is estimated to increase by 29 days (8% percent of a year). The reason for the small change is that the increase of life-time income induced is much smaller for the ψ-change than for the θ-change and, consequently, individuals hardly change their health behavior (line 2 in Table 1).

We now consider a change in education provoked by a change in α_1 (line 3 in Table 1). The parameter α_1 could be conceptualized as the impact of age on preserving the learned skills. A change of α_1 greatly modifies the wage-for-age curve and has a large impact on life-time income. As a result, the change in health behavior and longevity elicited by an additional year of education is "too large". One more year of education as a result of declining α_1 leads to 2.6 more years of longevity. The predicted education gradient is "too steep" in the sense that there is too little variation of education needed in order to fully explain the variation in longevity. A reasonable interpretation is thus that *small* variation in α_1 may have contributed to the education gradient in addition to the variation of in the return to education.

Besides the notion that α_1 varies at a given time across individuals, a secular decline of α_1 may have contributed to an increasing education gradient over time. The decline of farming and industrial production and the rise of the service sector has led to a decline in the importance of motor skills and a rise in the importance of crystallized abilities. In the year 2000, the Reference American is no longer a farmer or industrial laborer but rather a salesman or a consultant. Since crystallized abilities decline later in life, the incentive to acquire more education rises, which in turn elicits more healthy behavior and increases longevity.

The final potential mechanism consists of a change in the rate of productivity growth. Productivity growth devalues the costs of not working today and increases the benefit of working tomorrow. At a higher rate of growth, it becomes less costly to delay entry into work-life in favor of additional education. Inspecting (8) we see also that, once individuals are working, productivity growth is akin to increasing experience on the job. Consequently, individuals spend more on health and indulge less in unhealthy behavior. As shown in the final row of Table 1, an increase of productivity growth from 1.0 to 1.84 percent per year triggers one more year of education and, subsequently, behavioral changes, that enable a person to live about 1.7 years longer.

At the aggregate level, this is an interesting result since it motivates a causal effect from economic growth to education and longevity, while the literature in economic growth, to date, has focussed on causality in the oppositive direction (e.g. Cervelatti and Sunde, 2005). The problem is, however, that aggregate productivity growth varies within a limited range, which makes it impossible to motivate a large and further increasing education gradient. In order to exploit productivity growth to rationalize the education gradient, one should consider occupation-specific growth rates of productivity. In this context, the model predicts that people are motivated to acquire more education and live healthier lives when they are occupied in a high-growth sector of the economy. The other parameters of the model are not capable to motivate the gradient, neither via stand-alone variation nor jointly in combination with variation of other parameters. This leaves the idiosyncratic return to education as the most plausible source of variation behind the education gradient.

5.3. **Robustness of Results.** In this subsection, we consider the robustness of the main result with respect to alternative specifications of the model. In any experiment θ is raised from 0.141 to 0.150, i.e. the return to education at 13.5 years of education is raised from 6.8 to 7.2 percent. The first two rows, case 1 and 2, document that the education gradient operates independently from income. Results are shown when, ceteris paribus, the annual wage per unit of human capital is assumed to be 50 percent higher or lower. These income variations have very strong effects on longevity, which rises by 4 years or, respectively, falls by 3 years compared to the basic run. The gradient, however, that is the gain in longevity that is associated with one more year of education, is almost the same as for the benchmark run.

Table 2: The Education Gradient: Robustness Checks

Case	Δs	ΔT	$\Delta u/u$	$\Delta h/h$	comment
1) $\Delta\bar{w} = +50\%$	$+1$	$+0.47$	-11	$+5.6$	a richer individual
2) $\Delta\bar{w} = -50\%$	$+1$	$+0.50$	-6.3	$+3.0$	a poorer individual
3) $k_0 = 6\bar{w}$	$+1$	$+0.37$	-8.7	$+4.2$	inheritance (parental support)
4) $k_0 = 6\bar{w}$, $\bar{k} = 6\bar{w}$	$+1$	$+0.36$	-8.7	$+4.2$	inheritance and bequest
5) $\beta = 5$ ($\epsilon_q = -0.3$)	$+1$	$+0.56$	-7.9	$+3.9$	lower price elast. of unhealthy good
6) $B = 4 \cdot 10^{-7}$, $\omega = 1.3$	$+1$	$+0.56$	-12	$+4.4$	unit consumption more unhealthy
7) $B = 1 \cdot 10^{-7}$, $\omega = 1.5$	$+1$	$+0.48$	-8.0	$+4.9$	unit consumption less unhealthy
8) $p = 2$, $A = 0.00192$	$+1$	$+0.51$	-9.5	$+4.7$	higher price of health
9) $p = 2$, $\bar{w} = 20,400$	$+1$	$+0.41$	-11	$+5.4$	higher price of health

In all cases the experiment increases θ from 0.14 to 0.15. All other parameters from benchmark case (Figure 1).

During the education period, the Reference American accumulates debt of about \$ 150,000. This relatively high accumulation of debt at young ages is an unreasonable artefact originating from the assumption that the Reference American does not benefit from supporting parents (or public subsidies). In order to accommodate this criticism, case 3 endows the person with an inheritance of $6\bar{w}$, which is about four times the average annual labor income, a sum which can be regarded as sufficient to finance 4 or 5 years of voluntary education. The inheritance reduces the gradient by about one third to 0.37. The reason is that financial capital, ceteris paribus, reduces the incentive to protect human capital. A wealthy person prefers to finance larger parts of consumption and health expenditure in old age by returns on capital rather than human capital and savings from labor income. Case 4 also requires that the person leaves a bequest of $6\bar{w}$. It demonstrates that the mechanism runs mainly through the inheritance received rather than through the bequest left.

Case 5 in Table 2 adjusts the taste for unhealthy goods such that the price elasticity of demand equals -0.3. Such a value is observed at the lower end of estimates for the price elasticity of cigarettes. Higher education has somewhat less impact on health behavior but the health gradient widens a bit because the less educated spend relatively more on unhealthy goods when demand is less price elastic.

With case 6 and 7 we investigate the character of the unhealthy good. Case 6 assumes that the good is more unhealthy than cigarettes. B is raised by factor 2. The scale parameter ω is adjusted to 1.3 such that the calibration continues to produce education, expenditure, and longevity from the basic run. The greater unhealthiness of consumption of small quantities of the good implies a higher incentive for the better educated to stay away from this good. The experiment consequently predicts a somewhat larger education gradient.

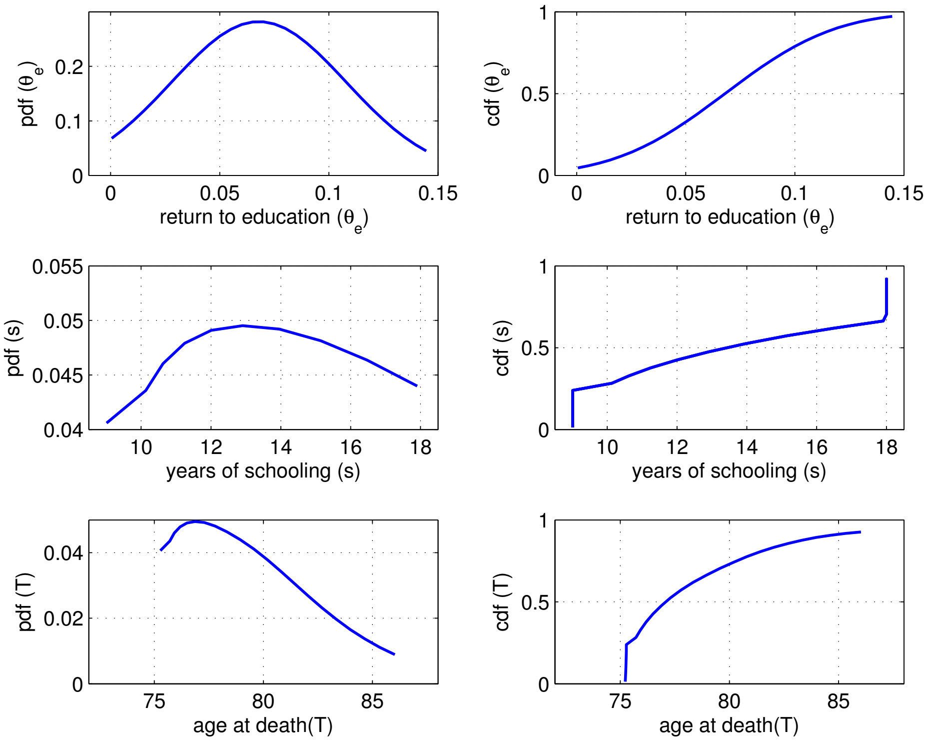
Inverting the result from above, the model produces a smaller gradient if u is generally less unhealthy. This is confirmed by case 7 in which B has been reduced by about a factor of 2 and the scale parameter has been adjusted to 1.5. The new parameter values reflect a good for which consumption of small quantities entails relatively small effects on health, whereas large excess consumption has severe consequences. It could perhaps be thought of as alcohol. The predicted education gradient is 0.43 and thus a bit smaller than for the basic run (representing cigarettes). The reason is that better educated individuals have less incentive to stay away from consuming the good.

Finally, we consider the impact of the price of health. Setting $p = 2$ implies that the cost of health provisions doubles compared to the benchmark case. In order to fit the Reference American, at least one parameter has to be re-calibrated such that life-span of the benchmark case is restored. Two natural candidates for the calibration are the productivity of health expenditure A and the level of income $\bar{w}$. Case 8 considers the adjustment via higher A. In this case, the predicted change of health and health behavior is virtually the same as for the benchmark run. The outcome is intuitive because the individual is compensated for the higher price of health by higher efficiency of health expenditure. The re-calibration via higher income is shown in case 9. As a result, the health gradient is smaller, and individuals spend slightly more on health but individuals respond more noticeably to the higher price of health by reducing unhealthy consumption.

5.4. **Health Inequality.** In this section we investigate the extent to which the proposed mechanism contributes to the explanation of health inequality. For this purpose, we investigate a society that differs by their return to education (and nothing else). We built the calibration on Harmon et al. (2003) who explicitly consider dispersion of the return to education at the individual level. Assuming it is normally distributed, they estimate a standard deviation of 4.2 percent. Relying on this study is attractive because

Harmon et al. estimate the mean of the return to education to be exactly at the value of our benchmark calibration (at 6.8 percent) such that we need no changes at all for the calibrated model. Other studies (referenced in Harmon et al., 2003) arrive at quantitatively similar estimates. A standard deviation of 4.2 is huge since it implies that, optimally, a part of the population should acquire no education at all (based on their low individual return) and another part should obtain only compulsory education.

Figure 4: Health Inequality and the Return to Education



The figure shows the implied distribution of education and longevity when the return to education according to the empirical distribution $N(0.068, 0.042)$ is fed into the model.

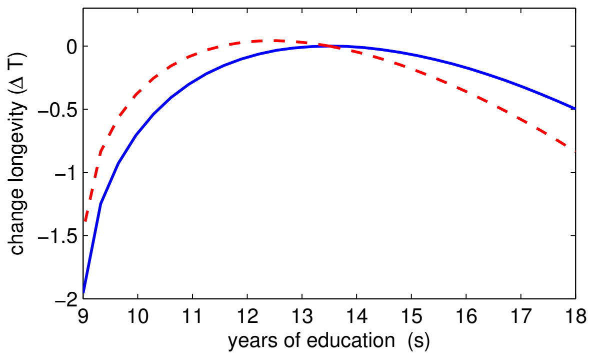
In order to implement the dispersion correctly we need to take into account that we assumed decreasing returns to education (in contrast to Harmon et al., 2003). We calibrate around 13.5 years of education (the level of education received by the Reference American) such that $0.068{=}\theta\xi$ with $\xi = 13.5^{-\psi} = 0.48$ and use the fact that $var(\xi\theta) = \xi^2 var(\theta)$. The estimated standard deviation for our setup is thus 0.083. The upper panels of Figure 4 show the implied partial and cumulative distributions of the return to education.

When the return distribution is fed into the model, we obtain the distribution of education as shown in the center panel of Figure 4. Here, we have truncated the distribution at the compulsory level and at 18 years of education (nobody acquires more than a masters degree). As shown in the cumulative distribution on the right-hand side, this means that a substantial part of the population acquires no more than 9 years of compulsory education and another part educates at the maximum level.

The bottom panel shows the implied distribution of age at death, which varies from 75 to 88 years. Notice that the upper branch of the cdf is not truncated because there is variety in income among the best-educated due to their idiosyncratic returns to education. The standard deviation in lifespan generated by the model is 3.9. The empirical standard deviation, estimated from life tables, for white American men in the year 2000 is 13.9 (Sasson, 2016). This means that the proposed channel explains about 28 percent of the observed variance.

5.5. **Endogenous vs. Exogenous Education.** In this subsection, I re-investigate the problem under the condition that education is not individually chosen but exogenously given. I show that the education gradient under exogenous education is relatively small after individuals achieved basic education and that it eventually turns negative. Endogenous education, driven by the idiosyncratic return to education, appears to be a more powerful explanation of the observable education gradient.

Figure 5: Longevity Gain: Exogenous Education



The figure shows the change of longevity compared to the benchmark (75.6 years) when the individual is forced to spend s years on education. Solid line: Benchmark calibration. Dashed line: $\theta = 0.13$ (instead of 0.14).

The solid line in Figure 5 shows results for the Reference American when problem (1) – (3) is solved under the constraint that the individual has to spend s years on education. For a better understanding, recall that the calibration of the benchmark model implied that the unconstrained Reference American acquires voluntarily 13.5 years of education (at a θ of 0.14 and a return to education of 6.8 percent). The model predicts that the Reference American would live 2 years less if he would be forced to acquire only 9 years of education. At this low level of education, the health gradient is relatively steeply increasing. Yet, a maximum is naturally reached at 13.5 years of education. If enforced education is further increasing, the health gradient turns negative. The reason is that life-time income assumes a maximum at 13.5 years of education. Extending the education period further reduces life-time income (although it increases human capital) such that the individual spends less on health provisions. The predicted decrease is relatively mild but the important point is that with the Reference American's already high level of education it becomes more difficult, and eventually impossible to generate a positive education gradient in a model based on exogenous education.

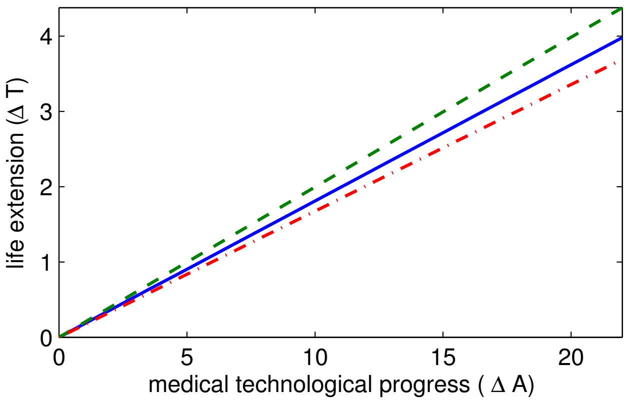
The red (dashed) line shows the result when θ is kept at 0.13 (instead of 0.14). In this case, the education gradient starts declining already after 12 years of education. Forcing this person to take up additional education beyond high school would reduce life-time income and longevity. These numerical experiments rationalize why the compulsory education gradient is large only when average education is low. In this case, it is likely that extending schooling is optimal from the individual's perspective. At a low level of education, school attendance is for most children not driven by cognitive and non-cognitive abilities but constrained by other factors, like distance to school, child labor, credit constraints, or leisure preferences. The model thus supports the empirical finding of a relatively large health effect from compulsory schooling at low levels of schooling (e.g. Lleras-Muney, 2005). At higher levels of education, however, it is hard to motivate an education gradient by further increasing compulsory education.

These results also rationalize why studies of monozygotic twins find relatively small health gains from education (Fujiwara and Kawachi, 2009; Lundborg, 2013). The reason is that monozygotic twins are likely to be endowed with similar cognitive ability, which trumps the length of the education period in the determination of human capital and health behavior. As shown in Figure 5, there is little variation in predicted longevity for persons sharing the same θ's. For example, for persons with a θ of 0.14 who attend school between 10 and 18 years set by compulsory education, their life span varies by less than a year although education varies by 8 years. In comparison, if schooling is motivated by idiosyncratic abilities and is endogenously chosen, the benchmark model (Figure 2) predicts that 8 more years of education leads to a longevity gain of four more years. These results are in line with the strong empirical association observed between cognitive ability and health and longevity (Deary, 2008; Der et al., 2009; Calvin et al., 2011).

5.6. **Medical Technological Progress and the Health Gradient.** In the final experiment, I investigate the impact of medical technological progress on life-time extension and in particular on the education gradient. It has been hypothesized that the observable secular increase of the education gradient may have its origin in technological progress because better educated persons have better access and more resources to utilize technological advances to their benefit (Cutler et al., 2010). Better educated persons are, ceteris paribus, richer and demand more health services in order to protect or repair their human capital. We thus expect that they benefit more from medical technological progress compared with their less educated counterparts.

The results presented in Figure 6 confirm this expectation. The figure shows longevity gains originating from increasing efficiency of health expenditure A; ΔA is measured in percent of the benchmark run. If medical technology advances at an annual rate of 1 percent, the level of A is 20 percent higher after about

Figure 6: Medical Technological Progress and the Health Gradient



The figure shows the gain in longevity for alternative progress of medical technology (ΔA). Blue (solid): benchmark run (13.5 years of education, $\theta = 0.1$). Green (dashed): 4 years more of education ($\theta = 0.12$). Red (dash-dotted): 7.2 years more of education ($\theta = 0.14$). The longevity gain is measured relative to a person's own initial life-span for both types.

18 years. The solid line shows the predicted longevity for the Reference American (endowed with a θ of 0.14, i.e. a return of education of 6.8 percent at 13.5 years of education). The dashed line shows implied longevity for a person endowed with $\theta = 0.16$ who attends school for 2 more years and experiences a return to education of 7.4 percent. The dash-dotted line reflects the longevity gain of a person with $\theta = 0.12$ who attends school for 1.8 years less, at a rate of return of 6.0 percent. Although everybody experiences an increase in longevity, the predicted gain is higher for the better educated. When medical technology advances by 20 percent, the longevity gap between a high school graduate and a college graduate widens by about one year.

6. Conclusion

This study has proposed a new view on the education gradient. It has assumed away any explanation based on attitudes, non-cognitive skills, and allocative or productive inefficiency of the uneducated. Instead it has asked how large a gradient can be motivated by optimal decisions on education and health behavior of individuals who know how their behavior affect their future health status and the time of death. The theory has been firmly built on insights from modern gerontology which allowed a reliable calibration for a "Reference American". It predicts that a person whose return to education (cognitive and non-cognitive skills) motivate one more year of education, spends more on health and less on unhealthy behavior such that the person lives about half a year longer. This means that the theory explains about half of the observable education gradient. The theory motivates an almost linear education gradient, in line with the empirical observation (Cutler and Lleras-Muney, 2010).

The theory suggests that idiosyncratic differences in abilities and endogenous choice of education are more powerful explanations of the gradient than forced exogenous education. It predicts, in line with the empirical observation from monozygotic twins (Fujiwara and Kawachi, 2009; Lundborg, 2013), that for similar cognitive ability, the length of the education period contributes little to the explanation of longevity. The theory has been utilized to explain why the education gradient increases over time through ongoing medical technological progress. The reason is that well educated persons demand relatively more health services and thus benefit more from health innovations than less educated persons.

The study has focussed on the human life cycle from young adulthood onwards. At this age, taking cognitive and non-cognitive skills as approximately given may be an appropriate simplification. But since these skills are malleable at younger ages (see e.g. Heckman, 2006), the present study also highlights the importance of childhood development for later in life. Equipped with a high return to education, individuals are not only motivated to educate for longer and earn more labor income but also to lead a healthier life and to live longer. Targeting fundamental skills at an early age is thus not only a promising policy concerning adult socioeconomic status but also with regard to health behavior and longevity.

References

Acemoglu, D., Finkelstein, A., and Notowidigdo, M.J., 2013, Income and health spending: evidence from oil price shocks, *Review of Economics and Statistics* 95, 1079-1095.

Arking, R., 2006, *The Biology of Aging: Observations and Principles*, Oxford University Press, Oxford.

Barro, R., G. Mankiw and X. Sala-i-Martin, 1995. Capital mobility in neoclassical models of growth. *American Economic Review* 85, 103–115

Ben-Porath, Y., 1967, The production of human capital and the life cycle of earnings, *Journal of Political Economy* 75, 352–365.

Bijwaard, G.E., van Kippersluis, H., and Veenman, J., 2015, Education and Health: The Role of Cognitive Ability, *Journal of Health Economics* 42, 29-43.

Bils, M. and Klenow, P.J., 2000, Does schooling cause growth?, *American Economic Review* 90, 1160–1183.

BLS, 2002, Consumer Expenditures in 2000, United States Department of Labor, Bureau of Labor Statistics, Report 958.

BLS, 2011, State average annual pay for 2000 and 2001, United States Department of Labor, Bureau of Labor Statistics (`http://www.bls.gov/news.release/annpay.t01.htm`).

Brunello, G., Fort, M., Schneeweis, N., and Winter-Ebmer, R., 2016, The causal effect of education on health: What is the role of health behaviors, *Health Economics* 25(3), 314336.

Calvin, C. M., Deary, I. J., Fenton, C., Roberts, B. A., Der, G., Leckenby, N., and Batty, G. D., 2011, Intelligence in youth and all-cause-mortality: systematic review with meta-analysis, *International Journal of Epidemiology* 40, 626-644.

Card, D., 1999, The causal effect of education on earnings, *Handbook of Labor Economics* Vol. 3, 1801–1863, Elsevier, Amsterdam.

Case, A. and Deaton, A., 2005, Broken down by work and sex: how our health declines, in: David Wise (ed.), *Analyses in the Economics of Aging*, University of Chicago Press, Chicago.

Cervellati, M. and Sunde, U., 2005. Human capital formation, life expectancy, and the process of development, *American Economic Review* 95, 1653–1672.

Chaloupka, F.J. and Warner, K.E., 2000, The economics of smoking, *Handbook of Health Economics* Vol. 1, Elsevier, Amsterdam, 1539–1627.

Chetty, R., 2006, A new method of estimating risk aversion, *American Economic Review* 96, 1821-1834.

Conti, G. and Heckman, J. and Urzua, S., 2010, The education-health gradient, *American Economic Review* 100, 234–38.

Contoyannis, P. and Jones, A.M., 2004, Socio-economic status, health and lifestyle, *Journal of Health Economics* 23, 965–995.

Cutler, D.M. and Lleras-Muney, A., 2006, Education and health: Evaluating theories and evidence, NBER Working Paper 12352.

Cutler, D.M., Lleras-Muney, A. and Vogl, T., 2011, Socioeconomic status and health: dimensions and mechanisms, in: Glied, S. and Smith, P.C. (eds), *The Oxford Handbook of Health Economics*, Oxford University Press, Oxford, pp. 124-163.

Cutler, D.M. and Lleras-Muney, A., 2010, Understanding differences in health behaviors by education, *Journal of Health Economics* 29, 1–28.

Cutler, D.M., Lange, F., Meara, E., Richards, S., and Ruhm, C.J., 2010, Explaining the rise in educational gradients in mortality, NBER Working Paper 15678.

Dalgaard, C.J. and Strulik, H., 2014, Optimal aging and death, *Journal of the European Economic Association* 12, 672-701.

Dalgaard, C.J. and Strulik, H., 2015, The economics of health demand and human aging: Health capital vs. health deficits, Working Paper, University of Goettingen.

Deary, I.L., 2008, Why do intelligent people live longer? *Nature* 456, 175-176.

Department of Transportation (2014), Treatment of the Value of Preventing Fatalies and Injuries in Preparing Economic Analysis (`www.dot.gov/policy`.

Der, G., Batty, G.D. and Deary, I.J., The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth, *Intelligence* 37, 573-580.

Ehrlich, I. and Chuma, H., 1990, A model of the demand for longevity and the value of life extension, *Journal of Political Economy* 98, 761-82

Elo, I.T. and Preston, S.H., 1996, Educational differentials in mortality: United States, 1979-1985, *Social Science & Medicine* 42, 47–57.

French, E., 2005, The effects of health, wealth, and wages on labour supply and retirement behaviour, *Review of Economic Studies* 72, 395-427.

Fries, J.F., 1980, Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303, 130-135.

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

Galama, T.J. and van Kippersluis, H., 2015, A theory of education and health, CESR-Schaeffer Working Paper No. 2015-001.

Gavrilov, L.A. and Gavrilova, N.S., 1991, *The Biology of Human Life Span: A Quantitative Approach*, Harwood Academic Publishers, London.

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

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

Grossman, M., 2006, Education and nonmarket outcomes, in: *Handbook of the Economics of Education* Vol. 1, 577–633, Elsevier, Amsterdam.

Hall, R.E. and Jones, C.I., 1999, Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* 114, 83-116.

Hall, R.E. and Jones, C.I., 2007, The value of life and the rise in health spending. *The Quarterly Journal of Economics* 122, 39–72

Harmon, C., Hogan, V., and Walker, I. (2003). Dispersion in the economic return to schooling. Labour economics, 10(2), 205-214.

Harttgen, K., Chatterji, S., Kowal, P., Strulik, H., and Vollmer, S., 2013, Patterns in disability and frailty in older adults: Results from the Survey of Health, Ageing and Retirement in Europe (SHARE), *PlosOne* Vol. 8 (10).

Hazan, M., 2009, Longevity and lifetime labor supply: Evidence and implications, *Econometrica* 77, 1829–1863.

Heckman, J.J. and Vytlacil, E., 2001, Identifying the role of cognitive ability in explaining the level of and change in the return of schooling, *Review of Economics and Statistics* 83, 1–12.

Heckman, J.J., 2006, Skill formation and the economics of investing in disadvantaged children, *Science* 312, 1900–1902.

Heckman, J.J., Humphries, J.E., and Veramendi, G., 2016, Returns to Education: The Causal Effects of Education on Earnings health and smoking, Working Paper, University of Chicago.

Hokayem, C., and Ziliak, J.P., 2014, Health, Human Capital, and Life Cycle Labor Supply. *American Economic Review* 104(5), 127-131.

Hong, K., Savelyev, P., and Tan, K.T., 2015, Understanding the Mechanisms Linking Cognitive Skills, Socioemotional Skills, and College Education with Longevity, Working Paper, Vanderbilt University.

Jaeckle, R., and Himmler, O. (2010). Health and Wages Panel data estimates considering selection and endogeneity. Journal of Human Resources 45(2), 364-406.

Jorgenson, D.W., Ho, M.S. and Stiroh, K.J., 2008, A retrospective look at the US productivity growth resurgence, *Journal of Economic Perspectives* 22 (1), 3–24.

Kaestner, R. and Callison, K., 2011, Adolescent cognitive and noncognitive correlates of adult health, *Journal of Human Capital* 5, 29–69.

Katz, L.F. and David, H., 1999, Changes in the wage structure and earnings inequality, *Handbook of Labor Economics* Vol. 3, 1463–1555.

Keehan, S. P, H. Lazenby, M. Zezza, and Catlin, A.C., 2004, Age estimates in the National Health Accounts. *Health Care Financing Review* 1, 1–16.

Kenkel, D.S., 1991, Health behavior, health knowledge, and schooling, *Journal of Political Economy* 99, 287–305.

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

Lundborg, P., 2013, The health returns to schooling – what can we learn from twins, *Journal of Population Economics* 26, 673-701.

Mackenbach, J.P., Stirbu, I., Roskam, A.J.R., Schaap, M.M., Menvielle, G., Leinsalu, M., and Kunst, A.E., 2008, Socioeconomic inequalities in health in 22 European countries, *New England Journal of Medicine* 358, 2468–248.

McFadden, D., 2005, Comment, in: David Wise (ed.), *Analyses in the Economics of Aging*, University of Chicago Press, Chicago.

Meara, E., White, C., and Cutler, D.M., 2004, Trends in medical spending by age, 1963–2000, *Health*

Affairs 23, 176–183.

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

Mitnitski, A.B., Mogilner, A.J., MacKnight, C., and Rockwood, K., 2002a, The accumulation of deficits with age and possible invariants of aging, *Scientific World* 2, 1816-1822.

Mitnitski, A.B., Mogilner, A.J., MacKnight, C., and Rockwood, K., 2002b, The mortality rate as a function of accumulated deficits in a frailty index, *Mechanisms of Ageing and Development* 123, 1457-1460.

Mitnitski, A., Song, X., Skoog, I., Broe, G.A., Cox, J.L., Grunfeld, E., and Rockwood, K., 2005, Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality, *Journal of the American Geriatrics Society* 53, 2184–2189.

Mitnitski, A., Bao, L., Rockwood, K., 2006, Going from bad to worse: a stochastic model of transitions in deficit accumulation, in relation to mortality, *Mechanisms of Ageing and Development* 127, 490–493.

Murnane, R.J., Willett, J.B., and Levy, F., 1995, The growing importance of cognitive skills in wage determination, *Review of Economics and Statistics* 77, 251–266.

Murphy, K. M. and Welch, F., 1990, Empirical age-earnings profiles. *Journal of Labor Economics* 8, 202-229

Murphy, K.M. and Topel, R.H., 2006, The value of health and longevity, *Journal of Political Economy* 114, 871–904.

Preston, S.H., Glei, D.A., and Wilmoth, J.R., 2010, Contribution of smoking to international differences in life expectancy, in: E.M. Crimmins, S.H. Preston, and B. Cohen (eds.), *International Differences in Mortality at Older Ages: Dimensions and Sources*, Washington, D.C.: National Academies Press, pp. 105-131.

Psacharopoulos, G., 1994, Returns to investment in education: A global update, *World Development* 22, 13251343.

Richards, H., and Barry, R., 1998. U.S. life tables for 1990 by sex, race, and education, *Journal of Forensic Economics* 11, 9–26.

Rockwood, K. and Mitnitski, A.B., 2007, Frailty in relation to the accumulation of deficits, *Journals of Gerontology Series A: Biological and Medical Sciences* 62, 722

Sasson, I., 2016, Trends in life expectancy and lifespan variation by educational attainment: United States, 19902010. *Demography*, forthcoming.

Schaie, K.W., 1994, The course of adult intellectual development, *American Psychologist* 49, 304-313.

Strulik, H., 2015a, A closed-form solution for the health capital Model, *Journal of Demographic Economics* 81(3), 301-316.

Strulik, H., 2015b, Frailty, Mortality, and the Demand for Medical Care, *Journal of the Economics of Ageing* 6, 5-12

Strulik, H. and Werner, K., 2016, 50 is the new 30 Long-run trends of schooling and retirement explained by human aging, *Journal of Economic Growth* 21, 165-187.

Turner, C., Tamura, R., Muholland, S., and Baier, S., 2007, Education and income of the states of the

United States: 1840-2000, *Journal of Economic Growth* 12, 101-158.

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

Zellner, A., 1969, On the aggregation problem: A new approach to a troublesome problem. In: M. Beckmann and H.P. Künzi (eds.) : *Economic Models, Estimation and Risk Programming: Essays in honor of Gerhard Tintner*, Springer Verlag, Berlin.

6.1. **Setup of the Problem.** Integrating (3) provides the following solution.

$$D(t) = D(0)\exp(\mu t) - \int_0^t \mu a \exp(\mu(t-v))\mathrm{d}v - \mu A \int_0^t h(v)^\gamma \exp(\mu(t-v))\mathrm{d}v$$
$$+ \mu B \int_0^t u(v)^\omega \exp(\mu(t-v))\mathrm{d}v. \tag{A.1}$$

Integrating (2) and using $x \equiv c + \beta u$ we get (A.2)

$$k(t) = k(0)\exp(rt) + \int_s^R \exp(r(t-v))w(v)H(s,v)\mathrm{d}v - \int_0^t \exp(r(t-v))x(v)\mathrm{d}v$$
$$- \int_0^t \exp(r(t-v))(q-\beta)u(v)\mathrm{d}v - \int_0^t \exp(r(t-v))ph(v)\mathrm{d}v. \tag{A.2}$$

Using (A.1) and (A.2), the initial conditions $D(0) = D_0$, $k(0) = k_0$, and the terminal conditions $D(T) = \bar{D}$, $k(T) = \bar{k}$, the Lagrangian associated with problem (1)- (4) is given by

$$\max_{d,h,s,T} L = \int_0^T \mathrm{e}^{-\rho t} \frac{x^{1-\sigma}}{1-\sigma}\ \mathrm{d}t$$
$$+ \phi \left\{ k_0 + \mathrm{e}^{-rs} \int_s^R \mathrm{e}^{-rt} w(t)H(s,t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}x(t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}ph(t)\mathrm{d}t - \int_0^T \mathrm{e}^{-rt}(q-\beta)u(t)\mathrm{d}t - \bar{k}\mathrm{e}^{-rT} \right\}$$
$$+ \lambda \left\{ D_0 - \mu a \int_0^T \mathrm{e}^{-\mu t}\mathrm{d}t - \mu A \int_0^T h(t)^\gamma \mathrm{e}^{-\mu t}\mathrm{d}t + \mu B \int_0^T u(t)^\omega \mathrm{e}^{-\mu t}\mathrm{d}t - \bar{D}\mathrm{e}^{-\mu T} \right\}. \tag{A.3}$$

Using (4), $w(t)H(s,t)$ in (A.3)is determined as (A.4).

$$w(t)H(t,s) = \bar{w}\exp(g_w t)\left\{ \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t\right] - \delta \exp(\alpha_2 t) \right\}. \tag{A.4}$$

Solution. The first order conditions for consumption, health expenditure, and unhealthy consumption are:

$$0 = \mathrm{e}^{-\rho t}x^{-\sigma} - \phi \mathrm{e}^{-rt} \tag{A.5}$$
$$0 = -\phi \mathrm{e}^{-rt}p - \lambda\mu A\gamma h^{\gamma-1}e^{-\mu t}. \tag{A.6}$$
$$0 = -\phi \mathrm{e}^{-rt}(q-\beta) + \lambda\mu B\omega u(t)^{\omega-1}\mathrm{e}^{-\mu t}. \tag{A.7}$$

Differentiating (A.5) with respect to time, we get (11), differentiating (A.6) with respect to time we get (12), and differentiating (A.7) with respect to time, we get (13). Next, solving (A.6) for λ and using the result to substitute λ in (A.7) provides (10) in the text.

The first order condition for optimal schooling is $\partial L/\partial s = 0$, that is

$$0 = \int_0^R \frac{\partial}{\partial s}\mathrm{e}^{-rt}w(t)H(s,t)\mathrm{d}t - \mathrm{e}^{-rt}w(s)H(s,s),$$

requiring that the gain from a marginal extension of education, the first term on the right hand side, equals the income lose from a marginal extension of education, the second term. Inserting $\partial H(s,t)/\partial s = (\theta \frac{s^{1-\psi}}{1-\psi} - \eta)\exp(\theta \frac{s^{1-\psi}}{1-\psi} + \eta(t-s) - \alpha_1 t)$, $w(t) = w(s)\exp(g_w(t-s))$, and $H(s,s) = \exp(\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s) - \delta\exp(\alpha_2 s)$, the optimal schooling condition becomes

$$(\theta s^{-\psi} - \eta)\exp\left[\theta \frac{s^{1-\psi}}{1-\psi} + (r-\eta-g_w)s\right] \int_s^R \mathrm{e}^{(g_w - r + \eta - \alpha_1)t}\mathrm{d}t = \exp\left[\theta \frac{s^{1-\psi}}{1-\psi} - \alpha_1 s\right] - \delta \mathrm{e}^{\alpha_2 s}.$$

Solving the integral provides (8) in the text.

Two conditions have to be fulfilled at the optimal T. The first condition is that $D(T) = \bar{D}$. Evaluating (A.1) at T and employing constant growth rates of h and u according to (12) and (13), this can be expressed as:

$$\bar{D} = D_0 \exp(\mu T) - \mu a \int_0^T \exp(\mu(T-t)\mathrm{d}t - \mu A \int_0^T h(0)^\gamma \exp(\gamma g_h t) \exp(\mu(T-t))\mathrm{d}t$$
$$+ \mu B \int_0^T u(0)^\omega \exp(\omega g_u t) \exp(\mu(T-t))\mathrm{d}t.$$

Solving the integrals provides (16) in the text.

The second condition for optimal death is that the Lagrangian evaluated at T assumes the value of zero, that is, using (A.3) and the Euler equations (10)-(12):

$$0 = \left(\frac{x(T)^{1-\sigma}-1}{1-\sigma}\right)\exp(\rho T) - \xi \bar{D}^\nu \exp(\rho T)$$
$$+ \phi\left[-\exp(-rT)x(T) - (q-\beta)\exp(-rT)u(T) - p\exp(-rT)h(T) + r\exp(-rT)\bar{k}\right]$$
$$+ \lambda\left[-\mu\exp(-\mu T) - \mu A h(T)^\gamma \exp(-\mu T) + \mu B u(T)^\omega \exp(-\mu T) + \mu\bar{D}\exp(-\mu T)\right]$$

Inserting from (A.5)–(A.7) that $\phi\exp(-rT) = x(T)^{-\sigma}\exp(-\rho T)$, that $\lambda\exp(-\mu T) = -\phi\exp(-rT)$ $\cdot ph(t)^{1-\gamma}/(\mu A\gamma)$, and that $\lambda\exp(-\mu T) = \phi\exp(-rT)(q-\beta)u(t)^{1-\omega}/(\mu B\omega)$ provides (17) in the text.

Using the Euler conditions (A.5)–(A.7) the budget constraint (A.2) can be written as:

$$0 = k(0) + w(s,R) - \int_0^T d(0)\exp((g_d - r)t) - p\int_0^T h(0)\exp((g_h - r)t)$$
$$- (q-\beta)\int_0^T u(0)\exp((g_u - r)t) - \bar{k}\exp(-rT).$$

Solving the integrals provides (14) in the text. Finally, human wealth $w(s,R)$ is obtained as

$$\int_s^R \exp(-rt)w(t)H(s,t)\mathrm{d}t = \bar{w}\exp\left[\theta\frac{s^{1-\psi}}{1-\psi} - \eta s\right]\int_s^R \mathrm{e}^{(\eta+g_w-r-\alpha_1)t}\mathrm{d}t - \bar{w}\delta\int_s^R \mathrm{e}^{\alpha+g_w-r)t}\mathrm{d}t.$$

Solving the integrals provides the final building block for (15) in the text.

Hamilton Approach. Alternatively, the model can be solved by using optimal control theory. The current value Hamiltonian associated with problem (1)-(4) is given by

$$\mathcal{J} = \frac{x^{1-\sigma}-1}{1-\sigma} + \lambda_D\mu\left[D - a - Ah^\gamma - Bu^\omega\right] + \lambda_k\left[\chi wH + rk - ph - d - (q-\beta)u\right].$$

The first order conditions and costate equations are:

$$0 = x^{-\sigma} - \lambda_k \tag{A.8}$$
$$0 = -\lambda_D B\omega u^{\omega-1} - \lambda_k(q-\beta) \tag{A.9}$$
$$0 = -\lambda_D \mu A\gamma h^{\gamma-1} - \lambda_k p \tag{A.10}$$
$$\lambda_D\mu = \lambda_D\rho - \dot{\lambda}_D \tag{A.11}$$
$$\lambda_k r = \lambda_k\rho - \dot{\lambda}_k. \tag{A.12}$$

From (A.9) and (A.10) follows (5) in the text. Differentiating (A.8) with respect to age and using (A.12) provides (6) in the text. Differentiating (A.10) with respect to time and using (A.11) and (A.12) provides (7) in the text. Differentiating (A.9) with respect to time and using (A.11) and (A.12) provides (8) in the text.