A Critique of an Aspect of Grossman's Model of Demand for Health Care

Pius C. Eze

1 Department of Economics, Godfrey Okoye University, Enugu, Nigeria.

Abstract

The model of demand for health care developed by Michael Grossman is generally accepted by health economists as the standard. But in spite of its general acceptance and its longevity, Grossman's model has some troublesome features. One such feature is the inadequacy of the human capital model for analyzing the demand for curative medical care. Given that preventive care is utilized in anticipation of illness, and its utilization and accruing benefits occur in different and sequenced time periods, preventive care is rightly viewed as a capital investment (as Grossman does) with current expenditures yielding future benefits. On the other hand, curative health care is distinguished by the fact that a person gets ill first and then seeks the care in order to cure (or, at least, treat) existing illness, and the utilization of curative care and the resulting care are presumed contemporaneous, curative care is more like the standard consumer goods such as food; in which case, any future benefit is an added benefit. From this analogy, it can be argued that curative care is better suited to the standard consumer model rather than the human capital model. Grossman's model neglects the distinction between curative care and preventive care, and instead generalizes the human capital model of demand for health care as if all health care is preventive. Also, Grossman's model ignores illness processes entirely as if all health decreases are due to aging or (natural) health depreciation. The present essay attempts to correct these faults.

1. Introduction

This essay conceptualizes health status in terms of illness, and measures illness by its severity. Illness at any moment is a health deviation, with its severity measured by the difference between two health levels. One level is a personal health benchmark which indicates the person is healthy, and the other is the person's actual health level at the moment. A person is considered healthy if and only if his or her actual level of health equals the personal benchmark level of health. Grossman's model of demand for health care neglects these facts by ignoring the distinction between curative health care and other types of care. At least for curative medical care, the appropriate concept of a person's health status is a measure of how ill the person is or the severity of the person's illness. The severity of a person's illness or how ill a person is is obtainable by comparing the person's level of health when the person is sick with the same person's level of health when the person is healthy. The illness concept of health status formalized in this paper contrasts with health as a capital stock.

An issue facing health care analysts and health care managers is that health itself is a latent variable, not directly observable nor directly measurable. This issue introduces the practical problem of how to measure health status which is a key variable in determining health care utilization policies, in planning health care access or in comparing the quality of health care delivery institutions. The health literature uses both mortality as well as morbidity measures of health status but Arrow (1963) and subsequent health economics literature prefer morbidity measures, in comparison with mortality measures. The preference for morbidity is possibly explained by the fact that morbidity measures can capture aspects of the quality of life of the individual as well as the population (Culyer and Newhouse, 2000). For example, Arrow (1963) identifies illness as a key variable motivating the demand for (or utilization of) medical care. The present essay extends the health economics literature by conceptualizing health status in terms of illness, thereby formalizing morbidity measures of health status.

Illness itself can be formalized as a fall and a deviation and can be measured by its severity which is the magnitude of the person's health deviation. The deviation of the person's level of health when the person is sick from a personal benchmark level at which the person is considered healthy is a real number, termed severity of illness (SOI). The reasoning is that, at least for curative medical care, the appropriate concept of health status is a measure of how ill a person is. This illness concept of health status is in agreement with
common usage but differs from the usage in the health economics literature where health is viewed as a capital stock. Using illness to conceptualize health status is neither new nor uncommon. For example, severity of illness and case-mix indices such as the United States Medicare’s diagnosis related groupings (DRGs) suppose that the sicker a person is, the more medical resources the person utilizes, ceteris paribus (see, for example, Horn and Sharkey, 1983; Eze and Wolfe, 1998).

The present paper focuses on curative medical care in order to sharpen the concepts. This follows a suggestion by Wolfe and van der Gaag (1981) that measures of health status are best designed for specific purposes. The question is: What is the most appropriate concept of health status in a model of demand for medical care? The answer offered in this essay is that, at least for curative medical care such as hospital inpatient care, illness is the most appropriate concept of health status. Health status is conceptualized in terms of illness and illness is measured by its severity, defined simply as the magnitude of a health deviation. The remainder of the paper is as follows. Section 2 presents an overview of the relevant literature with a focus on Grossman’s model. Section 3 conceptualizes and formalizes illness and severity of illness. Section 4 illustrates a link between severity of illness and medical care utilization and formalizes the concept of medical care need. The conclusion is in section 5 and section 6 has the references. An illustrative Figure is presented within the paper.

2. Overview of the Relevant Literature

The model of demand for health care developed by Grossman (1972) is generally accepted by health economists and others as the standard. See, for example, World Bank’s World Development Report: Investing in Health Handbook of Health Economics (edited by Culyer and Newhouse, 2000), and Mwabu (2007). In spite of its general acceptance and its longevity, Grossman’s model has features health economists find troublesome.

One troublesome feature is the inadequacy of the human capital model for analyzing the demand for curative medical care, such as hospital inpatient care. In order to appreciate the inadequacy, consider that the demand for health care can have preventive as well as curative aspects or dimensions. Preventive care is such that the utilization of care and the accruing benefits occur in different and sequenced time periods. In that sense, preventive care is rightly viewed as a capital investment (as Grossman does) with current expenditures yielding future benefits (Soares, 2014).

On the other hand, curative health care is distinguished by the fact that a person gets ill first and then seeks the care in order to cure (or, at least, treat) the illness (Arrow 1963). The incidence of illness and the utilization of curative care may be sequenced, but the utilization and the resulting cure are presumed contemporaneous. In such cases, any future benefit is an added benefit. In that sense, curative care is like the standard consumer goods such as food. For example, a person gets hungry first and then seeks food in order to satisfy (or ‘cure’ or ‘treat’) the hunger. Both the consumption or ingestion of food and the resulting satisfaction are presumed to occur in the same time period. From this analogy, it can be argued that curative care is better suited to the standard consumer theory rather than the human capital model. Grossman’s model neglected Arrow’s distinction between curative care and preventive care, and instead (a) generalizes the human capital model of demand for health care as if all health care is preventive; (b) ignores illness processes entirely as if all health decreases are (in essence) due to aging or (natural) health depreciation.

Another troublesome feature of Grossman’s model of demand for health care is related to the above discussion and concerns the conceptualization and measurement of health status. To see this, consider that a person’s health can deteriorate due to aging, on the one hand, or due to illness or injury, on the other hand. In general, the illness process differs from the aging process, and the responses to them differ also. First, illness at any moment can be conceptualized as a health deviation from a personal health benchmark level. On the other hand, in this context, aging is the ‘natural depreciation’ of the benchmark itself over appropriate segments the person’s lifecycle. Second, illness can be assumed random, neither inevitable nor natural (heredity apart), whereas aging is natural. That is, unlike illness, aging is undergone by everyone at some point in time regardless of the person’s other characteristics (health status, for example). Third, health care and natural healing can bring about complete cure or stoppage of illness-induced health reductions, whereas age-induced health reductions presumably are more difficult to restore, at least, at the present level of technology. Grossman’s model ignores illness completely and, instead, conceptualizes health status as a capital stock that depreciates or ages.

In discussing how to conceptualize and measure the explanatory variable health status, it is possible to distinguish personal health status from population health status. Prior to Arrow (1963), most writings on health care were by non-economists and focused mostly on population health (Aaron, 1981) rather than personal health, and was measured by various forms of mortality and morbidity rates. That was probably due to the nature of health care itself prior to the twentieth century (Folland et al. 1992). Arrow’s neoclassical methodology differed markedly from those used by previous health care analysts. For example, Arrow (1963) introduced individual demand for health care derivable from individual constrained optimization as well as the individual health status that motivates the demand. Arrow’s focus on illness and its influence on individual medical care utilization is, in a sense, a conceptualization of health status in terms of illness. Following Arrow,
health economists have tended to favor morbidity (illness) measures of health status compared to mortality measures (quality adjusted life years notwithstanding) which have little meaning for the individual, but also, possibly due to the fact that morbidity measures can take into account the quality of life (such as pain and suffering) of both the individual and the population; even though morbidity measures are not as accurate or as easy to interpret as mortality measures. Possibly for these reasons, the concern of health economists is how best to conceptualize and measure illness (morbidity). See, for example, Culyer and Newhouse (2000), van der Gaag and Perlman (1981).

The solid line in the Figure 1 (in the appendix) is a possible graphical representation of Grossman’s (1972) model of health stock with an initial endowment of health stock. The solid line is intended to show a lifecycle pattern of a person’s health status if the person is healthy throughout his or her lifetime. According to Grossman, along the solid line, reductions in the person’s level of health are due to aging (or natural depreciation) over the person’s lifetime. Even a healthy person ages. As such Grossman ignores illness and its processes.

Some authors recognize the inadequacy of natural health depreciations as a health process. For example, Phelps (2013) suggests that a typical person’s health stock could look something like the hashed line in Figure 1 (adapted from Phelps 2013, Figure 2.4). According to Phelps, declines in a person’s health can result from ‘aging’ as well as from ‘random events of illness and injury’. As such, Phelps would distinguish between declines along the solid line and declines along the hashed line. The hashed line can be generalized to accommodate random incidences, random intensities and random durations of illness. Note also that the solid line in Figure 1 is, more or less, the trend line of the hashed line. The hashed line coincides with the solid line only at points in time the person is healthy.

Other attempts to provide more realistic conceptualization health status in terms of illness include Williams’ (1981) suggestion that “there is a threshold [above] which a society considers someone as to all intent and purpose healthy”. For Harris and Kohn (2015), the relevant health variable in the demand for medical care is the deviation of contemporaneous health from a ‘reference health’ defined as “an average of previous health states” (p.1) or “the level of health to which the individual is accustomed” (p.3). Berndit and Williamson (1973) suggests a hypothetical health meter with which to measure how ill a person is, with calibrations in an “ordinal scale ranging from the non-impairment level (asymptomatic with no medical care need) to the total impairment level (death)”. According to Bush (1973) a “patient deviates from some norm of well-being, correctible or not, and we can talk about that deviation” as the conceptual measure of the magnitude of the person’s ill-health. Deb and Trivedi (1997) sort latent populations into the ‘healthy’ and the ‘ill’. Phelps and Newhouse (1974) expand their health production equation such that the contemporaneous health status equals an initial health status minus a random illness variable plus a health production function, which is a function of medical care usage; medical care itself is a function of illness and initial health status. The present paper uses the ideas contained in Figure 1 to combine, extend and operationalize the suggestions from Wolfe and van der Gaag, Williams, Phelps, Berndit and Williamson, Bush and others. The question is how to conceptualize and measure illness in terms of illness, and how to formalize and measure illness.

3. Formalization of Severity of Illness (SOI)

This section formalizes illness and severity of illness (SOI). When is a person sick and when is a person healthy? Imagine that, in any time period, each person has a level of health at which the person is considered healthy. The person’s level of health when healthy can be considered a personal benchmark. Imagine also that each person has a health index or health variable that measures the person’s actual level of health (whether sick or healthy). Berndit and Williamson (1973) imagine a hypothetical health meter with which to measure how ill a person is. To calibrate such a health meter, imagine that person i has a level of health H_i^* that indicates he or she is healthy. H_i^* is a personal benchmark in the sense that H_i^* can differ from H_j^*, for another person j ≠ i. A person’s health benchmark is a datum. It depends on personal characteristics (such as chronic health and genetic conditions, gender, etc.), and is related to the natural processes of growth, aging and dying (Editors, 1973). It is unaffected by contemporaneous illnesses or medical care usage, especially in the short run. One of the uses of medical history is to help determine the person’s health benchmark (van der Ven and van der Gaag, 1982). A health concept similar to the health benchmark is referred to as ‘normal health’ by Williamson (1981), ‘reference or accustomed health’ by Harris and Kohn (2015).

Let H_i denote person i’s actual level of health whether he or she is healthy or ill. For completeness, it can be assumed that a person cannot be better than healthy:  H_i ≤ H_i^*. With these definitions, illness is the fall or deviation of person i’s actual health level H_i from its benchmark level H_i^*. The severity of person i’s illness s_i is defined as the magnitude of this health deviation. Symbolically, the severity of person i’s illness is defined by the following identity:

\[ s_i = H_i^* - H_i \]  

As long as H_i = H_i^*, person i is considered healthy even if his H_i^* is low.

Conceptualizing health status in terms of illness recognizes that a person is considered healthy or otherwise relative to himself or herself and not relative to other standards. In general, the severity of each
person’s illness is ascertained by comparing that person’s actual level of health with the (benchmark) level of health of the same person when healthy. Then the cure of a person’s illness is the restoration of the person’s actual level of health to the personal benchmark level with no attempt to create someone else out of him or her. The sicker the person is, the larger the person’s illness severity is, by definition. The cure of a person’s illness corresponds to a reduction of his or her severity to zero.

In order to generalize this model, consider Person i’s lifetime, 0 ≤ t ≤ T, where t is a measure of time, for example, age. The solid line in Figure 1 shows \( H_i(t) \), the benchmark function, indicating that Person i is healthy. The hashed line shows Person i’s actual level of health at any time \( H_i(t) \). The severity of a person’s illness at any time is the momentary magnitude of the deviation of the hashed line from its trend line (the solid line). Symbolically, at any time t,

\[
\begin{align*}
\text{if Person i is healthy.} & \quad H_i(t) = H_i^*(t) \\
\text{if Person i is sick.} & \quad H_i(t) < H_i^*(t)
\end{align*}
\]

The severity of Person i’s illness at any time t is denoted by \( s_i(t) \), defined as the difference between \( H_i^*(t) \) and \( H_i(t) \):

\[
\begin{align*}
\text{if Person i is healthy.} & \quad s_i(t) = 0 & \text{3.5} \\
\text{if Person i is sick.} & \quad s_i(t) > 0 & \text{3.6}
\end{align*}
\]

As suggested by Grossman, Person i is born with an initial level of health \( H_i(t=0) = H_0 = H_i(t=0) \) (initial health stock) and dies at \( H_i(t=T) = H_d = H_i(t=T) \). In between birth and death, Person i’s health benchmark \( H_i^*(t) \) depreciates due to aging (as asserted by Grossman). Also, in between birth and death, Person i’s actual health (level) at any time period, \( H_i(t) \), can deviate from its benchmark (level) \( H_i^*(t) \) due to illness or injury. In diagnosing Person i’s illness, health professionals attempt to determine \( s_i(t) \). Person i is considered healthy if \( s_i(t) = 0 \). He or she is considered sick if and only if \( s_i(t) > 0 \). For example, in the Figure 1, the person is sick at \( t = t_1 \) but healthy at \( t = t_2 \) because \( s(t_1) > 0 \) while \( s(t_2) = 0 \). This is so even though \( H(t_1) > H(t_2) \). The sicker the person is, the larger his or her \( s_i(t) \) is. The cure of Person i’s illness corresponds to a reduction of \( s_i(t) \) to zero.

For Grossman, a person is healthier than another if he or she has a higher health stock. Using our terminology, Grossman would say that Person i is healthier than Person j if \( H_i^*(t) > H_j^*(t) \). In comparison, the model developed in this paper recognizes that a person is considered healthy or otherwise relative to himself or herself. If inter-personal comparisons must be made, then Person i is sicker than Person j at time t if \( H_i^*(t) - H_i(t) > H_j^*(t) - H_j(t) \).

4. Relation between Severity of Illness and Medical Care Usage

Consider a single time period. Let \( s_0 \) represent a measure of how ill a person is at the beginning of the period. The purpose of curative medical care is to reduce \( s_0 \) possibly to zero. Let \( M \) be the amount of medical care the person utilizes within the period. Let \( h(M) \) represent an index of the effectiveness of medical care, \( h(M) \geq 0 \). In the health economics literature, \( h(M) \) is termed a health production function embodying existing medical technology as well as relevant personal characteristics (Grossman 1972, Pauly 1980). Jack (1999) mistakenly refers to the inverse of \( h(M) \) as health status (p.58) and asserts that "the only effect of illness is to increase the effective price of health that an individual faces" (p.61).

Medical care need can be operationalize as follows. A person’s medical care need is the amount of medical care required to reduce \( s_0 \) to zero. Imagine that there is an objectively (or scientifically) determined amount of medical care, \( M^* \), required to reduce \( s_0 \) to zero. \( M^* \) is the person’s medical care need (Boulding, 1966) that can be obtained by solving the following identity:

\[
s_0 - h(M^*) = 0 \quad \text{4.1}
\]

Note that the sicker a person is, the more his or her medical care need, all else equal. If medical care were free, the sick person would choose to utilize medical care at the need level, all else equal (Aaron 1981). If, on the other hand, medical care is not free then, possibly because of the consumer’s budget constraint (including time costs), the optimal amount of medical care (actually) utilized may be less than his or her medical care need. In which case the illness may or may not be fully cured within the given period. If, at the end of the time period, the person is not fully cured, some level of the illness remains. Let \( s \) represent the severity of the person’s illness within the period. Then the technical relationship between the severity of the person’s illness and the person’s medical care usage is represented generally as follows:

\[
s = s_0 - h(M) \geq 0. \quad \text{4.2}
\]
Equation 4.2 states that how ill a person is in a time period (for example, at the middle or end of the period) depends, all else equal, on the initial health conditions and on the effectiveness of curative care utilized in that period. In a static, non-stochastic model, it is assumed that \( s = 0 \) if \( s_0 = 0 \). Note that in the relationship between health and medical care as presented in Equation 4.2, medical care \( M \) is utilized in order to cure existing illness \( s_0 \). Compare Equation 4.2 to Equation 4.3 where medical care \( M \) counteracts the depreciation in health (Muurinen 1982):

\[
H = H_0(1 - \delta) + h(M) \tag{4.3}
\]

where \( 0 \leq \delta \leq 1 \) is the rate of health depreciation and \( H_0 \) is the person’s initial health stock.

All medical care utilization can be categorized in terms of preventive and curative care. But such dichotomy is often difficult to establish in practice because ‘preventive’ and ‘curative’ refer to motivations for medical care use. Lillard discuss some of the difficulties involved in a curative-preventive dichotomy. As a result, empirical models need to consider different categorizations of medical care. For example, a trichotomy consisting of inpatient care, outpatient care and self-medication can permit the following hypothesis: All else equal, illnesses of ‘large enough’ severities require hospital inpatient care; illnesses of ‘medium’ severities require only ambulatory, outpatient or other non-hospital services while illnesses of ‘small’ severities require self-medication. Healthy persons and persons with illnesses of negligible severities require no medical care or medical attention at all. Symbolically,

\[
s = \begin{cases} 
0 & \text{if } s = 0 \\
M_0 & \text{if } 0 < s \leq s_0 \\
M_1 & \text{if } s_0 < s \leq s_1 \\
M_2 & \text{if } s > s_1 
\end{cases} \tag{4.4}
\]

\[
\text{selfmedication} = \begin{cases} 
0 & \text{if } s = 0 \\
M_0 & \text{if } 0 < s \leq s_0 \\
M_1 & \text{if } s_0 < s \leq s_1 \\
M_2 & \text{if } s > s_1 
\end{cases} \tag{4.5}
\]

\[
\text{outpatient} = \begin{cases} 
0 & \text{if } s = 0 \\
M_0 & \text{if } 0 < s \leq s_0 \\
M_1 & \text{if } s_0 < s \leq s_1 \\
M_2 & \text{if } s > s_1 
\end{cases} \tag{4.6}
\]

Where \( 0 < S_0 < S_1 < S_d \). Once the severity of a person’s illness rises to the death level \( s_d \), the person dies and all medical therapies cease, \( M = 0 \). This suggests limits to the assumed monotonicity in medical care usage. That is, medical care can actually become zero for high enough severities. \( 0 < M_0 < M_1 < M_2 \) are the levels of medical care corresponding to the various levels of severity. It can be conjectured that hospital inpatient care is curative; but self-medication or outpatient care can be either preventive or curative or both. Wagstaff (1986) and van de Ven and van der Gaag (1982) include self-medication is their empirical analyses. Institutional and informational constraints may necessitate a person’s contemporaneous use of more than one type of medical care.

In some cases, a patient must utilize both outpatient care and inpatient care only because he has to visit an outpatient facility or an emergency room in order to be admitted to a hospital. In such cases, error terms in econometrics equations may be correlated. Lastly, the relation between severity of a person’s illness and the person’s medical care usage can be simultaneous (see Wolfe and van der Gaag, 1981) because a person’s severity of illness can be endogenous. Measures of severity of illness found in data sets can be ex-ante (before medical care is utilized), contemporaneous, or ex post. For predicting medical care usage, ex ante SOI seems the appropriate choice, while ex-post SOI seems more appropriate for measuring health outcomes of medical interventions.

5. Conclusion

The model of demand for health care developed by Grossman (1972) has some features health economists find troublesome. The present paper advocates the use of illness to conceptualize health status, and thereby highlights problems with Grossman’s conceptualization and measurement of health status. For example, severity-of-illness and case-mix indices such as the United States Medicare’s diagnosis related groupings (DRGs) suppose that, unlike Grossman’s model, more severe illnesses require larger medical resources to ‘cure’, ceteris paribus. The concept of health status developed in this paper is with respect to medical care utilization. If a person’s illness is neither curable nor controllable with medical care, the person is not sick and requires no medical care, even though such a person may be sick with regards to other types of care (nursing care, for example). Also, according to Weisbrod, incurable diseases are expensive mainly due to ‘half-way technologies’.
References

Bibliography