

The Grossman model after 40 years

Peter Zweifel

Published online: 6 September 2012
© Springer-Verlag 2012

Abstract This editorial presents a critical review of the health model pioneered by Michael Grossman (MGM) in 1972 [8]. It argues that whereas the MGM has great charm for economists, it fails to achieve acceptance by interested laypersons and policy makers. The main reasons for this failure are: (1) the assumption of a long and fixed planning horizon, (2) a fixed ratio between individuals healthcare expenditure and the cost of their own health-enhancing efforts regardless of their state of health, and (3) their presumed ability to restore the state of health deemed optimal at a speed that does not depend on their state of health. An alternative formulation emphasizing the stochastic nature of health production is sketched that conceptually provides solutions to these three problems. In addition, it permits discarding a popular medical argument that seems to undermine the very basis of welfare analysis applied to health by claiming preferences to be unstable: “As long as you are healthy, you don’t give a damn, but as soon as you are sick, you are prepared to sacrifice everything to restore your health.” The editorial concludes by outlining a research program that may help health economists break away from their MGM fixation.

Review of the Grossman model

When it was published in 1972 in the *Journal of Political Economy*, Michael Grossmans model (MGM) constituted a major breakthrough in health economics [8]. However, the acronym MGM already suggests that the model amounts to something like the Hollywood dream factory

Metro-Goldwyn-Mayer: much elegance, very inspiring, but of limited relevance to the real world. It also found a Shake-spearian herald in Adam Wag-staff, who formulated the MGM as an optimal problem controller and complemented it with simple functional relationships for empirical testing [17]. The original MGM hypothesis has remained the same since its beginning. Individuals simultaneously optimize two assets over their life cycle: health and wealth. The return to health is nonfinancial, in the guise of healthy days; it contributes to wealth because additional healthy time can be used to generate more labor income. Holding a stock of health entails the usual capital user cost consisting of depreciation (increasing with age to reflect worsening health), cost of investment, and change in value (e.g., a housing asset whose value increases in times of inflation). The return to wealth is financial, amounting to a rate of interest. The dynamically optimal path of the two assets calls for a certain amount of investment both in health (health-enhancing efforts and medical care, respectively) and wealth (saving), both being derived demands, as known in production theory. However, as evidenced in Wagstaff [17], it is a long way from the basic optimum condition to an empirically testable formulation.

A first review of the MGM appeared in the inaugural issue of the *Journal of Health Economics* [15]. Grossman provided a comprehensive 30-year review in the *Handbook of Health Economics* [9]. In passing, it may be noted that he devoted more than one page to a somewhat angry rebuff of the criticisms raised in the textbook by Zweifel and Breyer, Chap. 3.3 [19] (ZB henceforth); see also Zweifel, Breyer, and Kifmann, Chap. 3 [20] (ZBK henceforth). It will come as little surprise that these critical points will be reiterated below; however, the MGM will be done justice by first reviewing its theoretical modifications and empirical applications.

P. Zweifel (✉)
Zurich, Switzerland
e-mail: peter.zweifel@econ.uzh.ch

One refinement has been the introduction of uncertainty. Chang [4] had already recognized that portfolio optimization methods could be applied to health as one of several assets, with risk aversion entering the picture. As risk-averse individuals typically buy insurance, Liljas [13] extended MGM to take into account health insurance. As could be expected, the main implications of MGM remained unchanged, in particular, the prediction of a positive relationship between the (permanent, desired) stock of health and the derived demand for health care. An important generalization has been provided by researchers from Lund (Sweden), who by now can be considered the true MGM believers in Europe. A shortcoming of previous formulations is that the individual had been considered in isolation, whereas decision making with regard to health [and the third asset in question, wisdom (education), one might note] often occurs within the family. Jacobson [10] introduced the corresponding combined production possibility frontier, whereas Bolin et al. [2] explored the possibility of strategic interaction between spouses with the possibility of one free-riding on the other.

In his empirical test of MGM, Wagstaff [17] noted that health status was negatively correlated with the demand for medical care. This constitutes a fundamental contradiction to the basic MGM hypothesis, which states that health care is a derived demand with regard to the desired stock of health. Yet the layperson's experience is that you see the doctor when you are sick, implying a negative relationship between health and health care. However, health stock as usually measured may not reflect the (permanent, desired) stock of health as defined in the MGM. Arguably, the most comprehensive attempt to distil this latent health status from a cross section of observable indicators was undertaken by Leu and Doppmann [12] (cited in ZB, Chap. 4.4.1 and ZBK, Chap. 4.4) using Swiss data. However, the authors did not pursue their project any further, causing them to be little cited. Their approach was to distinguish between latent variables (permanent health, permanent income) and indicators that are all subject to measurement errors [Linear Structural Relations (LISREL), also known as Multiple Indicator Multiple Indicator Cause (MIMIC)]. Despite all these efforts, medical inputs (visits to physicians, hospitals, and spas) continued to be negatively related to latent health, flying in the face of the "derived-demand" hypothesis of the MGM, according to which (permanent, i.e. long-run optimal) health stock should be positively related to inputs (in particular, medical care). Wagstaff [18] sought to shore up the MGM by introducing partial adjustment as it is known from macroeconomic investment functions: investments of a given period bridge only part of the gap between desired and actual capital stock, with the speed of adjustment constituting a decision variable in principle. Wagstaff [18], indeed, finds that

Danes aged >41 years have a lower speed of adjustment than younger ones. However, the real issue is that this speed depends crucially on health status; when the gap between desired and actual health status is wide, failure to cover it quickly entails high opportunity cost (just think of being rushed to the nearest hospital with sirens on).

Moreover, in Wagstaff [18], the LISREL estimates in fact contradict the MGM. Whereas outpatient visits (OUTPAT) serve as the benchmark indicator of investment in the previous period I_{t-1} , with a coefficient of +1.00, other indicators of I_{t-1} , such as general practitioner (GP) visits and hospital days (sic!) have significantly negative coefficients among those <41 years. Most crucially, according to the partial adjustment model of investment, I_{t-1} should be positively related to inherited health stock H_{t-1} , at least among the elderly, whose speed of adjustment likely is lower than the rate of depreciation on H_{t-1} (see Eq. 10 in [18]). However, consider the two findings below for the older part of the sample (which hold true with minimal changes for the younger as well),

1. H_t is positively related to H_{t-1} with coeff. = 0.849 ($t = 16$);
2. H_{t-1} is negatively related to I_{t-1} (indicated mainly by OUTPAT) with coeff. = -0.253 ($t = -4.7$).

Logically, this implies

3. H_t is negatively related to I_{t-1} , and in view of (1), H_{t-1} is negatively correlated with I_{t-1} as well, i.e., better health goes along with less, not more, investment and hence medical care.

Therefore, introducing lags and partial investment fails to remedy the crucial shortcoming of MGM, i.e., the negative correlation between health status and medical care. It is amazing that neither the reviewers nor the editors of *Health Economics* recognized this. Or was there collusion between the journal and the author, serving their shared interest in keeping the MGM bandwagon rolling? *Honni soit qui mal y pense...*

This author pleads guilty of joining the MGM bandwagon, too (thank heaven in a little-cited piece [16]!). The panel data analyzed consisted of Swiss health insurance records complemented by subjective health status and socioeconomic characteristics measured in 1989 and 1992 (sample I, $N = 477$) and in 1981 and 1992 (sample II, $N = 212$). In sample I, Tobit estimates indicate that health care expenditure (HCE) if at all decreases with subjective health (coeff. = -0.206, s.e. = 0.518) when age is controlled for, whereas in the very long-run sample II, it does increase (coeff. = 0.933, s.e. = 0.413). Three educational indicators (vocational, college, university), while having the predicted negative sign in four out of six cases (reflecting increased individual productivity permitting less

reliance on medical care) consistently fall (far) short of statistical significance. Despite these contradictions, the conclusion reads: “the use of panel data ... combined with accounting for the distributional peculiarities characterizing the demand for medical care ... permit a much more *successful* (italics added) testing of the Grossman model than was hitherto possible”.

In sum, even after 40 years of effort, the main criticisms of the MGM still stand:

1. A long planning horizon of fixed length: Dynamic optimization makes sense only if it extends over several periods. Long planning horizons undoubtedly apply when it comes to assessing wealth and wisdom (skills, augmented by education). In the case of health, the MGM view is close to untenable. Just about everyone has experienced a spell of very bad health, causing his or her planning horizon to shrink to days or even minutes (when the rescue vehicle after an accident rushes the injured to the nearest hospital). The MGM denies the inherently stochastic nature of health as the outcome of a production process. And, indeed, there is incidental evidence supporting this criticism. Although Benitez-Silva and Ni [1] pay lip service to MGM (as do many), they find that among 38 % of their respondents, variations in expected longevity (derived from the longitudinal US Health and Retirement Study, presumably an excellent indicator of changes in permanently desired health stock) fail to accord with changes in self-reported health status. But of course, the authors do not point to the contradiction with MGM but conclude, “These potentially serious problems raise doubts regarding the use and interpretation of the computed health changes Our empirical results suggest that self-reported health changes are a preferred measure of health dynamics.” Yet, this advice had already been followed by Nocera and Zweifel [16] but with limited success (see above).
2. A fixed ratio between individuals HCE and the cost of their own health-enhancing efforts in the production of health: Admittedly, this is not part of the core dynamic optimum condition. However, in their attempt at deriving testable implications, authors have used the Cobb-Douglas production function, following the lead of Wagstaff [17]. This type of production function has been discarded in general production theory mainly because it imposes a unitary elasticity of substitution. Given optimization, this implies that the optimal ratio of expenditure on two inputs is equal to the ratio of their exponents in the Cobb-Douglas function ([14], Chap. 3.D). In the present context, individuals are predicted to incur HCE and bear the cost of their own health-enhancing efforts in a fixed proportion because

the exponents of the Cobb-Douglas function are exogenously given. However, only economists are willing to accept the implication stating that bad health leaves HCE relative to the cost of one's own effort unaffected. In the advent of sickness, any layperson would expect medical care to become the crucial input, pretty much replacing one's own efforts. Galama and Kapteyn [6] address this issue at least indirectly. They introduce a threshold value beyond which extra medical care is counterproductive, causing relative productivities to become state dependent. Even Bolin and Lindgren [3] recently left the path of true MGM believers by admitting deviations from the optimal level of the individual's preventive efforts that go along with worsened health. They study the “implications for equilibrium and stability”—issues that do not sit well with the MGM.

3. Ability to restore the state of health deemed optimal under all circumstances: Taking this literally, one would have to ask why individuals die at all—unless, of course, they suddenly view a zero stock of health optimal. Attempts at suicide reflect such a view. Now Grossman et al. [7] (Grossman the pediatrician this time), examining suicidal attempts among young Navajo Indians, find that a family history of such attempts is associated with an odds ratio of <2.3 ; i.e., it makes attempted suicide an estimated 2.3-fold more likely *ceteris paribus*. Apparently, strategic interactions within the family as the producer of health—the MGM extension by Bolin et al. [2]—often result in a zero-health Nash equilibrium. On the other hand, Corman and Grossman [5] (Grossman the economist again this time) investigated cases in which strategic interactions (at least among accountable individuals) are not possible. They found that the availability of abortion serves to decrease neonatal mortality rates in the United States. This can be interpreted as parents reducing the health stock of unwanted children to zero within a few months (so much for immediate adjustment to the desired stock of health!). It may also be of interest to note that the MGM is not alluded to in that article at all.

An alternative formulation: stochastics and state dependence

Already in ZB, Chap. 3.4 (but see also ZBK, Chap. 3.4) an alternative to the MGM was proposed. Here, individuals are much less the masters of their fate than in the MGM world. Being inserted in a stochastic process consisting of a string of healthy and sick days, all they have is a (small) influence on the probability of being sick (π) the following

day. If currently healthy, they can exert preventive effort costing their time to increase the probability of being healthy tomorrow $(1-\pi)$. If currently sick, they have to rely on medical care to achieve this. Therefore, both their consumption possibilities (C) and $(1-\pi)$ become state dependent. Individuals are assumed to value consumption (which requires goods and time as inputs) and chances of being healthy tomorrow, whereas medical care does not have utility itself.

In the longer term (weeks, up to a year, say), the binary distribution determines expected waiting time until there is a change from healthy to sick. This waiting time is given by $1/\pi$. Therefore, given decreasing marginal returns, there is a critical value for time spent on prevention beyond which π does not decrease sufficiently anymore to generate net healthy time. One therefore obtains the well-known bell-shaped boundary of production possibilities shown in Fig. 1. It is representative of the overall population because the majority of people currently are in good health (notwithstanding the medical adage: “a healthy person is one that has not been diagnosed enough”).

Juxtaposing this frontier with indifference curves (steeply sloped to reflect the importance of health) already has an important implication. At least in countries in which citizens are not rationed in their access to healthcare services, health is never an investment good but always a consumption good. The optimum of Fig. 1 cannot lie on the increasing portion of the frontier, where more investment in health also permits to increase consumption. Rather, it necessarily lies beyond the peak, indicating a trade-off between health and consumption. This insight also casts doubt on the relevance of the popular pure investment variant of the MGM.

State dependence can come in on the preference and/or the constraint side. Economists usually cling to the

constraint side, arguing that state-dependent preferences can be used to explain practically anything, which would rob microeconomics of its empirical content. Interestingly, health economists tend to accept their doctors thinly veiled criticism, “As long as you are healthy, you don’t give a damn; but as soon as you are sick, you are prepared to sacrifice everything to restore your health.” This is nothing but a claim suggesting that when it comes to health, individual preferences are unstable—which, of course, precludes a welfare judgment on anything medical because the ex-ante valuation of a healthy person will always be incompatible with the ex-post valuation of the same person as a patient. In the ZB/ZBK model, state dependence enters on the production side, and in two ways:

1. When healthy, individuals are more productive both in terms of C and $(1-\pi)$. As to consumption possibilities, the labor income earned through work is assumed to outweigh the fact that part of the available time is devoted to work. And individuals can always sacrifice consumption possibilities in favor of prevention to come rather close to the limit of 365 healthy days a year if they so desire.
2. When sick, individuals receive an exogenous income in the guise of social security benefits, which is assumed to be far lower than their regular labor income. More time could be available for consumption because there is no need to go to work. However, this must be balanced against time spent on travelling to see the physician, waiting in the practice, and recovering (especially when in hospital).

In the whole, Fig. 2 indicates that consumption possibilities shrink in both dimensions when the individual falls ill.

It contains two additional assumptions:

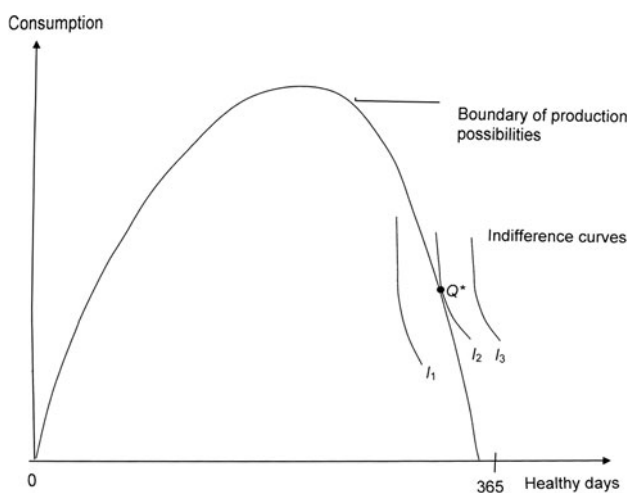


Fig. 1 Production possibilities and preference for health chances

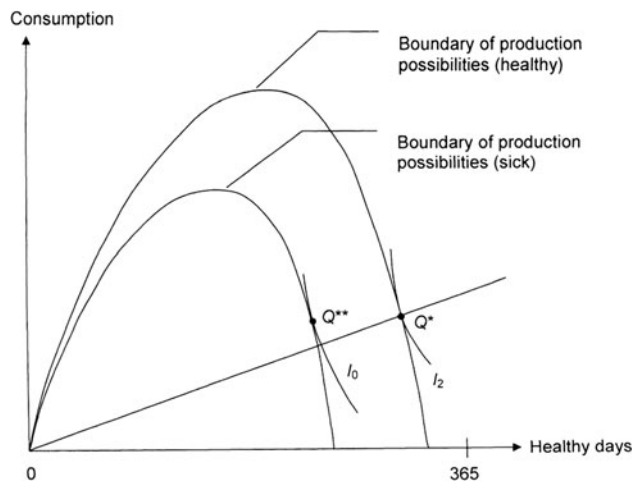


Fig. 2 State dependence in the production of consumption and healthy time

A1 The loss of productive capabilities is more marked in terms of health (healthy days) than in terms of consumption services.

A2 Preferences are homothetic (otherwise, no definite predictions are possible).

With A1 and A2 in hand, it can be easily shown that individuals behave in a way as though health (and hence medical care) is more important to them when sick rather than healthy, although their preferences are not state dependent. To see this, consider optimum Q^* in the state healthy. Along the ray through the origin, indifference curves have the same slope due to homotheticity. However, in view of assumption A1, the production possibility frontier in the sick state has steeper slope almost everywhere than its counterpart pertaining to the healthy state. Therefore, the tangency condition for an optimum can only be satisfied at a point such as Q^{**} , where the tangent runs steeper than at Q^* . However, the slope of this tangent indicates the relative value of the two goods in question. Evidently, the relative value of healthy days is higher in the sick than the healthy state. Therefore, there is no need for health economists to accept the doctor's claim that health preferences are unstable. This is a second insight that the MGM does not provide.

Admittedly, these advantages come at a price. First, the ZB/ZBK model is black and white because it contains only two states of health. Contrary to the continuous health stock in the MGM, generalizing to more than two states is by no means easy. Second, in the healthy state, demand for medical care is simply zero. In this regard, Galama and Kaptain [6], with their limit value, are more flexible (although specifying the limit where medical care does not contribute to health anymore may be difficult). But then, the ZB/ZBK formulation is certainly more palatable to interested laypersons and policy makers because of its message: "As long as I am healthy, I can do something to maintain my health; but as soon as I fall ill, I have to rely on the doctor."

Suggestions for future research

The ZB/ZBK model also points toward possibilities for fruitful future research. One is the substitutability of an individual's own preventive effort by medical care. Having only one input in a given state of health, the ZB/ZBK health production function seems to exclude substitutability. However, the more preventive effort in the healthy state, the higher $(1/\pi)$, hence the longer on expectation the duration of the healthy state during which no health care services are required. Conversely, more medical care restores good health quicker, resulting in a longer string of

healthy days. As noted by Zweifel and Manning [21], there is precious little empirical evidence on these relationships.

Another area of research is the corrective action of individuals suggested by the ZB/ZBK model. As the principal target variable is the chance of being healthy $[(1-\pi^*) < 1]$; this is behind Figs. 1, 2], individuals continually over- and undershoot their target by spending another healthy day $[(1-\pi) = 1]$ or another sick day $[(1-\pi) = 0]$. In the case of overshooting, one would expect skimping on preventive effort as the natural response. The prediction therefore is that the longer the string of healthy days, the lower preventive effort *ceteris paribus*.

Finally, modelling the distribution of HCE with its highly positive third and fourth moments has attracted much effort recently (see, e.g., [11]). However, this distribution reflects decisions taken by patients and health care providers. On the patient side, it should be related to the stochastic process governing the production of health. The longer the previous string of healthy days undermining the incentive for prevention, the greater presumably the degree of undershooting, calling for quick adjustment and hence an intensive use of healthcare services and particularly high HCE ("haste makes waste," as it were). Therefore, the distribution characteristics of HCE need to be related to the nature of the stochastic process prior to the advent of sickness.

As a final remark, this editorial was triggered by a Scandinavian friend complaining, with an eye on Lund University: "The Grossman model has misled a whole generation of health economists." While this likely amounts to an overstatement, there is something to be gained by breaking away from the MGM fixation. If this editorial should help in this process, it has achieved its objective.

References

1. Benitez-Silva, H., Ni, H.: Health status and the health dynamics in an empirical model of expected longevity. *Eur. J. Health Econ.* **27**(3), 564–584 (2008)
2. Bolin, K., Jacobson, L., Lindgren, B.: The family as the health producer—when spouses act strategically. *J. Health Econ.* **21**(3), 475–495 (2002)
3. Bolin, K., Lindgren, B.: The double-faceted nature of health investments—implications for equilibrium and stability in a demand-for-health framework, NBER Working Paper No. w17789, (2012)
4. Chang, F.-R.: Uncertainty and investment in health. *J. Health Econ.* **15**(3), 369–376 (1996)
5. Corman, H., Grossman, M.: Determinants of neonatal mortality rates in the U.S.A.—a reduced form model. *J. Health Econ.* **4**(3), 213–236 (1985)
6. Galama, T., Kapteyn, A.: Grossmans missing health threshold. *J. Health Econ.* **30**(5), 1044–1056 (2011)

7. Grossman, D.C., Milligan, B.C., Deyo, R.A.: Risk factors for suicide attempts among Navajo adolescents. *Am. J. Public Health* **81**(7), 870–874 (2001)
8. Grossman, M.: On the concept of health capital and the demand for health. *J. Polit. Econ.* **80**(2), 223–255 (1972)
9. Grossman, M.: The human capital model. In: Culyer, A.J., Newhouse J.P. (eds.), *Handbook of Health Economics*, vol. 1A, pp. 347–408. Elsevier, Amsterdam (2000)
10. Jacobson, L.: The family as a producer of health—an extended Grossman model. *J. Health Econ.* **19**(5), 611–637 (2000)
11. Jones, A., Lomas, J., Rice, N.: Applying beta-type size distributions to healthcare cost regressions, working papers 11/31. HEDG, University of York, York (2011)
12. Leu, R., Doppmann, R.J.: Die Nachfrage nach Gesundheit und Gesundheitsleistungen (The demand for health and healthcare services). In: Gäfgen, G. (ed.) *Ökonomie des Gesundheitswesens*, pp. 161–175. Duncker & Humblot, Berlin (1986)
13. Liljas, B.: The demand for health with uncertainty and insurance. *J. Health Econ.* **17**(2), 153–170 (1998)
14. Mas-Colell, A., Whinston, M.D., Green, J.R.: *Microeconomic Theory*. Oxford University Press, New York (1995)
15. Muurinen, J.-M.: Demand for health: a generalized Grossman model. *J. Health Econ.* **1**(1), 5–26 (1982)
16. Nocera, S., Zweifel, P.: The demand for health: an empirical test of the Grossman model using panel data. In: Zweifel, P. (ed.) *Health, the medical profession, and regulation, developments in health economics and public policy*, vol. 6, pp. 35–49. Kluwer, Boston (1998)
17. Wagstaff, A.: The demand for health: some new empirical evidence. *J. Health Econ.* **5**(3), 195–233 (1986)
18. Wagstaff, A.: The demand for health: an empirical reformulation of the Grossman model. *Health Econ.* **2**, 189–198 (1993)
19. Zweifel, P., Breyer, F.: *Health Economics*. Oxford University Press, New York (1997)
20. Zweifel, P., Breyer, F., Kifmann, M.: *Health Economics*, 2nd edn. Springer, Heidelberg (2009)
21. Zweifel, P., Manning, W.G.: Moral hazard and consumer incentives in health care. In Culyer, A.J., Newhouse J.P. (eds.), *Handbook of Health Economics*, vol. 1A, pp. 409–459. Elsevier, Amsterdam (2000)