

WORKING P A P E R

Grossman's Missing Health Threshold

TITUS GALAMA, ARIE KAPTEYN

WR-684

May 2009

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.



RAND LABOR AND POPULATION

GROSSMAN'S MISSING HEALTH THRESHOLD

Titus J. Galama*

Arie Kapteyn*

May 5, 2009

Abstract

We present a generalized solution to Grossman's model of health capital (1972), relaxing the widely used assumption that individuals can adjust their health stock instantaneously to an "optimal" level without adjustment costs. The Grossman model then predicts the existence of a health threshold above which individuals do not demand medical care. Our generalized solution addresses a significant criticism: the model's prediction that health and medical care are positively related is consistently rejected by the data. We suggest structural and reduced form equations to test our generalized solution and contrast the predictions of the model with the empirical literature.

Keywords: health, demand for health, health capital, medical care, labor

JEL Codes : I10, I12, J00, J24

*RAND Corporation; This research was supported by the National Institute on Aging, under grants R01AG030824, P30AG012815 and P01AG022481. We are grateful to Arthur van Soest, Eddy van Doorslaer, Hans van Kippersluis, Erik Meijer, Raquel Fonseca, Pierre-Carl Michaud and James Hosek for useful discussions and comments.

1 Introduction

Grossman's model of health capital (1972a, 1972b, 2000) is considered a breakthrough in the economics of the derived demand for medical care. In Grossman's human capital framework individuals demand medical care (e.g., invest time and consume medical goods and services) for the consumption benefits (health provides utility) as well as production benefits (healthy individuals have greater earnings) that good health provides. The model has been employed widely to explore a variety of phenomena related to health, medical care, inequality in health, the relationship between health and socioeconomic status, occupational choice, etc (e.g., Muurinen and Le Grand, 1985; Case and Deaton 2005; Cropper 1977).

Yet the Grossman model has also received significant criticism. For example, the model has been criticized for its simplistic deterministic nature (e.g., Cropper 1977, Dardanoni and Wagstaff 1987), for not determining length of life (e.g., Ehrlich and Chuma, 1990), for allowing complete health repair (Case and Deaton 2005), and for its formulation in which medical investment in health has constant returns which is argued to lead to an unrealistic "bang-bang" solution (e.g., Ehrlich and Chuma, 1990). The criticism has led to theoretical and empirical extensions of the model (often by the same authors who provided the criticism), which to a large extent address the issues identified.¹ For an extensive review see Grossman (2000) and the work referenced therein.

However, there is one most significant criticism that thus far has not satisfactorily been addressed. Zweifel and Breyer (1997; p. 62) reject the Grossman model's central proposition that the demand for medical care is derived from the demand for good health: "*... the notion that expenditure on medical care constitutes a demand derived from an underlying demand for health cannot be upheld because health status and demand for medical care are negatively rather than positively related ...*" In a review of the empirical literature Zweifel and Breyer conclude that the model's prediction that health and medical care should be positively related (healthy individuals consume more medical goods and services) is consistently rejected by the data. For example, Cochrane et al. (1978) find in a study of various determinants of mortality across various countries that indicators of medical care usage are positively related to mortality. And more specifically, Wagstaff (1986) and Leu and Gerfin (1992), in estimating structural and reduced form equations of the Grossman model, find that measures of medical care are negatively correlated with measures of health and that the relationships are highly significant.²

It is of importance that this criticism be addressed. Dismissal of the central proposition of the Grossman model essentially amounts to rejecting the model itself. And a model of health and medical care should at a minimum predict the correct sign of the relationship between the two.

Several authors have sought to explain the consistently negative relation between health and

¹With the exception perhaps of the "bang-bang" solution and for allowing complete health repair, which we will discuss briefly in this work.

²Numerous other studies do not specifically test Grossman's structural and reduced form equations, but broadly test similar relations between measures of health and measures for the demand for medical goods and services, controlling for relevant demographic and other characteristics. These studies find similar results. See section 4 for a discussion.

medical care in empirical studies. For example, Grossman argues that the observed negative relation could be attributed to biases that arise if the conditional demand function is estimated with health treated as exogenous (Grossman 2000; p. 386).

Muurinen and Le Grand (1985), in attempting to explain the positive relation between mortality and medical care usage found by Cochrane et al. (1978), suggest that the negative relation between indicators of health and of medical care (apart from suggesting that medical care is actually harmful) could be explained by differences in socioeconomic status. Individuals with fewer resources derive relatively higher production benefits from their health stock. They thus would have relatively greater usage of the stock (i.e., higher rates of health deterioration) which would require higher medical care to compensate for health losses. But if health cannot be completely repaired due to the increased use-intensity they would have inferior health states. High mortality would then be positively correlated with use of health services.

Wagstaff (1986) provides a detailed discussion of potential reasons why estimates of the Grossman model may lead to a negative relation between measures of medical care usage and measures of health. On the one hand, one might argue that the coefficients determined in Wagstaff (1986) and similar analyses are not reliable estimates of the model's parameters. For example, Wagstaff suggests that in moving from the theoretical to the empirical model inappropriate assumptions may have been introduced (see Wagstaff 1986 for details). Or the identification of medical care with market inputs may insufficiently characterize health inputs if non-medical inputs are important in the production of health. On the other hand, one may take the estimates at face value and seek explanations in terms of the underlying model. Interestingly, Wagstaff (1986) suggests that, contrary to what is assumed in Grossman's theoretical work, the negative relationship may reflect a non-instantaneous adjustment of health capital to its "optimal" value.³ This, Wagstaff argues, may be the result of a constraint on medical care or be due to the existence of adjustment costs. Wagstaff finds in subsequent analysis (Wagstaff 1993) that a reformulation of Grossman's empirical model with non-instantaneous adjustment is not only more consistent with Grossman's theoretical model but also with the data.

Indeed, in earlier theoretical work building on a simplified version of the Grossman model (Galama et al. 2009) we concluded that the widely employed assumption in the Grossman literature that any health "excess" or "deficit" can be adjusted instantaneously and at no adjustment cost may be too restrictive. Any "excess" in health capital cannot rapidly dissipate as individuals with "excessive" health can at best decide not to consume medical care.⁴ As a consequence their health deteriorates at the natural deterioration rate $d(t)$ (i.e., non instantaneous) until health reaches Grossman's "optimal" level. Thus an individual's health is not always at the predicted "optimal"

³Throughout this paper we will refer to Grossman's solution for the optimal health level as "optimal" health (using quotation marks) to reflect the fact that the Grossman solution is not always the optimal solution. Grossman's solution is optimal only in the absence of corner solutions. In this work we explore corner solutions in which individuals do not consume medical care for periods of time. The Grossman solution is then strictly speaking not the optimal solution.

⁴In other words medical care is restricted to be non-negative and the situation where individuals do consume medical care represents a corner solution.

level. While the widely employed assumption that an individual's health follows Grossman's solution for the "optimal" path allows one to derive simple model predictions for empirical validation (and indeed this may be the primary reason for its use), it is otherwise unnecessary and is not demanded by theory. Importantly, Wagstaff's (1993) work suggests that individuals do not adjust their health stocks instantaneously. In other words, not only is there no theoretical basis for the assumption, empirical evidence suggests the assumption is not valid.

In this paper we relax the widely used assumption that individuals can adjust their health stock to Grossman's "optimal" level instantaneously. We do not restrict an individual's health path to Grossman's "optimal" solution but allow for corner solutions where the optimal response for healthy individuals is to not consume medical goods and services for some period of time. We then find that the Grossman model predicts a substantially different pattern of medical care over the life-time than previously was assumed. Healthy individuals initially do not demand medical care till their health has deteriorated to a certain threshold level given by Grossman's "optimal" health. Subsequently their health evolves as the Grossman solution for the "optimal" path as individuals begin to demand medical care. In other words, Grossman's "optimal" health level is in fact a "health threshold" rather than an "optimal" trajectory. This simple pattern potentially addresses the most damning criticism: we find that the Grossman model predicts that healthy individuals (those above the threshold) do not consume medical care, but the unhealthy (at the threshold) do. Grossman's model thus predicts that healthy individuals demand less medical care, not the opposite, in agreement with the empirical literature.

Our working hypothesis is that a significant share of the population is healthy for much of their life. In our definition the healthy do not demand medical care. This would help explain the observed negative relation between measures of health and measures of medical care. Further, as we will see, this hypothesis can explain a number of other empirical facts.

A consequence of the assumption that a significant share of the population is healthy for much of their life, combined with the threshold nature of the demand for medical care, is that health investment in the Grossman model is to be strictly interpreted as medical care. It is the type of health investment (own time inputs and purchases of goods and services in the market) that individuals engage in when they are unhealthy and seek to "repair" their health. The Grossman literature sometimes views health investment as including a wide range of other types of investments, such as: preventive care (e.g., medical check ups), healthy dieting, and sports / exercise. Strictly speaking, the Grossman model does not contain the concept of healthy or unhealthy consumption nor of preventive care. In contrast to medical care, individuals engage in such activities when they are healthy as well as when they are unhealthy. In other words, these types of health investment are not part of the current formulation of the Grossman model where health investments take place only when individuals are unhealthy. The Grossman model, however, does offer an alternative way to include such health investments, by slowing the deterioration rate. For example, Case and Deaton (2005) model the effect of healthy consumption (e.g., healthy dieting, sports / exercise) as slowing and unhealthy consumption (e.g., smoking, excessive alcohol consumption) as accelerating the rate of deterioration. Preventive care may

operate in a similar manner. Here we consider these extensions as beyond the scope of the current paper.

As mentioned before, we are motivated by the lack of a theoretical justification in the Grossman literature for employing the assumption that health is always at Grossman's "optimal" level (see Galama et al. 2009) and by Wagstaff's (1993) empirical analysis that suggests the assumption is not valid. A further motivation comes from the observation that the above attempts to explain the observed negative relationship between measures of health and measures of medical care do not pass the principle of Occam's razor when compared to the simple explanation put forward here that individuals cannot adjust their health stocks instantaneously (Wagstaff 1986, 1993; Galama et al. 2009). Our proposed explanation is the simplest in that we adopt the Grossman model as is and make one fewer assumption than is commonly made in the Grossman literature.

The aim of this paper is to investigate the solutions and predictions of the Grossman model without restricting the solutions to Grossman's so-called "optimal" solution by allowing for corner solutions. We proceed as follows. In section 2, we reformulate the Grossman model in continuous time allowing for corner solutions, solve the optimal control problem and derive first-order conditions for consumption and health. In section 3 we present structural form and reduced form solutions for health, medical care and consumption to enable empirical testing of our reformulation of the Grossman model. In section 4 we contrast the predictions of our generalized solution of the Grossman model with the traditional solution and with the empirical literature. We conclude in section 5 and provide detailed derivations in the Appendix.

2 General framework: the full Grossman model

We present the original human-capital model of the derived demand for health by Grossman (Grossman 1972a, 1972b, 2000) in continuous time (see also Wagstaff, 1986; Wolfe, 1985; Zweifel and Breyer, 1997; Ehrlich and Chuma, 1990). Health is treated as a form of human capital (health capital) and individuals derive both consumption (health provides utility) and production benefits (health increases earnings) from it. The demand for medical care is a derived demand: individuals demand "good health", not the consumption of medical care. In the original formulation of the Grossman model (Grossman 1972a, 1972b, 2000) health yields an output of healthy time and consumption and medical care constitute both own-time inputs and goods or services purchased in the market. Simplified versions of the Grossman model have been presented by Case and Deaton (2005) who assume consumption and production benefits are functions of health rather than healthy time, Wolfe (1985) who assumes health does not provide utility, and Case and Deaton (2005) and Wagstaff (1986) who do not include time inputs into the production of consumption nor in the production of medical care. For an excellent review of the basic concepts of the Grossman model see Muurinen and Le Grand (1985).

Individuals maximize the life-time utility function

$$\int_0^T U\{C(t), s[H(t)]\}e^{-\beta t} dt, \quad (1)$$

where T denotes total life time, β is a subjective discount factor and individuals derive utility $U\{C(t), s[H(t)]\}$ from consumption $C(t)$ and from reduced sick time $s[H(t)]$. Sick time is assumed to be a function of health $H(t)$. Time t is measured from the time individuals begin employment. Utility decreases with sick time $\partial U(t)/\partial s(t) \leq 0$ and increases with consumption $\partial U(t)/\partial C(t) \geq 0$. Sick time decreases with health $\partial s(t)/\partial H(t) \leq 0$. Further we assume diminishing marginal benefits: $\partial^2 U(t)/\partial^2 s(t) \geq 0$ and $\partial^2 U(t)/\partial^2 C(t) \leq 0$.

The objective function (1) is maximized subject to the following constraints:

$$\dot{H}(t) = I(t) - d(t)H(t), \quad (2)$$

$$\dot{A}(t) = \delta A(t) + Y\{s[H(t)]\} - p_X(t)X(t) - p_m(t)m(t), \quad (3)$$

and we have initial and end conditions: $H(0)$, $A(0)$ and $A(T)$ are given.

$\dot{H}(t)$ and $\dot{A}(t)$ in equations (2) and (3) denote time derivatives of health $H(t)$ and assets $A(t)$. Health (equation 2) can be improved through medical health investment $I(t)$ (medical care) and deteriorates at the “natural” health deterioration rate $d(t)$. Using equation (2) we can write $H(t)$ as a function of medical care $I(t)$ and initial health $H(0)$,

$$H(t) = H(0)e^{-\int_0^t d(s)ds} + \int_0^t I(x)e^{-\int_x^t d(s)ds} dx. \quad (4)$$

Assets $A(t)$ (equation 3) provide a return δ (the interest rate), increase with income $Y\{s[H(t)]\}$ and decrease with purchases in the market of goods $X(t)$ and medical goods and services $m(t)$ at prices $p_X(t)$ and $p_m(t)$, respectively. Income $Y\{s[H(t)]\}$ is assumed to be a decreasing function of sick time $s[H(t)]$.

Integrating equation (3) over the life time we obtain the life-time budget constraint

$$\int_0^T p_X(t)X(t)e^{-\delta t} dt + \int_0^T p_m(t)m(t)e^{-\delta t} dt = A(0) - A(T)e^{-\delta T} + \int_0^T Y\{s[H(t)]\}e^{-\delta t} dt. \quad (5)$$

The left-hand side of (5) represents life-time consumption of market goods and life-time consumption of medical goods and services, and the right-hand side represents life-time financial resources in terms of life-time assets and life-time earnings.

Goods $X(t)$ purchased in the market and own time inputs $\tau_C(t)$ are used in the production of consumption $C(t)$. Similarly medical goods and services $m(t)$ and own time inputs $\tau_I(t)$ are used

in the production of medical care $I(t)$. The efficiencies of production are assumed to be a function of the consumer's stock of knowledge E (an individual's human capital exclusive of health capital [e.g., education]) as it is generally believed that the more educated are more efficient consumers of medical care (see, e.g., Grossman 2000),

$$I(t) = I[m(t), \tau_I(t); E], \quad (6)$$

$$C(t) = C[X(t), \tau_C(t); E]. \quad (7)$$

The total time available in any period $\Omega(t)$ is the sum of all possible uses $\tau_w(t)$ (work), $\tau_I(t)$ (medical care), $\tau_C(t)$ (consumption) and $s[H(t)]$ (sick time),

$$\Omega(t) = \tau_w(t) + \tau_I(t) + \tau_C(t) + s[H(t)]. \quad (8)$$

In this formulation one can interpret $\tau_C(t)$, the own-time input into consumption $C(t)$ as representing leisure.

Income $Y\{H[s(t)]\}$ is taken to be a function of the wage rate $w(t)$ times the amount of time spent working $\tau_w(t)$,

$$Y\{H[s(t)]\} = w(t) \left\{ \Omega(t) - \tau_I(t) - \tau_C(t) - s[H(t)] \right\}. \quad (9)$$

So far we have simply followed Grossman's formulation in continuous time. See Wagstaff (1986), Wolfe (1985), Zweifel and Breyer (1997), and Ehrlich and Chuma (1990) for similar formulations. Our formulation differs however in one crucial respect from prior work: we explicitly impose the constraint that medical care is non-negative for all ages and allow for corner solutions in which individuals do not demand medical care ($I(t) = 0$).

2.1 Periods where individuals do not demand medical care: $I(t) = 0$

It is commonly assumed that any initial "excess" in health capital can be shed and any "deficit" can be repaired over a small period of time and at negligible cost. In other words, individuals are capable of ensuring that their health is at a certain desirable or "optimal" level (e.g., Grossman 1972a, 1972b, 2000; Case and Deaton 2005; Muurinen 1982; Wagstaff 1986; Zweifel and Breyer 1997, Ehrlich and Chuma 1990; Ried 1998).⁵ This assumption is not necessarily always stated explicitly. The literature generally assumes that there are no corner solutions. In making this assumption the literature restricts the solution to Grossman's "optimal" solution. While this allows one to derive simple model predictions for empirical validation, it is unnecessary.

It is useful to view medical health investment $I(t)$ as encompassing activities related to health repair (e.g., purchases of medical goods and services and own-time inputs) and to view health-damaging environments (e.g., work and living environments, etc) as affecting the rate $d(t)$

⁵While many authors realize that medical health investments cannot be negative (i.e. that corner solutions exist), the literature has not fully explored the implications of this constraint.

at which health capital deteriorates (see, e.g., Wagstaf 1986; Case and Deaton 2005). Similar to Grossman (1972a, 1972b, 2000) we treat the health deterioration rate $d(t)$ as strictly exogenous.

Healthy individuals, those with health levels above the “optimal” level, may desire to substitute health capital for more liquid capital. In other words, individuals may wish to “sell” their health. But, as equation (4) shows individuals cannot “choose” health optimally. Instead they can consume medical care (medical health investment) $I(t)$ optimally. But medical care $I(t)$, viewed as health-promoting cannot be traded (individuals cannot “sell” health through negative medical health investment) and is therefore positive for all ages $I(t) \geq 0$. As a result health cannot deteriorate faster than the health deterioration rate $d(t)$. This corresponds to the corner solution $I(t) = 0$.

Thus, we have the following optimal control problem: the objective function (1) is maximized with respect to the control functions $C(t)$ and $I(t)$ and subject to the constraints (2 and 3). The Lagrangean or generalized Hamiltonian (see, e.g., Seierstad and Sydsaeter 1987) of this problem is:

$$\begin{aligned} \mathfrak{J} = & U\{C(t), s[H(t)]\}e^{-\beta t} + q_H(t)\{I(t) - d(t)H(t)\} \\ & + q_A(t)\{\delta A(t) + Y\{s[H(t)]\} - p_X(t)X(t) - p_m(t)m(t)\} + q_I(t)I(t), \end{aligned} \quad (10)$$

where $q_H(t)$ is the adjoint variable associated with the differential equation (2) for health $H(t)$, $q_A(t)$ is the adjoint variable associated with the differential equation (3) for assets $A(t)$, and $q_I(t)$ is a multiplier associated with the condition that health investment is non negative, $I(t) \geq 0$.

2.2 First-order conditions

The first-order condition for maximization of (1) with respect to consumption, subject to the conditions (2) and (3) is (see the Appendix for details)

$$\partial U(t)/\partial C(t) = q_A(0)\pi_C(t)e^{(\beta-\delta)t}, \quad (11)$$

where the Lagrange multiplier $q_A(0)$ is the shadow price of life-time wealth (see, e.g., Case and Deaton 2005) and $\pi_C(t)$ is the marginal cost of consumption $C(t)$

$$\pi_C(t) \equiv \frac{p_X(t)}{\partial C(t)/\partial X(t)} = \frac{w(t)}{\partial C(t)/\partial \tau_C(t)}. \quad (12)$$

The first-order condition for maximization of (1) with respect to health, subject to the conditions (2) and (3) is (see the Appendix for details)

$$\begin{aligned} \frac{\partial U(t)}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)} &= q_A(0) \left\{ \pi_I(t) \left[d(t) + \delta - \tilde{\pi}_I(t) \right] - \frac{\partial Y(t)}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)} \right\} e^{(\beta-\delta)t} + \left[\dot{q}_I(t) - q_I(t)d(t) \right] e^{\beta t} \\ &\equiv q_A(0) \left[\pi_H(t) - \varphi_H(t) \right] e^{(\beta-\delta)t} + \left[\dot{q}_I(t) - q_I(t)d(t) \right] e^{\beta t}, \end{aligned} \quad (13)$$

where $\pi_I(t)$ is the marginal cost of medical health investment $I(t)$ (see equation 10 in Grossman 2000)

$$\pi_I(t) \equiv \frac{P_m(t)}{\partial I(t)/\partial m(t)} = \frac{w(t)}{\partial I(t)/\partial \tau_I(t)} \quad (14)$$

$\tilde{\pi}_I(t) \equiv \dot{\pi}(t)/\pi(t)$, $\pi_H(t)$ is the user cost of health capital at the margin,

$$\pi_H(t) \equiv \pi_I(t) [d(t) + \delta - \tilde{\pi}_I(t)], \quad (15)$$

and $\varphi_H(t)$ is the marginal production benefit of health

$$\varphi_H(t) \equiv \frac{\partial Y(t)}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)}. \quad (16)$$

Note that we have to impose that the user cost of health capital at the margin exceeds the marginal production benefits of health $\pi_H(t) > \varphi_H(t)$. Without this condition, the consumption of medical care would finance itself by increasing wages by more than the user cost of health. As a result of this, consumers would choose infinite medical care paid for by infinite earnings increases to reach infinite health.

Equations (11) and (13) describe the first-order conditions for the constrained optimization problem. Equation (11) is similar to equation 4a by Wagstaff (1986) and equation 6 by Case and Deaton (2005). Equation (13) is similar to equations 13, 1-13 and 11 of Grossman (1972a), (1972b) and (2000), respectively, equation 4b by Wagstaff (1986), equation 3.5 of Zweifel and Breyer (1997), and equation 6 by Case and Deaton (2005), for $q_I(t) = 0$ (i.e., $I(t) > 0$).⁶ The essential difference between our results and those of fore mentioned authors is in the term $q_I(t)$ which is non-vanishing for $I(t) = 0$.

2.3 Grossman's solutions for consumption and health

The first-order condition (13) contains an expression in the multiplier $q_I(t)$ which is non-vanishing ($q_I(t) \neq 0$) for corner solutions in which individuals do not demand medical care ($I(t) = 0$). Let's first focus on the solution where $q_I(t) = 0$. This special case corresponds to the solutions found by Grossman (1972a, 1972b, 2000). The first-order condition (13) determines the "optimal" level of health for the "traditional" Grossman solution.

Denoting Grossman's "optimal" solutions for consumption, consumption goods, medical care, medical goods and services, own time input into the production of consumption, own time input into the production of medical care, sick time and health by $C_*(t)$, $X_*(t)$, $I_*(t)$, $m_*(t)$, τ_{C_*} , τ_{I_*} , $s_*(t)$, and $H_*(t)$, we have:

$$\partial U(t)/\partial C_*(t) = q_{A_*}(0)\pi_{C_*}(t)e^{(\beta-\delta)t}, \quad (17)$$

⁶Various other authors have presented first-order conditions for the Grossman model. The list provided here is not exhaustive.

and,

$$\begin{aligned} \frac{\partial U(t)}{\partial s_*(t)} \frac{\partial s_*(t)}{\partial H_*(t)} &= q_{A_*}(0) \left\{ \pi_{I_*}(t) \left[d(t) + \delta - \tilde{\pi}_{I_*}(t) \right] - \frac{\partial Y(t)}{\partial s_*(t)} \frac{\partial s_*(t)}{\partial H_*(t)} \right\} e^{(\beta-\delta)t} \\ &\equiv q_{A_*}(0) \left[\pi_{H_*}(t) - \varphi_{H_*}(t) \right] e^{(\beta-\delta)t}. \end{aligned} \quad (18)$$

The first-order condition (17) determines the level of consumption. It requires the marginal benefit of consumption to equal the product of the shadow price of life-time wealth $q_{A_*}(0)$, the marginal cost of consumption $\pi_{C_*}(t)$, and a time varying exponent that either grows or decays with time, depending on the difference $\beta - \delta$ between the time preference rate β and the interest rate δ . Increasing lifetime resources will lower $q_{A_*}(0)$ ⁷ and hence increase consumption. The marginal cost of consumption $\pi_{C_*}(t)$ increases with the price $p_{X_*}(t)$ of consumption goods $X_*(t)$ and with wages $w(t)$, and decreases with the efficiency of consumption goods in producing consumption, $\partial C_*(t)/\partial X_*(t)$ and with the efficiency of time inputs $\tau_{C_*}(t)$ in producing consumption, $\partial C_*(t)/\partial \tau_{C_*}(t)$ (see equation 12). Since the marginal benefit of consumption $\partial U(t)/\partial C_*(t)$ is a decreasing function of consumption $C_*(t)$, higher prices of consumption goods $p_{X_*}(t)$, higher wages $w(t)$ and lower efficiencies $\partial C_*(t)/\partial X_*(t)$ and $\partial C_*(t)/\partial \tau_{C_*}(t)$ ⁸ lower the equilibrium level of consumption $C_*(t)$.

The marginal benefit of health (equation 18) equals the product of the shadow price of life-time wealth $q_{A_*}(0)$, the user cost of health capital at the margin $\pi_{H_*}(t)$ minus the marginal production benefits of health $\varphi_{H_*}(t)$, and a time varying term with exponent $-(\beta - \delta)t$. Since the marginal benefit of health $[\partial U(t)/\partial s_*(t)][\partial s_*(t)/\partial H_*(t)]$ is a decreasing function in health $H_*(t)$, lower lifetime resources (higher $q_{A_*}(0)$), higher user cost of health capital $\pi_{H_*}(t)$ and lower production benefits of health $\varphi_{H_*}(t)$ will lower the level of health $H_*(t)$. The user cost of health capital (see equations 14 and 15) increases with the price $p_{m_*}(t)$ of medical goods/services, with wages $w(t)$, the health deterioration rate $d(t)$ and the rate of return on assets δ (reflecting an opportunity cost). The user cost of health capital decreases with the efficiency of medical goods/services in producing medical care, $\partial I(t)/\partial m_*(t)$, the efficiency of time input $\tau_{I_*}(t)$ in producing medical care, $\partial I_*(t)/\partial \tau_{I_*}(t)$, and with $\tilde{\pi}_{I_*}(t)$, the rate of relative change in the marginal cost of medical care π_{I_*} . The marginal production benefit of health $\varphi_{H_*}(t)$ (equation 16) increases with the extent to which health increases earnings $[\partial Y(t)/\partial s_*(t)][\partial s_*(t)/\partial H_*(t)]$.

A lower price of medical goods/services thus increases health. This is pertinent in a cross-country comparison, but also when comparing across the life-cycle, for instance if health care is subsidized for certain age groups (like Medicare in the U.S.) Also, more efficient medical care will lead to greater health. Efficiency can explain variations within a country (if for instance individuals with a higher education level are more efficient consumers of medical care, Goldman and Smith, 2002) or across countries (if health care is more efficient in one country than in another).

⁷This result can be obtained by substituting the solutions for consumption, health, and medical care in the budget constraint (equation 5) and solving for $q_{A_*}(0)$. See, for example, Galama et al. (2009).

⁸I.e., where large increases in $X_*(t)$ and/or $\tau_{C_*}(t)$ result in an insignificant increase in $C_*(t)$.

2.4 Corner solutions

We allow for corner solutions in which individuals do not demand medical care $I(t) = 0$. This situation occurs when individuals have initial health endowments $H(0)$ that are greater than Grossman's "optimal" level of health $H_*(0)$.

We follow a simple intuitive approach. The corner solution is associated with a non-vanishing Lagrange multiplier $q_I(t)$. The solution for consumption is still provided by the first-order condition (11) as this condition is independent of the Lagrange multiplier $q_I(t)$. The solution for medical care is simply

$$I(t) = 0. \quad (19)$$

We do not need to use the first-order condition (13) to obtain the solution for health. Using equation (4) and $I(x) = 0$ we have

$$H(t) = H(0)e^{-\int_0^t d(s)ds}. \quad (20)$$

In other words, in the absence of medical care health deteriorates at the natural deterioration rate $d(t)$. The corner solution is fully determined by equations (11), (19) and (20).

3 Empirical model

The Grossman literature assumes that an individual's health follows Grossman's "optimal" health path, $H_*(t)$ (e.g., Grossman 1972a, 1972b, 2000; Case and Deaton 2005; Muurinen 1982; Wagstaff 1986; Zweifel and Breyer 1997, Ehrlich and Chuma 1990; Ried 1998). In other words, the literature assumes that either the initial health endowment $H(0)$ is at or very close to Grossman's "optimal" health stock $H_*(0)$ or that individuals find this health level desirable and are capable of rapidly dissipating or repairing any "excess" or "deficit" in health.

Corner solutions, where individuals do not demand medical care ($I(t) = 0$), occur when individuals are healthy, i.e. $H(t) > H_*(t)$. Health then deteriorates at the natural deterioration rate $d(t)$ (see equation 20) until it reaches Grossman's level $H(t) = H_*(t)$. Individuals then begin to demand medical care $I(t) > 0$. In other words, the Grossman solution for the "optimal" health stock represents a health "threshold" instead. In our generalized solution of the Grossman model, $H_*(t)$ is the minimum health level individuals "demand" to be economically productive (production benefits of health) or satisfied (consumption benefits of health). Individuals only consume medical care when they are "unhealthy" (health levels at the threshold) and not when they are "healthy" (health levels above the threshold).

Wolfe (1985) assumes an initial surplus of health and is, to the best of our knowledge, the only researcher who has attempted to explore the consequences of corner solutions in Grossman's model in some detail. Wolfe employs a simplified Grossman model where health (or, alternatively, reduced sick time as in Grossman's original formulation) does not provide utility. Wolfe interprets

the onset of “... a discontinuous mid-life increase in health investment ...” with retirement. We however do not associate the discontinuous increase in medical health investment with retirement but with becoming unhealthy (health levels at the health threshold leading to consumption of medical care to improve health). We allow the onset of medical health investment to take place anytime during the life of individuals, including allowing for the possibility that the onset never occurs. While Wolfe (1985) provides a convincing argument that high initial health endowments are plausible⁹, we simply assume that initial health $H(0)$ can take any positive value (including values below the health threshold).

We distinguish three scenarios as shown in figure 1. We show the simplest case in which the health threshold $H_*(t)$ is constant across age (e.g., for constant user cost of health capital $\pi_{H_*}(t) = \pi_{H_*}(0)$, constant production benefits of health $\varphi_{H_*}(t) = \varphi_{H_*}(0)$ and for $\beta = \delta$; see equations 17 and 18) but the scenarios are valid for more general cases. Scenarios A and B begin with initial health $H(0)$ greater than the initial health threshold $H_*(0)$ and scenario C begins with initial health $H(0)$ below the initial health threshold $H_*(0)$. In scenario A health $H(t)$ reaches the health threshold $H_*(t)$ during life (before the age of death T) at age t_1 . In scenario B health $H(t)$ never reaches the health threshold $H_*(t)$ during the life of the individual. In scenario C individuals begin working life with health levels $H(0)$ below the initial health threshold $H_*(0)$.

In scenarios A and B the solution for health is determined by the corner solution presented in section 2.4 for young ages (scenario A) or all ages (scenario B). In scenario A, after health reaches the threshold level the solutions are determined by the “traditional” Grossman solution. In scenarios A and B we do not have to assume that individuals adjust their health to reach the health threshold.

In contrast, in scenario C we follow the traditional Grossman model and assume that an individual is able to adjust his/her health level to reach the health threshold (“optimal” health). Individuals will invest initial assets $A(0)$ to improve initial health $H(0)$ such that initial health equals the initial health threshold $H(0) = H_*(0)$. These solutions have been criticized by Ehrlich and Chuma (1990) as being unrealistic “bang-bang” solutions; the adjustment takes place instantaneously. It is, however, not necessary to assume that the adjustment is instantaneous as individuals will have had ample time to consume medical care before they enter the labor force. There is also naturally an adjustment cost associated with these medical investments in the sense that such individuals begin their work life with fewer assets as a result of the purchase of medical care in the market before they entered the labor force. In other words, by the time individuals enter the labor force their health has gradually reached the health threshold and the adjustment cost is

⁹On the grounds that “... the human species, with its goal of self-preservation, confronts a different problem than the individual who seeks to maximize utility. The evolutionary solution to the former may entail an excessive health endowment in the sense that an individual might prefer to have less health and to be compensated with wealth in a more liquid form ...” In other words, humans may have been endowed with “excessive” health as a result of our evolutionary history which required good physical condition to hunt and gather food, defend ourselves, survive periods of hunger etc. Today’s demands on human’s physical condition are essentially based on the utility of good health and on economic productivity, which in an increasingly knowledge-intensive environment may be significantly smaller than in pre-historic times.

reflected in reduced assets. The health of such individuals will then continue to evolve along the health threshold (the “optimal” health path).

Further, as mentioned before, our working hypothesis is that most individuals are healthy for most of their life (health levels above the health threshold). A consequence of this is that scenario C, where initial health is below the initial health threshold, is less relevant for our discussion. That is, we do not disagree with Ehrlich and Chumas criticism of the Grossman model. The formulation could benefit from a more realistic incorporation of medical technology (allowed to instantaneously take effect in the Grossman model) or from diminishing returns to medical care so that a consumer doesn’t demand such investment all at once (the solution Ehrlich and Chuma offer; see also Case and Deaton, 2005). For the purpose of the current research such extensions would complicate the model and provide relatively little benefit.

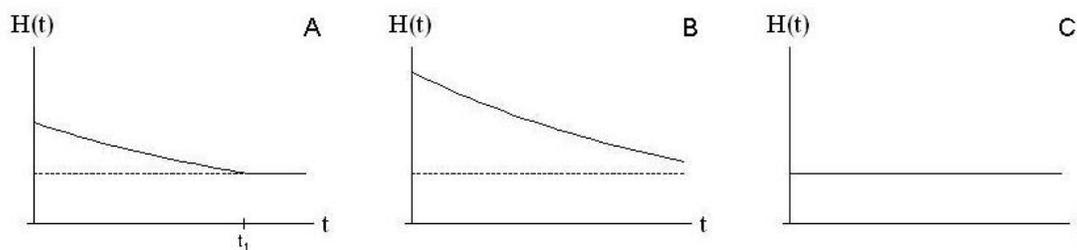


Figure 1: Three scenarios for the evolution of health. t_1 in scenario A denotes the age at which health (solid line) has evolved towards the threshold health level (dotted line).

Following Grossman (1972a, 1972b, 2000) and Wagstaff (1986) we derive structural and reduced form equations for empirical testing. Empirical tests of Grossman’s model in the empirical literature have been based on estimating two sub-models (1) the “pure investment” model in which the restriction $[\partial U(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0$ is imposed and (2) the “pure consumption” model in which the restriction $[\partial Y(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0$ is imposed. To allow comparison with previous research we adopt the same restrictions and explore the same two sub-models. As Wagstaff (1986) notes equation (18) can be transformed into a linear estimating equation with the restriction $[\partial U(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0$ or $[\partial Y(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0$, but this is not the case for the more general model. In addition, without imposing these restrictions analytical solutions for health, medical care and consumption cannot be obtained without making further assumptions. Lastly, the two sub models represent two essential characteristics of health: health as a means to produce (investment) and health as a means to provide utility (consumption). We now discuss each sub-model in turn.

3.1 Pure investment model

In the following we follow Grossman (1972a, 1972b, 2000). We impose

$$[\partial U(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0, \quad (21)$$

assume that sick time is a power law in health

$$s(t) = \beta_0 + \beta_1 H(t)^{-\beta_2}, \quad (22)$$

where β_1 and β_2 are positive constants (e.g., Wagstaff 1986).¹⁰ We thus have $[\partial Y(t)/\partial s(t)][\partial s(t)/\partial H(t)] = \beta_1 \beta_2 w(t) H(t)^{-(\beta_2+1)}$. We further assume that medical health investment (medical care) is produced by combining own time and medical goods/services according to a Cobb-Douglas constant returns to scale production function

$$I(t) = \mu_I(t) m(t)^{1-k_I} \tau_I(t)^{k_I} e^{\rho_I E}, \quad (23)$$

where $\mu_I(t)$ is an efficiency factor, $1 - k_I$ is the elasticity of medical care $I(t)$ with respect to medical goods/services $m(t)$, k_I is the elasticity of medical care $I(t)$ with respect to health time input $\tau_I(t)$, and ρ_I determines the extent to which education E improves the efficiency of medical care $I(t)$. Further, the ratio of the marginal product of medical care with respect to medical goods/services $\partial I(t)/\partial m(t)$ and the marginal product of medical care with respect to own time investment $\partial I(t)/\partial \tau_I(t)$ equals the ratio of the price of medical goods/services $p_m(t)$ to the wage rate $w(t)$ (representing the opportunity cost of time; see equation 14)

$$\frac{\partial I(t)/\partial m(t)}{\partial I(t)/\partial \tau_I(t)} = \frac{p_m(t)}{w(t)} = \frac{1 - k_I}{k_I} \frac{\tau_I(t)}{m(t)}. \quad (24)$$

Lastly, we follow Wagstaff (1986) and Cropper (1981) and assume the health deterioration rate $d(t)$ to be of the form

$$d(t) = d_\bullet e^{\beta_3 t + \beta_4 \mathbf{X}(t)}, \quad (25)$$

where $d_\bullet \equiv d(0)e^{-\beta_4 \mathbf{X}(0)}$ and $\mathbf{X}(t)$ is a vector of environmental variables (e.g., working and living conditions, hazardous environment, etc) that affect the deterioration rate. The vector $\mathbf{X}(t)$ may include other exogenous variables that affect the deterioration rate, such as education (Muurinen, 1982).

3.1.1 Health threshold

Structural form equations The structural form equation for the health “threshold” (Grossman’s solution for “optimal” health) is as follows (see the Appendix for details)

$$\begin{aligned} \ln H(t) &= \beta_5 + \epsilon(1 - k_I) \ln w(t) - \epsilon(1 - k_I) \ln p_m(t) + \epsilon \rho_I E - \epsilon(\beta_3 + \beta_6)t - \epsilon \beta_4 \mathbf{X}(t) \\ &- \epsilon \ln d_\bullet - \epsilon \ln \{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\}, \end{aligned} \quad (26)$$

¹⁰But note that negative values can be allowed as long as $\beta_1 \beta_2 > 0$

where $\epsilon \equiv (\beta_2 + 1)^{-1}$, the constant $\beta_5 \equiv \epsilon \ln(\beta_1 \beta_2) + \epsilon \ln[k_I^{k_I}(1 - k_I)^{(1-k_I)}] + \epsilon \ln \mu_I(0)$, and we allow medical technology $\mu_I(t) = \mu_I(0)e^{-\beta_6 t}$ to depend on age (e.g., the efficiency of medical goods/services $m(t)$ and own time inputs $\tau_I(t)$ in improving health could diminish with age).¹¹ It is customary to assume that the term $\ln d_\bullet$ in equation (26) is an error term with zero mean and constant variance $\xi_1(t) \equiv -\ln d_\bullet$ (as in Wagstaff, 1986, and Grossman 1972a, 1972b, 2000) and that the term $\ln[1 + \delta/d(t) - \tilde{\pi}_I(t)/d(t)]$ (the last term in equation 26) is small or constant (see, e.g., Grossman 1972a, 2000),¹² or that it is time dependent $\ln[1 + \delta/d(t) - \tilde{\pi}_I(t)/d(t)] \propto t$ (e.g. Wagstaff, 1986). We do not have to make these assumptions as in our generalized solution of the Grossman model the rate of deterioration $d(t)$ is observable for those times that individuals do not demand medical care (i.e., for corner solutions). While we assume that the last term in equation (26) is small, our formulation allows us to estimate and test this common assumption.

The demand for health (equation 26) thus increases with wages $w(t)$ and with education E and decreases with prices $p_m(t)$ and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4 \mathbf{X}(t)$). The relation with age t is ambiguous. To ensure that health declines with age, it is commonly assumed that health deterioration increases with age, $\dot{d}(t) > 0$ (i.e. that $\beta_3 > 0$).¹³ But since wages $w(t)$ generally increase with years of experience (e.g., Mincer 1974) it is possible that the health threshold initially increases with age t .

The structural equation for the “optimal” consumption of medical goods/services is as follows

$$\begin{aligned} \ln m(t) &= \beta_7 + \ln H(t) + k_I \ln w(t) - k_I \ln p_m(t) - \rho_I E \\ &+ (\beta_3 + \beta_6)t + \beta_4 \mathbf{X}(t) + \ln d_\bullet + \ln[1 + \tilde{H}(t)d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)}], \end{aligned} \quad (27)$$

where $\beta_7 \equiv -\ln \mu_I(0) - k_I \ln[k_I/(1 - k_I)]$. It is customary to assume that the last term in equation (27), $\ln[1 + \tilde{H}(t)/d(t)] = \ln[1 + \tilde{H}(t)d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)}]$, is small and can be ignored (Grossman 1972b) or treated as an error term (Wagstaff 1986). This would require that the effective rate of change in health $\dot{H}(t)$ is smaller than $d(t)H(t)$. This assumption is perhaps not unreasonable if medical care is efficient and slows down the effective health decline $\dot{H}(t)$. Note, once more that in our generalized solution of the Grossman model $d(t)$ can be observed during times when corner solutions hold. The last term in equation (27) can thus be estimated. For small $\tilde{H}(t)/d(t)$, we have $\ln[1 + \tilde{H}(t)/d(t)] \sim \tilde{H}(t)/d(t)$.

Equation (27) predicts that Grossman’s “optimal” demand for medical goods/services and Grossman’s “optimal” demand for health are positively related. This is the crucial prediction

¹¹For example, elderly and frail patients may not be able to cope with certain aggressive chemotherapy regimens. Note also that advances in medical technology could be modeled by an increasing $\mu_I(0)$ with time (e.g., $\mu_I(0)$ increases with subsequent cohorts).

¹²This would require that the real interest rate δ and changes in the ratio of the price of medical goods/services and the efficiency of medical goods/services in producing medical care $\pi_I(t) = p_m(t)/[\partial I(t)/\partial m(t)]$ are much smaller than the health deterioration rate $d(t)$ or that changes in the interest rate and in $\tilde{\pi}_I(t)$ follow the same pattern as changes in $d(t)$ (so that the term is approximately constant).

¹³Assuming that the efficiency of medical care decreases with age $\beta_6 > 0$ provides an alternative means to achieve the same result.

which empirical studies consistently reject. Further, the demand for medical goods/services increases with wages $w(t)$ and the health deterioration rate (terms d_\bullet, β_3 and $\beta_4 \mathbf{X}(t)$), and decreases with education E and prices $p_m(t)$.

The literature usually focuses on the equations for health (26) and medical care (27), but note that equation (11) provides a condition for consumption $C(t)$ as well, which, after making some reasonable assumptions, can be utilized to obtain expressions for consumption goods $X(t)$ (see the Appendix for details). The budget constraint (equation 5) then provides the solution for assets $A(t)$.

Reduced form equations Wagstaff (1986) notes that one way of overcoming the unobservability of health capital is to estimate reduced-form demand functions for health and medical goods/services. Combining (26) and (27) and eliminating any expression in health $H(t)$ we find (see the Appendix for details):

$$\begin{aligned} \ln m(t) &= \beta_8 + [k_I + \epsilon(1 - k_I)] \ln w(t) - [k_I + \epsilon(1 - k_I)] \ln p_m(t) \\ &- (1 - \epsilon) \rho_I E - \epsilon[\beta_3 - (1 - \epsilon)\beta_6] t - \epsilon \beta_4 \mathbf{X}(t) - \epsilon \ln d_\bullet \\ &- \epsilon \ln \{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\} \\ &+ \ln \{ \epsilon(1 - k_I) [\tilde{w}(t) - \tilde{p}_m(t)] - \epsilon(\beta_3 + \beta_6) - \epsilon \beta_4 \partial \mathbf{X}(t) / \partial t + d_\bullet e^{\beta_3 t + \beta_4 \mathbf{X}(t)} + \epsilon O(t) \}, \end{aligned} \quad (28)$$

where $\beta_8 \equiv \beta_5 + \beta_7$ and

$$O(t) = \frac{\tilde{d}(t) [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6] + k_I \left[\frac{\dot{\tilde{w}}(t)}{\tilde{w}(t)} - \tilde{w}(t)^2 \right] + (1 - k_I) \left[\frac{\dot{\tilde{p}}_m(t)}{\tilde{p}_m(t)} - \tilde{p}_m(t)^2 \right]}{[d(t) + \delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]}, \quad (29)$$

which we assume to be small (of the order $\tilde{d}(t) \times \delta, \tilde{d}(t) \times \tilde{w}(t)$, etc).

The demand for medical goods/services (equation 28) increases with wages $w(t)$ and the efficiency of medical care (term β_6), and decreases with prices $p_m(t)$, education E , and the health deterioration rate (terms d_\bullet, β_3 and $\beta_4 \mathbf{X}(t)$).¹⁴

3.1.2 Corner solution

We have (using equations 20 and 25)

$$\ln H(t) = \ln H(0) - d_\bullet \int_0^t e^{\beta_3 s + \beta_4 \mathbf{X}(s)} ds, \quad (30)$$

and

$$m(t) = 0. \quad (31)$$

Note that during periods in which the corner solutions hold it is in principle possible to determine the rate of deterioration d_\bullet empirically. Hence we do not have to assume that the term $\ln d_\bullet$ in equations (26) and (27) is an error term.

¹⁴For $0 < \epsilon < 1$.

3.1.3 Regime switching

The time t_1 when health has deteriorated to the “threshold” level must satisfy the following condition (given by equating 26 with 30):

$$\begin{aligned}
\ln H(t_1) &= \beta_5 + \epsilon(1 - k_I)\ln w(t_1) - \epsilon(1 - k_I)\ln p_m(t_1) + \epsilon\rho_I E - \epsilon(\beta_3 + \beta_6)t_1 - \epsilon\beta_4 \mathbf{X}(t_1) \\
&- \epsilon \ln d_\bullet - \epsilon \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t_1 - \beta_4 \mathbf{X}(t_1)} [\delta - k_I \tilde{w}(t_1) - (1 - k_I) \tilde{p}_m(t_1) - \beta_6]\} \\
&= \ln H(0) - d_\bullet \int_0^{t_1} e^{\beta_3 s + \beta_4 \mathbf{X}(s)} ds
\end{aligned} \tag{32}$$

The model thus implies a switch of regimes at time t_1 . Before t_1 the evolution of health is given by equation (30), whereas after t_1 it is given by (26). Empirically, this would generate a switching regression model with endogenous switching. Once health hits the “optimal” path, the process governing health switches from (30) to (26). Similarly, before t_1 the demand for medical goods/services is given by equation (31), whereas after t_1 it is given by (27) or, alternatively, by (28).

3.2 Pure consumption model

In the following we follow Wagstaff (1986). We impose

$$[\partial Y(t)/\partial s(t)][\partial s(t)/\partial H(t)] = 0. \tag{33}$$

To convert (18) into estimable equations we have to specify a functional form for the utility function.

3.2.1 Utility specification

Grossman (1972a, 1972b, 2000) formulates his model in terms of sick time¹⁵ and assumes that sick time $s(t)$ is a function of health $H(t)$; $s(t) = s[H(t)]$. An alternative formulation is provided by Case and Deaton (2005). Case and Deaton formulate a simplified Grossman model in which utility and income are functions of health $H(t)$ directly, rather than indirectly through sick-time $s(t)$ which in turn is assumed to be a function of health $s(t) = s[H(t)]$ (as in Grossman 1972a, 1972b, 2000). Following Case and Deaton we write utility $U\{C(t), s[H(t)]\} = U\{C(t), H(t)\}$ and income $Y\{s[H(t)]\} = Y[H(t)]$ as functions of health $H(t)$ instead of sick time $s(t)$. Essentially both formulations are equivalent except that Case and Deaton’s formulation is more general, allowing for example for earnings to be influenced not only by reductions in sick time but also increased worker efficiency resulting from good health. And, at any time we can revert back to the original specification in terms of sick time if deemed desirable.

¹⁵One possible reason for this formulation is that the NORC data set the author employed in empirical testing of the model contained information on sick days

We begin by noting that (see the first-order conditions 11 and 13)

$$\frac{\partial U(t)}{\partial H(t)} = \pi_C(t)^{-1} [\pi_H(t) - \varphi_H(t)] \frac{\partial U(t)}{\partial C(t)} + [\dot{q}_I(t) - q_I(t)d(t)] e^{\beta t}. \quad (34)$$

In other words, the marginal benefit of health $\partial U(t)/\partial H(t)$ is given by the function $\pi_C(t)^{-1}[\pi_H(t) - \varphi_H(t)]$ times the marginal benefit of consumption $\partial U(t)/\partial C(t)$ and an additional expression in $q_I(t)$. For Grossman's solutions we have $q_I(t) = 0$ and the additional term vanishes.

Equation (34) suggests that the marginal utility of health $\partial U(t)/\partial H(t)$ and the marginal utility of consumption $\partial U(t)/\partial C(t)$ are functions of both health $H(t)$ and consumption $C(t)$. To allow for this we specify the following constant relative risk aversion (CRRA) utility function:

$$U[C(t), H(t)] = \frac{1}{1-\rho} [C(t)^\zeta H(t)^{1-\zeta}]^{1-\rho}, \quad (35)$$

where ζ ($0 \leq \zeta \leq 1$) is the relative "share" of consumption versus health and ρ ($\rho > 0$) the coefficient of relative risk aversion.

The functional form for the utility function can account for the observation that the marginal utility of consumption declines as health deteriorates (e.g., Finkelstein, Luttmer and Notowidigdo, 2008). The authors find that a one-standard deviation increase in the number of chronic diseases is associated with an 11 percent decline in the marginal utility of consumption relative to this marginal utility when the individual has no chronic diseases (the 95 percent confidence interval ranges between 2 percent and 17 percent). This would rule out the strongly separable functional form for the utility function employed by Wagstaff (1986), where the marginal utility of consumption is independent of health. While we follow Wagstaff (1986) in most of the derivations we do not adopt his utility specification.

3.2.2 Health threshold

Structural form equations The structural equation for the health "threshold" (Grossman's solution for "optimal" health) is as follows (see the Appendix for details)

$$\begin{aligned} \ln H(t) &= \beta_9 + \ln X(t) + \ln p_X(t) - k_I \ln w(t) - (1 - k_I) \ln p_m(t) \\ &+ \rho_I E - (\beta_3 + \beta_6)t - \beta_4 \mathbf{X}(t) - \ln d_\bullet \\ &- \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\}, \end{aligned} \quad (36)$$

where $\beta_9 \equiv \ln \mu_I(0) - \ln(1 - k_C) + \ln[k_I^k (1 - k_I)^{(1-k_I)}] + \ln[(1 - \zeta)/\zeta]$. The health threshold thus increases with consumption goods $X(t)$, prices for consumption goods $p_X(t)$, and education E and decreases with wages $w(t)$, prices of medical goods/services $p_m(t)$, and the health deterioration rate (terms d_\bullet, β_3 and $\beta_4 \mathbf{X}(t)$). The last term is generally assumed to be small and can be estimated in our formulation.

The structural form equation for medical goods/services is the same as for the pure investment model (equation 27) and is repeated for convenience:

$$\begin{aligned} \ln m(t) &= \beta_7 + \ln H(t) + k_I \ln w(t) - k_I \ln p_m(t) - \rho_I E \\ &+ (\beta_3 + \beta_6)t + \beta_4 \mathbf{X}(t) + \ln d_\bullet + \ln[1 + \tilde{H}(t)d_\bullet^{-1}e^{-\beta_3 t - \beta_4 \mathbf{X}(t)}], \end{aligned} \quad (37)$$

where $\beta_7 \equiv -\ln \mu_I(0) - k_I \ln[k_I/(1 - k_I)]$.

Combining equation (36) with (37) and eliminating any expression in health $H(t)$ we find (see the Appendix for details):

$$\begin{aligned} \ln m(t) &= \beta_{12} + \ln X(t) + \ln p_X(t) - \ln p_m(t) - \ln d_\bullet - \beta_3 t - \beta_4 \mathbf{X}(t) \\ &- \ln\{1 + d_\bullet^{-1}e^{-\beta_3 t - \beta_4 \mathbf{X}(t)}[\delta - k_I \tilde{w}(t) - (1 - k_I)\tilde{p}_m(t) - \beta_6]\} \\ &+ \ln\{d_\bullet e^{\beta_3 t + \beta_4 \mathbf{X}(t)} - (1 - k_I)\tilde{p}_m(t) - k_I \tilde{w}(t) - (\beta_3 + \beta_6) - \beta_4 \partial \mathbf{X}(t)/\partial t \\ &+ \tilde{X}(t) + \tilde{p}_X(t) + \mathcal{O}(t)\}, \end{aligned} \quad (38)$$

where $\beta_{12} \equiv \beta_7 + \beta_9$, and the expression for $\mathcal{O}(t)$ is provided by equation (29).

Reduced form equations Note that the health threshold (equation 36) is expressed directly as a function of consumption goods $X(t)$. This relation is different from the one found by Wagstaff (1986; his equation 12), which is the result of our choice for the functional form of the utility function (equation 35). Wagstaff (1986) finds that health $H(t)$ is a function of the shadow price of life-time wealth $q_A(0)$. We can obtain a similar reduced form expression to the one found by Wagstaff (1986) by using the first-order condition (11) and making some reasonable assumptions to obtain an expression for consumption good $X(t)$. We then find (see the Appendix for details):

$$\begin{aligned} \ln H(t) &= \beta_{10} - \chi(1/\rho\chi - 1)(1 - k_C) \ln p_X(t) - \chi(1 - k_I) \ln p_m(t) \\ &- \chi[k_I + (1/\rho\chi - 1)k_C] \ln w(t) + \chi[\rho_I + (1/\rho\chi - 1)\rho_C]E \\ &- \chi[(\beta_3 + \beta_6) + (1/\rho\chi - 1)\beta_{11} + (\beta - \delta)/\rho\chi]t - \chi\beta_4 \mathbf{X}(t) - \chi \ln d_\bullet + \ln q_A(0)^{-1/\rho} \\ &- \chi \ln\{1 + d_\bullet^{-1}e^{-\beta_3 t - \beta_4 \mathbf{X}(t)}[\delta - k_I \tilde{w}(t) - (1 - k_I)\tilde{p}_m(t) - \beta_6]\}, \end{aligned} \quad (39)$$

where

$$\begin{aligned} \beta_{10} &\equiv \chi \ln \mu_I(0) + \chi(1/\rho\chi - 1) \ln \mu_C(0) + \chi \ln[k_I^{k_I}(1 - k_I)^{(1 - k_I)}] \\ &+ \chi(1/\rho\chi - 1) \ln[k_C^{k_C}(1 - k_C)^{(1 - k_C)}] + \chi \ln[(1 - \zeta)/\zeta] + \ln \zeta^{1/\rho}, \end{aligned}$$

and

$$\chi \equiv \frac{1 + \rho\zeta - \zeta}{\rho}, \quad (40)$$

and we allow the efficiency of consumption to depend on age $\mu_C(t) = \mu_C(0)e^{-\beta_{11}t}$.

An expression for the shadow price of life-time wealth $q_A(0)$ in equation (39) can be obtained by using the life-time budget constraint (equation 5), substituting the solutions for consumption,

health, and medical care and solving for $q_A(0)$ (see, for example, Galama et al. 2009). The shadow price of life-time wealth $q_A(0)$ is found to be a complicated function of life-time wealth (assets, life-time income), wages $w(t)$, prices $p_m(t)$, $p_X(t)$, education E and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4\mathbf{X}(t)$). Wagstaff (1986) provides a simple approximation for the shadow price of life-time wealth $q_A(0)$ (his equations 15 and 16) which may be easier to use in empirical testing of the model.

Assuming that both medical goods / services $m(t)$ and time input $\tau_I(t)$ increase medical care suggests $0 \leq k_I \leq 1$, and if education E increases the efficiency of medical care then $\rho_I > 0$ (see equation 23). Similarly we have $0 \leq k_C \leq 1$ and $\rho_C > 0$ (see equation 62). Finkelstein, Luttmer and Notowidigdo (2008) provide evidence that the marginal utility of consumption declines as health deteriorates. Assuming further diminishing marginal benefits of health $\partial^2 U(t)/\partial^2 H(t) < 0$ we find $1 < \chi < 1 + 1/\rho$ (and hence $0 < \rho < 1$ and $1/\rho\chi > 1$).

For these parameter values we find that the health threshold (equation 39) increases with education E , life-time wealth $q_A(0)^{-1/\rho}$, and decreases with the price of consumption goods $p_X(t)$, the price of medical care $p_m(t)$, wages $w(t)$, and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4\mathbf{X}(t)$). The health threshold could increase or decrease with age depending on the sign of $\chi(\beta_3 + \beta_6) + \chi(1/\rho\chi - 1)\beta_{11} + [(\beta - \delta)/\rho]$ and on the evolution of wages $w(t)$ with years of experience (e.g., Mincer, 1974).

Combining equation (37) with (39) we find:

$$\begin{aligned}
\ln m(t) &= \beta_{13} - \chi(1/\rho\chi - 1)(1 - k_C) \ln p_X(t) - [k_I + \chi(1 - k_I)] \ln p_m(t) \\
&- \chi[(1 - 1/\chi)k_I + (1/\rho\chi - 1)k_C] \ln w(t) + \chi[(1 - 1/\chi)\rho_I + (1/\rho\chi - 1)\rho_C]E \\
&- \chi[(1 - 1/\chi)(\beta_3 + \beta_6) + (1/\rho\chi - 1)\beta_{11} + (\beta - \delta)/\rho\chi]t \\
&- (\chi - 1)\beta_4\mathbf{X}(t) - (\chi - 1) \ln d_\bullet + \ln q_A(0)^{-1/\rho} \\
&- \chi \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\}, \tag{41}
\end{aligned}$$

where $\beta_{13} \equiv \beta_7 + \beta_{10}$. The demand for medical goods/services (equation 41) increases with education E , life-time wealth $q_A(0)^{-1/\rho}$, and decreases with the price of consumption goods $p_X(t)$, the price of medical goods/services $p_m(t)$, wages $w(t)$, and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4\mathbf{X}(t)$). The health threshold could increase or decrease with age depending on the sign of $\chi(1 - 1/\chi)(\beta_3 + \beta_6) + \chi(1/\rho\chi - 1)\beta_{11} + [(\beta - \delta)/\rho]$ and on the evolution of wages $w(t)$ with years of experience (e.g., Mincer, 1974).

3.2.3 Corner solution

The solutions are given by the corner solutions (30) and (31) derived in section 2.4.

3.2.4 Regime switching

The time t_1 when health has deteriorated to the “threshold” level must satisfy the following condition (given by equating 36 or 39 with 30):

$$\begin{aligned}
\ln H(t_1) &= \beta_9 + \ln X(t_1) + \ln p_X(t_1) - k_I \ln w(t_1) - (1 - k_I) \ln p_m(t_1) \\
&+ \rho_I E - (\beta_3 + \beta_6)t_1 - \beta_4 \mathbf{X}(t_1) - \ln d_\bullet \\
&- \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t_1 - \beta_4 \mathbf{X}(t_1)} [\delta - k_I \tilde{w}(t_1) - (1 - k_I) \tilde{p}_m(t_1) - \beta_6]\} \\
&= \beta_{10} - \chi(1/\rho\chi - 1)(1 - k_C) \ln p_X(t_1) - \chi(1 - k_I) \ln p_m(t_1) \\
&- \chi[k_I + (1/\rho\chi - 1)k_C] \ln w(t_1) + \chi[\rho_I + (1/\rho\chi - 1)\rho_C] E \\
&- \chi[(\beta_3 + \beta_6) + (1/\rho\chi - 1)\beta_{11} + (\beta - \delta)/\rho\chi] t_1 - \chi\beta_4 \mathbf{X}(t_1) - \chi \ln d_\bullet + \ln q_A(0)^{-1/\rho} \\
&- \chi \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t_1 - \beta_4 \mathbf{X}(t_1)} [\delta - k_I \tilde{w}(t_1) - (1 - k_I) \tilde{p}_m(t_1) - \beta_6]\} \\
&= \ln H(0) - d_\bullet \int_0^{t_1} e^{\beta_3 s + \beta_4 \mathbf{X}(s)} ds. \tag{42}
\end{aligned}$$

Similar to the previous discussion for the pure investment model, the model thus implies a switch of regimes at time t_1 . Before t_1 the evolution of health is given by equation (30), whereas after t_1 it is given by (36) or by (39). Empirically, this would generate a switching regression model with endogenous switching. Once health hits the optimal path, the process governing health switches from (30) to (36), or alternatively to (39). Similarly, before t_1 medical care is given by equation (31), whereas after t_1 it is given by (37) or alternatively (38) or (41).

4 Model Predictions

The Grossman model has been tested in a number of empirical studies on a variety of datasets from different countries (Grossman 1972a; Wagstaff 1986, 1993; Leu and Doppman 1986; Leu and Gerfin 1992; van Doorslaer 1987; Van de Ven and van der Gaag, 1982; Erbsland, Ried and Ulrich 2002; Gerdtham et al. 1999; Gerdtham and Johannesson 1999).¹⁶ Despite the large

¹⁶Grossman (1972a) employs the 1963 health interview survey conducted by the National Opinion Research Center (NORC) of the U.S. civilian noninstitutionalized population. Grossman employs measures of sick time and self-reported health and restricts the dataset to individuals with positive sick time. Wagstaff (1986) employs the 1976 Danish Welfare Survey (DWS) and uses principal components analysis (PCA) to derive a smaller number of health components from a long list of health indicators. Wagstaff also uses the wealth of DWS measures of work environment and use-related health depreciation. Measures of medical care employed are general practitioner visits, weeks in hospital and number of complaints for which medicine are taken. Wagstaff (1993) employs the Danish Health Survey (DHS) and uses a latent variable health model (multiple indicators multiple causes; MIMIC). Leu and Doppman (1986) employ a latent health variable, latent earnings and latent transfer income model based on Socio-medical indicators for the population of Switzerland (SOMIPOPS) data combined with the Swiss income and wealth study (SEVS). General practitioner consultations, hospital days and sick days are used as measures of medical care. Leu and Gerfin (1992) employ the same datasets as Leu and Doppman (1986) but follow a different methodology (health is a latent variable but no other latent variables are employed). Van Doorslaer (1987) estimates a latent health and latent medical knowledge

variety in methodologies and the diversity in cultural and institutional environments these datasets represent, the studies are broadly in agreement with one another and confirm the predictions of the Grossman model for the demand for *health*. Health is found to increase with income (wages, life-time earnings), and education, and decreases with age, the price of medical goods/services, being single, and with environmental factors, such as, physically and mentally demanding work environments, manual labor, psychological stress factors.¹⁷

While reduced form estimates of the demand for *medical care* are generally in agreement with the predictions of the Grossman model, this is not true for structural estimates (see Wagstaff 1986). Structural estimates allow for direct testing of the relationship between health (most often a latent health variable is employed) and medical care. The most noticeable feature of such structural estimates is the consistently negative relationship between health and medical care (healthy individuals do not go to the doctor). But this relationship is predicted to be positive in the traditional solution of the Grossman model (see equation 27; those who consume more medical care are healthier). Further, the negative relationship between health and medical care is found to be the most statistically significant of any relationship between medical care and any of the independent variables (see, e.g., Grossman 1972a; Wagstaff 1986, 1993; Leu and Doppman 1986; Leu and Gerfin 1992; van Doorslaer 1987; Van de Ven and van der Gaag, 1982; Erbsland, Ried and Ulrich 2002).

We assume that each of the scenarios A, B and C occur in reality (see Figure 1). In other words, that there exist *healthy* individuals who consume medical care during some part of their life (scenario A; initial health above the initial health threshold and the threshold reached during life), *very healthy* individuals who never consume medical care (scenario B; initial health well above the initial health threshold and the threshold never reached), and *ill* individuals who consume medical care their entire life (scenario C; initial health at the health threshold). We do not a-priori know the distribution of healthy, very healthy and ill individuals in the population but if a statistically significant share of individuals have initial health endowments $H(0)$ above the initial health threshold $H_*(0)$ (scenarios A and B) then empirical tests should be able to distinguish between

variable model to the Health Interview Survey of the Belgian National Health Research Project on Primary Health Care conducted in 1976 among the Dutch-speaking (Flemish) population. Van de Ven and van der Gaag (1982) employ a MIMIC model with latent health and data from a health-care survey among 8000 households in The Netherlands. Erbsland, Ried and Ulrich (2002) use data from the German Socio Economic Panel (SOEP). They use a model with a latent health and a latent environment variable and restrict the analysis to the working population and to those with a positive demand for medical services. Gerdtham et al. (1999) use a rating scale and a time-trade off method to obtain measures of health as well as a self-reported measure of health from data collected in Uppsala County in Sweden. Gerdtham and Johannesson (1999) use a self-reported health measure from 1991 data of the Level of Living Survey (LNU), a random sample of the Swedish population. Both Gerdtham et al. (1999) and Gerdtham and Johannesson (1999) provide estimates of the demand for health but no structural estimates for the demand of medical care.

¹⁷In addition, these studies find that health increases with healthy behavior (sports, healthy eating and sleeping habits) and decreases with being overweight and with smoking. Females are found to be in lower health. And, moderate alcohol consumption is found to have a positive or negligible impact on health (e.g., Gerdtham et al. 1999, Leu and Doppman 1986). Since the effect of consumption (healthy and unhealthy forms) on health as well as health behaviors (exercise, sleeping habits) and gender differences are not part of the Grossman model we do not discuss these here.

the interpretation of the Grossman model advocated here (represented by the joint occurrences of scenarios A, B and C) and the interpretation adopted in the literature (represented by scenario C only).

In the following we will contrast the predictions of our interpretation of the Grossman model with the more generally held interpretation and with empirical observations from the literature.

4.1 Similarities

The predictions for the demand for health and for medical care for unhealthy individuals (those individuals whose health is at the threshold) in our generalized solution of the Grossman model are, with the exception of some minor differences in formulation, the same as for the original solution of the Grossman model. Those predictions have largely been verified in the empirical literature, with the exception of the relation between the demand for health and the demand for medical care (see for details the earlier discussion and references therein). We summarize our predictions in Table 1.

Our generalized solution of the Grossman model broadly replicates the predictions of the traditional solution of the Grossman model. This can be seen as follows. Since the empirical literature has not distinguished between healthy and unhealthy individuals (a concept introduced in this work) a mixture of healthy and unhealthy individuals will have been included in the samples investigated. If at any time the proportion of unhealthy individuals (those whose health is at the health threshold and who behave according to the traditional Grossman solution) is significant this could produce the observed relationships, with the exception of the relation between health and medical care. The reason that the relationship between health and medical care is different stems from the significantly different behavior between healthy and unhealthy individuals. The healthy do not consume medical care while the unhealthy do. If both healthy and unhealthy individuals are included in a sample this would produce the observed strong negative relationship between measures of health and measures of medical care. At the same time, if we can restrict the sample to the unhealthy, we should observe the positive relationship between health and medical care as predicted by Grossman.¹⁸

As Table 1 shows we expect health to decrease with the price of medical goods/services $p_m(t)$, unhealthy environmental factors $(\mathbf{X}(t), d_\bullet, \beta_3)$, and increase with education E ¹⁹ and with the efficiency of medical care $-\beta_6$. The relation with age t is ambiguous as wages $w(t)$ increase with working experience (e.g., Mincer, 1974) potentially countering the “aging” variables β_3, β_6 ,

¹⁸Note that in retirement there is no production benefit from health as income, a pension / savings, is independent of the health status of the individual. Whether individuals demand less health as a result is unclear. The increased availability of leisure could reduce or increase the demand for health depending on whether leisure is a substitute or complement of health (see for a discussion Galama et al. 2009). Given potential differences in the demand for health between workers and retirees it may be necessary to distinguish between workers and retirees to potentially establish the positive relationship between health and medical care.

¹⁹Note that education could possibly enter through lowering the rate of health deterioration $d(t)$ in addition, or as an alternative, to increasing the efficiency of medical care; see, e.g., Muurinen (1982)

	Health			Medical care
	PI ^a	PC	PC (Full) ^b	PI/PC
	Eq. 26	Eq. 36	Eq. 39	Eqs. 27, 37
Health $H(t)$	n/a	n/a	n/a	+
Wages $w(t)$	+	-	-	+
Price of medical goods/services $p_m(t)$	-	-	-	-
Education E	+	+	+	-
Age t	?	?	?	?
(Un)healthy environment $\mathbf{X}(t), d_{\bullet}, \beta_3$	(-)+	(-)+	(-)+	(+)-
Consumption good $X(t)$	n/a	+	n/a	n/a
Price of consumption good $p_X(t)$	n/a	+	-	n/a
Life-time wealth $q_A(0)^{-1/\rho}$	n/a	n/a	+	n/a
Efficiency of medical care $-\beta_6$	+	+	+	-

Table 1: Relationships between the health threshold (Health) and various model variables and between the demand for medical care (Medical care) and various model variables for both the pure investment (PI) and the pure consumption (PC) models. Equation numbers (Eq.) refer to the structural form equations in section 3.

^a Relations are valid for $\epsilon = 1/(1 + \beta_2) > 0$.

^b For plausible parameter choices. Precise relationships and conditions under which relations are valid are provided in section 3.

$\beta - \delta$. The effect of wages $w(t)$ is unclear, with a positive effect on health in the pure investment (PI) and a negative effect on health in the pure consumption (PC) model. Do note however that the predictions for the PC model have less predictive power than for the PI model. The structural form equation (36) includes consumption good $X(t)$, an endogenous variable, which in turn is a function of exogenous variables, such as wages $w(t)$, the price of medical goods/services $p_m(t)$, education E , etc. The inclusion of consumption good $X(t)$ in the structural form equation may distort the relationships between health and the exogenous variables. While the structural form equation (39) does not suffer from this problem, the predictions shown in the table depend on assumptions about model parameters (see table note b in the Table 1). In addition, the shadow price of life-time wealth $q_A(0)$ is a complicated function of various exogenous variables over the life cycle. Equation (39) thus suffers from a similar lack of transparency.

With regard to the demand for medical care, Table 1 shows that we expect the demand for medical goods/services to decrease with the price of medical goods/services $p_m(t)$, education E and the efficiency of medical care $-\beta_6$, and to increase with health $H(t)$, wages $w(t)$ and unhealthy environmental factors ($d_4, \beta_3, \mathbf{X}(t)$). The predictions for the PI and PC models are the same. As discussed earlier the positive relationship between health and medical care is expected to be observable only if the sample can be restricted to unhealthy individuals.

4.2 Differences

In addition to the above predictions of our generalized solution of the Grossman model that are the same as in the traditional solution of the model, there are a number of distinctly different predictions. Those are discussed in detail below. We denote the predictions of our interpretation of the Grossman model by *Health threshold*, the more generally held interpretation by “*Optimal*” *stock* and the empirical observations from the literature by *Empirical literature*.

1. Medical care and health are negatively correlated if measured across healthy and unhealthy individuals

“Optimal” stock: Health and medical care are positively correlated (see equations 27 and 37), i.e. individuals who consume more medical care are healthier.

Health threshold: Healthy individuals ($H(t) > H_*(t)$) do not consume medical care, while unhealthy individuals ($H(t) = H_*(t)$) do. I.e. healthy individuals do not go to the doctor much, do not take much medicine, are not found to stay often in hospitals. Measured across a sample of healthy and unhealthy individuals we expect unhealthy individuals to consume more medical care than healthy individuals.

Empirical literature: As discussed earlier the most striking feature of structural form estimates of the demand for *medical care* (see, e.g., Grossman 1972a; Wagstaff 1986, 1993; Leu and Doppman 1986; Leu and Gerfin 1992; van Doorslaer 1987; Van de Ven and van der Gaag 1982; Erbsland, Ried and Ulrich 2002) is the persistent and highly statistically significant negative relation found between measures of health and measures of medical

care. The studies employ a variety of methodologies and a variety of datasets representing different cultural and institutional settings in a number of different countries (Europe and U.S.), yet their findings are largely in agreement with one another. None of these studies separate a healthy from an unhealthy population and hence we expect to observe a strong negative correlation between health and medical care if the population consists of both healthy and unhealthy individuals.²⁰

2. Healthy people do not consume medical care

“Optimal” stock: In the standard solution of the Grossman model individuals consume medical care at all ages.

Health threshold: In our generalized solution healthy individuals (individuals whose health $H(t)$ is above the threshold $H_*(t)$) do not consume medical goods/services, i.e. we would expect some fraction of the population at any given time to not consume medical goods/services.

Empirical literature: We would expect that healthy people pay few visits to the doctor (perhaps only to prevent illness, such as for a “health check up”) and that they do not require much medical care (hospital stays, use medicine, etc). For example, Wagstaff (1986) observes that 48% of the 1976 Danish Welfare Survey (DWA) sample he employed recorded zero general practitioner visits and 46.5% recorded zero weeks in hospital.

3. Effective health deterioration slows when individuals reach the health threshold

“Optimal” stock: In the standard solution of the Grossman model health evolves as Grossman’s “optimal” health stock, i.e. we do not expect to see discontinuous changes in the evolution of health.

Health threshold: Healthy people ($H(t) > H_*(t)$) do not consume medical goods/services and their health deteriorates at the “natural” deterioration rate $\dot{H}(t) = -d(t)H(t)$. When, as a result of health deterioration their health reaches the health threshold $H(t) = H_*(t)$ (i.e., they have become unhealthy by our definition) they begin to consume medical goods/services and their health deteriorates at a lower effective rate $\dot{H}(t) = I(t) - d(t)H(t)$. If medical care improves one’s health (e.g., medical care is effective), we expect to observe slower effective health deterioration $\dot{H}(t)$ or even health improvement when individuals reach the health threshold and begin to consume medical goods/services).²¹

²⁰Grossman (1972a) however selected a sub sample of the NORC dataset by restricting the data to those individuals that reported positive sick time and Erbsland, Ried and Ulrich (2002) restricted the sample to individuals reporting positive demand for health services. Interestingly Grossman (1972a) shows the least statistically significant negative relation between health and medical outlays of all the studies (t-stat of -5.84 [see table 7 OLS estimates]). Erbsland, Ried and Ulrich (2002) report t-values of around -10 for three measures of medical care usage. Other studies, on the other hand, report values of at least -10 and up to -90. Perhaps the restriction of the samples to individuals that report positive sick time or positive medical care partially limited the sample to unhealthy respondents.

²¹Note the distinction between the effective health deterioration rate $\dot{H}(t)$ and the “natural” health deterioration rate $d(t)$.

Empirical literature: Van Kippersluis et al. (2008) examine inequality in self-reported health (SRH) as a function of income in 11 European countries. The authors transform the ordinal SRH information onto a cardinal scale using utility scores for the SRH categories taken from the 2001 Canadian Community Household Survey (CCHS). The authors find a remarkable consistency in the pattern of health with age. In most countries health deteriorates gradually from early adulthood until around age 50 after which it generally levels off before accelerating rapidly after age 70. The authors find this middle-age plateau (ages 50-70) rather puzzling, but it would be consistent with a slowing of the decline in health resulting from increased medical care as the average individual reaches a health threshold. After age 70, as terminal illnesses set in, health again declines rapidly.

Smith (2004, 2007) uses self-reported health (SRH) status from the National Health Interview Survey (NHIS) and PSID to show how disparity in health between low- and high-income individuals (the so-called socio-economic status [SES]-health gradient) increases with age till about age 60 after which the disparity narrows (see Van Doorslaer et al. 2008 for an excellent review of the literature on the SES-health gradient over the life cycle). The percentage of individuals reporting excellent or very good health status declines rapidly till age 60 for the first income quartile households (lowest income) and then remains fairly constant out till age 90. The 2nd to 4th income quartiles however show a more gradual decline.

Similarly, Case and Deaton (2005) present several plots of self-reported health (SRH) status from the NHIS as a function of age. Women and men in the bottom income quartile show a rapid deterioration in SRH between ages 20 and 60 after which the SRH curve flattens significantly (see their figure 2). Again we see no evidence for a flattening of SRH with increasing age for the upper income quartile (in fact we see gradually deteriorating SRH status). This suggests that high SES individuals reach a health threshold much later (their SRH deteriorates slower) than low SES individuals. As a result they see no need to consume medical goods/services even at late ages and their effective health deterioration does not slow with age.

Van Kippersluis et al. (2009) find similar results for the Netherlands using a rich dataset based on the Health Interview Surveys and administrative data from Statistics Netherlands (CBS). The data allows the authors to study SRH as well as mortality, to disentangle the effect of ageing from that of cohort effects and to use actual (not reported) income from tax files. The authors find the pattern of the SES-health gradient over the life cycle in the Netherlands to be remarkably similar to that in the U.S., despite significant differences in the two countries' institutions.

Wagstaff (1993) fits an empirical reformulation of the Grossman model to two data subsets, those aged under 41 and those aged over 41. The author finds that for the over 41s the rate of effective health deterioration $\dot{H}(t)$ is lower than for the under 41s (the estimated relationship is $H_t \propto 0.849H_{t-1}$ for the over 41s [table 2b in Wagstaff 1993] and $H_t \propto 0.687H_{t-1}$ for the

under 41s [table 2a in Wagstaff 1993]). Further, the fit is better for the over 41s ($R^2 = 0.595$) than for the under 41s ($R^2 = 0.394$). Since we expect that an older population will have relatively more individuals with health levels at or near the health “threshold” we would expect this population to provide a better fit to the “traditional” solution of the Grossman model.

So, perhaps older individuals, and in particular low income individuals, are slowing their effective health deterioration $\dot{H}(t)$ in late age by consuming medical goods/services as a threshold model would predict.²²

4. Effective health deterioration and medical care are negatively correlated

“Optimal” stock: According to the structural form equation (26) we find $\dot{H}(t) \propto -\epsilon(\beta_3 + \beta_6)H(t)$ (assuming variation in wages $w(t)$, prices $p_m(t)$ and environment $\mathbf{X}(t)$ is slow). Thus, high effective health deterioration requires that $\beta_3 + \beta_6$ is large and/or that health $H(t)$ is large ($\epsilon > 0$ is required to reproduce other empirical findings; see note *a* in Table 1). The model then predicts that medical goods/services $m(t) \propto H(t)e^{(\beta_3 + \beta_6)t}$ are also high and increase exponentially with age (see equation 27). This would produce a positive correlation between effective health deterioration and medical goods/services.

Health threshold: Measured across healthy and unhealthy individuals we expect to observe that healthy individuals will have rapid health deterioration ($\dot{H}(t) = -d(t)H(t)$) and low demand for medical care ($I(t) = 0$; they do not consume medical goods/services) while unhealthy individuals will be characterized by low effective health deterioration rates ($\dot{H}(t) = I(t) - d(t)H(t)$) and high demand for medical care ($I(t) > 0$). This would produce a negative correlation between effective health deterioration and the consumption of medical goods/services.

Empirical literature: The discussion under item 3 suggests that individuals may slow their effective health deterioration as they age and begin to consume medical care. Further research is needed to empirically test this prediction.

5. Medical care increases discontinuously when individuals become unhealthy

“Optimal” stock: In the standard solution of the Grossman model health evolves as the “optimal” health stock and individuals consume medical care continuously, i.e. there is no switching of dynamics and we do not expect to see discontinuous changes in medical care.

Health threshold: Healthy people ($H(t) > H_*(t)$) do not consume medical care. When, as

²²At these high ages SRH may suffer from selection effects. Unhealthy individuals may have higher mortality and drop out of the sample in higher numbers than healthy individuals. Further, SRH status suffers from framing bias, that is, individuals compare their health with a reference of what constitutes good health in their respective age group. In other words, they may be answering the question “Considering my age I am in good/bad health” instead of “I am in good/bad health”. Both effects would either reduce the significance of the observed flattening of SRH or could provide an alternative explanation for the observation.

a result of health deterioration their health reaches the health threshold $H(t) = H_*(t)$ (i.e., they have become unhealthy by our definition) they begin to consume medical care.

Empirical literature: The literature has, as far as we know, not tested this prediction before. The empirical test is described in Section 3. Some moderate support for the notion that the dynamics of healthy and unhealthy individuals are significantly different comes from the following observation. Grossman noted in his original work (Grossman 1972a; chapter V, p. 56) that over two thirds of the NORC sample he used in empirical testing of his model, reported no sick days. He notes that “... *Since the characteristics of these two groups [reporting sick days and no sick days] are very similar, it is difficult to explain the behavior of the [group that had no sick days]. Put differently, the two groups essentially represent “two different samples,” and problems arise when the data are pooled ...*”²³

6. Blue collar workers let their health deteriorate faster and to lower levels than white collar workers

“Optimal” stock: Blue and white collar workers²⁴ consume medical care at all times. Blue collar workers (see equation 26) have lower levels of health, assuming lower wages $w(t)$, lower levels of education E , higher “natural” deterioration rates $d(t)$ (i.e. higher values of d_0 , β_3 , and β_4 and assuming $\epsilon > 0$, $0 < k_I < 1$ and $\rho_I > 0$; see equation 25). The “traditional” solution of the Grossman model is unclear about the effective health deterioration rate $\dot{H}(t)$ for blue versus white collar workers.²⁵

Health threshold: In scenario A, initially while blue and white collar workers are healthy (health above the “threshold”), a blue collar worker’s health deteriorates faster than that of a white collar worker, assuming blue collar workers have higher health deterioration rates $d(t)$ as a result of physically demanding work and working environments that are more detrimental to health (see equation 30). The health of blue collar workers deteriorates to lower levels as their health threshold is lower (see discussion above under “*Optimal*” stock and equation 26). Once workers reach the health threshold it is unclear what the nature of differences (if any) is for the effective health deterioration rate $\dot{H}(t)$ for blue versus white

²³Strictly speaking we distinguish healthy from unhealthy individuals by whether they are above (healthy) or below (unhealthy) the health threshold and whether they do not (healthy) or do (unhealthy) consume medical care. But the number of sick days is assumed to be a function of health, and healthy individuals are expected to report relatively fewer sick days than unhealthy individuals.

²⁴Blue collar workers are broadly defined as individuals who generally have 1) lower levels of education, 2) lower wages, and 3) perform “hard” labor (e.g., construction). White collar workers on the other hand generally 1) are more educated, 2) earn higher wages, and 3) perform “light” jobs (e.g., office workers). As a result of “hard” labor and worse working environments blue collar workers are believed to be characterized by higher “natural” health deterioration rates $d(t)$ than white collar workers (e.g., Case and Deaton, 2005; Muurinen and Le Grand, 1985).

²⁵Assuming wages $w(t)$, medical prices $p_m(t)$ and environmental variables $\mathbf{X}(t)$ are relatively constant with age t we have $\dot{H}(t) \propto -\epsilon(\beta_3 + \beta_6)H(t)$. From this it is not immediately obvious that the effective health deterioration rate would be different for blue versus white collar workers, though β_3 (the exponential rate of decay of $d(t)$; see equation 25), may be higher for blue collar workers.

collar workers (see discussion above under “*Optimal*” stock).

Empirical literature: Case and Deaton (2005) investigate the rate of change in self reported health by occupation using data from the NHIS. The authors find that those who are employed in manual occupations have worse health than those who work in professional occupations and that the health effect of occupation operates at least in part independently of the personal characteristics of the workers. Cutler et al. (2008) present similar results using mortality as an indicator of health. Van Kippersluis et al. (2009) present similar results using the self reported health status of Dutch working males.

Further, as discussed earlier under item 3, the health of women and men in the bottom income quartile deteriorates much faster than that of the top income quartile. It is much harder to assess from the self-reported health measures presented in Smith (2004, 2007), Case and Deaton (2005) and Van Kippersluis et al. (2009) whether blue collar workers let their health deteriorate to lower levels of health, though generally speaking blue collar workers are found to be in worse health than white collar workers (e.g., Case and Deaton, 2005; Smith 1999, 2004, 2007, Van Kippersluis et al. 2008, 2009; as well as the evidence provided by the aforementioned studies that estimated the demand for health and found health to increase with, e.g., education, wages and to decrease with, e.g., physically demanding work). Similar patterns hold for other measures of socioeconomic status, such as education and wealth and other indicators of health, such as disability, and mortality (e.g., van Doorslaer et al. 2008).

7. The relationship between education and health is expected to be positive and differs for healthy and unhealthy individuals

“Optimal” stock: Health and education are positively related if the efficiency of medical care increases with education (equation 26 for $\rho_I > 0$). If health deterioration $d(t)$ decreases with education E , i.e., education is part of the vector $\mathbf{X}(t)$ of environmental variables that affect the deterioration rate, then the education component of β_4 ($\beta_{4,E}$) is negative and hence higher levels of education E through their affect on the deterioration rate $d(t)$ increase the level of health. The effect is similar to the presumed increased efficiency of medical care usage through education, $\rho_I > 0$, and both effects cannot be separated in the “traditional” solution of the Grossman model (the term in the structural form equation is $\ln H(t) \propto \epsilon(\rho_I - \beta_{4,E})E$). There is no difference between healthy and unhealthy individuals as in the “traditional” solution of the Grossman model this distinction is not made.

Health threshold: In scenario A, initially while individuals are healthy any relationship between health and education (see equation 30) works only through the effect (if any) of education on the rate of deterioration $d(t)$ and we have $\ln H(t) \propto -e^{\beta_{4,E}} (\beta_{4,E} < 0)$. When individuals have reached the health “threshold” both pathways (through the presumed increased efficiency of medical care usage and through any affect on the rate of deterioration $d(t)$) are relevant and we have the same relationship as for the “optimal” stock: $\ln H(t) \propto$

$$\epsilon(\rho_I - \beta_{4,E})E.$$

Empirical literature: A positive association between education and health has been established in the empirical literature (see, e.g., the evidence provided by the aforementioned studies that estimated the demand for health and found health to increase with education). To the best of our knowledge the literature has not yet made an attempt to test the interpretation of the Grossman model advocated here, i.e., to distinguish between healthy and unhealthy individuals and test differences in their respective relationships between health and education. The empirical test is described in Section 3.

5 Discussion

We have presented arguments for a generalized solution of the Grossman model (Grossman 1972a, 1972b). Our generalized solution of the Grossman model can deal with an important criticism of the model: that the model's prediction that health and medical care are positively related is consistently rejected by the data (e.g., Zweifel and Breyer 1997, p. 62). We find that this prediction is based on the widely used and unnecessary assumption that the health stock is always at Grossman's solution for "optimal" health. There is no theoretical basis for this assumption and empirical evidence suggests it is not valid. Removing this widely used restriction and allowing for the existence of corner solutions where individuals do not consume medical care, we find that the Grossman model predicts the existence of a health threshold.

We have contrasted the predictions of the generalized solution of the Grossman model advocated here with the empirical literature. Our generalized solution replicates the predictions of the traditional Grossman model (which have largely been verified in the empirical literature) with the exception of the problematic prediction that health and medical care should be positively correlated (which has been rejected in the empirical literature). As with the traditional solution of the Grossman model (a special case of our generalized solution) we broadly expect health to decrease with the cost of medical goods/services and with environmental factors that are detrimental to health (e.g., working conditions) and to increase with education. The effect of income is unclear as different sub models predict a different relation with health. With regard to the demand for medical care, we expect medical care to decrease with the cost of medical goods/services $p_m(t)$ and with education, and to increase with wages and with environmental factors that are detrimental to health.

In addition, our generalized solution of the Grossman model produces a number of predictions that are different from the traditional solution of the Grossman model. First, it replicates the observed negative relation between health and medical care as in our generalized solution of the Grossman model healthy individuals (whose health is above the health threshold) do not consume medical care while the unhealthy (at the threshold) do. Second, we find that individuals do not consume medical care at all times as healthy people do not consume medical care. Basically our generalized solution of the Grossman model predicts the intuitively natural behavior that

healthy individuals do not go to the doctor or stay in hospital while the unhealthy do (except for preventive care or as a result of a sudden health shock, both phenomena are currently not part of the Grossman model). Third, we find that effective health deterioration slows as individuals reach the health threshold and begin to consume medical care. Fourth, our generalized solution of the Grossman model predicts that the effective health deterioration rate $\dot{H}(t)$ (the net effect of “aging” and medical care) will be smaller for individuals who consume more medical care. Fifth, we predict that the consumption of medical care increases discontinuously as healthy individuals begin to consume medical care once their health reaches the health threshold. Sixth, our generalized solution of the Grossman model can account for the observation that blue collar workers tend to have faster rates of effective health deterioration $\dot{H}(t)$ than white collar workers (e.g., Case and Deaton 2005). Lastly, because the model distinguishes between healthy and unhealthy individuals who behave differently, the model allows for a number of tests that are not possible in the traditional interpretation of the Grossman model. For example, Muurinen (1982) has argued that education improves health through lowering the natural health deterioration rate $d(t)$ (aging) and not just (or perhaps not at all) through improving the efficiency of an individual’s consumption of medical care (Grossman 1972a, 1972b). Since the first pathway (lowering the deterioration rate) operates only for healthy individuals and both pathways operate for unhealthy individuals it should in theory be possible to establish empirically the relative importance of both pathways. Also, while the natural deterioration rate $d(t)$ is not directly observable in the traditional interpretation of the Grossman model, it is directly observable in our interpretation as individuals who are healthy let their health deteriorate at exactly this rate (assuming good empirical measures of health status are available).

A review of the empirical literature suggests that our generalized solution of the Grossman model can account for a greater number of observations than can the traditional solution. Ultimately though, the model needs to be verified in direct empirical testing. To this end we have provided detailed structural and reduced form equations for the pure consumption and pure investment models for both the healthy and unhealthy phases of life. Empirically, the proposed model is a switching regression model with endogenous switching. Once health hits the health threshold, the process governing health and medical care switches.

The corner solutions presented in this work contribute to better describing the behavior of individuals whose health is above the threshold level for parts of the life cycle (the *healthy* and the *very healthy*). However, for those individuals whose health is at the threshold over the life cycle (the *ill*) we have simply adopted the assumption commonly made in the Grossman literature that individuals are able to adjust their health to a desirable level. This assumption may be less severe though in the case of the *ill*. It is, for example, not necessary to assume that the adjustment is instantaneous as individuals will have had ample time to consume medical care before they enter the labor force. There is also naturally an adjustment cost associated with these investments in the sense that such individuals begin their work life with fewer assets as a result of the purchase of medical care in the market before they entered the labor force.

Natural extensions of the model would be to include uncertainty and health shocks (e.g., to

address the criticism by Cropper 1977, Dardanoni and Wagstaff 1987), to revisit the assumption of complete health repair (e.g., the criticism by Case and Deaton 2005), to revisit the unrealistic so-called “bang-bang” solutions that the model produces when an individual’s health is initially below the threshold (the *ill*; the criticism by Ehrlich and Chuma, 1990), to include length of life as a decision variable (endogenous T ; e.g., Ehrlich and Chuma, 1990), to include healthy and unhealthy behaviors such as unhealthy consumption (e.g., smoking), healthy consumption (e.g., dieting; see Case and Deaton, 2005) and preventive care, and to explore the solutions in which the decision to perform “hard” labor is endogenous (see, e.g., Case and Deaton, 2005). Following Cropper (1981) and Wagstaff (1986) we have assumed that the natural deterioration rate $d(t)$ is exogenously determined by environmental factors such as, e.g., working conditions, hazardous environment, etc. The model thus assumes that blue collar workers have no choice but to perform hard labor and face worse living, working and schooling environments. But, as Case and Deaton (2005) argue, individuals may accept risky and unhealthy work environments, in exchange for higher pay.

References

- Case, A., and A. Deaton (2005), "Broken Down by Work and Sex: How our Health Declines", in: David A. Wise (ed.) (2005), *Analyses in the Economics of Aging*, The University of Chicago Press, Chicago, pp. 185-212.
- Cochrane, A.L., A.S St Leger, and F. Moore (1978), "Health service 'input' and mortality 'output' in developed countries", *J. Epid. Communit. Hlth* 32, pp. 200-205
- Cropper, M. L. (1977), "Health, Investment in Health, and Occupational Choice", *The Journal of Political Economy*, Vol. 85, No. 6, pp. 1273-1294.
- Cropper, M. L. (1981), "Measuring the Benefits from Reduced Morbidity", *The American Economic Review*, Vol. 71, No. 2, Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association., pp. 235-240.
- Ehrlich, I. and Chuma, H. (1990), "A Model of the Demand for Longevity and the Value of Life Extension", *The Journal of Political Economy*, Vol. 98, No. 4, pp. 761-782.
- Erbsland, M., Ried, W. and Ulrich, V. (2002), "Health, Health Care, and the Environment: Econometric Evidence from German Micro Data", in *Econometric Analysis of Health Data*, John Wiley & Sons, Ltd, pp. 25-36.
- Finkelstein, A., Luttmer, E.F.P. and Notowidigdo, M.J. (2008), "What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption", NBER Working Paper 14089.
- Galama, T.J., Kapteyn, A., Fonseca, F., Michaud, P.C. (2009), "Grossman's Health Threshold and Retirement.", RAND Working Paper, WR-658.
- Gerdtham, U.G., Johannesson, M., Lundberg, L., Isacson D. (1999), "The demand for health: results from new measures of health capital", *European Journal of Political Economy*, Volume 15, Number 3, pp. 501-521.
- Gerdtham, U.G., and Johannesson, M. (1999), "New estimates of the demand for health : results based on a categorical health measure and Swedish micro data", *Social science & medicine*, Vol. 49, No. 10, pp. 1325-1332.
- Goldman, D.P. and J.P. Smith (2002), "Can Patient Self-Management Help Explain the SES Health Gradient?", *PNAS* 99, pp. 10929-10934.

- Grossman, M. (1972a), "The Demand for Health-A theoretical and Empirical Investigation", New York: National Bureau of Economic Research.
- Grossman, M. (1972b), "On the Concept of Health Capital and the Demand for Health", *The Journal of Political Economy*, Vol. 80, No. 2, pp. 223-255.
- Grossman, M. (2000), "The Human Capital Model", in *Handbook of Health Economics*, Volume 1, pp. 347-408, A.J. Culyer and J.P. Newhouse (Eds.), Elsevier Science.
- Leu, R.E. and Doppman, R.J. (1986), "Gesundheitszustandsmessung und Nachfrage nach Gesundheitsleistungen", in: E. Wille (Hrsg.) *Informations- und Planungsprobleme in öffentlichen Aufgabenbereichen*, Frankfurt am Main/Bern/New York: Lang 1986, pp. 1-90.
- Leu, R.E. and Gerfin, M. (1992), "Die Nachfrage nach Gesundheit - Ein Empirischer Test des Grossman-Modells" (Demand for health - an empirical test of the Grossman model), in: Oberender, P. (ed.), *Steuerungsprobleme im Gesundheitswesen*, Baden-Baden: Nomos, pp. 61-78.
- Mincer, J.A. (1974), "Schooling, Experience, and Earnings", Columbia University Press.
- Muurinen, Jaana-Marja (1982), "Demand for health: a generalized Grossman model", *Journal of Health Economics* 1, pp. 5-28.
- Muurinen, Jaana-Marja, and Le Grand, Julian (1985), "The Economic Analysis of Inequalities in Health", *Soc. Sci. Med*, Vol. 20, No. 10, pp. 1029-1035.
- Seierstad, A. and K. Sydsaeter (1987), "Optimal Control Theory With Economic Applications", in *Advanced Textbooks in Economics*, Volume 24, Series Editors: C.J. Bliss and M.D. Intriligator, Elsevier, North Holland.
- Smith, J.P. (1999), "Healthy bodies and thick wallets", *Journal of Economic Perspectives* 13(2): pp. 145-166.
- Smith, J. (2004), "Unraveling the SES-health connection", in *Aging, Health, and Public Policy: Demographic and Economic Perspectives*, a supplement to *Population and Development Review* 30: pp. 108-132.
- Smith, J.P. (2007), "The impact of socioeconomic status on health over the life course", *Journal of Human Resources* 42(4): pp. 739-764.

Ried, Walter (1998), "Comparative dynamic analysis of the full Grossman model", *Journal of Health Economics* 17, pp. 383-425.

Van Doorslaer, E.K. (1987), "Health, knowledge and the demand for medical care", Assen, 171 pp., ISBN 90-232-2335-7.

Van Doorslaer, E.K., Van Kippersluis, H., O'Donnell, O., Van Ourti, T. (2008), "Socioeconomic Differences in Health over the Life Cycle: Evidence and Explanations.", Netspar Panel paper prepared for the Netspar Panel Day 16 October, 2008, Tilburg

Van Kippersluis, H., Van Ourti, T., O'Donnell, O. and Van Doorslaer, E. (2008) "Health and income across the life cycle and generations in Europe", Tinbergen Institute discussion paper series 08-009/3, Erasmus University Rotterdam.

Van Kippersluis, H., O. O'Donnell, E. van Doorslaer and Van Ourti, T., (2009) "Socioeconomic Differences in Health over the Life Cycle in an Egalitarian Country.", Tinbergen Institute discussion paper series 09-006/3, Erasmus University Rotterdam.

Van de Ven, W. and van der Gaag, J. (1982), "Health as an unobservable: A MIMIC-model of the demand for health care.", *Journal of Health Economics*, Vol. 1, pp. 157-183.

Wagstaff, A. (1986), "The demand for health: Some new empirical evidence. ", *Journal of Health Economics*, Vol. 5, pp. 195-233.

Wagstaff, A. (1993), "The demand for health: An empirical reformulation of the Grossman model", *Health Economics*, Vol. 2, pp. 189-198.

Wolfe, John R. (1985), "A Model of Declining Health and Retirement", *The Journal of Political Economy*, Vol. 93, No. 6, pp. 1258-1267.

Zweifel, P., Breyer, F., (1997), *Health Economics*, Oxford University Press, New York.

A Appendix: derivations

A.1 First-order conditions

Associated with the Lagrangian (equation 10) we have the following conditions:

$$\begin{aligned}
 \dot{q}_A(t) &= -\partial\mathfrak{L}(t)/\partial A(t) \Rightarrow \\
 \dot{q}_A(t) &= -\delta q_A(t) \Leftrightarrow \\
 q_A(t) &= q_A(0)e^{-\delta t},
 \end{aligned} \tag{43}$$

$$\begin{aligned}
 \dot{q}_H(t) &= -\partial\mathfrak{L}(t)/\partial H(t) \Rightarrow \\
 \dot{q}_H(t) &= q_H(t)d(t) - \frac{\partial U(t)}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)} e^{-\beta t} \\
 &\quad - q_A(0)e^{-\delta t} \frac{\partial Y[H(t)]}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)},
 \end{aligned} \tag{44}$$

$$\begin{aligned}
 \partial\mathfrak{L}(t)/\partial X(t) &= 0 \Rightarrow \\
 \partial U(t)/\partial C(t) &= q_A(0) \left(\frac{p_X(t)}{\partial C(t)/\partial X(t)} \right) e^{(\beta-\delta)t} \\
 &\equiv q_A(0)\pi_C e^{(\beta-\delta)t},
 \end{aligned} \tag{45}$$

$$\begin{aligned}
 \partial\mathfrak{L}(t)/\partial\tau_C(t) &= 0 \Rightarrow \\
 \partial U(t)/\partial C(t) &= q_A(0) \left(\frac{w(t)}{\partial C(t)/\partial\tau_C(t)} \right) e^{(\beta-\delta)t} \\
 &\equiv q_A(0)\pi_C e^{(\beta-\delta)t},
 \end{aligned} \tag{46}$$

$$\begin{aligned}
 \partial\mathfrak{L}(t)/\partial m(t) &= 0 \Rightarrow \\
 q_H(t) + q_I(t) &= q_A(0) \left(\frac{p_m(t)}{\partial I(t)/\partial m(t)} \right) e^{-\delta t} \\
 &\equiv q_A(0)\pi_I e^{-\delta t},
 \end{aligned} \tag{47}$$

$$\begin{aligned}
 \partial\mathfrak{L}(t)/\partial\tau_I(t) &= 0 \Rightarrow \\
 q_H(t) + q_I(t) &= q_A(0) \left(\frac{w(t)}{\partial I(t)/\partial\tau_I(t)} \right) e^{-\delta t} \\
 &\equiv q_A(0)\pi_I e^{-\delta t}.
 \end{aligned} \tag{48}$$

Equation (45) provides the first-order condition for maximization of (1) with respect to consumption, subject to the conditions (2) and (3). Using (47) to obtain an expression for

$\dot{q}_H(t)$ and substituting the results for $q_H(t)$ and $\dot{q}_H(t)$ in (44) we find the first-order condition for maximization of (1) with respect to health, subject to the conditions (2) and (3). The resulting first-order conditions are provided by equations (11) and (13) in section 2.

A.2 Structural and reduced form: pure investment model

We begin with the first-order condition for optimal health (18). We have (using equations 22 through 24)

$$\pi_I(t) = \frac{\partial Y(t)}{\partial s(t)} \frac{\partial s(t)}{\partial H(t)} [d(t) + \delta - \tilde{\pi}_I(t)]^{-1} \quad (49)$$

$$= \beta_1 \beta_2 w(t) H(t)^{-(\beta_2+1)} [d(t) + \delta - \tilde{\pi}_I(t)]^{-1} \quad (50)$$

$$= \frac{P_m(t)}{\partial I(t)/\partial m(t)} = \frac{e^{-\rho_I E}}{\mu_I(t) k_I^{k_I} (1 - k_I)^{(1-k_I)}} w(t)^{k_I} p_m(t)^{(1-k_I)}. \quad (51)$$

This leads to the structural form equation (26).

Now consider the equations for medical health investment (equations 2 and 23) and using (24),

$$\ln I(t) = \rho_I E + (1 - k_I) \ln m(t) + k_I \ln \tau_I(t) + \ln \mu_I(t) \quad (52)$$

$$= \rho_I E + \ln m(t) + k_I \ln p_m(t) - k_I \ln w(t) + \ln \mu_I(t) + k_I \ln [k_I / (1 - k_I)] \quad (53)$$

$$= \ln [\dot{H}(t) + d(t) H(t)] \quad (54)$$

$$= \ln d(t) + \ln H(t) + \ln [1 + \tilde{H}(t)/d(t)]. \quad (55)$$

This leads to the structural form equation (27).

Using (26) and (27) we find

$$\begin{aligned} \ln m(t) &= \beta_8 + [k_I + \epsilon(1 - k_I)] \ln w(t) - [k_I + \epsilon(1 - k_I)] \ln p_m(t) \\ &- (1 - \epsilon) \rho_I E + (1 - \epsilon) \ln d_\bullet + (1 - \epsilon)(\beta_3 + \beta_6)t + (1 - \epsilon)\beta_4 \mathbf{X}(t) \\ &- \epsilon \ln \{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\} \\ &+ \ln [1 + \tilde{H}(t)/d(t)], \end{aligned} \quad (56)$$

where $\beta_8 \equiv \beta_5 + \beta_7$.

Combining equations (53) and (54) we find:

$$\dot{H}(t) + d(t) H(t) = \mu_I(t) [k_I / (1 - k_I)]^{k_I} m(t) p_m(t)^{k_I} w(t)^{-k_I} e^{\rho_I E}, \quad (57)$$

the solution of which is

$$H(t) = e^{\rho_I E} [k_I / (1 - k_I)]^{k_I} \int_0^t \mu_I(x) m(x) p_m(x)^{k_I} w(x)^{-k_I} e^{-\int_x^t d(s) ds} dx. \quad (58)$$

We then have

$$1 + \tilde{H}(t)/d(t) = \frac{\mu_I(t)m(t)p_m(t)^{k_I}w(t)^{-k_I}}{d(t) \int_0^t \mu_I(x)m(x)p_m(x)^{k_I}w(x)^{-k_I}e^{-\int_x^t d(s)ds} dx}. \quad (59)$$

Substituting equation (59) into equation (56) and differentiating the result with respect to time t we find the reduced form expression (28).

While the literature largely focuses on the relations for health $H(t)$ and medical goods/services $m(t)$ the model does allow for the derivation of relations for consumption goods $X(t)$ and assets $A(t)$. In the pure investment model we have $\partial U(t)/\partial H(t) = 0$, i.e. utility $U(t)$ is independent of health $H(t)$. We assume a simple functional form for the utility function:

$$U[C(t)] = \frac{C(t)^{1-\rho}}{1-\rho}. \quad (60)$$

The first-order condition (equation 11) then leads to:

$$C(t)^{-\rho} = q_A(0)\pi_C(t)e^{(\beta-\delta)t}. \quad (61)$$

Grossman (1972a, 1972b, 2000) assumes that medical health investment is produced by combining time and medical goods/service according to a Cobb-Douglass constant returns to scale production function (see equation 23). A similar assumption can be made that consumption is produced by combining time τ_C and consumption goods $X(t)$ as follows:

$$C(t) = \mu_C(t)X(t)^{1-k_C}\tau_C(t)^{k_C}e^{\rho_C E}, \quad (62)$$

where $\mu_C(t)$ is an efficiency factor, $1 - k_C$ is the elasticity of consumption $C(t)$ with respect to consumption goods $X(t)$, k_C is the elasticity of consumption $C(t)$ with respect to time input $\tau_C(t)$, and ρ_C determines the extent to which education E improves the efficiency of consumption $C(t)$.

Further the ratio of the marginal product of medical care with respect to medical goods/services $\partial I(t)/\partial m(t)$ and the marginal product of medical care with respect to own-time investment $\partial I(t)/\partial \tau_I(t)$ equals the ratio of the price of medical goods/services $p_m(t)$ to the wage rate $w(t)$ (representing the opportunity cost of time; see equation 24). Similarly, the ratio of the marginal product of consumption with respect to consumption goods $\partial C(t)/\partial X(t)$ and the marginal product of consumption with respect to time inputs $\partial C(t)/\partial \tau_C(t)$ equals the ratio of the price of consumption good $p_X(t)$ to the wage rate $w(t)$ (see equation 12). We then have

$$\pi_I(t) = \frac{p_m(t)}{\partial I(t)/\partial m(t)} = \frac{p_m(t)^{1-k_I}w(t)^{k_I}e^{-\rho_I E}}{\mu_I(t)k_I^{k_I}(1-k_I)^{(1-k_I)}}, \quad (63)$$

$$\pi_C(t) = \frac{p_X(t)}{\partial C(t)/\partial X(t)} = \frac{p_X(t)^{1-k_C}w(t)^{k_C}e^{-\rho_C E}}{\mu_C(t)k_C^{k_C}(1-k_C)^{(1-k_C)}}. \quad (64)$$

Assuming the Cobb-Douglas constant returns to scale production function for medical health investment (equation 23) and for consumption (equation 62) we obtain the following expressions for consumption goods $X(t)$ and medical goods/services $m(t)$

$$X(t) = (1 - k_C) \frac{\pi_C(t)}{p_X(t)} C(t), \quad (65)$$

$$m(t) = (1 - k_I) \frac{\pi_I(t)}{p_m(t)} [\dot{H}(t) - d(t)H(t)]. \quad (66)$$

Using equations (61, 64, and 65) we find

$$\begin{aligned} \ln X(t) &= \beta_{13} - [k_C + (1 - k_C)/\rho] \ln p_X(t) + k_C[(\rho - 1)/\rho] \ln w(t) \\ &\quad - \rho_C[(\rho - 1)/\rho] E - [(\beta - \delta)/\rho] t + \ln q_A(0)^{-1/\rho}, \end{aligned} \quad (67)$$

where $\beta_{13} \equiv \ln(1 - k_C) - [(\rho - 1)/\rho] \ln[k_C^{k_C} (1 - k_C)^{(1 - k_C)}] - [(\rho - 1)/\rho] \ln \mu_C(t)$.

It is straightforward though tedious to derive an expression for the shadow price of life-time wealth $q_A(0)$, using the life-time budget constraint (5), the expression for sick time $s[H(t)]$ (equation 22), income $Y[H(t)]$ (equation 9), consumption good $X(t)$ (the above equation), health $H(t)$ (equation 26), and medical goods/services $m(t)$ (equation 27). $q_A(0)$ is then found to be a complicated function of life-time wealth (assets, life-time income), wages $w(t)$, prices $p_m(t)$, $p_X(t)$, education E and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4 \mathbf{X}(t)$). The expression itself is not very insightful and is hence not reproduced here.

A.3 Structural and reduced form: pure consumption model

Using the utility specification (35), the first-order conditions (11) and (13), and equation (34) we find

$$\begin{aligned} \frac{\partial U[C(t), H(t)]}{\partial C(t)} &= \zeta C(t)^{\zeta - \rho\zeta - 1} H(t)^{1 - \zeta - \rho + \rho\zeta} \\ &= q_A(0) \pi_C(t) e^{(\beta - \delta)t} \end{aligned} \quad (68)$$

$$\begin{aligned} \frac{\partial U[C(t), H(t)]}{\partial H(t)} &= (1 - \zeta) C(t)^{\zeta - \rho\zeta} H(t)^{-\zeta - \rho + \rho\zeta} \\ &= q_A(0) [\pi_H(t) - \varphi_H(t)] e^{(\beta - \delta)t} + [\dot{q}_I(t) - q_I(t)d(t)] e^{\beta t} \\ &= \pi_C(t)^{-1} [\pi_H(t) - \varphi_H(t)] \frac{\partial U[C(t), H(t)]}{\partial C(t)} + [\dot{q}_I(t) - q_I(t)d(t)] e^{\beta t} \end{aligned} \quad (69)$$

The solution for the health threshold (Grossman's solution for "optimal" health) follows from combining equation (68) with (69), assuming $\varphi_H(t) = 0$ (pure consumption) and using $q_I(t) = \dot{q}_I(t) = 0$. We then find:

$$\begin{aligned} \ln H(t) &= \ln C(t) + \ln \left(\frac{1 - \zeta}{\zeta} \right) + \ln \pi_C - \ln \pi_I - \ln d(t) \\ &\quad - \ln[1 + \delta/d(t) - \tilde{\pi}_I(t)/d(t)]. \end{aligned} \quad (70)$$

Combining equations (25, 63, 64, 65 and the above expression) leads to the structural form equation (36). Further, combining equations (65, 68, 69 and 70) we find:

$$H(t) = q_A(0)^{-1/\rho} \zeta^{1/\rho} \left(\frac{1-\zeta}{\zeta} \right)^\chi \pi_C(t)^{-\chi(1/\rho\chi-1)} \pi_I(t)^{-\chi} d(t)^{-\chi} [1 + \delta/d(t) - \tilde{\pi}_I(t)/d(t)]^{-\chi} \quad (71)$$

which leads to the structural form equation (39).

As in the pure investment model one can find an expression for the shadow price of life-time wealth $q_A(0)$ for the pure consumption model, using the life-time budget constraint (5), the expression for income $Y[H(t)]$ (equation 9), consumption good $X(t)$ (equation 65), health $H(t)$ (equation 39), and medical goods/services $m(t)$ (equation 37). As in the pure investment model the expression is found to be a complicated function of life-time wealth (assets, life-time income), wages $w(t)$, prices $p_m(t)$, $p_X(t)$, education E and the health deterioration rate (terms d_\bullet , β_3 and $\beta_4 \mathbf{X}(t)$).

Combining equation (36) with (37) we find:

$$\begin{aligned} \ln m(t) &= \beta_{12} + \ln X(t) + \ln p_X(t) - \ln p_m(t) + \ln[1 + \tilde{H}(t)/d(t)] \\ &- \ln\{1 + d_\bullet^{-1} e^{-\beta_3 t - \beta_4 \mathbf{X}(t)} [\delta - k_I \tilde{w}(t) - (1 - k_I) \tilde{p}_m(t) - \beta_6]\}, \end{aligned} \quad (72)$$

where $\beta_{12} \equiv \beta_7 + \beta_9$.

Substituting equation (59) into equation (72) and differentiating the result with respect to time t we find the reduced form expression (38).