

**ACCOUNTING FOR FETAL ORIGINS:
HEALTH CAPITAL VS. HEALTH
DEFICITS**

Carl-Johan Dalgaard
Casper Worm Hansen
Holger Strulik

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Accounting for Fetal Origins: Health Capital vs. Health Deficits

Carl-Johan Dalgaard*, Casper Worm Hansen[†] and Holger Strulik**

October 2019.

Abstract. The fetal origins hypothesis has received considerable empirical support, both within epidemiology and economics. The present study compares the ability of two rival theoretical frameworks in accounting for the kind of path dependence implied by the fetal origins hypothesis. We argue that while the conventional health capital model is irreconcilable with fetal origins of late-in-life health outcomes, the more recent health deficit model can generate shock amplification consistent with the hypothesis. We also develop a theory of ontogenetic growth in utero and during childhood, unify it with the theory of adult aging, and discuss the transmission of early-life shocks to late-life health deficit accumulation.

Keywords: Fetal Origins; Health Capital; Health Deficits; Ontogenetic Growth; In Utero Development.

JEL: I10, J13, D91.

*Department of Economics, University of Copenhagen, CAGE and CEPR, London. Contact info: Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen, Denmark; email: carl.johan.dalgaard@econ.ku.dk.

[†] Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen, Denmark; email: Casper.Worm.Hansen@econ.ku.dk.

** University of Goettingen, Department of Economics, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de

1. INTRODUCTION

Half a century ago epidemiologists would tend to view the fetal state as a protected one. Since then epidemiological evidence has been accumulating that this appears not to be the case, which has spawned the fetal origins hypothesis. The fetal origins hypothesis suggests that morbidities *in utero* may cause epigenetic changes in the fetus that instigate morbidities late-in-life though without being directly visible for most of the life course (e.g., Almond and Currie, 2011a).

Within economics, research has considerably strengthened the case that *in utero* (or early-in-life) shocks indeed appear to impact on late-in-life health (e.g., Almond, 2006; Van den Berg et al., 2006; Almond and Mazumder, 2011; Lin and Lui, 2014; Bhalotra and Rawlings, 2011). In addition, research has demonstrated effects beyond late-in-life health, including human capital and labor market outcomes (e.g. Bleakley, 2007; Almond et al., 2009; Nelson, 2010; Bhalotra and Venkataramani, 2016); welfare dependence (Almond, 2006; Oreopoulos et al., 2008), and even investment behavior (Cronqvist et al., 2016). From a broader perspective, the fetal origins hypothesis thus seems to be a promising avenue through which to gain further insights into the causes and intergenerational transmission of inequality (e.g., Currie, 2011).

From a theoretical perspective, however, the fetal origins hypothesis poses a problem for the current workhorse model within health economics: the Grossman (1972a) model. At the heart of the model lies the concept of “health capital”, which is a stock that depreciates but can be augmented by health investments akin to physical capital accumulation. Herein lies the key problem: The fact that health depreciation is assumed proportional to the stock of health implies that the late-in-life health stock becomes largely unaffected by initial conditions, such as those prompted by *in utero* shocks, as the initial conditions “depreciate away” over the life course. The problem is further aggravated by the need to assume an accelerated rate of health depreciation with the passing of time in order for the Grossman model to account for mortality.¹

These difficulties are avoided, however, in the framework of health deficits as developed in Dalgaard and Strulik (2014). Based on research in the natural sciences the process of aging is conceptualized as a gradual loss of redundancy in the human body, which causes increasing frailty and ultimately death. This conceptualization of aging is anchored in the biological literature,

¹Assuming that the depreciation rate is age-dependent has other counterfactual implications. For example, Zweifel et al. (1999) demonstrate that among the elderly health expenditure is not predicted by chronological age once “time remaining until death” is controlled for. This suggests that health status, and not the year on the birth certificate, is what matters to health investments.

and can be given strong micro-foundations. Moreover, it leads to a law of motion for human frailty, which depends on physiological parameters and health expenditures. While the process of increasing frailty, measured by health deficits, is accelerating with age it may be slowed by health investments. In the context of the issue at hand the theory implies that health deficits accumulate exponentially over the life course, a prediction that has been repeatedly verified in research within gerontology, which means that small differences in initial conditions between individuals are *amplified* with the passing of time. The current paper demonstrates that this type of framework is well positioned to explain life course dynamics associated with the fetal origins hypothesis (and shocks in early childhood) in the context of long-run health outcomes as well as other socioeconomic outcomes. Indeed, the basic model has been adapted to study, among other things, the link between health and education (Strulik, 2018), years in retirement (Dalgaard and Strulik, 2017), the gender-gap in mortality (Schünemann et al., 2017b), and the health gap between married and unmarried individuals (Schünemann et al., 2019).

The health deficit model of human aging applies from early adulthood to death such that early-life health shocks are only indirectly represented as (small) initial health deficits in adulthood. In order to directly discuss the origin of shocks and their transmission during child development, we also develop a model of ontogenetic growth. This model derives child growth from conception to adulthood from first principles in thermodynamics, i.e. the energy needs to create and maintain body cells. We show that growth dynamics imply a natural separation of childhood into two distinct periods; an early period, in which shocks are amplified by child growth; and a later period, in which shocks are dampened by child growth. This provides a micro-foundation of the frequent assumption in economic models of child development that such distinct periods exist (Heckman, 2007; Almond and Currie, 2011a,b). The model naturally connects with the model of human aging by conceptualizing the period of child growth as the build-up of redundancy of (functioning) body cells and by conceptualizing the aging period as depletion of redundancy. We thus arrive at a unified theory of human development from conception to death.

The paper proceeds as follows. In the next section we compare basic versions of the health capital model and the health deficit model in their ability to account for the fetal origins hypothesis and elaborate on their testable implications. In Section 3 we provide a profound discussion of the impact of initial conditions on lifetime health outcomes within the health deficit model where investments are optimally determined. In Section 4, we set up the model of ontogenetic

growth and derive its implication for the transmission of health and nutrition shocks during childhood and their implications for health deficit accumulation in old age. Section 5 concludes.

2. BASIC MODELS

2.1. Health Capital Accumulation. The survey by Almond and Currie (2011a) provides an illustration of the inability of the Grossman (1972a) model to account for fetal origins. The illustration has the following law of motion for health capital as the main ingredient:

$$H_t = (1 - \delta)H_{t-1} + I_t, \quad H_0 \text{ given, } H > \underline{H} \quad (1)$$

in which H is the stock of health capital, δ is a constant rate of health capital depreciation, I represents health investment, and \underline{H} is a hypothesized lower boundary for health beyond which individuals expire. Repeated substitution leaves us with the following expression for the stock of health at time t :

$$H_t = (1 - \delta)^t H_0 + \sum_{i=0}^{t-1} (1 - \delta)^i I_{t-(i+1)}.$$

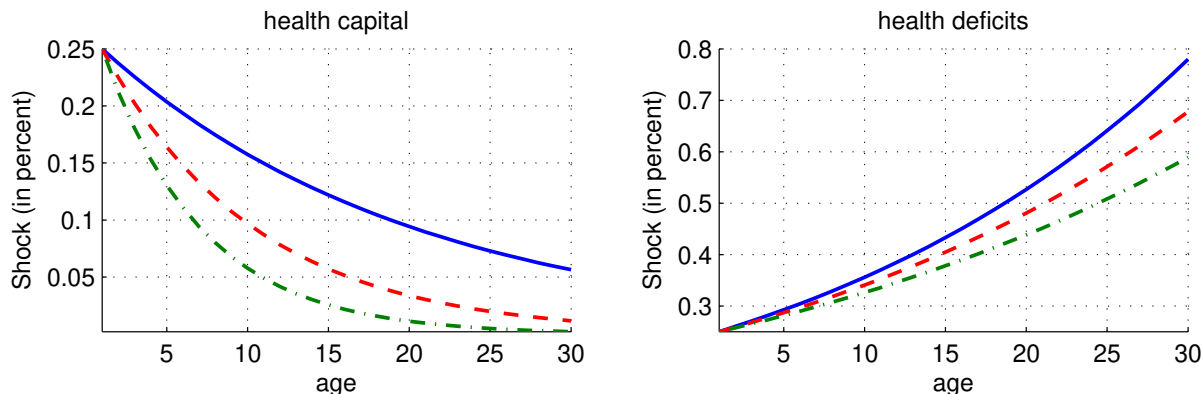
The key observation to make is that shocks *in utero* that influence initial health, H_0 , depreciate away with the passing of time. In general, events in the past are far less important to current health than recent events. This is an inevitable consequence of the basic assumption in the Grossman model that health depreciates in proportion to the stock of health. In principle, the model therefore imposes that healthy individuals age faster than unhealthy (or elderly) individuals, *ceteris paribus*. Consequently, initial conditions will be of little consequence later in life.

The panel on the left hand side of Figure 1 provides some numerical illustrations of this point, replicating Figure 1 of Almond and Currie (2011a). It shows how an initial shock, which creates a 25 percent deviation in initial health to a reference individual, depreciates with age for three different rates of health capital depreciation. At five percent depreciation the initial 25 percent deviation is melted down to about a five percent deviation at age 30. At 15 percent, initial differences are basically equalized at age 30.²

²More formally, Figure 1 shows the impact on the long run relative level of health of two individuals (1 and 2, say) after one is hit by a shock at time zero:

$$d(H_t^1/H_t^2) = (1 - \delta)^t dH_0^1$$

FIGURE 1: SHOCK PERSISTENCE BY AGE: HEALTH CAPITAL VS. HEALTH DEFICITS ACCUMULATION



The figure shows how persistent a 25 percent negative shock to the birth endowment would be given alternative annual depreciation rates. Left: blue (solid) line: 5 percent depreciation; red (dashed) line: 10% depreciation; green (dash-dotted): 15% depreciation. Right: blue (solid) line: $\mu = 0.04$; red (dashed): $\mu = 0.035$; green (dash-dotted): $\mu = 0.03$ ($E = 0.02$ and $D_0 = 0.02$).

As it turns out, the Grossman model actually holds a stronger prediction than what is indicated by the experiment conducted in Figure 1. Observe that the absolute difference in health capital between two individuals ($i = 1, 2$ respectively) with different initial conditions (i.e., different H_0), in the absence of health investments, is given by

$$H_t^1 - H_t^2 = (1 - \delta)^t (H_0^1 - H_0^2).$$

Hence, the Grossman model implies a stronger version of non-persistence than convergence in relative health levels: Namely, *absolute* convergence in health levels between individuals with different initial conditions, holding investments fixed.

So far we have assumed a constant rate of health depreciation, δ . Naturally, in the Grossman model the depreciation rate is not constant, as it theoretically would enable individuals to “live forever” contingent on sufficient health capital investments (see e.g. Grossman (1972a), section III). Instead the depreciation rate is assumed to increase with the passing of time as the individual ages. Obviously, this only serves to strengthen the prediction that initial health shocks lose significance with the passing of time. In this case, depreciation of initial health differences become faster than geometric.

Notice that $\delta > 0$ is sufficient to generate convergence. This implies that convergence is also obtained in refinements of the health capital model. For example, we could allow for individual-specific and/or stochastic depreciation rates, as briefly discussed in Grossman (1972b).

In Appendix B, we discuss a variant of the health capital model where early-life shocks imply a greater depreciation rate of health capital. This leads to the prediction that the health impact of early-life shocks is amplified at young ages but preserves the feature of convergence of initial health differences in old age.

The exercise conducted in Figure 1 and the subsequent discussion set health investments to zero. This may seem to open the door to a simple way of reconciling the Grossman model with fetal origins, namely through investments. However, as pointed out by Almond and Currie (2011a, p. 158) in the context of the illustration depicted in Figure 1:

If investments in all periods subsequent to the shock are affected by the shock, then prenatal exposures could be important for adult health in the Grossman (1972a) framework. However, the fetal origins literature posits an important and persistent biological effect of the prenatal period – that is, holding investments fixed.

It is important to appreciate that fetal origins, from the point of view of the medical literature, involve a specific mechanism. The early literature argued that environmental shocks would “program” the fetus with a predisposition towards various diseases, like coronary heart disease (e.g., Barker, 1995). Today a widespread view is that early-in-life shocks affect late-in-life health outcomes due to epigenetic changes. That is, changes in hereditary traits brought on by environmental influence (e.g., Gluckman and Hanson, 2004; Wu et al., 2004; Hilakivi-Clark and De Assis, 2006; Dolinoy et al. 2007; Waterland and Michels, 2007; Sinclair et al, 2007; Thompson and Einstein, 2010). Animal trials have been instrumental in providing proof of the principle of the fetal origin’s hypothesis (McMullen and Mostyn, 2009, for a review). Accordingly, in order to fully account for the fetal origins hypothesis a theory would have to allow for an influence from initial conditions on long run health outcomes *conditional* on investments, for purely biological reasons. The Grossman model does not allow for such a line of influence, as seen above.

Finally, it should be observed that while the standard Grossman model does not suggest that initial conditions influence subsequent investments, the more recent work by Heckman (2007) does. The theory of human capability formation creates dynamic complementarities by assuming that health investments happen at two (or more) distinct periods in life such that health outcomes are produced with the distinct health investments as inputs. Since negative early-in-life shocks, or low initial investments, reduce the productivity of future investments,

early-in-life events can have very persistent effects. The end result is not a given, of course. At present, the evidence in favor of dynamic complementarities seems to be largely descriptive in nature. Moreover, while some studies find that parental investments reinforce shocks, implying persistence in early-in-life shocks through investments, other studies find that parents act in a compensatory fashion.³

When dynamic complementarities are introduced into the Grossman model the resulting framework can generate persistence, which is broadly consistent with the fetal origins hypothesis. As should be clear, however, initial conditions only influences eventual outcomes *via* investments. As a consequence, the “Grossman-Heckman” framework cannot account for an impact of initial conditions on late-in-life outcomes holding investments fixed, and by extension it cannot account for the fetal origins hypothesis as it is conventionally understood in the medical literature. Combining the “Heckman mechanism” with our deficit model would not have this drawback, as will be clear from the discussion to follow.

2.2. Health Deficit Accumulation. The health deficit model of Dalgaard and Strulik (2014), in its simplest form, can be written as

$$D_t - D_{t-1} = \mu(D_{t-1} - E), \text{ for } D < \bar{D}, D_0 \text{ given}, \quad (2)$$

where D denotes health deficits and it is measured by the health deficit index as the relative number of health deficits that a person has out of a long list of potential deficits. Accordingly, the index is defined on a 0 to 1 scale, and aging (declining health status) occurs as the index gradually traverses towards one. In general, individuals with a higher deficit index are to be considered more frail, and thus physiologically older. In practise the process of deficit accumulation continues until an upper boundary for deficits, \bar{D} , is reached at which point the individual expires. The parameter μ is the “natural” rate of aging, and E is an “environmental constant”. Equation (2) derives from the literature on gerontology and the underlying parameters have been estimated with great precision (see Dalgaard and Strulik, 2014). For example, empirical estimates suggest that μ is between 0.03 and 0.045, depending on gender and country of origin

³See Currie and Almond (2011b), Almond and Mazumder (2013), and Almond et al. (2018) for a detailed discussion of the Heckmann (2007) model and reviews of the parental investment literature. In Section 4 we propose a model of child development that endogenously generates a division of childhood in two distinct periods with shock amplification in early childhood and shock dampening in adolescence.

(Mitnitski et al. 2002a).⁴ By extension, it is worth noting that in contrast to the Grossman model where the object of interest – health capital – is an unobserved variable, the object of interest in the present model – health deficits – is empirically observable.

Repeated substitution of (2) leads to

$$D_t = (1 + \mu)^t D_0 - \sum_{i=0}^{t-1} \mu^{1+i} (1 + \mu)^i E_{t-(i+1)}.$$

Since $(1 + \mu)^t$ grows with increasing age t , an inherent feature of the health deficit model is that early-in-life shocks that influence the initial relative number of deficits, D_0 , are amplified over time. This creates a force of divergence: initially unhealthier individuals accumulate health deficits faster than initially healthy individuals.

The panel on the right hand side of Figure 1 provides a numerical illustration of this feature. As in the previous section we study the impact a health shock that creates a 25% deviation in initial deficits relative to a reference individual. The deviation from benchmark increases over time. For $\mu = 0.04$, the initial 25% deviation has reached 80% percent at the age of 30.⁵

As noted above, the interpretation of E in the natural science literature is that of “environmental” influences. While some such influence can be external to individuals (such as pollution), E may also be influenced by deliberate health investments. By increasing E such investments will serve to slow down the process of deficit accumulation and thus provide the prospect of a longer life. In the illustration in Figure 1, the level of E is ignored so as to provide a clean comparison with the properties of the Grossman model in the absence of health investments. Nevertheless it is also of interest to understand the consequences of allowing for optimal health investments in the presence of shocks to initial deficits within the deficit model. The next section therefore studies the impact from initial deficits in the original health deficit model.

As a final remark on the properties of the basic deficit model its worth observing that it also holds radically different implications from the Grossman model in terms of the evolution of *absolute* health differences. Comparing the absolute difference in health deficits between two

⁴The exponential nature of health deficit accumulation as been confirmed in a variety of studies for samples from different populations, see e.g. Shi et al. (2011); Harttgen et al. (2013); Mitnitski and Rockwood (2013, 2016); Abeliatsky and Strulik (2018a).

⁵More formally, Figure 1 shows the impact on the long run relative level of health deficits of two individuals (1 and 2, say) after one is hit by a shock in utero (time zero):

$$d(D_t^1/D_t^2) = (1 + \mu)^t dD_0^1$$

individuals ($i = 1, 2$ respectively) with different initial conditions (i.e., different D_0), in the absence of health investments ($E = 0$), is given by:

$$D_t^1 - D_t^2 = (1 + \mu)^t (D_0^1 - D_0^2).$$

Hence, initial differences in health deficits are amplified and the model thus predicts absolute divergence in health holding investments fixed.

2.3. Testable Implications. The theoretical discussion clarifies how an empirical researcher can discriminate between the health capital model and the health deficit model. Indirectly, any evidence of late-life health repercussions that can be causally related to early-life health shocks rejects the health capital model and confirms the health deficit model. Direct evidence, however, would employ a measure of health that can be tracked over time for individuals and ideally it would be based on the same metric to measure health deficits as the theoretical model, i.e. the deficit index. For a given health shock, the health deficit framework predicts that (health) outcomes between affected and non-affected individuals *amplify* throughout adulthood. In contrast, the Grossman model (even augmented with dynamic complementarities) predicts the opposite; namely that the initial shock is diminishing *during adulthood*. These alternative predictions are clearly visualized in Figure 1.

As a result, a discriminating test between the two models is to check if the health effects of an environmental shock in utero (or during childhood) are increasing during adulthood. This is, of course, somewhat data demanding in that it requires that the affected and non-affected individuals are observed more than once in adulthood (i.e., a panel of individuals). Yet first evidence along these lines is available. Abeliasky and Strulik (2018b) show that individuals exposed to hunger episodes during childhood diverge year-by-year during adulthood, in terms of health deficits, from comparable non-exposed individuals. Abeliasky and Strulik (2019) investigate a very mild health shock, namely the season of birth. They show that individuals born in autumn develop less health deficits than those born in spring and that the difference in health deficits gets larger with advancing age. This corresponds with the finding of Doblhammer and Vaubel (2001) that individuals born in autumn live longer than individuals born in spring.

3. SHOCK PERSISTENCE IN THE HEALTH DEFICIT MODEL

We begin by rewriting equation (2) for a continuous notion of age and separating E into a “real” environmental constant a and the impact of health investment on health deficit accumulation:

$$\dot{D}(t) = \mu(D(t) - a - Ah(t)^\gamma). \quad (3)$$

Here, the parameters $A > 0$ and $0 < \gamma < 1$ reflect the state of the health technology, and h is health investment. While A refers to the general power of health expenditure in maintenance and repair of the human body, the parameter γ specifies the degree of decreasing returns of health expenditure. The larger γ the larger the relative productivity of cost-intensive high-technology medicine in maintaining and repairing deteriorated human bodies. Bad health promotes death such that individuals die when \bar{D} health deficits have been accumulated.

Individuals are interested only in maximizing their lifetime utility from consumption:

$$\int_{\tau}^T e^{-\rho(t-\tau)} u(c(t)) dt, \quad (4)$$

with $u(c) = (c^{1-\sigma} - 1)/(1-\sigma) + b$ for $\sigma \neq 1$ and $u(c) = \log(c) + b$ for $\sigma = 1$. Here σ is the inverse of the elasticity of intertemporal substitution and ρ is the rate of time preference. Allowing for death to be a stochastic event and considering health as an element in the utility function leads to some further interesting results but does not change the basic insight on the accumulation of health deficits (see Strulik, 2015a, Schuenemann et al., 2017a). We thus focus on the simpler model here.

Besides spending income on final goods, individuals may save or borrow at a net interest rate r . Individuals take all prices as exogenously given. The law of motion for individual wealth k is thus given by (5):

$$\dot{k}(t) = w + rk(t) - c(t) - ph(t), \quad (5)$$

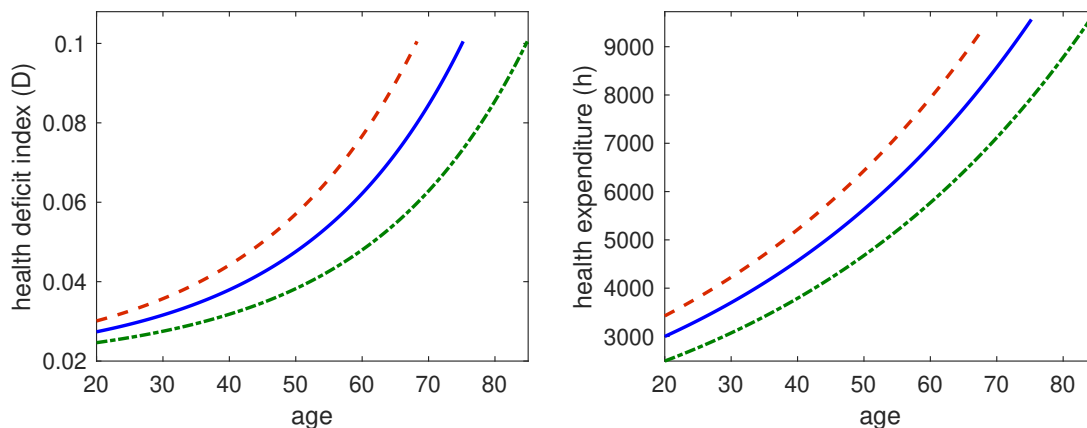
in which w is the (annual) wage, r is the interest rate, and p is the price of health goods. The problem is to maximize (4) subject to the accumulation equations (3) and (5), the initial conditions $D(\tau) = D_\tau$, $k(\tau) = k_\tau$, and the terminal conditions $k(T) = \bar{k}$, $D(T) = \bar{D}$. At the very basic level the problem is to trade off the benefits and costs of health investments over the life cycle. The benefits consists in, by slowing down the process of aging, a longer life which allows for more consumption along the extensive margin. However, by increasing health investments,

individuals forego consumption in the current period. In Appendix A we provide the details on the analytical solution of this free terminal value problem. Here, we present the numerical solution of a calibrated version.

We take the calibration of the model for an average 20 years old male U.S. American in the year 2000 from Dalgaard and Strulik (2014). This means that we set the rate of aging μ to 0.043, which is the rate of health deficit accumulation estimated for Canadian men by Mitnitski et al. (2002a). Rockwood and Mitnitski (2007) stress the similarity of their results for U.S. and Canadian populations but they do not report the detailed results for the U.S. analysis. We set the interest rate to 6 percent, following Barro et al. (1995), and we set $\gamma = 0.19$ to capture the growth of health spending at an annual rate of two percent over the life cycle (Keehan et al., 2004). From the estimates of Mitnitski et al. (2002a) we set $D(0) = 0.0274$ as the relevant initial value at age 20 and $\bar{D} = 0.1005$, i.e. 55.2 years later since the life-expectancy of a 20 year old U.S. American in the year 2000 was 55.2 years. We set $a = 0.013$ such that the model predicts a life-expectancy at age 20 of 42 years for $A = 0$ (corresponding to the life expectancy in the late 19th century when adult life expectancy was only modestly affected by medical technology). We set $\rho = r$ such that the age-consumption profile is constant over the life as obtained by Browning and Ejrnæs (2009) for childless households. We take GDP per worker in the U.S. in the year 2000 (PPP\$ 77,003) and assume a capital share of 1/3, which implies an annual labor income (in international dollars) of \$ 51,335. We normalize $p = 1$ and set $\sigma = 1$ in order to obtain a value of life at age 20 consistent with the estimate of Murphy and Topel (2006). Finally, we estimate $A = 0.00139$ such that the individual dies with deficits \bar{D} at age 75.2, according to the life-expectancy of 20 years old U.S. Americans in the year 2000.

In Figure 2 we replicate the benchmark run of Dalgaard and Strulik (2014), represented by blue (solid lines). We then look at an individual that is initially 10 percent less healthy than the Reference American, represented by red (dash-dotted) lines, and an individual that is initially 10 percent healthier than the Reference American. These differences in initial health deficits at age 20 can be thought of as resulting from negative health shocks earlier in life (and perhaps in utero). In line with observations, the model predicts that unhealthier individuals spend more on health (panel on the right hand side). But the calibrated health technology is not powerful enough to equalize initial health differences. In fact, initial health differences get amplified over time: as individuals age, the vertical distance between the individuals' deficit trajectories gets

FIGURE 2: INITIAL HEALTH AND HEALTH DEFICIT ACCUMULATION



Blue (solid) lines replicate results for the Reference American in Dalgaard and Strulik (2014). Green (dashed) lines: individual with 10 percent less initial health deficits at age 20. Red (dash-dotted) lines: individual with 10 percent more initial health deficits at age 20.

larger, see the panel on the left-hand side of Figure 2. The underlying reason for this pattern is that initial deficits influence the effectiveness of health investments: the greater the health deficits the smaller the impact of a given amount of health investments in prolonging life. In this sense the model involves dynamic complementarities akin to those found in the human capability theory (Heckman, 2007).⁶

4. A THEORY OF CHILD DEVELOPMENT AND ITS IMPACT ON LATE-LIFE HEALTH DEFICITS

4.1. A Simple Theory of Ontogenetic Growth. As is evident, the health deficit model tracks the evolution of health deficits, health expenditure, and consumption over the life cycle of adults, starting from about the age of 20. Hence, in a strict sense, the analysis does not fully capture fetal origins. Fetal origins are “only” represented as initial values of health deficits in adulthood $D(0)$. In order to directly discuss the origin of shocks and their transmission during childhood, we propose a model of ontogenetic growth, which provides the missing link between biological aging of adults and the fetal origins of late-life health. There are at least two main differences between health development of children and adults. (i) There is no rational health investment from the side of the fetus or child. Instead the fetus receives nutrients and perhaps

⁶The simple health deficit model of Dalgaard and Strulik (2014) focusses on health expenditure and consumption as the only behavioral variables. There exist a couple of further developments using the health deficit approach that extend the basic model by further behavioral choices and we explored the role of fetal origins (i.e. of higher initial health deficits) in some of these models as well. For example, the model of Dalgaard and Strulik (2017) predicts that individuals who start out with more initial health deficits retire earlier. The model of Strulik (2018) augmented by endogenous retirement predicts that individuals who start out with more initial health deficits achieve less education and earn less lifetime income.

negative health shocks from the mother, and children receive nutrition and health investments from their parents. (ii) There is body growth. Child development can be understood as the accumulation of body cells such that an initially nonviable fetus gets healthier over time. Here we propose a model of ontogenetic growth that is naturally unified with the model of human aging through the notion of redundancy. Growth in utero and in childhood can be understood as the build-up of body-cell redundancy and organ reserve. Aging, in gerontology, is understood as depletion of redundancy in (functioning) body cells, i.e. as accelerated loss of organ reserve such that individuals become increasingly frail. It has been estimated that initially, as a young adult, the functional capacity of human organs is tenfold higher than needed for survival (Fries, 1980). Gavrilov and Gavrilova (1991) provide a micro-foundation of human aging from reliability theory, understood as the gradual loss of functionality of elements (body cells), which causes the loss of organ functionality and, eventually, death. By combining the periods of childhood (build-up of redundancy) and adulthood (depletion of redundancy) we arrive at a unified theory of human development from conception to death.⁷

Growth of humans and other mammals is appropriately described by the von Bertalanffy equation (von Bertalanffy, 1957; Kooijman, 2000; West et al., 2001). The equation originates from an energy balance stating that the energy consumed is equal to the energy used for the creation and maintenance of body cells such that $B = bm + e_c \dot{m}$, in which B is the energy consumed, m is the number of body cells, b is the energy needed to maintain a cell and e_c is the energy needed to create a cell. Assume (without loss of generality) that a cell has unit weight and that $e_c = 1$, then the number of cells and thus the weight of the fetus evolves as:

$$\dot{m} = B - bm. \tag{6}$$

Energy flow per unit of time is given by:

$$B = \min \left\{ am^\beta, \bar{B} \right\}, \tag{7}$$

⁷A related, yet different issue is whether and how shocks in early life affect parental child investments and whether and to which degree these investments can remediate the effects of the shock (Almond and Currie, 2011b; Almond and Mazumder, 2013). Here, we implicitly consider a reduced-form measure of shocks that includes already parental adjustment behavior (e.g. the shock comprises an infection as well as the impact of parental investment on the cure of the infection). Dalgaard and Strulik (2015; 2016) consider a simplified version of child growth in an overlapping generations model of optimal child investments and fertility. They use the model to explain, among others things, why children in larger families are smaller and why children grow to become taller adults in richer societies.

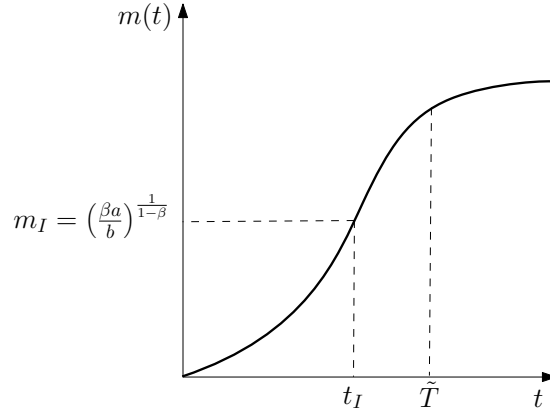
in which \bar{B} is an energy supply constraint. If $B = am^\beta$, energy supply is unconstrained and equation (7) states that the metabolic rate scales with the body size of the fetus (or child). This allometric relationship between energy consumption B and body mass m is known as Kleiber’s Law (Kleiber, 1932). The scaling parameter β is estimated with high precision as $3/4$ for mammals and almost all terrestrial animals, yielding the famous “mouse-to-elephant curve” (Brody, 1945). West et al. (1997) provide a microfoundation of the scaling law by showing that organisms, viewed as energy transporting networks that minimize energy dissipation, fulfil Kleiber’s law with $\beta = 3/4$. The fact that $\beta < 1$ implies that larger bodies are more energy-efficient in the sense that they need less energy to maintain a body cell.

Inserting $B = am^\beta$ into (6) and solving the resulting Bernoulli differential equation provides the famous von Bertalanffy equation for body size (von Bertalanffy, 1957):

$$m(t) = \left\{ a/b - \left[a/b - m(\tau)^{1-\beta} \right] e^{-(1-\beta)b(t-\tau)} \right\}^{\frac{1}{1-\beta}}, \quad (8)$$

in which τ is the initial time and $m(\tau)$ is the initial body mass. For $\tau = 0$ the equation describes ontogenetic growth from the time of conception. West et al. (2001) provide a microfoundation for the von Bertalanffy equation for $\beta = 3/4$.

FIGURE 3: ONTOGENETIC HUMAN GROWTH



Graphical representation of the von Bertalanffy equation; (t_I, m_I) : inflection point; \tilde{T} : age at end of growth.

Empirically, the von Bertalanffy equation is a good approximation of body growth if food supply is abundant and it is frequently used to describe ontogenetic growth of humans and other animals (Kooijman, 2000; Karkach, 2006). Graphically, the von Bertalanffy equation has a sigmoid shape, as shown in Figure 3. The figure shows the full trajectory of human growth from conception to adulthood. If growth would never stop, body size would converge towards

$m = (a/b)^{1-\beta}$, as can be read off from (8) for $t \rightarrow \infty$. However, human growth is determinate, which means that it stops at a certain age $t = \tilde{T}$. While \tilde{T} is idiosyncratic it is typically reached a few years after sexual maturity. An implication that we note for later purpose is that $m(\tau) < (a/b)^{1/(1-\beta)}$ for any arbitrary initial body size $m(\tau)$.

The sigmoid shape implies that there exists an inflection point (at age t_I) at which the shape of the growth curve changes from convex to concave. The existence of the inflection point requires no additional assumption but follows as result from the simple thermodynamics underlying (8). This provides a natural explanation for why shocks in utero and in early childhood are particularly damaging for human development: they occur at an age $t < t_I$, for which initial shocks are amplified during ontogenetic growth. In order to develop this feature in more detail, consider the differential equation $\dot{m} = am^\beta - bm$ that follows from (6) and (7) in the case of no energy supply restrictions. Taking the second derivative we obtain $\ddot{m} = \beta am^{\beta-1} - b$. Noticing that the age-growth pattern is convex for $\ddot{m} > 0$, we conclude a convex growth for

$$m(t) < m_I \equiv \left(\frac{\beta a}{b} \right)^{\frac{1}{1-\beta}}, \quad (9)$$

in which m_I is body size at the inflection point (see Figure 3). Inserting m_I for $m(t)$ in (8) and solving for age, we obtain an analytic solution for age at the inflection point (t_I) and conclude that human growth follows a convex pattern at ages

$$t < t_I \equiv \frac{1}{(1-\beta)b} \log \left[\left(\frac{a/b - m(\tau)^{1-\beta}}{a/b} \right) \frac{1}{1-\beta} \right] + \tau, \quad (10)$$

in which $m(\tau)$ is initial body size at age $\tau < t_I$. Since $\ddot{m} > 0$ for $t < t_I$, shocks (due to nutrition constraints, accidents, or infections) that occur at ages smaller than t_I are amplified during ontogenetic growth, whereas shocks at ages larger than t_I are dampened.

In order to see the condition for shock amplification more clearly, consider a shock that changes body size at time τ and compute how it translates into a change in body size at age t , $t > \tau$:

$$\frac{\partial m(t)}{\partial m(\tau)} = \left\{ a/b - \left[a/b - m(\tau)^{1-\beta} \right] e^{-(1-\beta)b(t-\tau)} \right\}^{\frac{\beta}{1-\beta}} e^{-(1-\beta)b(t-\tau)} m(\tau)^{-\beta}.$$

Next, compute how this effect changes with advancing age of the individual:

$$\begin{aligned} \frac{\partial^2 m(t)}{\partial m(\tau) \partial t} &= \frac{\beta}{1-\beta} x^{\frac{\beta}{1-\beta}-1} \left[a/b - m(\tau)^{1-\beta} \right] (1-\beta) b e^{-(1-\beta)b(t-\tau)} e^{-(1-\beta)b(t-\tau)} m(\tau)^{-\beta} \\ &\quad - x^{\frac{\beta}{1-\beta}} (1-\beta) b e^{-(1-\beta)b(t-\tau)} m(\tau)^{-\beta}, \end{aligned} \quad (11)$$

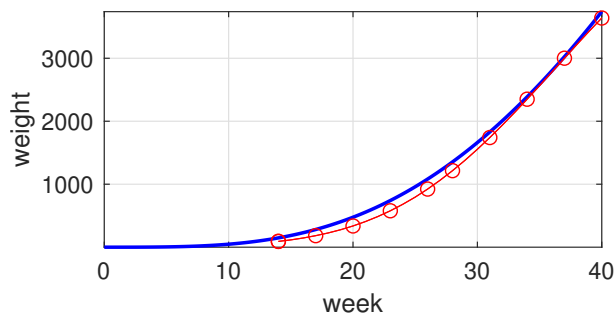
with $x \equiv a/b - [a/b - m(\tau)^{1-\beta}] e^{-(1-\beta)b(t-\tau)}$. Generalizing the computational experiments from Section 3, we conclude that initial differences between individuals of different body size at time t (individuals who experienced different shocks) are amplified with age if $\frac{\partial^2 m(t)}{\partial m(\tau) \partial t} > 0$ and that initial differences are dampened with age if $\frac{\partial^2 m(t)}{\partial m(\tau) \partial t} < 0$. For $\frac{\partial^2 m(t)}{\partial m(\tau) \partial t} = 0$ the initial difference is preserved as individuals get older. Solving $\frac{\partial^2 m(t)}{\partial m(\tau) \partial t} = 0$ for t , we obtain after some algebra, that initial differences are amplified for ages $t < t_I$ (see Appendix C for details).

In related studies, Heckman (2007) and Almond and Currie (2011a,b) propose a model of childhood development inspired from production theory in economics. There, childhood is arbitrarily divided into two periods and child investments in the two periods combine via a CES production function to produce health or human capital of the child. For specific assumptions on the parameters, the result can be achieved that shocks matter more in the first period in the sense that a shock in the first period cannot be completely reversed by an equally sized shock in the opposite direction in the second period. Here, we consider development of the child to be continuous in time with no arbitrary separation into periods and without imposing an economic production function. We then derive from simple thermodynamic principles of cell maintenance and creation an inflection point that endogenously separates childhood into two periods. In the first period, shocks are amplified during ontogenetic growth while in the second period shocks are dampened.

4.2. In Utero Development. Figure 4 shows a calibration of the model for in-utero development when energy supply is unconstrained and there are no shocks. Parameters are $a = 1.08$, $b = 0.07$, $\beta = 3/4$ and $m(0) = 0.0001$. Body weight is measured in grams. The blue solid line shows the model prediction for body weight by week of gestation, the red dotted line shows the data for American boys from Kiserud et al. (2017). The age growth curve is clearly convex during in-utero growth, implying that shocks during this period will be amplified as the child develops.

The model offers two distinct possibilities to introduce impediments to child development: energy constraints and cell-damaging shocks. We first consider cell-damaging shocks. For these considerations it is helpful to recall that the microfounded model of human aging (Gavrilov and Gavrilova, 1991) derives the Gompertz-Makeham law of human aging from the assumption that not all elements of the body are perfectly functioning at the initial age of adulthood. Here, we use the fact that the beginning of adulthood coincides with the end of childhood and associate

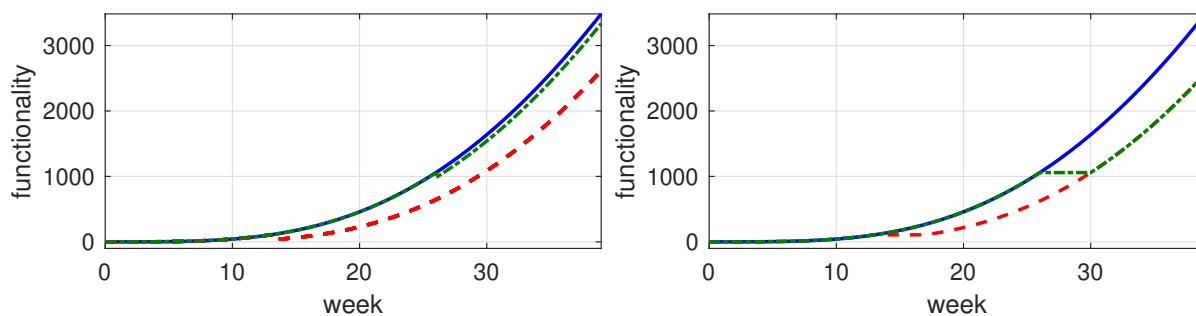
FIGURE 4: ONTOGENETIC HUMAN GROWTH IN UTERO



Solid blue line: model; dotted red line: data from Kiserud et al. (2017) for male fetal weight during gestation.

body cells with the “elements” in Gavrilov and Gavrilova (1991). We measure functionality of the body by the total mass of functioning body cells. Health shocks (like accidents or infections) are then conceptualized as exogenous damage or destruction of body cells. In Figure 5 we apply this methodology to the calibrated model. The solid line re-iterates the unshocked trajectory of in-utero growth (for which created body cells are equal to functioning body cells). In the panel on the left hand side we consider the loss of functionality of 75g of body cells at the beginning of the second trimester (red dashed lines) and at the beginning of the third trimester (green dash-dotted lines). Obviously, since growth is convex in utero, shocks of a given size matter more when they happen earlier in the gestation period and the distance of functionality between the shocked embryo and the unshocked embryo increases during the in-utero growth period.

FIGURE 5: SHOCKS AND FUNCTIONALITY DURING IN-UTERO GROWTH



Solid blue lines: benchmark model (unhindered growth). Left: functionality shock (accident) in form of loss of 75g of functioning body cells at the beginning of the second trimester (red dashed lines) and at the beginning of the third trimester (green dashed-dotted lines). Right: nutritional constraint binding at $\bar{B} = bm(\tau)$ for 4 weeks (Ramadan) with onset at $\tau = 13$ (red dashed lines) and $\tau = 26$ (green dash-dotted lines).

Next consider energy constraints. If food is not abundant, fetal growth is constrained by maternal supply of energy (or, more generally, nutrients), $B = \bar{B}$. Naturally, if the fetus is energy-constrained, a larger share of energy is needed for maintenance and less is available for

cell creation such that growth is retarded in these periods. Suppose energy supply in utero is proportional to the metabolic rate of the mother and that the mother’s metabolic rate scales with her size according to Kleiber’s law such that $\bar{B} \propto M^{3/4}$ where M is the size of the mother and $M^{3/4}$ is the metabolic rate of the mother (Kleiber, 1932; West et al., 1997). We then conclude that, *ceteris paribus*, fetuses with larger mothers are less likely (or less frequently) energy-constrained and are thus larger at birth. More generally, we conclude from equations (6)–(8) that (i) fetuses who are less frequently energy-constrained in utero are bigger at any time in utero and thus also at birth, (ii) individuals who are born too early, i.e. at low t , are smaller at birth.

As an example application, consider a Muslim mother who observes the Ramadan during pregnancy (Almond and Mazumder, 2011). We model this in a drastic sense such that for the four weeks during Ramadan there is no child growth and consider an onset of Ramadan at the beginning of the second or third trimester. Formally, we normalize the nutritional shock such that the nutritional constraint binds at $\bar{B} = bm(\tau)$ for 4 weeks with $\tau = 13$ and $\tau = 26$. The panel on the right hand side of Figure 5 shows the results. The solid line reiterates again the growth trajectory for unconstrained energy supply. Dashed lines show results for Ramadan onset at $\tau = 13$ and dash-dotted lines for $\tau = 26$. The results can be interpreted in two different ways. First, the model predicts that the mother’s observation of Ramadan during pregnancy leads to lower birth weight. Second, the model predicts that the newborn child has a lower number of functioning body cells (compared to unhindered growth) and is thus of inferior health. The figure also shows the limitations of the small model. Since all body cells are treated equally, there is no impact of the timing of Ramadan during pregnancy. In order to explain a greater impact of early life nutrition shocks one would need to differentiate between the importance of body cells and introduce a mechanism why energy constraints have more severe impact early in the gestation period.⁸ Notice the difference between body growth and functionality in the two scenarios. In the scenario on the left hand side of Figure 5 (accidents, infections) the body grows unhindered from an energy perspective and the shock reduces functionality of existing cells. In

⁸Alternatively, one could argue that the energy constraint (7) is more likely to be binding if the fetus is already large. This feature would imply that the last trimester is more important for body growth and nutritional constraints. Such an outcome would be consistent with the finding of Doblhammer and Vaupel (2001) and Abeliatsky and Strulik (2019) that individuals born in autumn (for whom fresh fruits have been abundant in summer) age slower than individuals born in spring.

the scenario on the right, energy restrictions impede body growth and thus the full development of functioning cells during ontogenetic growth.

4.3. Childhood Development. The theory of in utero development can be linked with the theory of aging, which applies to full-grown adults (starting at about age 20) by a period of growth in childhood. This leads to an encompassing theory of ontogenetic growth from conception to adulthood. Child growth after birth can also be approximated by the von Bertalanffy equation but less accurately than during gestation. The reason is that there are periods of growth spurts and growth slowdowns in childhood and the simple equation is not able to capture these details precisely. The supply of energy (nutrients) in childhood is no longer constrained by the metabolic rate of the mother but by the provision of food by the parents. In Dalgaard and Strulik (2015; 2016) we developed a theory of human growth where nutritional investments were conceptualized as a choice variable of parents. Here, for simplicity, we take again the position of the developing child for whom nutrition is exogenous and the energy flow is thus given by differential equation (6) and constraint (7).

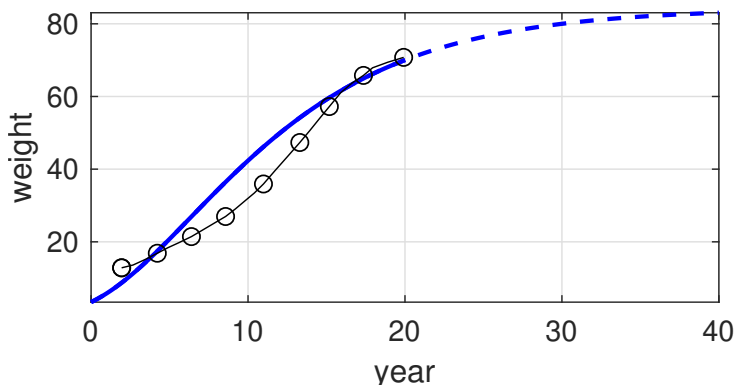
For child growth, we measure t in units of years and body mass in kilograms. Figure 6 shows a calibration for $a = 1.52$, $b = 0.5$, $\beta = 3/4$, and birth weight $m(0) = 3.4$ kg (solid line). Dots show the actual average weight of US boys in the year 2000 according to CDC (2000). The end of the solid line at age 20 indicates the end of the determinate growth period. The dashed line indicate how growth would proceed if growth would be indeterminate and would never stop. We see that the growth trajectory is concave in middle childhood and adolescence (after about age 7) but that the concavity is not strong (compared to convexity in utero). The trajectory from age 7 to 20 is almost linear. This means that there is small amplification of shocks in early childhood until about age 7 followed by a period of small dampening of shocks until the onset of adulthood.

Like in utero development, the development of child health is potentially affected in two distinct ways. (i) The nutritional constraint \bar{B} is occasionally binding, e.g. the child experiences a hunger period. (ii) Adverse shocks (infections and accidents) reduce the number of functioning cells. We could also imagine a combination of both, e.g. if fighting infections needs energy and thus reduces the energy available for body growth. This could be conceptualized as an increase in the energy needed to maintain a body cell, i.e. a larger b during the infection period.

We next investigate how adverse shocks affect the development of functioning body cells at different ages. Functionality is again measured by the mass of functioning body cells. Specifically, we consider shocks conceptualized as loss of functionality of 3 kg of body cells at age 1 (red dashed lines) and at age 5 (green dashed-dotted lines). These are drastic shocks resulting, for example, from a severe accident or illness. This is admittedly a crude measure that does not take into account any difference of importance of specific cells (e.g. eye-cells vs. bone-cells). We solve the model as calibrated above twice: once for the health shocked and once for unhindered (optimal) childhood development. We then compute at any age the difference in body functionality between the shocked and the unshocked growth trajectory. In the panel on the left hand side of Figure 7, the solid (blue) lines represent differential development when the shock hits at age 1 and dashed (red) lines consider the shock at age 5. Shocks are amplified during child growth when the growth trajectory is convex, which is in the calibrated model until about age 7. After the inflection point, differences in functionality are dampened with further child development. In the application, this means that the child does not recover from the shock in early childhood. At age 20 (when development stops) the loss of functionality is about the same as at age 1 (with an interim phase of higher damage). The shock at age 5 is closer to the inflection point and is thus amplified for only a short period and then dampened, implying that the developing child recovers about 50% of the loss from the shock.

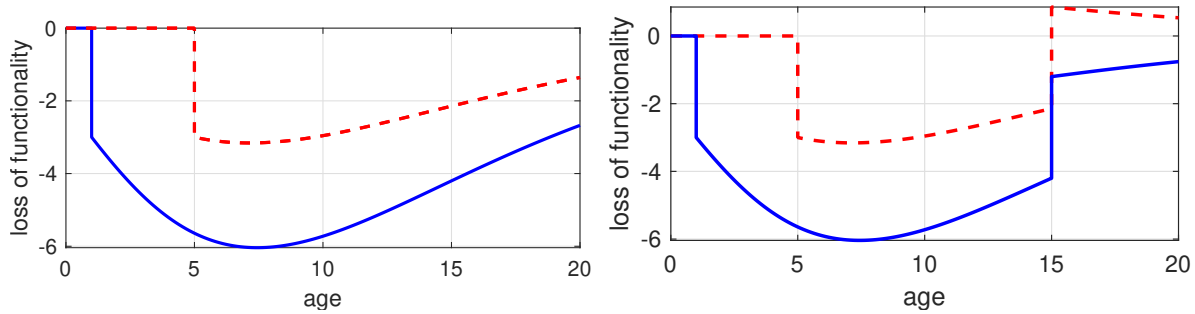
We finally consider whether and to which degree child investments in the second stage of child development can make up for or overturn the consequences from of shocks in early childhood. This exercise is inspired by the comparative static exercise performed in Almond and Currie

FIGURE 6: ONTOGENETIC GROWTH IN CHILDHOOD AND ADOLESCENCE



Solid line: model; dots: weight for age data for boys from CDC (2000); dashed line: hypothetical growth if growth were indeterminate (solution of (8) for age > 20).

FIGURE 7: LOSS OF FUNCTIONALITY AND REPAIR DURING CHILDHOOD GROWTH



Left panel: functionality shock (accident) in form of loss of 3 kg of functioning body cells at age 1 (solid blue lines) and at age 5 (red dashed lines). Loss of functionality is measured in terms of the difference in functioning body cells between the shocked child and an otherwise identical unshocked child. Right panel: as for left panel plus repair shock of the same size at age 15.

(2011b, Section 2.2.1) using the Heckman (2007) model. Specifically, we consider a positive shock in cell functionality at age 15 of the same absolute size as the negative shock experienced at age 1 or 5, respectively. Investigating such an admittedly naive repair shock is nevertheless interesting since it reveals non-obvious insights on the importance of the timing of shocks for child development. Results are shown in the panel on the right hand side of Figure 7. We see that the shock at age 5 is almost fully remediated by investments in adolescence. The shock at age 1, in contrast, is only incompletely repaired. As argued above, the inflection point can be considered as a natural subdivision of childhood (into an amplification period and a dampening period of shocks). Knowledge about whether a shock occurred in the earlier or later period, however, is not enough to infer whether it could be compensated with late-life investments. Instead, the timing is important. As a rule, the earlier shocks occur in childhood (including in-utero development) the less can they be remediated by investment in late childhood. Child development conceptualized as ontogenetic growth according von Bertalanffy is more complicated than the simple two-period model of Heckman (2007) suggests.

4.4. From Conception to Death: Ontogenetic Development, Aging, and Late-Life Health. Finally, we fit the pieces together. Based on insights from biology (Fries, 1980; Gavrilov and Gavrilova, 1991), we consider the period of ontogenetic growth in utero and in childhood as build-up of redundancy (of functioning body cells) and the adult period as depletion of redundancy which leads to loss of functioning, accumulation of health deficits, and eventually death. We connect the periods of ontogenetic growth and aging by assuming that health deficits at age 20 are inversely proportional to the mass of functioning body cells at age 20. Specifically, let the outcome of unhindered (optimal) ontogenetic growth in terms of cell functionality be

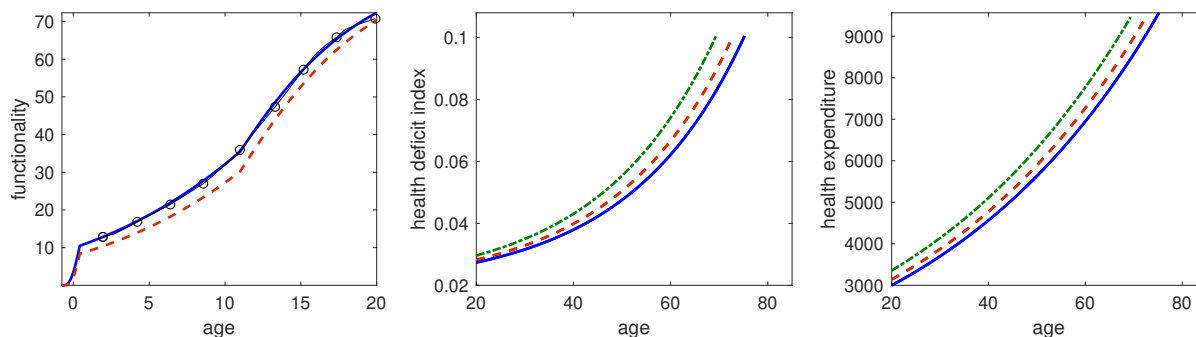
denoted by $F(20)^*$ and actual functionality be denoted by $F(20)$. Then, actual health deficits at age 20 fulfil $D(20)/D(20)^* = F(20)^*/F(20)$, in which $D^*(20)$ are initial health deficits given optimal child development.

For in-utero growth we use the von Bertalanffy equation as calibrated in Section 4.2. For childhood growth, the function calibrated in Section 4.3 does not fit the data very well. We thus follow the usual approach in empirical models of childhood growth models and estimate the equation piecewise (Karkach, 2006). It turns out that a subdivision in two periods leads already to a good fit of the data. The inflection point is obtained at age 11, suggesting that humans grow differently in early childhood and in adolescence. The best fit of the data is reached when the in-utero growth equation continues to hold for 22 weeks (0.4 years) after birth, suggesting that the period of rapid (strongly convex) growth is continued for a while after birth. We solve the model for time measured since conception. In representing the results, however, we follow the convention and convert time since conception to age and birth is associated with age zero. This means that conception takes place at age -0.75 (9 months before birth).

The solid (blue) line in the panel on the left-hand side of Figure 8 shows the matched curves when there are no shocks during ontogenetic growth. The black dots reiterate the data from Figure 6. The dash (red) line shows the trajectory for an otherwise identical individual who experienced an in-utero shock at the start of the second trimester that results in the loss of 100g of functioning body cells. The shock is greatly amplified during the gestation period, leading to a loss in birth weight of 1400g (birth weight 2.0 kg instead of 3.4 kg when unshocked). After birth, the shock is mildly amplified further until age 11 and then it is mildly dampened during adolescence. At age 20, the individual who experienced the in-utero shock, has accumulated 97% of the potential (optimal) cell functionality, implying that the dis-functionality and thus health deficit index is 3.0 percent greater than that of an un-shocked individual.

The center-panel of Figure 8 shows the aging process. The solid (blue) line reflects aging without in-utero shock, as calibrated in Dalgaard and Strulik (2014), see Figure 2. The dashed (red) line shows aging given the in-utero shock. The small initial difference is gradually amplified and the health distance between both individuals is greatest in old-age. Given the small shock, however, the distance is not great. The micro-foundation of aging, however, does not necessarily imply an exactly inversely proportional relation between health deficits and body functionality. It implies only a monotonous relationship (Gavrilov and Gavrilova, 1991). Generalizing, we next

FIGURE 8: CHILD DEVELOPMENT AND ADULT AGING



Left panel: Blue (solid) line unshocked (optimal) growth in utero and childhood. black dots: data according to CDC (2000). Red (dashed) line: functionality shock in form of loss of 100 g of functioning body cells in the 13th week of gestation. Middle panel: aging of the benchmark individual (solid blue line) and of the individual with shock experience in utero. Red (dashed) line: functioning body cells at 20 and health deficits at 20 are inversely proportional, $D(20)/D(20)^* = F(20)^*/F(20)$. Green (dash-dotted) line: monotonous transformation, $D(20)/D(20)^* = 1.02F(20)^*/F(20)$; Right panel: associated health expenditure.

assume that $D(20)/D(20)^* = \kappa F(20)^*/F(20)$ and set $\kappa = 1.05$ (instead of 1.00), implying that a three percent difference in functionality translates to a 8 percent difference in initial health deficits. The dash-dotted (green) line shows the implied aging trajectory. The divergence in health deficits is now better visible.

The panel on the right hand side of the figure shows the associated optimal health expenditure. Individuals who start out with more health deficits take countermeasures to reduce their faster aging but these investments do not fully remediate the initial disadvantage. In other words, the health difference between individuals diverges *despite* optimal countermeasures. Without the additional health investments of the shocked individuals, divergence would be greater.

A remaining question is whether individuals are able to observe slight deteriorations of health (as assumed by the health deficit model). Alternatively, these damages may go unnoticed in young and middle adult age and are only diagnosed in old age when the associated health deficits become sufficiently visible. Given such a visibility threshold, health deficit become earlier diagnosed in old age if they were initially larger. These perceptibility issues may explain why aging related health problems are diagnosed in old age although they are de facto always present and originate from sub-optimal development in young age and perhaps in utero. While health deficits are slowly accumulating (from e.g. mild hypertension to difficulties running or lifting weight to more severe cardiovascular problems), the empirical association between early-life health shocks and late life-life health outcomes is in many studies only observed when health

problems became sufficiently severe. In this sense, the health deficit model, fills the “invisible” gap from early-life health shocks to health outcomes in old age.

Finally, notice that the health deficit model predicts that individuals who achieved optimal development in childhood eventually experience the same health deficit index as the individuals with unfavorable health- or nutrition-shocks in early life. This is a manifestation of the generality of human aging understood as “intrinsic, cumulative, progressive, and deleterious loss of function that eventually culminates in death” (Arking, 2006). The shocked individuals, however, experience these health deficits earlier.

5. CONCLUSION

An influential strand of literature within health economics has over the last decade provided convincing evidence in favor of the fetal origins hypothesis: *in utero* shocks have the ability to influence late-in-life outcomes. Relevant outcomes involve both health issues as well as a range of socio-economic outcomes. In this study we have argued that the current workhorse model of health economics, the Grossman (1972a) model, is incapable of accounting for such effects. Indeed, since the notion of health – health capital– is analogous to physical capital, the model posits that health status depreciates more when the health status of individuals is high and less when the health status is low. These features imply, as demonstrated above, that the Grossman model generates the prediction that individuals with different initial conditions, prompted by *in utero* shocks, converge in health status during life, holding investments fixed. Convergence in health status in the aftermath of early-in-life shocks occurs both in a relative and in an absolute sense. This prediction is strengthened if one allows the health depreciation rate to grow over time, as it is required for the Grossman model to be reconcilable with the fact of mortality. It is possible to generate important persistence in health outcomes through investments; for example, by introducing the human capability theory of Heckman (2007). But since the fetal origins hypothesis asserts an impact from *in utero* influences *conditional* on investments, the Grossman model remains irreconcilable with the hypothesis.

In contrast, this paper demonstrates that the health deficit model offers radically different predictions. At its core the model conceptualizes aging as a continual process of loss of function – increasing frailty – that culminates in death. The notion of frailty is captured by way of the deficit index: as humans age (health declines) the relative fraction of potential age-related health

conditions climbs steadily upward. This underlying process, which can be slowed down by health investments, is exponential in nature. By implication, small differences in initial conditions at young age are amplified during life. The exponential nature of increasing deficits during life has been confirmed repeatedly by empirical work within gerontology. Overall, the deficit model seems well positioned to account for the type of dynamics implied by the fetal origins hypothesis.

Finally we have linked our biologically-founded model of aging and health deficit accumulation (originating from a loss in redundancy in organ reserve) with a biologically-founded model of ontogenetic growth (conceptualized as the build-up of redundancy in organ reserve). We have shown how the ontogenetic growth model is derived from the energy needs for cell creation and maintenance that fulfil a simple law of energy conservation and Kleiber's scaling law for human metabolism. The model of ontogenetic growth naturally divides childhood in two distinct periods, an early period of shock amplification and a later period of shock dampening growth. We have shown how in utero shocks are transmitted during childhood towards the onset of health deficit accumulation in adulthood and how they affect aging and health in old age. Taken together, we have thus proposed a new model of human development from conception to death, which motivates the fetal origins hypothesis of late-life health deficits.

REFERENCES

- Abeliansky, A. and Strulik, H. (2018a). How we fall apart: Similarities of human aging in 10 European countries. *Demography* 55, 341-359.
- Abeliansky, A. and Strulik, H. (2018b). Hungry children age faster. *Economics and Human Biology* 29, 211-22.
- Abeliansky, A. and Strulik, H. (2019). Season of birth, health, and aging. *Economics and Human Biology*, forthcoming.
- Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *Journal of Political Economy* , 114(4), 672-712.
- Almond, D., and Currie, J. (2011a). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives* 25(3), 153-72.
- Almond, D., and Currie, J. D. (2011b). Human capital development before age five. In: *Handbook of Labor Economics*, Vol. 4, pp. 1315-1486. Elsevier, Amsterdam.
- Almond, D., and Mazumder, B. (2011). Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy. *American Economic Journal: Applied Economics*, 3(4), 56-85.
- Almond, D., and Mazumder, B. (2013). Fetal origins and parental responses. *Annual Review of Economics* 5(1), 37-56.
- Almond, D., Edlund, L. and Palme, M. (2009). Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *Quarterly Journal of Economics*, 124(4): 1729-72.
- Almond, D., Currie, J. and Duque, V. (2018). *Journal of Economic Literature* 56(4), 1360-1446.
- Arking, R. (2006). *The Biology of Aging: Observations and Principles*, Oxford University Press, Oxford.
- Barker, D.J. (1995). Fetal origins of coronary heart disease. *British Medical Journal* 311(6998), 171-174.
- Barro, R., G. Mankiw and X. Sala-i-Martin (1995). Capital mobility in neoclassical models of growth. *American Economic Review* 85, 103-115
- Bhalotra, S., and Rawlings, S.B. (2011). Intergenerational persistence in health in developing countries: The penalty of gender inequality?. *Journal of Public Economics* 95(3), 286-299.
- Bhalotra, S., and Venkataramani, A. (2016). Shadows of the Captain of the Men of Death: Early Life Health, Human Capital Investment and Institutions. *Working paper*.
- Brody, S., 1945, *Bioenergetics and Growth*, Van Nostrand-Reinhold, New York.
- Browning, M. and Ejrnæs, M. (2009). Consumption and children. *Review of Economics and Statistics* 91, 93-111.

- CDC (2000). Growth Charts: United States, Boys 2 – 20 years. National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion. <http://www.cdc.gov/growthchar>
- Cronqvist, H., Previtro, A., Siegel, S., & White, R. E. (2016). The fetal origins hypothesis in finance: Prenatal environment, the gender gap, and investor behavior. *Review of Financial Studies* 29(3), 739-786.
- Currie, J. (2011). Inequality at Birth: Some Causes and Consequences. *American Economic Review* 101(3), 1-22.
- Dalgaard, C.-J. and Strulik, H. (2014). Optimal aging and death: Understanding the Preston Curve, *Journal of the European Economic Association* 12, 672-701.
- Dalgaard, C.J., and Strulik, H. (2015). The physiological foundations of the wealth of nations. *Journal of Economic Growth* 20(1), 37-73.
- Dalgaard, C.J., and Strulik, H. (2016). Physiology and development: Why the West is taller than the rest. *Economic Journal* 126, 2292–2323.
- Dalgaard, C.-J. and Strulik, H. (2017). The genesis of the golden age: Accounting for the rise in health and leisure. *Review of Economic Dynamics*, 24, 132-151.
- Doblhammer, G., and Vaupel, J. W. (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences* 98(5), 2934-2939.
- Dolinoy, D.C., Das, R., Weidman, J.R., and Jirtle, R.L. (2007). Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatric Research* 61, 30R-37R.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3), 130-135.
- Gavrilov, L.A. and Gavrilova, N.S. (1991), *The Biology of Human Life Span: A Quantitative Approach*. Harwood Academic Publishers, London.
- Gluckman, P.D. and Hanson, M.A. (2004). Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. *Pediatric Research*, 56(3), 311-317.
- Grossman, M. (1972a). On the concept of health capital and the demand for health. *Journal of Political Economy* 80(2), 223-255.
- Grossman, M. (1972b). *The Demand for Health: A Theoretical and Empirical Investigation*. New York: Columbia University Press for the National Bureau of Economic Research.
- Harttgen, K., Kowal, P., Strulik, H., Chatterji, S., and Vollmer, S. (2013). Patterns of frailty in older adults: comparing results from higher and lower income countries using the Survey of Health, Ageing and Retirement in Europe (SHARE) and the Study on Global AGEing and Adult Health (SAGE). *PLOS One* 8(10), e75847.
- Hilakivi-Clarke, L., & De Assis, S. (2006). Fetal origins of breast cancer. *Trends in Endocrinology & Metabolism*, 17(9), 340-348.

- HCCI (2012). 2007-2010 Children’s Health Care Spending Report, Health Care Cost Institute (<http://www.healthcostinstitute.org/childrensreport>, accessed January, 7th, 2016.)
- Heckman, J.J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences* 104(33), 13250-13255.
- Karkach, A.S. (2006). Trajectories and models of individual growth. *Demographic Research* 15, 347-400.
- Keehan, S. P, H. Lazenby, M. Zezza, and A.C. Catlin (2004). Age Estimates in the National Health Accounts. *Health Care Financing Review* 1, 1–16.
- Kiserud, T., Piaggio, G., Carroli, G., Widmer, M., Carvalho, J., Jensen, L. N., ... and Benachi, A. (2017). The World Health Organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS medicine* 14(1), e1002220.
- Kleiber, M., 1932, Body size and metabolism, *Hilgardia* 6, 315-353.
- Kooijman, S.A.L.M. (2000). *Dynamic energy and mass budgets in biological systems*. Cambridge University Press.
- Lin, M.J. and Liu, E.M. (2014). Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *Journal of Health Economics* 37: 152-163.
- McMullen, S., and Mostyn, A. (2009). Animal models for the study of the developmental origins of health and disease. *Proceedings of the Nutrition Society*, 68(03), 306-320.
- 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.B, Song, X., and Rockwood, K. (2013). Assessing biological aging: the origin of deficit accumulation. *Biogerontology* 14(6), 709-717.
- Mitnitski, A.B, and K. Rockwood. (2016). The rate of aging: the rate of deficit accumulation does not change over the adult life span. *Biogerontology*, 17(1), 199–204.
- Murphy, K.M., and Topel, R.H. (2006). The value of health and longevity. *Journal of Political Economy* 114(5), 871-904.
- Nelson, R.E. (2010). Testing the fetal origins hypothesis in a developing country: evidence from the 1918 influenza pandemic. *Health Economics*, 19(10), 1181-1192.
- Oreopoulos, P., Stabile, M., Walld, R. and Roos, L. (2008). “Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins.” *Journal of Human Resources*, 43(1): 88–138.

- 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, 722-727.
- Rockwood, K., Mitnitski, A., and Howlett, S. E. (2015). Frailty: Scaling from Cellular Deficit Accumulation?, in: Theou, O. and Rockwood, K. (eds.) *Frailty in Aging.*, Basel: Karger, 1-14.
- Shi, J., Song, X., Yu, P., Tang, Z., Mitnitski, A., Fang, X., and Rockwood, K. Analysis of frailty and survival from late middle age in the Beijing Longitudinal Study of Aging, *BMC Geriatrics* 11(1), 17.
- Sinclair, K.D., Lea, R.G., Rees, W.D., and Young, L.E. (2007). The developmental origins of health and disease: current theories and epigenetic mechanisms. *Society of Reproduction and Fertility supplement*, 64, 425.
- Schüenemann, J., Strulik, H., and Trimborn, T. (2017a). Going from bad to worse: Adaptation to poor health health spending, longevity, and the value of life. *Journal of Economic Behavior & Organization* 140, 130-146.
- Schüenemann, J., Strulik, H., and Trimborn, T. (2017b). The gender gap in mortality: How much is explained by behavior? *Journal of Health Economics* 54, 79-90.
- Schüenemann, J., Strulik, H., and Trimborn, T. (2019). The marriage gap: Optimal aging and death in partnerships. *Review of Economic Dynamics*, forthcoming.
- Strulik, H. (2015a). Frailty, mortality, and the demand for medical care, *Journal of the Economics of Ageing* 6, 5-12 .
- Strulik, H. (2015b). The Health Hump. *Journal of Demographic Economics*, 2017, Vol. 83, 245-258.
- Strulik, H. (2018). The return to education in terms of wealth and health, *Journal of the Economics of Ageing* 12, 1-14.
- Thompson, R.F., and Einstein, F.H. (2010). Epigenetic basis for fetal origins of age-related disease. *Journal of Women's Health* 19(3), 581-587.
- Van Den Berg, G.J., Lindeboom, M. and Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review* 96(1), 290-302.
- von Bertalanffy, L. (1957). Quantitative laws in metabolism and growth. *Quarterly Review of Biology* 32(3), 217-231.
- Waterland, R.A., and Michels, K.B. (2007). Epigenetic epidemiology of the developmental origins hypothesis. *Annual Review of Nutrition* 27, 363-388.
- West G.B., J. H. Brown and B.J. Enquist, 1997, A general model of the origin of allometric scaling laws in biology, *Science* 276, 122-26.
- West, G.B., Brown, J.H., and Enquist, B. J. (2001). A general model for ontogenetic growth. *Nature* 413(6856), 628-31.

Wu, G., Bazer, F.W., Cudd, T.A., Meininger, C.J., and Spencer, T.E. (2004). Maternal nutrition and fetal development. *Journal of Nutrition* 134(9), 2169-2172.

Zweifel, P., Felder, S. and Meiers, M. (1999). Ageing of population and health care expenditure: a red herring? *Health Economics* 8, 485-496.

APPENDIX A: SOLUTION OF THE HEALTH DEFICIT MODEL

The maximization problem given in equations (3)–(5) together with the initial conditions $k(\tau) = 0$, $D(\tau) = D_\tau$ and the terminal conditions $k(T) = \bar{k}$ and $D(T) = \bar{D}$ constitutes a free terminal time problem of optimal control. The unknown terminal time is the age at death. The Hamiltonian associated with this problem reads

$$H = \frac{c^{1-\sigma} - 1}{1 - \sigma} + \lambda\mu(D - a - Ah^\gamma) + \phi(rk + w - c - ph).$$

For $\sigma = 1$ the first term is replaced by $\log(c)$. The first order conditions wrt. c and h and the co-state equations are

$$c^{-\sigma} - \phi = 0 \quad \Rightarrow \quad c^{-\sigma} = \phi \quad \Rightarrow \quad \sigma\dot{c}/c = -\dot{\phi}/\phi \quad (\text{A.1})$$

$$-\lambda\mu A\gamma h^{\gamma-1} - p\phi = 0 \quad (\text{A.2})$$

$$\lambda\mu = \lambda\rho - \dot{\lambda} \quad \Rightarrow \quad \mu - \rho = -\dot{\lambda}/\lambda \quad (\text{A.3})$$

$$\phi r = \phi\rho - \dot{\phi} \quad \Rightarrow \quad r - \rho = -\dot{\phi}/\phi. \quad (\text{A.4})$$

Equation (A.4) is the well known Euler equation requiring that the shadow price of consumption (ϕ) grows at the rate of the interest rate less the time preference rate. Analogously, the Euler equation (A.3) requires that the shadow price of health grows at the rate of health deterioration (μ) less the time preference rate.

Log-differentiate (A.2) wrt. time and insert (A.3) and (A.4) to obtain optimal growth of health expenditure:

$$\frac{\dot{\lambda}}{\lambda} - \frac{\dot{\phi}}{\phi} = (1 - \gamma)\frac{\dot{h}}{h} \quad \Rightarrow \quad -\mu + \rho + r - \rho = (1 - \gamma)\frac{\dot{h}}{h}.$$

Solving for the growth rate of health expenditure we obtain the ‘‘Health Euler equation’’

$$g_h \equiv \frac{\dot{h}}{h} = \frac{r - \mu}{1 - \gamma}. \quad (\text{A.5})$$

By inserting (A.3) into (A.1) we obtain the Ramsey rule

$$g_c \equiv \frac{\dot{c}}{c} = \frac{r - \rho}{\sigma}. \quad (\text{A.6})$$

The dynamic model can be solved without numerical integration since the differential equations can be solved analytically. We set (wolog) $\tau = 0$ and begin with noting that, because g_h is optimally constant according to (A.5), the differential equation (3) can be rewritten as $\dot{D} = \mu(D - a - Ah(0)^\gamma \exp(\gamma g_h t))$. Given $D(0) = D_0$ the solution at time T is:

$$D(T) = D_0 \exp(\mu T) - \mu Ah(0)^\gamma \exp(\mu T) \int_0^T \exp(\gamma g_h - \mu) dt + \mu a \exp(\mu T) \int_0^T \exp(-\mu t) dt.$$

At the time of expiry the boundary condition requires $D(T) = \bar{D}$. Solving the integrals in the above equation we get

$$\bar{D} = D_0 \exp(\mu T) - \frac{\mu Ah(0)^\gamma \exp(\mu T)}{g_D} [\exp(g_D T) - 1] - a [\exp(\mu T) - 1], \quad (\text{A.7})$$

where $g_D \equiv (\gamma r - \mu)/(1 - \gamma)$. Next, we integrate (3) and insert $k(0) = k_0$ and $k(T) = \bar{k}$ to obtain

$$\begin{aligned} \bar{k} &= k_0 \exp(rT) + w \exp(rT) \int_0^T \exp(-rt) dt \\ &\quad - c(0) \exp(rT) \int_0^T \exp[(g_c - r)t] dt - ph(0) \exp(rT) \int_0^T \exp[(g_h - r)t] dt. \end{aligned}$$

Divide by $\exp(rT)$. Note that $g_h - r = (\gamma r - \mu)/(1 - \gamma) \equiv g_D$ and solve the integrals to obtain

$$\bar{k} \exp(-rT) = k_0 - \frac{w}{r} [\exp((r)T) - 1] - \frac{c(0)}{g_c - r} [\exp((g_c - r)T) - 1] - \frac{ph(0)}{g_D} [\exp(g_D T) - 1]. \quad (\text{A.8})$$

The condition for optimal T requires that the Hamiltonian assumes the value of zero at T . Otherwise, it would have been optimal to live longer or die earlier. Also, at expiry $D(T) = \bar{D}$ and $k(T) = \bar{k}$. Thus the Hamiltonian at age T reads

$$0 = H(T) = u(c(T)) + \lambda(T)\mu [\bar{D} - a - Ah(T)^\gamma] + \phi(T) [r\bar{k} + w - c(T) - ph(T)].$$

Insert $\lambda(T)$ and $\phi(T)$ from (A.1) and (A.2) to get

$$0 = u(c(T)) - \frac{p}{c(T)^\sigma} \left[\frac{(\bar{D} - a)h(T)^{1-\gamma}}{\gamma A} - \frac{h(T)}{\gamma} - \frac{w + r\bar{k}}{p} + c(T) + h(T) \right]$$

where $u_T \equiv \log c(T)$ in the case of log-utility and $u_T \equiv [c(T) - 1]^{1-\sigma}/(1 - \sigma)$ otherwise. Noting that $c(T) = c(0) \exp(g_c T)$ and $h(T) = h(0) \exp(g_h T)$ this provides

$$\begin{aligned} 0 &= u_T - \frac{\exp(-\sigma g_c T)}{c(0)^\sigma} \times \\ &\quad \left\{ \frac{(\bar{D} - a)}{\gamma A} ph(0)^{1-\gamma} \exp((1 - \gamma)g_h T) - \frac{1 - \gamma}{\gamma} ph(0) \exp(g_h T) - w - r\bar{k} + c(0) \exp(g_c T) \right\}, \end{aligned} \quad (\text{A.9})$$

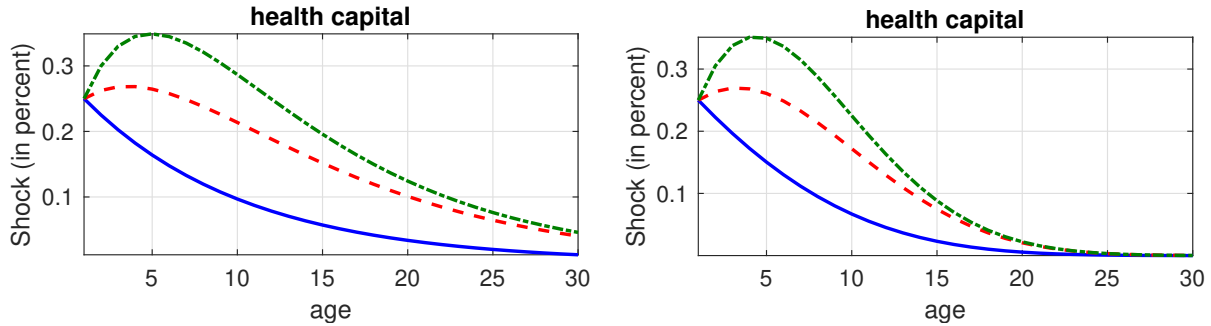
with $u_T \equiv \log(c(0)) + g_c T + b$ in the case of log-utility and $u_T \equiv [c(0) \exp(g_c T) - 1]^{1-\sigma}/(1 - \sigma) + b$ otherwise. The three equations (A.7) – (A.9) can be solved for the three unknowns: $c(0)$, $h(0)$, and T . Having found the optimal initial values and the optimal terminal time, the dynamic system consisting of (3), (5), (A.5), and (A.6) is fully specified and it can be solved for the optimal life-cycle trajectories of c, h, k and D .

APPENDIX B: HEALTH SHOCKS INCREASE HEALTH CAPITAL DEPRECIATION

Here, we consider an interesting refinement of the health capital model and show that it does not resolve the problem of convergence of initial health differences in old age. Specifically, we implement the assumption that the rate of depreciation of the stock of health capital is a positive function of the initial health shock. To see the implications by way of example, let us reconsider the 25% health shock from the computational experiment in the main text (and from Almond and Currie, 2011). The blue line in the panel on the left-hand side in Figure A.1. reiterates the result from Figure 1 in the main text. It shows the difference of health capital between an unshocked individual and an individual who experienced 25% health reduction at time 0. The initial health difference is depreciated away as the individuals grow older. Suppose now that health depreciates faster for individuals who experienced a negative initial health shock. Red

(dashed) lines show the case where the health shock implies an increase of the depreciation rate to 15%. Green (dash-dotted) lines repeat the experiment when the shock increases the depreciation rate to 20%. We see that this leads to an increase in the health difference between individuals in young adulthood, implying divergence and, temporarily, shock amplification. As individuals grow older, however, convergence sets in and health differences are depreciated away. This variant of the model thus predicts that early-life health shocks lead to large health impediments at young age but only to insignificant health impediments in old age. It fails to motivate fetal origins of late-life health deficits.

FIGURE A.1: HEALTH CAPITAL DEPRECIATION INCREASES WITH EXPERIENCE OF IN-UTERO SHOCK



The figure shows how a 25 percent negative shock at time 0 develops given alternative annual depreciation rates. Left: constant rate of depreciation; blue (solid) line: 10 percent depreciation; red (dashed) line: initial shock associated with increase of depreciation rate to 15%; green (dash-dotted) line: initial shock associated with increase of depreciation rate to 20%. Right: Same exercise with the additional assumption that health depreciation increases at a rate of 5 percent per year as individuals age.

The result can also be derived more formally, following the argument made in the main text (subsequent to equation (1)). The difference in health capital between two individuals (1 and 2) is given by

$$H_t^1 - H_t^2 = (1 - \delta^1)^t H_0^1 - (1 - \delta^2)^t H_0^2.$$

Let the individual who experienced the shock be indexed by 2, $H_0^2 < H_0^1$ and assume that the shock is associated with a higher rate of health capital depreciation, $\delta^1 < \delta^2$. Then, the difference $H_t^1 - H_t^2$ becomes larger for small positive t (compared to the case of shock-independent δ). For large t , however, the difference $H_t^1 - H_t^2$ converges to zero, irrespective of the difference in the δ 's. The reason is that the sufficient condition for convergence to zero is $0 < \delta < 1$, which is always fulfilled in the Grossman model, irrespective of any additional assumptions on the size of δ . It is an inherent feature of the Grossman model. This can be illustrated further by assuming that δ is increasing with age, an assumption frequently made in applications of the Grossman model.

In the panel on the left hand side of Figure A.1, we assume that the depreciation rate increases exponentially in age, $\delta^i(t) = \delta^i(0) \cdot e^{\nu t}$, $\nu > 1$ and $i = 1, 2$. We set $\nu = 0.05$ and assume that $\delta^1(0) = 0.1$ for the unshocked individual while for the individual who experienced the in-utero shock $\delta^2(0) = 0.15$ (red dashed lines) and $\delta^2(0) = 0.2$ (green dash-dotted lines). We see that, as explained in the main text, the positive age-dependence leads to faster convergence of initial health differences.

APPENDIX C: PROOF THAT SHOCKS ARE AMPLIFIED IFF $t < t_I$

In order to find the threshold for shock amplification, we solve (11) for $\frac{\partial^2 m(t)}{\partial m(\tau) \partial t} = 0$, i.e. we solve

$$\begin{aligned} & \left[\frac{a}{b} - m(\tau)^{1-\beta} \right] \frac{\beta}{1-\beta} e^{-(1-\beta)b(t-\tau)} - a/b + \left[a/b - m(\tau)^{1-\beta} \right] e^{-(1-\beta)b(t-\tau)} = 0 \\ \Rightarrow & \left[\frac{a}{b} - m(\tau)^{1-\beta} \right] \left(\frac{\beta}{1-\beta} + 1 \right) e^{-(1-\beta)b(t-\tau)} = \frac{a}{b} \\ \Rightarrow & \left[\frac{\frac{a}{b} - m(\tau)^{1-\beta}}{a/b} \right] \frac{1}{1-\beta} = e^{(1-\beta)b(t-\tau)}. \end{aligned}$$

Taking logs and solving for t provides t_I from (10) in the main text.