THE VALUE OF LIFE AND THE RISE IN HEALTH SPENDING*

ROBERT E. HALL AND CHARLES I. JONES

Over the past half century, Americans spent a rising share of total economic resources on health and enjoyed substantially longer lives as a result. Debate on health policy often focuses on limiting the growth of health spending. We investigate an issue central to this debate: Is the growth of health spending a rational response to changing economic conditions—notably the growth of income per person? We develop a model based on standard economic assumptions and argue that this is indeed the case. Standard preferences—of the kind used widely in economics to study consumption, asset pricing, and labor supply—imply that health spending is a superior good with an income elasticity well above one. As people get richer and consumption rises, the marginal utility of consumption falls rapidly. Spending on health to extend life allows individuals to purchase additional periods of utility. The marginal utility of life extension does not decline. As a result, the optimal composition of total spending shifts toward health, and the health share grows along with income. In projections based on the quantitative analysis of our model, the optimal health share of spending seems likely to exceed 30 percent by the middle of the century.

I. INTRODUCTION

The United States devotes a rising share of its total resources to health care. The share was 5.2 percent in 1950, 9.4 percent in 1975, and 15.4 percent in 2000. Over the same period, health has improved. Life expectancy at birth was 68.2 years in 1950, 72.6 years in 1975, and 76.9 years in 2000.

Why has this health share been rising, and what is the likely time path of the health share for the rest of the century? We present a framework for answering these questions. In the model, the key decision is the division of total resources between health care and nonhealth consumption. Utility depends on quantity of life—life expectancy—and quality of life—consumption. People value health spending because it allows them to live longer and to enjoy better lives.

* We are grateful to David Cutler, Amy Finkelstein, Victor Fuchs, Alan Garber, Michael Grossman, Emmett Keeler, Ron Lee, Joseph Newhouse, Tomas Philipson, David Romer, Robert Topel, the editors and referees, and participants at numerous seminars and NBER meetings for helpful comments. Jones thanks the Center for Economic Demography and Aging at Berkeley for financial support. Matlab programs that generate the numerical results in this paper are available at Jones's website. Contact information for the authors follows. Robert E. Hall: rehall@stanford.edu, http://stanford.edu/~rehall. Charles I. Jones: chad@econ.berkeley.edu, http://www.econ.berkeley.edu/~chad

The Quarterly Journal of Economics, February 2007

In our approach, standard preferences—of the kind economists use to study issues ranging from consumption to asset pricing to labor supply—are able to explain the rising share of health spending. As consumption increases, the marginal utility of consumption falls quickly. In contrast, extending life does not run into the same kind of diminishing returns. As we get older and richer, which is more valuable: a third car, yet another television, more clothing—or an extra year of life? There are diminishing returns to consumption in any given period and a key way we increase our lifetime utility is by adding extra periods of life.

Standard preferences imply that health is a superior good with an income elasticity well above one. As people grow richer, consumption rises but they devote an increasing share of resources to health care. Our quantitative analysis suggests these effects can be large: projections in our model typically lead to health shares that exceed 30 percent of GDP by the middle of the century.

Many of the important questions related to health involve the institutional arrangements that govern its financing—especially Medicare and employer-provided health insurance. One approach would be to introduce these institutions into our model and to examine the allocation of resources that results. We take an alternative approach. We examine the allocation of resources that maximizes social welfare in our model. We abstract from the complicated institutions that shape spending in the United States and ask a more basic question: from a social welfare standpoint, how much should the nation spend on health care, and what is the time path of optimal health spending?

The recent health literature has emphasized the importance of technological change as an explanation for the rising health share—for example, see Newhouse [1992]. According to this explanation, the invention of new and expensive medical technologies causes health spending to rise over time. Although the development of new technologies unquestionably plays a role in the rise of health spending, the technological explanation is incomplete for at least two reasons.

First, expensive health technologies do not need to be used just because they are invented. Although distortions in health insurance in the United States might result in overuse of expensive new technologies, health shares of GDP have risen in virtually every advanced country in the world, despite wide variation in systems for allocating health care [Jones 2003]. We investigate whether the social payoff associated with the use of new technologies is in line with the cost. Second, the invention of the new technologies is itself endogenous: Why is the United States investing so much in order to invent these expensive technologies? By focusing explicitly on the social value of extending life and how this value changes over time, we shed light on these questions.

We begin by documenting the facts about aggregate health spending and life expectancy, the two key variables in our model. We then present a simple stylized model that makes some strong assumptions but that delivers our basic results. From this foundation, we consider a richer and more realistic framework and develop a full dynamic model of health spending. The remainder of the paper estimates the parameters of the model and discusses a number of projections of future health spending derived from the model.

Our research is closely related to a number of empirical and theoretical papers. Our work is a theoretical counterpart to the recent empirical arguments of David Cutler and others that high levels and growth rates of health spending may be economically justified [Cutler et al. 1998; Cutler and McClellan 2001; Cutler 2004]. On the theoretical side, our approach is closest in spirit to Grossman [1972] and Ehrlich and Chuma [1990], who consider the optimal choice of consumption and health spending in the presence of a quality-quantity tradeoff. Our work is also related to a large literature on the value of life and the willingness of people to pay to reduce mortality risk. Classic references include Schelling [1968] and Usher [1973]. Arthur [1981], Shepard and Zeckhauser [1984], Murphy and Topel [2003], and Ehrlich and Yin [2004] are more recent examples that include simulations of the willingness to pay to reduce mortality risk and calculations of the value of life. Nordhaus [2003] and Becker, Philipson, and Soares [2005] conclude that increases in longevity have been roughly as important to welfare as increases in nonhealth consumption, both for the United States and for the world as a whole. Barro and Barro [1996] develop a model in which health investments reduce the depreciation rate of schooling and health capital; health spending as a fraction of income can then rise through standard transition dynamics.

We build on this literature in two ways. First and foremost, the focus of our paper is on understanding the determinants of the aggregate health share. The existing theoretical literature generally focuses on individual-level spending and willingness to pay to reduce mortality. Second, we consider a broader class of preferences for longevity and consumption. Many earlier papers specialize for their numerical results to constant relative risk aversion utility, with an elasticity of marginal utility between zero and one. This restriction occurs because these papers do not consider a constant term in flow utility. As we show later, careful attention to the constant is crucial for understanding the rising health share. In particular, when a constant is included, a standard utility function with an elasticity of marginal utility well above one is admissible. This property is the key to the rising health share in the model.

II. Basic Facts

We will be concerned with the allocation of total resources to health and other uses. We believe that the most appropriate measure of total resources is consumption plus government purchases of goods and services. That is, we treat investment and net imports as intermediate products. Similarly, we measure spending on health as the delivery of health services to the public and do not include investment in medical facilities. Thus we differ conceptually (but hardly at all quantitatively) from other measures that include investment in both the numerator and denominator. When we speak of consumption of goods and services, we include government purchases of nonhealth goods and services.

Figure I shows the fraction of total spending devoted to health care, according to the U.S. National Income and Product Accounts. The numerator is consumption of health services plus government purchases of health services and the denominator is consumption plus total government purchases of goods and services. The fraction has a sharp upward trend, but growth is irregular. In particular, the fraction grew rapidly in the early 1990s and flattened in the late 1990s. Not shown in the figure is the resumption of growth after 2000.

Figure II shows life expectancy at birth for the United States. Following the tradition in demography, this life expectancy measure is not expected remaining years of life (which depends on unknown future mortality rates), but is life expectancy for a hypothetical individual who faces the cross-section of mortality rates from a given year. Life expectancy has grown about 1.7 years per decade. It shows no sign of slowing over the fifty years

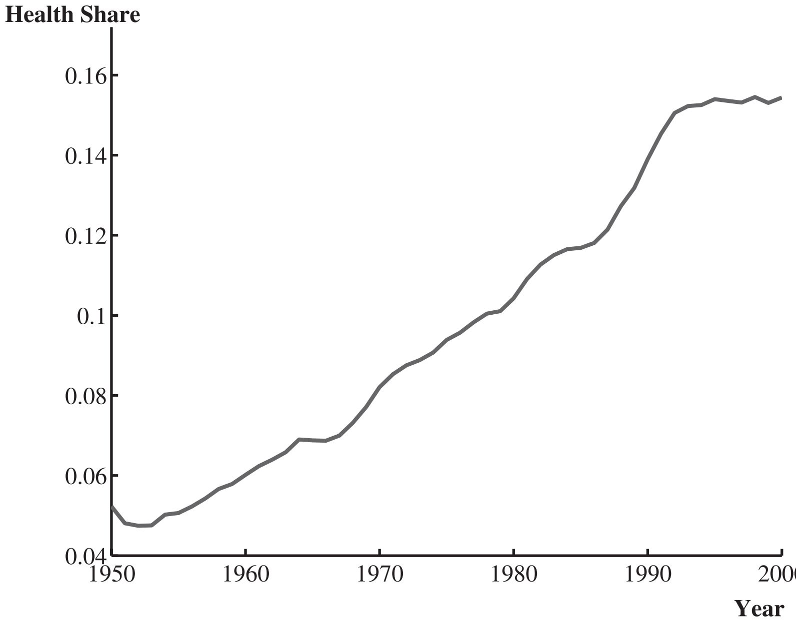

FIGURE I
The Health Share in the United States

Note: The numerator of the health share is consumption of health services plus government purchases of health services and the denominator is consumption plus total government purchases of goods and services. For further information on sources, see Section V.

reported in the figure. In the first half of the 20th century, however, life expectancy grew at about twice this rate, so a longer times series would show some curvature.

III. BASIC MODEL

We begin with a model based on the simple but unrealistic assumption that mortality is the same in all age groups. We also assume that preferences are unchanging over time, and income and productivity are constant. This model sets the stage for our full model, in which we incorporate age-specific mortality and productivity growth. As we will show in Section IV, the stark assumptions we make in this section lead the full dynamic model to collapse to the simple static problem considered here.

The economy consists of a collection of people of different

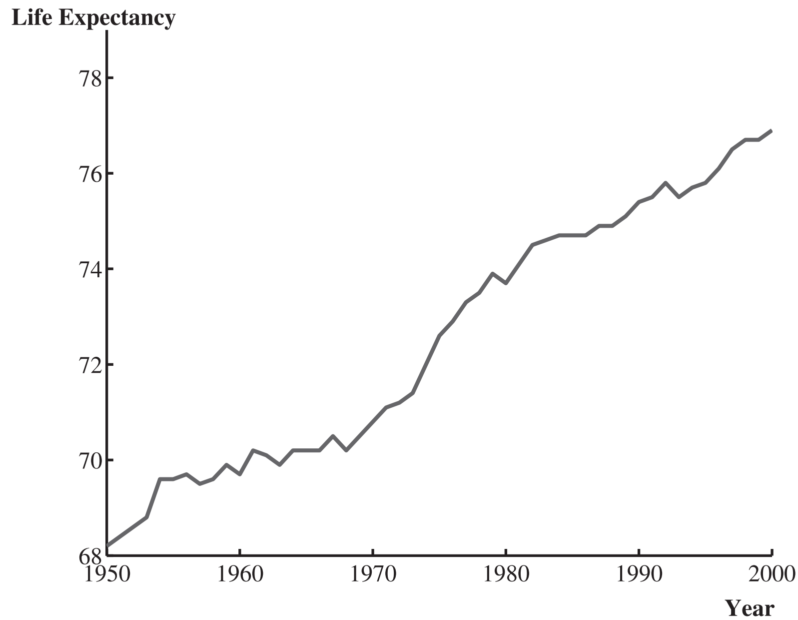

FIGURE II
Life Expectancy in the United States

Note: Life expectancy at birth data are from Table 12 of National Vital Statistics Report Volume 51, Number 3 "United States Life Tables, 2000," December 19, 2002. Center for Disease Control.

ages who are otherwise identical, allowing us to focus on a representative person. Let x denote the person's *health status.* The mortality rate of an individual is the inverse of her health status, $1/x$. Since people of all ages face this same mortality rate, x is also equal to life expectancy. For simplicity at this stage, we assume zero time preference.

Expected lifetime utility for the representative individual is

$$U(c, x) = \int_0^\infty e^{-(1/x)t} u(c) dt = xu(c). \tag{1}$$

That is, lifetime utility is the present value of her per-period utility $u(c)$ discounted for mortality at rate $1/x$. In this stationary environment, consumption is constant so that expected utility is the number of years an individual expects to live multiplied by per-period utility. We assume for now that period utility depends only on consumption; in the next section, we will introduce a quality-of-life term associated with health. Here and throughout the paper, we normalize utility after death at zero.

Rosen [1988] pointed out the following important implication of a specification of utility involving life expectancy: When lifetime utility is per-period utility, u, multiplied by life expectancy, the level of u matters a great deal. In many other settings, adding a constant to u has no effect on consumer choice. Here, adding a constant raises the value the consumer places on longevity relative to consumption of goods. Negative utility also creates an anomaly—indifference curves have the wrong curvature and the first-order conditions do not maximize utility. As long as u is positive, preferences are well behaved.[1]

The representative individual receives a constant flow of resources, y, that can be spent on consumption or health:

$$c + h = y. \tag{2}$$

The economy has no physical capital or foreign trade that permits shifting resources from one period to another.

Finally, a health production function governs the individual's state of health:

$$x = f(h). \tag{3}$$

The social planner chooses consumption and health spending to maximize the utility of the individual in (1) subject to the resource constraint (2) and the production function for health status (3). That is, the optimal allocation solves

$$\max_{c,h} f(h)u(c) \quad s.t. \quad c + h = y. \tag{4}$$

The optimal allocation equates the ratio of health spending to consumption to the ratio of the elasticities of the health production function and the flow utility function. With $s \equiv h/y$, the optimum is

1. Rosen also discussed the following issue: If the elasticity of utility rises above one for low values of consumption—as it can for the preferences we estimate in this paper—mortality becomes a good rather than a bad. A consumer would achieve a higher expected utility by accepting higher mortality and the correspondingly higher level of later consumption. Thus one cannot take expected utility for a given mortality rate as an indicator of the welfare of an individual who can choose a lower rate. This issue does not arise in our work, because we consider explicit optimization over the mortality rate. An opportunity for improvement of the type Rosen identified would mean that we had not maximized expected utility.

$$\frac{s^*}{1-s^*} = \frac{h^*}{c^*} = \frac{\eta_h}{\eta_c}, \tag{5}$$

where $\eta_h \equiv f'(h)\frac{h}{x}$, and $\eta_c \equiv u'(c)\frac{c}{u}$.

Now suppose we ignore the fact that income and life expectancy are taken as constant in this static model and instead consider what happens if income grows. The shortcut of using a static model to answer a dynamic question anticipates the findings of our full dynamic model quite well.

The response of the health share to rising income depends on the movements of the two elasticities in (5). The crux of our argument is that the consumption elasticity falls relative to the health elasticity as income rises, causing the health share to rise. Health is a superior good because satiation occurs more rapidly in nonhealth consumption.

Why is η_c decreasing in consumption? In most branches of applied economics, only marginal utility matters. For questions of life and death, however, this is not the case. We have normalized the utility associated with death at zero in our framework, and how much a person will pay to live an extra year hinges on the level of utility associated with life. In our application, adding a constant to the flow of utility, $u(c)$, has a material effect—it permits the elasticity of utility to vary with consumption.

Thus our approach is to take the standard constant-elastic specification for marginal utility but to add a constant to the level of utility. In this way, we stay close to the approach of many branches of applied economics that make good use of a utility function with constant elasticity for marginal utility. In finance, it has constant relative risk aversion. In dynamic macroeconomics, it has constant elasticity of intertemporal substitution. In the economics of the household, it has constant elasticity of substitution between pairs of goods.

What matters for the choice of health spending, however, is not just the elasticity of marginal utility, but also the elasticity of the flow utility function itself. With the constant term added to a utility function with constant-elastic marginal utility, the utility elasticity declines with consumption for conventional parameter values. The resulting specification is then capable of explaining the rising share of health spending.

With this motivation, we specify flow utility as

$$u(c) = b + \frac{c^{1-\gamma}}{1-\gamma}, \tag{6}$$

where γ is the constant elasticity of marginal utility. On the basis of evidence discussed later in the paper, we consider $\gamma > 1$ to be likely. In this case, the second term is negative, so the base level of utility, b, needs to be positive enough to ensure that flow utility is positive over the relevant values of c. The flow of utility $u(c)$ is always less than b, so the elasticity η_c is decreasing in consumption. More generally, any bounded utility function $u(c)$ will deliver a declining elasticity, at least eventually, as will the unbounded $u(c) = \alpha + \beta \log c$. Thus the key to our explanation of the rising health share—a marginal utility of consumption that falls sufficiently quickly—is obtained by adding a constant to a standard class of utility functions.

An alternative interpretation of the first-order condition is also informative. Let $L(c, x) \equiv U(c, x)/u'(c)$ denote the value of a life in units of output. Then, the optimal allocation of resources can also be characterized as

$$s^* = \eta_h \frac{L(c^*, x^*)/x^*}{y}. \tag{7}$$

The optimal health share is proportional to the value of a year of life L/x divided by per-capita income. If the flow of utility is given as in (6), it is straightforward to show that the value of a year of life satisfies

$$\frac{L(c, x)}{x} = bc^{\gamma} - \frac{c}{\gamma - 1}. \tag{8}$$

For $\gamma > 1$, the growth rate of the value of a life year approaches γ times the growth rate of consumption from above. Therefore, the value of a year of life will grow faster than consumption (and income) if γ is larger than 1. According to (7), this is one of the key ingredients needed for the model to generate a rising health share.

A rapidly declining marginal utility of consumption leads to a rising health share provided the health production elasticity, η_h, does not itself fall too rapidly. For example, if the marginal product of health spending in extending life were to fall to zero—say it was technologically impossible to live beyond the age of 100—then health spending would cease to rise at that point.

Whether or not the health share rises over time is then an empirical question: there is a race between diminishing marginal utility of consumption and the diminishing returns to the production of health. As we discuss later, for the kind of health production functions that match the data, the production elasticity declines very gradually, and the declining marginal utility of consumption does indeed dominate, producing a rising health share.

Finally, we can also generalize the utility function to $U(c, x)$ in place of $xu(c)$, so that lifetime satisfaction is not necessarily proportional to the length of the lifetime. The solution for this case is $s^*/(1 - s^*) = \eta_h \eta_x / \eta_c$, where $\eta_x \equiv U_x x / U$ is the elasticity of utility with respect to life expectancy. Our result, then, is that the health share rises when the consumption elasticity falls faster than the product of the production and life expectancy elasticities. As just one example $U(c, x) = x^{\alpha} u(c)$ delivers a constant η_x even with sharply diminishing returns to life expectancy (that is, α close to zero), so our main results are unchanged in this case.

The simple model develops intuition, but it falls short on a number of dimensions. Most importantly, the model assumes constant total resources and constant health productivity. This means it is inappropriate to use this model to study how a growing income leads to a rising health share, the comparative static results not withstanding. Still, the basic intuition for a rising health share emerges clearly. The health share rises over time as income grows if the marginal utility of consumption falls sufficiently rapidly relative to the joy of living an extra year and the ability of health spending to generate that extra year.

IV. The Full Dynamic Model

We turn now to the full dynamic model, allowing age-specific mortality and the associated heterogeneity, as well as growth in total resources and productivity growth in the health sector. This model also incorporates a quality-of-life component associated with health spending.

An individual of age a in period t has an age-specific state of health, $x_{a,t}$. As in the basic model, the mortality rate for an individual is the inverse of her health status. Therefore, $1 - 1/x_{a,t}$ is the per-period survival probability of an individual with health $x_{a,t}$.

An individual's state of health is produced by spending on health $h_{a,t}$:

$$x_{a,t} = f(h_{a,t};\, a,\, t). \tag{9}$$

In this production function, health status depends on both age and time. Forces outside the model that vary with age and time may also influence health status; examples include technological change and education.

The starting point for our specification of preferences is the flow utility of the individual, $u(c_{a,t}, x_{a,t})$. In addition to depending on consumption, flow utility depends on health status, $x_{a,t}$. Spending on health therefore affects utility in two ways, by increasing the quantity of life through a mortality reduction and by increasing the quality of life.

We assume this utility function takes the following form:

$$u(c_{a,t}, x_{a,t}) = b + \frac{c_{a,t}^{1-\gamma}}{1-\gamma} + \alpha\,\frac{x_{a,t}^{1-\sigma}}{1-\sigma}, \tag{10}$$

where γ, α, and σ are all positive. The first term is the baseline level of utility whose importance we stressed earlier.[2] The second term is the standard constant-elastic specification for consumption. We assume further that health status and consumption are additively separable in utility and that quality of life is a constant-elastic function of health status. Additive separability is of course a strong assumption. It implies that the marginal utility of consumption does not vary with health status and ultimately delivers the result that consumption itself will optimally be invariant to health status. We could relax this assumption in our framework and still obtain our main results. However, even the direction of the effect is unclear: Is the marginal utility of consumption higher or lower for sick people? One can easily think of reasons why it might be lower. On the other hand, the marginal utility of having a personal assistant or of staying in a nice hotel with lots of amenities might actually be higher for people with a lower health status.[3] Our separability assumption can be viewed as a natural intermediate case.

In this environment, we consider the allocation of resources

2. Previous versions of this paper considered the possibility that this intercept varied by age and time. In some of our estimation, we treated these $b_{a,t}$ terms as residuals that rationalized the observed health spending data as optimal. See Hall and Jones [2004] for more on this approach.

3. We thank a referee for this observation.

that would be chosen by a social planner who places equal weights on each person alive at a point in time and who discounts future flows of utility at rate β. Let $N_{a,t}$ denote the number of people of age a alive at time t. Then social welfare is

$$\sum_{t=0}^{\infty} \sum_{a=0}^{\infty} N_{a,t}\beta^t u(c_{a,t}, x_{a,t}). \tag{11}$$

The optimal allocation of resources is a choice of consumption and health spending at each age that maximizes social welfare subject to the production function for health in (9) and subject to a resource constraint we will specify momentarily.

It is convenient to express this problem in the form of a Bellman equation. Let $V_t(N_t)$ denote the social planner's value function when the age distribution of the population is the vector $N_t \equiv (N_{1,t}, N_{2,t}, \ldots, N_{a,t}, \ldots)$. Then the Bellman equation for the planner's problem is

$$V_t(N_t) = \max_{\{h_{a,t}, c_{a,t}\}} \sum_{a=0}^{\infty} N_{a,t}u(c_{a,t}, x_{a,t}) + \beta V_{t+1}(N_{t+1}) \tag{12}$$

subject to

$$\sum_{a=0}^{\infty} N_{a,t}(y_t - c_{a,t} - h_{a,t}) = 0, \tag{13}$$

$$N_{a+1,t+1} = \left(1 - \frac{1}{x_{a,t}}\right) N_{a,t}, \tag{14}$$

$$N_{0,t} = N_0, \tag{15}$$

$$x_{a,t} = f(h_{a,t}; a, t). \tag{16}$$

$$y_{t+1} = e^{g_y} y_t. \tag{17}$$

The first constraint is the economy-wide resource constraint. Note that we assume that people of all ages contribute the same flow of resources, y_t. The second is the law of motion for the population. We assume a large enough population so that the number of people aged $a + 1$ next period can be taken equal to the number aged a today multiplied by the survival probability. The third constraint specifies that births are exogenous and constant at N_0. The final two constraints are the production function for health and the law of motion for resources, which grow exogenously at rate g_y.

Let λ_t denote the Lagrange multiplier on the resource constraint. The optimal allocation satisfies the following first order conditions for all a:

$$u_c(c_{a,t}, x_{a,t}) = \lambda_t, \tag{18}$$

$$\beta \frac{\partial V_{t+1}}{\partial N_{a+1,t+1}} \frac{f'(h_{a,t})}{x_{a,t}^2} + u_x(c_{a,t}, x_{a,t}) f'(h_{a,t}) = \lambda_t, \tag{19}$$

where we use $f'(h_{a,t})$ to represent $\partial f(h_{a,t}; a, t)/\partial h_{a,t}$. That is, the marginal utility of consumption and the marginal utility of health spending are equated across people and to each other at all times. This condition together with the additive separability of flow utility implies that people of all ages have the same consumption c_t at each point in time, but they have different health expenditures $h_{a,t}$ depending on age.

Let $v_{a,t} \equiv \frac{\partial V_t}{\partial N_{a,t}}$ denote the change in social welfare associated with having an additional person of age a alive. That is, $v_{a,t}$ is the social value of life at age a in units of utility. Combining the two first-order conditions, we get

$$\frac{\beta v_{a+1,t+1}}{u_c} + \frac{u_x x_{a,t}^2}{u_c} = \frac{x_{a,t}^2}{f'(h_{a,t})}. \tag{20}$$

The optimal allocation sets health spending at each age to equate the marginal benefit of saving a life to its marginal cost. The marginal benefit is the sum of two terms. The first is the social value of life $\beta v_{a+1,t+1}/u_c$. The second is the additional quality of life enjoyed by people as a result of the increase in health status.

The marginal cost of saving a life is dh/dm, where dh is the increase in resources devoted to health care and dm is the reduction in the mortality rate. For example, if reducing the mortality rate by .001 costs \$2,000, then saving a statistical life requires $1/.001 = 1{,}000$ people to undertake this change, at a total cost of two million dollars. Our model contains health status x as an intermediate variable, so it is useful to write the marginal cost as $\frac{dh}{dm} = \frac{dh/dx}{dm/dx}$. Since health status is defined as inverse mortality, $m = 1/x$ so that $dm = dx/x^2$. In the previous example, we required $1/dm$ people to reduce their mortality rate by dm to save a life. Equivalently, setting $dx = 1$, we require x^2 people to increase their health status by one unit in order to save a statistical life. Since the cost of increasing x is $dh/dx = 1/f'(h)$, the marginal cost of saving a life is therefore $x^2/f'(h)$.

By taking the derivative of the value function, we find that the social value of life satisfies the recursive equation:

$$v_{a,t} = u(c_t, x_{a,t}) + \beta\left(1 - \frac{1}{x_{a,t}}\right) v_{a+1,t+1} + \lambda_t(y_t - c_t - h_{a,t}). \tag{21}$$

The additional social welfare associated with having an extra person alive at age a is the sum of three terms. The first is the level of flow utility enjoyed by that person. The second is the expected social welfare associated with having a person of age $a + 1$ alive next period, where the expectation employs the survival probability $1 - 1/x_{a,t}$. Finally, the last term is the net social resource contribution from a person of age a, her production minus her consumption and health spending.

The literature on competing risks of mortality suggests that a decline in mortality from one cause may increase the optimal level of spending on other causes, as discussed by Dow, Philipson, and Sala-i-Martin [1999]. This property holds in our model as well. Declines in future mortality will increase the value of life, $v_{a,t}$, raising the marginal benefit of health spending at age a.

IV.A. Relation to the Static Model

It is worth pausing for a moment to relate this full dynamic model to the simple static framework. With constant income y, a time- and age-invariant health production function $f(h)$, $\beta = 1$, and a flow utility function that depends only on consumption, the Bellman equation for a representative agent can be written as

$$V(y) = \max_{c,h} u(c) + (1 - 1/f(h))V(y) \quad \text{s.t.} \quad c + h = y. \tag{22}$$

Given the stationarity of this environment, it is straightforward to see that the value function is

$$V(y) = \max_{c,h} f(h)u(c) \quad \text{s.t.} \quad c + h = y, \tag{23}$$

the static model we developed earlier, restated in discrete time.

V. Quantitative Analysis

In the remainder of the paper, we estimate the parameters of our model and provide a quantitative analysis of its predictions.

We are conscious of uncertainty in the literature regarding the values of many of the parameters in our model. The calculations that follow should be viewed as illustrative and suggestive, and we have done our best to indicate the range of outcomes one would obtain with other plausible values of the parameters. We begin by describing the data we use, then proceed to estimating the parameter values, and finally conclude with solving the model.

We assume a period in the model is five years in the data. We organize the data into 20 five-year age groups, starting at 0–4 and ending at 95–99. We consider 11 time periods in the historical period, running from 1950 through 2000.

Data on age-specific mortality rates are taken from Table 35 of the National Center for Health Statistics publication *Health, United States 2004*. This source reports mortality rates every ten years, with age breakdowns generally in ten-year intervals. We interpolated by time and age groups to produce estimates for five-year time intervals and age categories. We also obtained data on age-specific mortality rates from accidents and homicides from this publication and from various issues of *Vital Statistics of the United States*. Our main approach treats mortality from accidents and homicides separately from nonaccident mortality. The distinction between the two categories is important mainly for older children and young adults, where health-related mortality is so low that the declines in accidents account for a substantial part of the overall trend in mortality. Our model deals only with nonaccident mortality, so we slightly underestimate the total contribution of rising health spending to declining mortality.

Data on age-specific health spending are taken from Meara, White, and Cutler [2004]. These data are for 1963, 1970, 1977, 1987, 1996, and 2000. Using the age breakdowns for these years, we distributed national totals for health spending across age categories, interpolated to our five-year time intervals.

National totals for health spending are from Table 2.5.5 of the revised National Income and Product Accounts of the Bureau of Economic Analysis, accessed at bea.gov on February 13, 2004 (for private spending) and Table 3.15 of the previous NIPAs, accessed December 2, 2003 (for government spending). The empirical counterpart for our measure of total resources per capita, y, is total private consumption plus total government purchases of goods and services, from the sources described above, divided by population.

VI. Estimating the Health Production Function

Our model has a set of parameters for the health production function and a set related to preferences. Both play a key role in the determination of optimal health spending. This section discusses the estimation of the health production function while the next section considers the estimation of the preference parameters.

We begin by assuming a functional form for the production of health status. We assume the inverse of the nonaccident mortality rate, $\tilde{x}_{a,t}$, is a Cobb-Douglas function of health inputs[4]:

$$\tilde{x}_{a,t} = A_a(z_t h_{a,t} w_{a,t})^{\theta_a}. \tag{24}$$

In this production function, A_a and θ_a are parameters that are allowed to depend on age. z_t is the efficiency of a unit of output devoted to health care, taken as an exogenous trend; it is the additional improvement in the productivity of health care on top of the general trend in the productivity of goods production. The unobserved variable $w_{a,t}$ captures the effect of all other determinants of mortality, including education and pollution.[5]

VI.A. Identification and Estimation

To explain our approach to identifying the parameters of this production function—A_a and θ_a—we introduce a new variable, $s_{a,t} \equiv h_{a,t}/y_t$, the ratio of age-specific health spending to income per capita. We rewrite our health production function as

$$\tilde{x}_{a,t} = A_a(z_t y_t \cdot s_{a,t} \cdot w_{a,t})^{\theta_a}. \tag{25}$$

The overall trend decline in age-specific mortality between 1950 and 2000 can then be decomposed into the three terms in parentheses. First is a trend due to technological change, $z_t y_t$. In our benchmark scenario, we assume technical change in the

4. The equation determining overall health status is therefore

$$x_{a,t} = f_{a,t}(h_{a,t}) = \frac{1}{m_{a,t}^{acc} + m_{a,t}^{non}} = \frac{1}{m_{a,t}^{acc} + 1/\tilde{x}_{a,t}},$$

where m^{acc} is the mortality rate from accidents and homicides and m^{non} is nonaccident mortality.

5. In principle, this specification allows nonaccident mortality rates to fall to zero with enough technical progress or health spending, potentially leading life expectancy to rise to arbitrarily high levels. In practice, this is not a serious concern for the time horizons we consider. Life expectancy in our simulations rises only to about eighty-one years by 2050.

health sector occurs at the same rate as in the rest of the economy, so that $z_t = 1$ is constant. Because y_t rises in our data at 2.31 percent per year, this is the rate of technical change assumed to apply in the health sector. In a robustness check, we assume technical change is faster in the health sector, allowing z_t to grow at one percent per year so that technical change in the health sector is 3.31 percent.

The second cause of a trend decline in age-specific mortality is resource allocation: as the economy allocates an increasing share of per capita income to health spending at age a, mortality declines. This effect is captured by $s_{a,t}$.

Third, unobserved movements of $w_{a,t}$ cause age-specific mortality to decline. We have already removed accidents and homicides from our mortality measure, but increases in the education of the population, declines in pollution, and declines in smoking may all contribute to declines in mortality.

The key assumption that allows us to identify θ_a econometrically is that our observed trends—technological change and resource allocation—account for a known fraction μ of the trend decline in age-specific mortality. For example, in our benchmark case, we assume that technical change and the increased allocation of resources to health together account for $\mu = 2/3$ of the decline in nonaccident mortality, leaving 1/3 to be explained by other factors. As a robustness check, we also consider the case where these percentages are 50-50, so that $\mu = 1/2$. We first discuss why this is a plausible identifying assumption and then explain exactly how it allows us to estimate θ_a.

A large body of research seeks to understand the causes of declines in mortality; see Cutler and Deaton [2006] for a recent survey. Newhouse and Friedlander [1980] is one of the early cross-sectional studies documenting a low correlation between medical resources and health outcomes. Subsequent work designed to solve the difficult identification problem (more resources are needed where people are sicker) have generally supported this finding [Newhouse 1993; McClellan, McNeil, and Newhouse 1994; Skinner, Fisher, and Wennberg 2001; Card, Dobkin, and Maestas 2004; Finkelstein and McKnight 2005]. This work often refers to "flat of the curve" medicine and emphasizes the low marginal benefit of additional spending. On the other hand, even this literature recognizes that certain kinds of spending—for example the "effective care" category of Wennberg, Fisher, and Skinner [2002] that includes flu vaccines, screening for breast and colon cancer, and drug treatments for heart attack victims—can have important effects on health. Goldman and Cook [1984] attribute 40 percent of the decline in mortality from heart disease between 1968 and 1976 to specific medical treatments; Heidenreich and McClellan [2001] take this one step further and conclude that the main reason for the decline in early mortality from heart attacks during the last twenty years is the increased use of medical treatments. Part of the increased use may result from improvements in technology [Cutler et al. 1998]. Skinner, Fisher, and Wennberg [2001] emphasize that technological advances have been responsible for "large average health benefits" in the U.S. population. Nevertheless, other factors, including behavioral changes, increased education, and declines in pollution, have certainly contributed to the decline in mortality [Chay and Greenstone 2003; Grossman 2005].

While it would be a stretch to say there is a consensus, this literature is generally consistent with the identifying assumption made here: that $\mu = 2/3$ of the trend decline in mortality is due to technological progress and the increased allocation of resources to health care. When applied to our estimation (as described further later), this identifying assumption leads to the following decomposition. Averaged across our age groups, 35 percent of the decline in age-specific mortality is due to technological change, 32 percent to increased resource allocation to health, and 33 percent (by assumption) to other factors. In our robustness check that assigns 50 percent to other factors, the split is 26 percent to technological change and 24 percent to increased resource allocation. When we allow technical change to be a percentage point faster in the health sector, 40 percent of the mortality decline is due to technical change, 27 percent to resource allocation, and 33 percent (by assumption) to unobserved factors.

How does our assumption that μ is known allow us to identify the parameters of the health production function? Take logs of (24) to get

$$\log \tilde{x}_{a,t} = \log A_a + \theta_a (\log z_t + \log h_{a,t} + \log w_{a,t}). \tag{26}$$

Our approach to identification is to construct a model whose disturbance is known not to have a trend. That orthogonality condition makes a time trend eligible as an instrumental variable—we apply GMM based on that condition.

If the unobserved component $w_{a,t}$ itself had no time trend, we would use the time trend as an instrument in estimating (26) directly. But our disturbance, $w_{a,t}$, surely does have a time trend: part of the reduction in mortality at a given age is due to factors other than technological change and increased resource reallocation. We use information about the contribution of the other factors to arrive at an equation where the time trend is a proper instrument.

We decompose the disturbance $w_{a,t}$ as

$$\log w_{a,t} = g_{w,a}t + \eta_{a,t}, \tag{27}$$

where $g_{w,a}$ is the age-specific trend in other determinants of mortality and $\eta_{a,t}$ is the random, nontrended part of the disturbance.

Combining (26) and (27) gives our estimating equation

$$\log \tilde{x}_{a,t} = \log A_a + \theta_a\,(\log z_t + \log h_{a,t} + g_{w,a}t) + \epsilon_{a,t}, \tag{28}$$

where the new disturbance $\epsilon_{a,t} \equiv \theta_a \eta_{a,t}$ is orthogonal to a linear trend. Therefore if we knew the value of $g_{w,a}$, we could use a linear time trend as an instrument to estimate θ_a.

Our assumption that we know μ allows us to compute $g_{w,a}$. Note that $1 - \mu$ is the fraction of trend mortality decline that is due to $w_{a,t}$. Therefore,

$$1 - \mu = \frac{g_{w,a}}{g_z + g_{h,a} + g_{w,a}}. \tag{29}$$

But if we know μ, then we know every term in this equation other than $g_{w,a}$ (g_z by assumption and $g_{h,a}$ from data), so we can use this equation to calculate the trend growth rate in $w_{a,t}$, and we are done.

We use GMM to estimate A_a and θ_a in (28). Our two orthogonality conditions are that $\epsilon_{a,t}$ has zero mean and that it has zero covariance with a linear time trend. Because health spending is strongly trending, the trend instrument is strong and the resulting estimator has small standard errors.[6]

Figure III shows the GMM estimates of θ_a, the elasticity of adjusted health status, $\tilde{x}$, with respect to health inputs, by age category. The groups with the largest improvements in health status over the fifty-year period, the very young and the middle-aged, have the highest elasticities, ranging from 0.25 to 0.40. The

6. The data we use in this estimation are the spending and mortality data discussed in Section V. For each age, we have data at five-year intervals for the period 1950 to 2000.

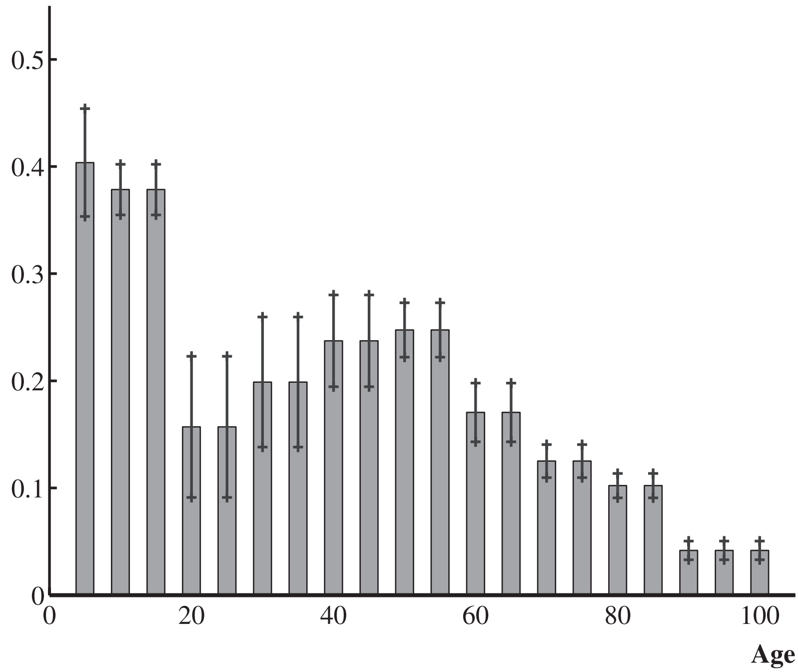

FIGURE III

Estimates of the Elasticity of Health Status with Respect to Health Inputs

Note: The height of each bar reports our estimate of the production function parameter θ_a, the elasticity of adjusted health status with respect to health inputs: $\tilde{x}_{a,t} = A_a\,(z_t h_{a,t} w_{a,t})^{\theta_a}$. The ranges at the top of the bars indicate $\pm$ two standard errors.

fact that the estimates of θ_a generally decline with age, particularly at the older ages, constitutes an additional source of diminishing returns to health spending as life expectancy rises. For the oldest age groups, the elasticity of health status with respect to health inputs is only 0.042.

Figure IV shows the actual and fitted values of health status for two representative age groups. Because the health technology has two parameters for each age—intercept and slope—the equations are successful in matching the level and trend of health status. The same is true in the other age categories.

VI.B. The Marginal Cost of Saving a Life

Our estimates of the health production function imply a value for the marginal cost of saving a life. Recall, from the

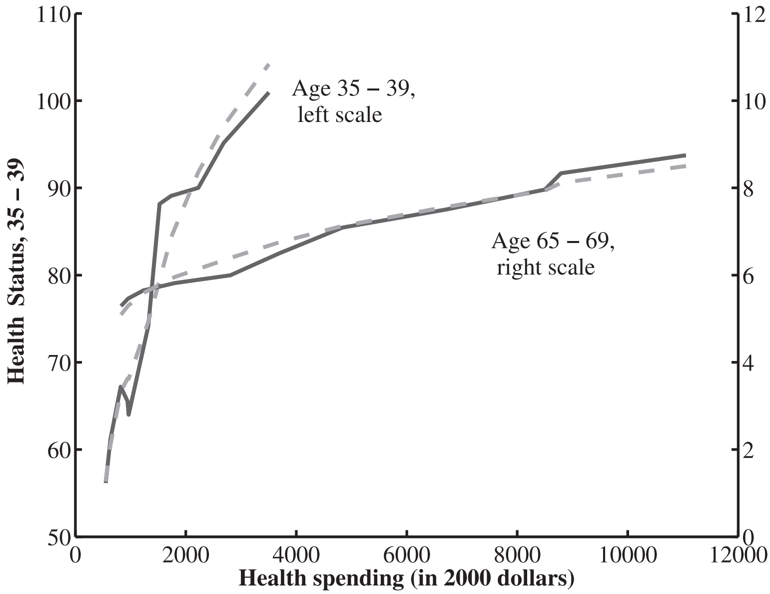

FIGURE IV
Goodness of Fit for the Health Technology

Note: The solid lines show data on health spending h on the horizontal axis and health status, x, on the vertical axis, for two age groups, 35–39 and 65–69, for the period 1950 through 2000. The dashed lines show the fitted values from the estimated production function in (28).

discussion surrounding (20), that this marginal cost is $x^2/f'(h)$. With our functional form for the health technology, the marginal cost of saving a life is $h\tilde{x}/\theta$.[7] Our work provides estimates of the value of life that can be compared with others derived either from other approaches on the cost side or from consumer choice involving mortality hazards, the demand side.

Table I shows this marginal cost of saving a life for various age groups. We can interpret these results in terms of the literature estimating the value of a statistical life (VSL). For example, the marginal cost of saving the life of a forty-year-old in the year 2000 was about \$1.9 million. In our robustness checks, this marginal cost reached as high as \$2.5 million (in the case where θ_a is

7. This expression has a nice interpretation: $\tilde{x}$ is the inverse of the nonaccident mortality rate, so it can be thought of as the number of living people per nonaccident death. h is health spending per person, so $h\tilde{x}$ is the total amount of health spending per death. The division by θ adjusts for the fact that we are interested in the marginal cost of saving a life, not the average.

TABLE I
THE MARGINAL COST OF SAVING A LIFE
(THOUSANDS OF YEAR 2000 DOLLARS)

Age	1950	1980	2000	Robust maximum, 2000	Per year of life saved, 2000	Growth rate, 1950–2000
0–4	10	160	590	(790)	8	7.8
10–14	270	2,320	9,830	(13,110)	152	7.2
20–24	1,170	3,840	8,520	(11,360)	155	4.0
30–34	500	2,120	4,910	(6,540)	108	4.6
40–44	160	740	1,890	(2,520)	52	4.9
50–54	70	330	1,050	(1,400)	39	5.4
60–64	50	280	880	(1,180)	47	5.9
70–74	40	280	790	(1,050)	67	6.2
80–84	40	340	750	(1,000)	125	6.1
90–94	50	420	820	(1,090)	379	5.6

Note: The middle columns of the table report estimates of the marginal cost of saving a life for various age groups. These estimates are calculated as $h\bar{x}/\theta$, using the estimates of θ given in Figure IV and using actual data on health spending and mortality by age. Standard errors for these values based on the standard errors of θ_a are small. The Robust Maximum column shows the maximum marginal cost we obtained in the various robustness checks described in the text; see Table II. The "Per Year of Life Saved" column divides the cost of saving a life by life expectancy at that age.

identified with the assumption that only 1/2 rather than 2/3 of declines in mortality are due to technical change and resource allocation). These numbers are at the lower end of the estimates of the VSL from the literature, which range from about two million to nine million dollars [Viscusi and Aldy 2003; Ashenfelter and Greenstone 2004; Murphy and Topel 2005]. If one believes the lower numbers, this suggests that health spending was at approximately the right level as a whole for this age group in 2000. Alternatively, of course, if one believes the higher estimates of the VSL from the literature, the calculation from Table I suggests that health spending for this group was too low.

The second-to-last column of the table provides an alternative view of the marginal cost of saving a life by stating the cost per year of life saved. It shows the cost of saving a statistical life in the year 2000, divided by life expectancy at each age. For example, the marginal cost of saving an extra year of life at age fifty is about $39,000. Interestingly, the cost of saving a life year in the youngest age category is only about $8,000, while the cost for saving a life year for the oldest ages rises to well above $100,000. These numbers are again typically below conventional estimates of the value of a year of life. Cutler [2004] reviews the literature and takes a rough value of $100,000 per year as reasonable. Murphy and Topel [2005] use theory to assign a six million dollar average value of life across ages and find life year values that are even higher. Taking our marginal cost estimates seriously then suggests the possibility that optimal health spending is substantially higher than actual spending. This finding will reappear later in our simulation results based on the full model.

VII. Estimating the Preference Parameters

Earlier we showed that the evolution of the optimal health share involves a race between diminishing returns to health spending and the diminishing marginal utility of consumption. Having estimated the parameters of the health technology, we turn in this section to finding values for the preference parameters: the curvature parameter γ, the discount factor β, the utility intercept b, and the quality of life parameters α and σ.

VII.A. Basic Preference Parameters

For the curvature parameter of the utility function, γ, we look to other circumstances where curvature affects choice. Large literatures on intertemporal choice [Hall 1988], asset pricing [Lucas 1994], and labor supply [Chetty 2006] each suggest that $\gamma = 2$ is a reasonable value. We explore alternative values ranging from near-log utility ($\gamma = 1.01$) to $\gamma = 2.5$. With respect to the discount factor, β, we choose a value that is consistent with our choice of γ and with a 6 percent real return to saving. Taking consumption growth from the data of 2.08 percent per year, a standard Euler equation gives an annual discount factor of 0.983, or, for the five-year intervals in our model, 0.918.

With these values for γ and β, we estimate the intercept of flow utility b to deliver a particular value of life for 35–39-year-olds in the year 2000 given the observed path of health spending.[8] As noted earlier, the empirical literature on the value of a statistical life encompasses a wide range of values, from a low of about two million dollars [Ashenfelter and Greenstone 2004] to highs of

8. For future values of health spending by age, we project the existing data forward at a constant growth rate. Until the year 2020, this growth rate is the average across the age-specific spending growth rates. After 2020 we assume spending grows at the rate of income growth. The rate must slow at some point; otherwise the health share rises above one. Our results are similar if we delay the date of the slowdown to 2050.

nine million dollars or more, as discussed in the survey of Viscusi and Aldy [2003]. Ashenfelter [2006] notes that the U.S. Department of Transportation uses a value of three million dollars in cost-benefit analysis. Murphy and Topel [2005] take as their benchmark a \$6.2 million dollar estimate used by the U.S. Environmental Protection Agency.

For our baseline case, we choose a value of three million dollars, somewhat at the lower end of the estimates. In robustness checks, we report results based on the higher values of four million and five million dollars. It will become clear why we choose the lower end of the range of estimates and how our results would change if even higher estimates were used.

VII.B. The Quality-of-Life Parameters

Our model emphasizes the tradeoff between consumption and quantity of life. As a robustness check, we also allow health spending to have a separate effect on the quality of life.

To calibrate the quality-of-life parameters α and σ—recall the utility function specified in (10)—we draw upon the extensive literature on quality-adjusted life years (QALYs); see Fryback et al. [1993] and Cutler and Richardson [1997]. This work focuses on the QALY weight, the flow utility level of a person with a particular disease as a fraction of the flow utility level of a similar person in perfect health. Surveys ask a range of people, including medical experts, what probability p of perfect health with probability $1 - p$ of certain death would make them indifferent to having a given health condition or what fraction of a year of future perfect health would make them indifferent to a year in that condition. Both of these measures correspond to the relative flow utility in our framework.

Cutler and Richardson [1997] estimate QALY weights by age. With newborns normalized to have a weight of unity, they find QALY weights of 0.94, 0.73, and 0.62 for people of ages 20, 65, and 85, in the year 1990. We use these weights to estimate α and σ based on the following two equations:

$$\frac{u(c_t,\, x_{20,t})}{.94} = \frac{u(c_t,\, x_{65,t})}{.73} = \frac{u(c_t,\, x_{85,t})}{.62},$$

for $t = 1990$. Because the value of life itself depends on these parameters, we simultaneously reestimate the utility intercept b to match the benchmark three million dollar value of life. The

TABLE II
PARAMETER VALUES FOR DIFFERENT SIMULATION RUNS

Baseline Parameter Values

Scenario	Utility curvature parameter γ	Empirical value of statistical life, 2000	Quality of life parameter α	Growth rate of z_t	Fraction of mortality trend from tech. and spending (μ)	Intercept in utility b
1	2	\$3 million	0	0	2/3	26.00

Robustness Checks

Scenario	Key change from baseline	Explanation of change	Intercept, b
2	$\gamma = 2.5$	Vary utility curvature parameter γ	22.12
3	$\gamma = 1.5$	″	30.53
4	$\gamma = 1.01$	″	131.87
5	VSL = \$4 million	Vary empirical value of life	34.35
6	VSL = \$5 million	″	42.70
7	$\alpha = 2.396$, $\sigma = 1.051$	Allow quality of life effect	66.27
8	$g_z = .01$	Vary production of health	25.96
9	$\mu = 1/2$	″	25.92

Note: The first section of the table shows the values of various key parameters in our baseline simulation. The remainder of the table shows how parameters are varied one at a time in our robustness checks. The last column of the table reports the estimated value of the intercept in the utility function, b, obtained by matching the specified value of life for 35–39-year-olds in the year 2000.

resulting estimates are $\alpha = 2.396$, $\sigma = 1.051$, and $b = 66.27$. With three equations and three unknowns, estimation is a matter of solving for the values, so there are no standard errors.

In addition to the QALY interpretation, these numbers can be judged in another way. They imply that a sixty-five-year-old would give up 82 percent of her consumption, and an eighty-five year-old would give up 87 percent of her consumption to have the health status of a 20 year-old. The intuition behind these large numbers is the sharp diminishing returns to consumption measured by γ. To explain what may seem to be a small difference in relative utilities of .94 versus .73 requires large differences in consumption. Health is extremely valuable.

VII.C. Summary of Parameter Choices

Table II summarizes our choices of parameter values, both for the benchmark case and the various robustness checks discussed earlier.

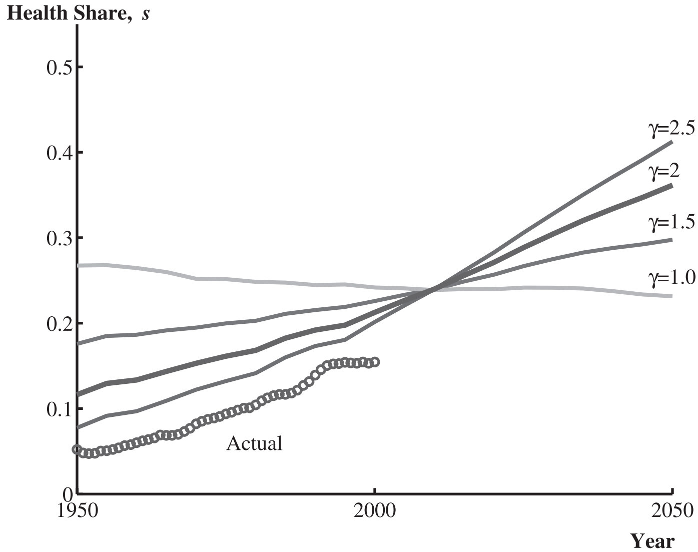

FIGURE V
Simulation Results: The Health Share of Spending

Note: Circles "o" show actual data for the health share. Solid lines show the models predictions under the baseline scenario ($\gamma = 2$) and for alternative choices of the utility curvature parameter. See Table II for other parameter values.

VIII. Solving the Model

We now solve the model over the years 1950 through 2050 for each of our nine scenarios. For the historical period 1950–2000, we take resources per person, y, at its actual value. For the projections into the future, we assume income continues to grow at its average historical rate of 2.31 percent per year. The details for the numerical solution of the model are available from either author's website.

Figure V shows the calculated share of health spending over the period 1950 through 2050 in the first four scenarios, those where γ is allowed to vary from 1.01 to 2.5. A rising health share is a robust feature of the optimal allocation of resources in the health model, as long as γ is not too small. As suggested in our simple model—for example, see (8)—the curvature of marginal utility, γ, is a key determinant of the slope of optimal health spending over time. If marginal utility declines quickly so that γ is high, the optimal health share rises rapidly. This growth in health spending reflects a value of life that grows faster than income. In fact, in the simple model, the value of a year of life is roughly proportional to c^{γ}, illustrating the role of γ in governing the slope of the optimal health share over time.

For near-log utility (where $\gamma = 1.01$), the optimal health share declines. The reason for this is the declining elasticity of health status with respect to health spending in our estimated health production technology (recall Figure III). In this case, the marginal utility of consumption falls sufficiently slowly relative to the diminishing returns in the production of health that the optimal health share declines gradually over time.[9]

Figure VI shows optimal health spending when other baseline parameter values are changed. The changes considered in this figure essentially change the *level* of optimal health spending, while the utility curvature parameter of the previous figure governs the slope. Allowing for a higher empirical value of life in the year 2000 or allowing quality of life to enter utility raises optimal health spending substantially. For example, with a five million dollar value of life, optimal health spending in the year 2000 is 28 percent of GDP, almost double the observed share.

On the other hand, allowing for more of the decline in trend mortality to be explained by factors other than rising health resources leads to a lower optimal health share. For example, allowing technical change in the health sector to be one percentage point faster than in the rest of the economy or reducing the share of mortality decline explained by technical change and resource allocation from 2/3 to 1/2 deliver relatively similar results. In both these cases, less decline in age-specific mortality is due to health spending, so the estimates of θ_a in the production function are smaller. Since health spending runs into sharper diminishing returns, the overall health share of spending is lower. These simulations suggest that the observed share in the year 2000 was close to optimal.

Optimal health shares lie within a fairly large range, reflecting the fairly substantial uncertainty that exists surrounding the key parameters of the model. Nevertheless, an interesting result of these simulations is that optimal health spending is invariably high. This is true for the year 2000 but also out into the future.

9. The careful reader might wonder why all of the optimal health shares intersect in the same year, around 2010. This is related to the fact that the utility intercept b is chosen to match a specific level for the value of life for 35–39 year olds and to the fact that our preferences feature a constant elasticity of marginal utility.

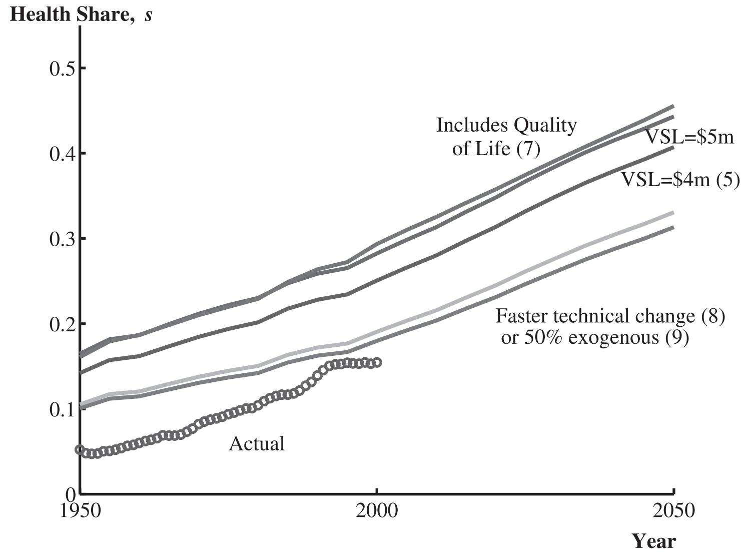

FIGURE VI
Robustness Checks: The Health Share of Spending

Note: Circles show actual data for the health share. Solid lines are predictions of the model under alternative scenarios (the scenario numbers in parentheses correspond to those reported in Table II). Scenarios 5 and 6 allow the empirical value of life in 2000 to be higher, at four and five million dollars. Scenario 7 allows quality of life terms to enter utility. Scenario 8 assumes that technical change in the health sector is 1 percentage point faster than in the rest of the economy. Scenario 9 assumes that 1/2 of the decline in age-specific mortality (rather than our baseline value of 2/3) is due to technological change and increased resource allocation.

For example, by 2050, optimal health spending as a share of GDP ranges from a low of 23 percent for the case of log utility to a high of 45 percent.

Figure VII examines the variation in health spending at the micro level in our baseline scenario. This figure shows actual and simulated health spending by age, for 1950, 2000, and 2050. A comparison of the results for the year 2000 shows that actual and optimal spending are fairly similar for most ages, with two exceptions. Optimal health spending on the youngest age group is substantially higher than actual spending: given the high mortality rate in this group, the marginal benefit of health spending is very high, as was shown earlier. Similarly, while optimal health spending generally rises until age 80, it declines after that point. It is worth noting in this respect that the underlying micro

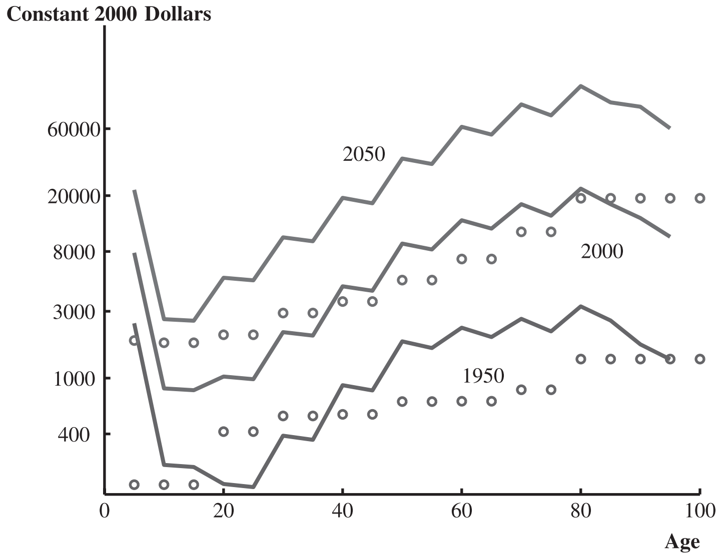

FIGURE VII
Health Spending by Age
Note: Circles denote actual data and solid lines show simulation results for the baseline scenario; see Table II for parameter values.

data we use for health spending groups were from all ages above 75 together.

Figure VIII shows actual and projected levels of life expectancy at birth for all nine of our simulation runs. The first thing to note in the figure is the overall similarity of the life expectancy numbers. Because there are such sharp diminishing returns to health spending in our health production function, relatively large differences in health spending lead to relatively small differences in life expectancy. A second thing to note is that the projected path does not grow quite as fast as historical life expectancy. The reason is again related to the relatively sharp diminishing returns to health spending that we estimate. If the historical rate of increase of 1.7 years per decade were to prevail, life expectancy would reach 85.5 years by 2050; instead it reaches about 81.5 years in our simulations. If anything, it appears our estimation of the health production function builds in too much diminishing returns, which tends to hold down health spending.

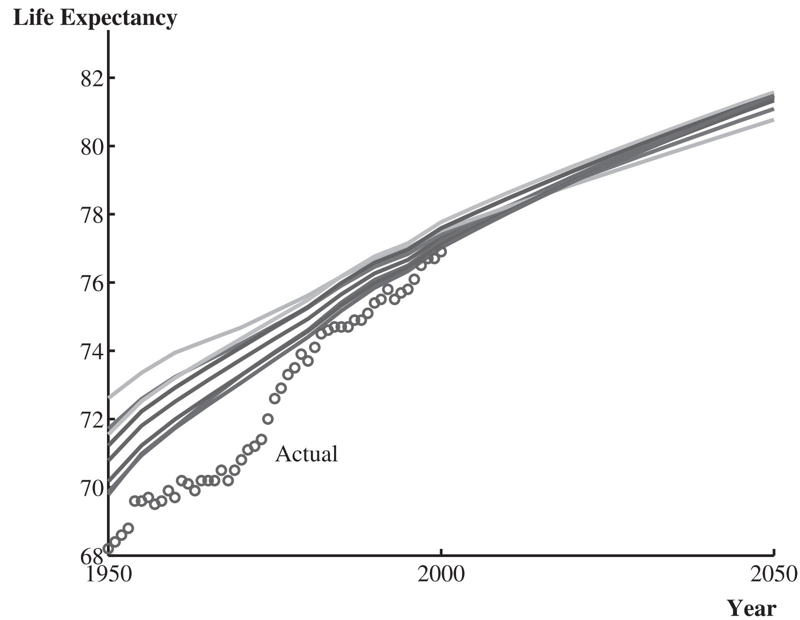

FIGURE VIII
Simulation Results: Life Expectancy at Birth

Note: See notes to Figures V and VI. Life expectancy is calculated using the cross-section distribution of mortality rates at each point in time.

IX. Concluding Remarks

A model based on standard economic assumptions yields a strong prediction for the health share. Provided the marginal utility of consumption falls sufficiently rapidly—as it does for an intertemporal elasticity of substitution well under one—the optimal health share rises over time. The rising health share occurs as consumption continues to rise, but consumption grows more slowly than income. The intuition for this result is that in any given period, people become saturated in nonhealth consumption, driving its marginal utility to low levels. As people get richer, the most valuable channel for spending is to purchase additional years of life. Our numerical results suggest the empirical relevance of this channel: optimal health spending is predicted to rise to more than 30 percent of GDP by the year 2050 in most of our simulations, compared to the current level of about 15 percent.

This fundamental mechanism in the model is supported empirically in a number of different ways. First, as discussed earlier, it is consistent with conventional estimates of the intertemporal elasticity of substitution. Second, the mechanism predicts that the value of a statistical life should rise faster than income. This is a strong prediction of the model, and a place where careful empirical work in the future may be able to shed light on its validity. Costa and Kahn [2004] and Hammitt, Liu, and Liu [2000] provide support for this prediction, suggesting that the value of life grows roughly twice as fast as income, consistent with our baseline choice of $\gamma = 2$. Cross-country evidence also suggests that health spending rises more than one-for-one with income; this evidence is summarized by Gerdtham and Jonsson [2000].

One source of evidence that runs counter to our prediction is the micro evidence on health spending and income. At the individual level within the United States, for example, income elasticities appear to be substantially less than one, as discussed by Newhouse [1992]. A serious problem with this existing evidence, however, is that health insurance limits the choices facing individuals, potentially explaining the absence of income effects. Our model makes a strong prediction that if one looks hard enough and carefully enough, one ought to be able to see income effects in the micro data. Future empirical work will be needed to judge this prediction. A suggestive informal piece of evidence is that exercise seems to be a luxury good: among people with sedentary jobs, high wage people seem to spend more time exercising than low wage people, despite the higher opportunity cost of their time.

As mentioned in the introduction, the recent health literature has emphasized the importance of technological change as an explanation for the rising health share. In our view, this is a proximate rather than a fundamental explanation. The development of new and expensive medical technologies is surely part of the process of rising health spending, as the literature suggests; Jones [2003] provides a model along these lines with exogenous technical change. However, a more fundamental analysis looks at the reasons that new technologies are developed. Distortions associated with health insurance in the United States are probably part of the answer, as suggested by Weisbrod [1991]. But the fact that the health share is rising in virtually every advanced country in the world—despite wide variation in systems for allocating health care—suggests that deeper forces are at work. A fully worked-out technological story will need an analysis on the preference side to explain why it is useful to invent and use new and expensive medical technologies. The most obvious explanation is the model we propose in this paper: new and expensive technologies are valued because of the rising value of life.

Viewed from every angle, our results support the proposition that both historical and future increases in the health spending share are desirable. The magnitude of the future increase depends on parameters whose values are known with relatively low precision, including the value of life, the curvature of marginal utility, and the fraction of the decline in age-specific mortality that is due to technical change and the increased allocation of resources to health care. Nevertheless, we believe it likely that maximizing social welfare in the United States will require the development of institutions that are consistent with spending 30 percent or more of GDP on health by the middle of the century.

HOOVER INSTITUTION, DEPARTMENT OF ECONOMICS, STANFORD UNIVERSITY, AND NATIONAL BUREAU OF ECONOMIC RESEARCH

UNIVERSITY OF CALIFORNIA AT BERKELEY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

REFERENCES

Arthur, W.B., "The Economics of Risk to Life," *American Economic Review,* LXXI (1981), 54–64.

Ashenfelter, Orley, "Measuring the Value of a Statistical Life: Problems and Prospects," NBER Working Paper 11916, 2006.

——, and Michael Greenstone, "Using Mandated Speed Limits to Measure the Value of a Statistical Life," *Journal of Political Economy,* CXII (2004), Part 2, S226–S267.

Barro, Robert J., and Jason R. Barro, "Three Models of Health and Economic Growth," Harvard University working paper, 1996.

Becker, Gary S., Tomas J. Philipson, and Rodrigo R. Soares, "The Quantity and Quality of Life and the Evolution of World Inequality," *American Economic Review,* XCV (2005), 277–291.

Card, David, Carlos Dobkin, and Nicole Maestas, "The Impact of Nearly Universal Insurance coverage on Health Care Utilization and Health: Evidence from Medicare," NBER Working Paper 10365, 2004.

Chay, Kenneth Y., and Michael Greenstone, "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *Quarterly Journal of Economics,* CXVIII (2003), 1121–1167.

Chetty, Raj, "A Bound on Risk Aversion Using Labor Supply Elasticities," *American Economic Review,* forthcoming, XCVI (2006).

Costa, Dora L., and Matthew E. Kahn, "Changes in the Value of Life, 1940–1980," *Journal of Risk and Uncertainty,* XXIX (2004), 159–180.

Cutler, David, *Your Money or Your Life: Strong Medicine for America's Health Care System* (New York, NY: Oxford University Press, 2004).

Cutler, David M., and Elizabeth Richardson, "Measuring the Health of the U.S. Population," *Brookings Papers on Economic Activity, Microeconomics* (1997), 217–282.

——, and Mark McClellan, "Is Technological Change in Medicine Worth It?" *Health Affairs,* XX (2001), 11–29.

——, Angus S. Deaton, and Adriana Lleras-Muney, "The Determinants of Mortality," *Journal of Economic Perspectives,* XX (2006), 97–120.
——, Mark B. McClellan, Joseph P. Newhouse, and Dahlia Remler, "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics,* CXIII (1998), 991–1024.
Dow, William H., Tomas J. Philipson, and Xavier Sala-i-Martin, "Longevity Complementarities Under Competing Risks," *American Economic Review,* LXXXIX (1999), 1358–1371.
Ehrlich, Isaac and Hiroyuki Chuma, "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy,* XCVIII (1990), 761–782.
——, and Yong Yin, "Explaining Diversities in Age-Specific Life Expectancies and Values of Life Saving: A Numerical Analysis," NBER Working Paper 10759, 2004.
Finkelstein, Amy and Robin McKnight, "What Did Medicare Do (And Was It Worth It)?" NBER Working Paper 11609, 2005.
Fryback, Dennis G., E.J. Dasbach, R. Klein, B.E. Klein, N. Dorn, K. Peterson, and P.A. Martin, "The Beaver Dam Health Outcomes Study: Initial Catalog of Health-State Quality Factors," *Medical Decision Making,* XIII (1993), 89–102.
Gerdtham, Ulf-G., and Bengt Jonsson, "International Comparisons of Health Expenditure: Theory, Data and Econometric Analysis," in *Handbook of Health Economics,* Anthony J. Culyer and Joseph P. Newhouse, eds. (Amsterdam: North-Holland, 2000).
Goldman, L., and E.F. Cook, "The Decline in Ischemic Heart Disease Mortality Rates: An Analysis of the Comparative Effects of Medical Interventions and Changes in Lifestyle," *Annals of Internal Medicine,* CI (1984), 825–836.
Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy,* LXXX (1972), 223–255.
——, "Education and Nonmarket Outcomes," in *Handbook on the Economics of Education,* Eric Hanushek and Finis Welch, eds. (Amsterdam: North-Holland, 2005).
Hall, Robert E., "Intertemporal Substitution in Consumption," *Journal of Political Economy,* XCVI (1988), 339–357.
——, and Charles I. Jones, "The Value of Life and the Rise in Health Spending," NBER Working Paper 10737, 2004.
Hammitt, James K., Jin-Tan Liu, and Jin-Long Liu, "Survival is a Luxury Good: The Increasing Value of a Statistical Life," Harvard University mimeo, 2000.
Heidenreich, Paul A., and Mark McClellan, "Trends in Treatment and Outcomes for Acute Myocardial Infarction: 1975-1995," *American Journal of Medicine,* CX (2001), 165–174.
Jones, Charles I., "Why Have Health Expenditures as a Share of GDP Risen So Much?" U.C. Berkeley working paper, 2003.
Lucas, Deborah, "Asset Pricing with Undiversifiable Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle," *Journal of Monetary Economics,* XXXIV (1994), 325–342.
McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse, "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables," *Journal of the American Medical Association,* CCLXXII (1994), 859–866.
Meara, Ellen, Chapin White, and David M. Cutler, "Trends in Medical Spending by Age, 1963–2000," *Health Affairs,* XXIII (2004), 176–183.
Murphy, Kevin M., and Robert Topel, "The Economic Value of Medical Research" in *Measuring the Gains from Medical Research: An Economic Approach,* Kevin M. Murphy and Robert Topel, eds. (Chicago: University of Chicago Press, 2003).
—— and ——, "The Value of Health and Longevity," NBER Working Paper 11405, 2005.
Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?," *Journal of Economic Perspectives,* VI (1992), 3–21.
——, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press, 1993).

——, and Lindy J. Friedlander, "The Relationship between Medical Resources and Measures of Health: Some Additional Evidence," *Journal of Human Resources,* XV (1980), 200–218.
Nordhaus, William D., "The Health of Nations: The Contribution of Improved Health to Living Standards" in *Measuring the Gains from Medical Research: An Economic Approach,* Kevin M. Murphy and Robert Topel, eds. (Chicago: University of Chicago Press, 2003).
Rosen, Sherwin, "The Value of Changes in Life Expectancy," *Journal of Risk and Uncertainty,* I (1988), 285–304.
Schelling, Thomas C., "The Life You Save May Be Your Own," in *Problems in Public Expenditure Analysis,* Samuel B. Chase, ed. (Washington D.C.: Brookings Institution, 1968).
Shepard, Donald S., and Richard J. Zeckhauser, "Survival versus Consumption," *Management Science,* XXX (1984), 423–439.
Skinner, Jonathan, Elliott Fisher, and John E. Wennberg, "The Efficiency of Medicare," NBER Working Paper 8395, 2001.
Usher, Daniel, "An Imputation to the Measure of Economic Growth for Changes in Life Expectancy," in *The Measurement of Economic and Social Performance,* M. Moss, ed. (New York: National Bureau of Economic Research, 1973).
Viscusi, W. Kip, and Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty,* XXVII (2003), 5–76.
Weisbrod, Burton A., "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment," *Journal of Economic Literature,* XXIX (1991), 523–552.
Wennberg, John E., Elliott S. Fisher, and Jonathan S. Skinner, "Geography and the Debate over Medicare Reform," *Health Affairs,* February 13 (2002), W96–W114.