Why Have Health Expenditures as a Share of GDP Risen So Much?

Charles I. Jones*

Department of Economics, U.C. Berkeley and NBER
E-mail: chad@econ.berkeley.edu
http://elsa.berkeley.edu/˜chad

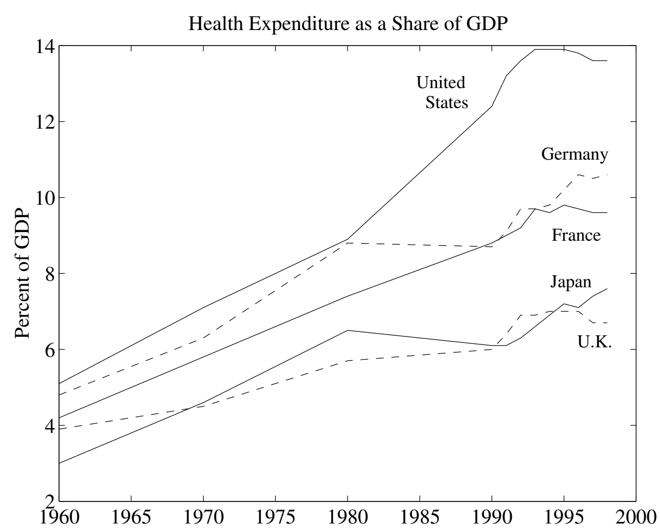
September 30, 2002 – Version 1.5

Aggregate health expenditures as a share of GDP have risen in the United States from about 5 percent in 1960 to nearly 14 percent in recent years. Why? This paper explores a simple explanation based on technological progress. Medical advances allow diseases to be cured today, at a cost, that could not be cured at any price in the past. When this technological progress is combined with a Medicare-like transfer program to pay the health expenses of the elderly, the model is able to reproduce the basic facts of recent U.S. experience, including the large increase in the health expenditure share, a rise in life expectancy, and an increase in the size of health-related transfer payments as a share of GDP.

Key Words: health economics, technological change
JEL Classification: I1, O40

* I would like to thank David Cutler, Ryan Edwards, Victor Fuchs, Ted Keeler, Ron Lee, Mark McClellan, Jim Poterba, David Romer, Emmanuel Saez, Alwyn Young, and participants of seminars at the San Francisco Fed, U.B.C., U.C. Berkeley, U.C. Davis, U.C. Riverside, and U.S.C. for very helpful comments. Meredith Beechey and Jesse Czelusta provided excellent research assistance. Financial support from a John M. Olin Foundation Faculty Fellowship and an Alfred P. Sloan Research Fellowship is gratefully acknowledged. I would also like to thank the Federal Reserve Bank of San Francisco for hosting me as a visiting scholar during portions of this research.

FIGURE 1. OECD Health Expenditures as a Share of GDP



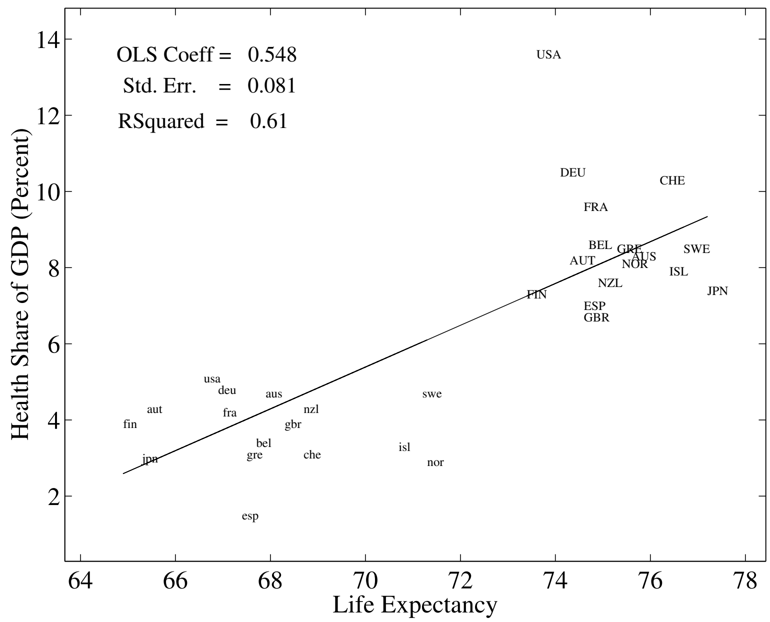
Source: OECD (2000), Table 16.

1. INTRODUCTION

In 1960, aggregate health expenditures in the United States were 5.1 percent of GDP; by 1997, the health expenditure share stood at 13.6 percent. Why have health expenditures as a share of GDP risen so dramatically?

One conventional explanation holds that U.S. government policy is responsible. Changes in transfer programs such as Medicare and Medicaid have possibly led to the increased spending on health. One problem with this explanation, however, is that similar increases in health expenditures as a share of GDP have been observed throughout the OECD, as shown in Figures 1 and 2. It is true, as is often noted, that the United States spends a larger fraction of its income on health than other countries, but the recent tendency of this expenditure ratio to increase seems to be a widespread

FIGURE 2. OECD Health Share vs. Life Expectancy, 1960 and 1997



Notes: Observations for 1960 are plotted by country code with lowercase letters; uppercase letters correspond to observations for 1997. Data from OECD (2000), Tables 7 and 16. Life Expectancy is for males only.

phenomenon. The typical change across the OECD is an increase in the health expenditure share from about 4 percent of GDP to about 8 percent.

Another possible explanation is the cost disease story of Baumol (1967). According to this explanation, health care, like other services, uses labor intensively and exhibits relatively low productivity growth. This leads to a rising relative price of health care and, depending on preferences, to a rising share of expenditures. Superficially, this argument appears somewhat plausible. Triplett and Bosworth (2000) report that measured labor productivity in health services actually declined at an annual rate of 2.2 percent between 1987 and 1997 while multifactor productivity declined at a rate of 2.6 percent, continuing negative trends that date back at least until 1973. They also report that the medical care component of the CPI grew at a rate of 6.5 percent between 1985 and 1995, while overall CPI inflation averaged only 3.6 percent.

On the other hand, the general cost disease explanation for health care rings hollow in the face of rapid technological advances such as the discovery of new drugs, diagnostic equipment, and medical procedures that appear to be occurring throughout the health sector. Moreover, these official statistics have been criticized in several recent papers. Case studies suggest that the CPI medical price index overstates inflation in medical care, with the implication that productivity growth is likely understated. Cutler, McClellan, Newhouse and Remler (1998) report that between 1983 and 1994, the quality-adjusted price of heart attack treatments increased at an annual rate of only 2.3 percent, substantially less than an input price index designed to mimic the medical care CPI methodology. In real terms, the treatment price actually declined at an annual rate of 1.1 percent. Shapiro, Shapiro and Wilcox (1999) examine the treatment price for cataracts between 1969 and 1994. While a CPI-like price index increased at an annual rate of 9.2 percent over this period, their alternative price index, only partially incorporating quality improvements, grew only 4.1 percent per year, falling relative to the total CPI at a rate of about 1.5 percent per year. Berndt, Bir, Busch, Frank and Normand (2000) estimate that the price of treating incidents of acute phase major depression declined in nominal terms by between 1.7 percent and 2.1 percent per year between 1991 and 1996.

Triplett (forthcoming) shows how changes in such price indexes can impact measured productivity. As an exercise, he applies a price index that shows zero inflation in medical care to the entire health sector. With this change, multifactor productivity in health services increases at a rate of 5.6 percent per year between 1987 and 1997, rather than declining at 2.6

percent. Collectively, these studies strongly suggest that health care is not in the same league as the performance of Mozart symphonies when it comes to technological progress, sharply calling into question a cost-disease based explanation.

Newhouse (1992) provides a nice survey of possible explanations for the rise in health expenditures as a share of GDP. In addition to changes in U.S. policy and the Baumol cost-disease story, Newhouse considers and rejects explanations driven by the increase in health insurance, increased income and health as a luxury good, and supplier-induced demand for medical care. For example the income effects explanation is problematic in large part because estimates of the income elasticity of demand for medical care in the United States, holding insurance constant, are only about 0.2 to 0.4.

After reviewing in detail these possible explanations, Newhouse finds that he is left with a large residual: these explanations account for only a minority of the increase in health expenditures, certainly less than half and perhaps less than a quarter. Instead, Newhouse argues that "I believe the bulk of the residual increase is attributable to technological change, or what might loosely be called the march of science and the increased capabilities of medicine" (p. 11).

This paper constructs a simple model to examine the effects of "the march of science" on health expenditures. Technological advances allow diseases and other health problems that could not be cured at any price in the past to be treated effectively today. Consider the development of MRIs, arthroscopic medical procedures, antibiotics, angioplasty, and drugs that treat depression. Before these and other technologies were discovered, health expenditure shares were low by default. The gradual discovery of treatments over time directly increases the expenditure share by permitting the desired health expenditures to occur.

Of course, a countervailing force is that technological progress continually reduces the cost of any given treatment, as the studies reviewed earlier suggested. MRIs and heart attack treatments — once adjusted for quality — are much cheaper today than a decade ago.

Another important component of the model is a Medicare-like transfer program. As modeled, this program features an endogenous tax rate that adjusts to ensure that the elderly live as long as is technologically feasible. This transfer system and technological progress turn out to be a powerful combination. Unchecked, this combination produces a rapidly growing health share that can eventually consume an arbitrarily large fraction of GDP. A critical parameter of the model then turns out to be the maximum transfer rate that society is willing to tolerate.

The paper is organized in the following way. Section 2 develops the basic model, while Section 3 adds the transfer program. Section 4 shows that the model is consistent with a range of evidence related to health spending and, for plausible parameter values, is able to match the large increase in the aggregate health spending share of GDP observed in the United States. Section 5 establishes the robustness of the results, and Section 6 offers some concluding remarks.

2. BASIC ACCOUNTING IN A SIMPLE MODEL

In this section we present a model meant to capture some of the facts related to health expenditures. We then use the model to conduct an accounting exercise to shed light on why health expenditures as a share of GDP might rise.

2.1. Setup

A rich model of health expenditures might take a Markov transition approach in which the state of each individual is indexed by age and health status. Health status might include a number of different diseases, as well as "healthy" and "dead." The Markov transition probabilities across health status are affected by age and by various treatments, which are in turn evolving over time because of technological progress. New treatments are very costly when first developed, but technological progress gradually reduces this cost and raises life expectancy.

In this approach, one could imagine simulating the model and then sorting the population at each date, ex post, according to life expectancy. On average, people close to dying would have high health expenditures, while those far from death would have low health expenditures, a prediction given empirical support by Fuchs (1984). In addition, health expenditures by people in the last year of life would be rising over time as those are exactly the people who would be most likely to use the expensive technologies at the frontier (almost by definition: people with one year of life remaining are most likely to be people affected by serious health problems that do not have easy cures).

While no substitute for writing down and exploring this rich model, this paper takes a more reduced-form approach as its point of departure. In particular, we assume that health expenditures vary across people at a point in time according to their remaining life expectancy:

$$h(\ell, t) = h(0, t)e^{-\theta\ell}, \tag{1}$$

where $h(\ell, t)$ is health expenditures by someone with life expectancy ℓ at time t. We assume $\theta > 0$; in fact, the model is written so that all parameters are assumed to be positive. Motivated by the discussion above, the cross-section of health expenditures is a decreasing function of life expectancy at any point in time. Notice that this formulation allows people of different ages to have the same life expectancy and the same health expenditures. For example, a pre-term infant with serious complications, a 50-year who just had a heart attack, and an 85-year old reaching the end of life could all have the same life expectancy. A simplifying assumption of the model is that they also have the same health expenditures.

The endpoint of this cross-sectional distribution, $h(0,t)$, represents health expenditures by a person at the end of life, and is given by

$$h(0,t) = \tilde{h}_0 e^{\tilde{\mu}\bar{a}(t)}, \tag{2}$$

where $\bar{a}(t)$ is average life expectancy in the population at time t. Health expenditures at the end of life are an increasing function of economy-wide life expectancy. Being in your last year of life in 1950 is different from being in your last year of life in 2002. Technological progress has allowed us to extend life by more than a decade over this half-century, but the medical advances that make this possible are costly. People who could not be kept alive in 1950 are able to live today because of these technological improvements. The parameter $\tilde{\mu}$ measures the rate at which these costs rise as life expectancy advances.

The discovery of new cures, treatments, and diagnostic techniques raises economy-wide life expectancy over time. As another simplification, we treat this technological progress as exogenous and assume that it results in a linear improvement in life expectancy:

$$\bar{a}(t) = \bar{a}_0 + \frac{t}{\alpha}. \tag{3}$$

That is, because of technological progress, life expectancy rises by 1 year every α years. White (2002) shows that this linear formulation provides an excellent fit of the data in high-income countries in the second half of the twentieth century with α approximately equal to 5.

It is helpful to combine equations (1), (2), and (3) to get

$$h(\ell,t) = h_0 e^{\mu t - \theta \ell}, \tag{4}$$

where $h_0 \equiv \tilde{h}_0 e^{\tilde{\mu}\bar{a}_0}$. At a given point in time, health expenditures decline across people as a function of life expectancy at rate θ. For a given life expectancy, health expenditures rise over time at rate $\mu \equiv \tilde{\mu}/\alpha$. While $\tilde{\mu}$ captures the rate at which health costs increase for each year that technological progress expands economy-wide life expectancy, μ is simply the annualized version of this rate: since life expectancy rises at a rate of $1/\alpha$ years every year, health costs in the last year of life rise annually at rate $\mu \equiv \tilde{\mu}/\alpha$.

In modern actuarial tables, life expectancy has a statistical meaning, different from the number of years a typical person might expect to live, which depends on future technological progress. To avoid making assumptions about the future, life expectancy is calculated as the expected length of life for a person born today and forever facing the mortality rates that prevail in the cross-section of ages today. With one change, we will give a similar interpretation to ℓ. The change is that ℓ measures the expected length of life assuming all technologically-available treatments are utilized. In fact, as we will see, some technologies may be so expensive that they will not be utilized and people will die who, at some cost, could live longer.

For the sake of simplicity, we assume that the population is uniformly distributed across life expectancy at each point in time, with $\ell \in [\ell^*(t), \bar{a}(t)]$. The endogenous variable $\ell^*(t)$ would be equal to zero if all available technologies were utilized. It will be greater than zero to the extent that some technologies are not used. For example, suppose that $\ell^* = 2$ at some time. In this case, people die when they, if all treatments were used, could live for two additional years. This might be an outcome if those additional treatments are extraordinarily expensive. We will define

$$a^*(t) \equiv \bar{a}(t) - \ell^*(t). \tag{5}$$

This can be thought of as actual life expectancy at birth in the population at time t, and, since the population is uniformly distributed across ℓ, it denotes the size (measure) of the population at time t. Note that the assumption that the population remains uniformly distributed across ℓ while life expectancy is growing means that there is population growth.

The timing of the model works as follows. Time is discrete. At the beginning of a period, people undertake their health expenditures, and then live for that period. People whose life expectancies in period t are in the range $[\ell^*(t), \ell^*(t)+1)$ die at the end of that period.

Finally, let $Y(t)$ denote aggregate income at time t so that $y(t) \equiv Y(t)/a^*(t)$ is per capita income. We assume $y(t) = y_0 e^{gt}$ so that per capita income grows exogenously at the constant rate g. At this point, we do not need to make any assumptions about the distribution of this income across people.

2.2. Analysis

With this setup, we are ready to solve for aggregate health expenditure as a share of GDP. First notice that total health expenditure, $H(t)$, is given by

$$H(t) \equiv \int_{\ell^*(t)}^{\bar{a}(t)} h(\ell, t) d\ell. \tag{6}$$

Dividing by GDP and recalling that $Y(t) = a^*(t)y(t)$ yields

$$\frac{H(t)}{Y(t)} = \frac{1}{a^*(t)} \int_{\ell^*(t)}^{\bar{a}(t)} \frac{h(\ell, t)}{y(t)} d\ell. \tag{7}$$

That is, the aggregate expenditure share is the average of the individual expenditure shares.[1]

[1] Notice that we are abusing the language a bit with this statement. Because we've made no assumptions about the distribution of income, the correct statement would be that the aggregate share is a *weighted* average of the individual expenditure shares, where

Using equation (4), it is straightforward to show that

$$\frac{H(t)}{Y(t)} = \frac{1}{\theta a^*(t)}\left(\frac{h(\ell^*,t)}{y(t)} - \frac{h(\bar{a}(t),t)}{y(t)}\right) \tag{8}$$

$$= \frac{1}{\theta a^*(t)}\frac{h(\ell^*,t)}{y(t)}(1 - e^{-\theta a^*(t)}). \tag{9}$$

The first line of this expression exploits a simple fact that the integral of an exponential process is proportional to the difference between its value at the two endpoints.[2] The second line takes advantage of the fact that this difference is proportional to the expenditure share by the person closest to death.

Finally, as long as $\theta a^*(t)$ is relatively large, a very good approximation is given by

$$\frac{H(t)}{Y(t)} \approx \frac{1}{\theta a^*(t)}\frac{h(\ell^*,t)}{y(t)}. \tag{10}$$

There are several things to note about this result. First, the aggregate health expenditure share is proportional to the largest quantity of health expenditures in the population divided by per capita GDP. Let us call this term, $h(\ell^*,t)/y(t)$, the health expenditure share at the end of life, because it reflects the health expenditures of individuals with the minimal level of life expectancy. It is slightly misleading to call it an expenditure *share* because it is per capita income in the denominator rather than any individual's income. Still, it is convenient to have a short phrase to describe this quantity.

---

the weights correspond to income shares. That is, we could multiply and divide the term inside the integral by $y(\ell,t)$. Rather than proceed in this way, it is convenient to leave the expression as $h(\ell,t)/y(t)$. We will speak of this as the health expenditure share by someone with life expectancy ℓ, but the income-weight caveat should be kept in mind.

[2]A more intuitive way to view this mathematical fact is to consider the average of an exponential process rather than the sum. If $x(t) = x_0 e^{gt}$, then $\bar{x}(t) \equiv \frac{1}{T}\int_0^T x(t)dt$ is equal to $1/gT \times (x(T) - x(0))$. Rearranging, we see that $g = 1/T \times (x(T) - x(0))/\bar{x}$. The exponential growth rate is equal to the percentage change in the process, where the change is taken relative to, not the starting or ending value, but to the exponential average itself.

Second, the factor of proportionality depends inversely on life expectancy, $a^*(t)$. Holding constant the expenditure share of people near death, an increase in economy-wide life expectancy reduces the aggregate expenditure share. We will call this the *dilution effect*. Recall that the aggregate spending share is the average of the individual spending shares. A higher level of life expectancy, with the highest expenditure share held constant, essentially means a larger measure of low-cost healthy people over which we are averaging. At any point in time, there are always five years worth of people who are five years away from dying. This is true whether economy-wide life expectancy is 60 years or 80 years, but in the latter case, these five years worth of people constitute a smaller fraction of the population.

Third, because of this dilution effect, growth in the aggregate expenditure share requires the expenditure share at the end of life to grow faster than life expectancy. Using equation (4), we can replace this expenditure share with the appropriate exponential terms to get

$$\frac{H(t)}{Y(t)} \approx \frac{1}{\theta a^*(t)} \frac{h_0}{y_0} \, e^{(\mu - g)t} e^{-\theta \ell^*}. \tag{11}$$

A necessary condition for H/Y to grow is that $\mu > g$, i.e. that health expenditures at a given level of life expectancy grow faster than per capita GDP.

Finally, we can use the approximation in equation (11) to get a rough sense of the numbers involved. As discussed in the introduction, between 1960 and 1997, the U.S. health expenditure share rose from 5.1 percent to 13.6 percent, that is by a factor of $13.6/5.1 = 2.7$. Similarly, life expectancy over this period rose from 66.6 years to 73.9 year, or by a factor of $73.9/66.6 = 1.11$, according to the data in Figure 2. To match these numbers, the expenditure share at the end of life must rise by a factor of $2.7 \times 1.11 = 3.0$. Assuming ℓ^* is constant and $g = .018$, this requires a value of $\mu = .047$. That is, health expenditures by people shortly before they die must have been rising at a rate of 4.7 percent per year.

By itself, this result is not surprising — it is simply a matter of accounting. If H/Y has been rising and if health expenditures are distributed exponentially across the population, then the health expenditure share at the end of life must have been rising, and we have simply calculated the rate at which this rise must occur.

The careful reader may now be bothered by something: if the expenditure share of decedents grows without bound, what prevents H/Y from growing beyond 100 percent? The answer, of course, must be that the expenditure share at the end of life cannot grow without bound: at some point, individual health expenditures relative to per capita income must level off. Since we have not discussed the economic forces governing an individual's choice of health expenditure, this result has been suppressed; it will be explored in detail in the next section.

Still, we can pause to see the basic implication of this bound. Suppose that $h(\ell^*, t)/y(t)$ reaches an upper bound and stops growing. From equation (10), the dilution effect will then dominate and H/Y declines, eventually asymptoting to zero. Recall from equation (7) that the aggregate expenditure share is simply the average of the individual expenditure shares. With the top expenditure share bounded, this average is continually diluted by the fact that a rising life expectancy introduces an increasing number of low-cost healthy individuals.

3. PAY-AS-YOU-GO HEALTH CARE FINANCING

To this point, we have made no economic assumptions about how health expenditures are determined at the individual level (at least none beyond our assumption on the distribution across all individuals). The individual expenditure decision is impacted significantly by public and private health insurance and by life cycle considerations. Rather than analyze this complicated choice, we make a number of special assumptions about the economics of individual decision-making. The goal is to build a model that is simple to analyze yet sophisticated enough to produce interesting results.

At some point, individuals approach the end of life and face large health expenses. We assume that these expenses are financed in the following way. First, individuals near the end of life have income and wealth that allows them to spend up to some multiple, σ, of disposable per capita income on health care. For simplicity, the level of this resource constraint is assumed to be the same across people. The parameter σ is a reduced-form way to capture several competing concerns. On the one hand, individuals may have access to health insurance, discussed below, so that they only need to pay some fraction of their health expenses, leaving more resources for consumption and bequests. This suggests that σ might be low. On the other hand, individuals know that they will face large health expenses near the end of life and may save for this purpose. This would make σ high. Clearly, σ could be greater or less than one, and in a richer model would not be invariant to the policy regime: changes in the health care system would change σ. Modeling these sophisticated effects is important. However, it is also complicated, and the results that one obtains are likely to be sensitive to the exact nature of private and public health insurance and to behavioral assumptions. We leave this valuable research to future work and instead focus on the simpler case where σ is a reduced-form parameter, with these caveats kept in mind.

In addition to their own resources, individuals in this model have access to a basic health insurance system. We will think of this as public health insurance, but there is no reason in the model why it could not be run privately. The insurance scheme is financed by a flat income tax (premium) at rate $\tau(t)$. It pays out benefits, $v(\ell, t)$, that depend on health status.

Because of rapid technological progress, the cost of keeping people alive in the last year of life rises more rapidly than per capita income ($\mu > g$). At some point, these costs outstrip an individual's ability to pay for treatments, so that even though the technology exists to prolong life, the cost is greater than the individual can afford. Call the date at which this occurs t^*, and notice that this date solves the following equation

$$h(0, t^*) = \sigma y(t). \tag{12}$$

Prior to date t^*, individuals can pay their own health expenses, and we will assume the insurance scheme does not operate, so that $\tau(t) = 0$ for $t < t^*$. Up until this date, aggregate health expenses are determined by equation (10) with $\ell^*(t) = 0$. That is, people live as long as is technologically possible.

After date t^*, this changes. The technological costs of keeping individuals alive as long as possible may exceed their resources. In the absence of a transfer scheme, such people would die immediately.

Transfers $v(\ell, t)$ are paid to these people in exactly the quantity needed to keep them alive until the budget is exhausted. That is

$$v(\ell, t) = h(\ell, t) - \sigma(1 - \tau(t))y(t), \quad \ell \in [\ell^*(t), \bar{\ell}(t)], \tag{13}$$

where $\bar{\ell}(t)$ denotes the life expectancy of the marginal person requiring a transfer, given implicitly by

$$h(\bar{\ell}(t), t) = \sigma(1 - \tau(t))y(t). \tag{14}$$

In the absence of transfers, all people with life expectancy less than $\bar{\ell}$ would die.

Finally, we assume that the transfer program operates with a balanced budget:

$$\tau(t)a^*(t)y(t) = \int_{\ell^*(t)}^{\bar{\ell}(t)} h(\ell, t) - h(\bar{\ell}(t), t)d\ell. \tag{15}$$

Notice that we've made an important shift in how we interpret $h(\ell,t)$. Previously, h represented actual expenses by people, and we did not concern ourselves with how these expenses were determined. Now we assume that $h(\ell,t)$represents expenditures necessary for life, at least at low ℓ, and we have explained how these necessary expenditures are financed.

We make the following additional assumptions about the insurance institution. After date t^*, transfer payments are used to keep people alive as long as is technologically feasible. That is, equation (15) can be thought of as determining the insurance premium $\tau(t)$ so that $\ell^*(t) = 0$. Finally, we assume that this premium/tax rate is capped at $\bar{\tau} < 1$. In general, there will be a third date t^{**} at which time this bound on the tax rate is binding. In this regime, equation (15) can be thought of as determining $\ell^*(t)$ so that the budget constraint is satisfied with $\tau(t) = \bar{\tau}$. That is, life expectancy is rationed by the willingness of society to pay the large health expenses of those near death.

We pause at this point to highlight the key endogenous variables and the key equations that determine them. These variables are $h(\ell,t), \bar{a}(t), a^*(t), \ell^*(t), \bar{\ell}(t), H(t)/Y(t)$, and $\tau(t)$. The equations that determine these endogenous values are (1), (3), (5), (14), (15), (10), and the conditions described above that set $\tau(t) = 0$ when $t \leq t^*$, $\ell^*(t) = 0$ for $t \leq t^{**}$, and $\tau(t) = \bar{\tau}$ when $t > t^{**}$. Notice that t^* is determined by equation (12) and t^{**} is determined as the date at which $\tau(t^{**}) = \bar{\tau}$ and $\ell^*(t^{**}) = 0$ in equation (15).

4. SIMULATING THE MODEL

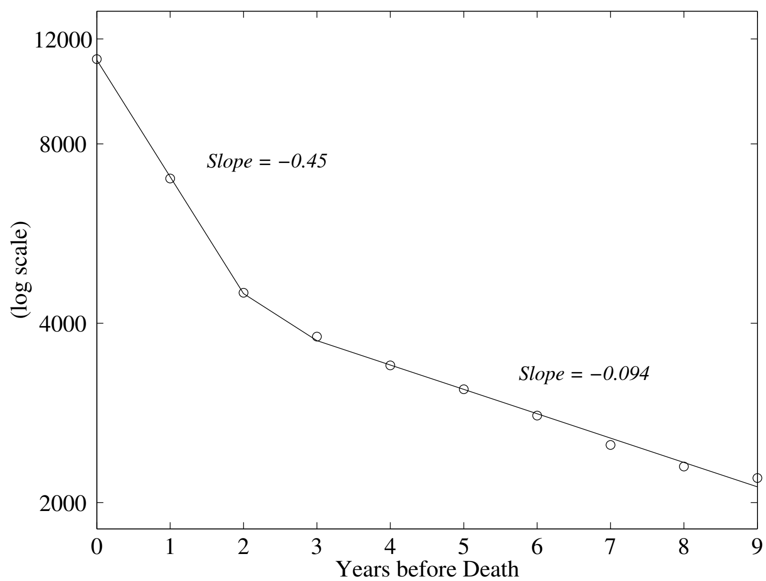
It is useful to pick some parameter values and simulate this model in order to see the kind of behavior that it can exhibit. Of course, the numbers in the simulation are purely suggestive and meant primarily to illustrate the basic forces at work in the model and the extent of their quantitative possibilities; one should not read too much into their exact values.

Values for some parameters are relatively easy to infer from various stylized facts. For the underlying growth rate of per capita income, g, we take a value of 1.8 percent, corresponding roughly to the growth rate of GDP per capita in the United States in the postwar period. The parameter α measures the number of years it takes technological progress to raise life expectancy by a year. In the United States, male life expectancy increased from 66.6 years in 1960 to 73.9 years in 1997, suggesting a value of α of $(1997 - 1960)/(73.9 - 66.6) = 5.07$.

Picking a value for θ gives us the opportunity to examine two different pieces of data. Recall that θ captures the rate at which health expenditures decline with life expectancy in a cross-section of the population. Miller (2001) reports average annual Medicare expenditures per enrollee for 129,166 randomly-chosen people who died in 1989 or 1990, using data from Lubitz, Beebe and Baker (1995). Data from the Continuous Medicare History Sample tracks their Medicare expenditures back to 1974, allowing Miller to look at Medicare expenditures versus years before death.[3] These numbers are plotted in Figure 3. While the model assumes this relation to be log-linear, the data suggest that it is not. Rather, a log-linear spline seems to fit the data better. Medicare expenditures rise at a rate of 9.4 percent per year for people 3 to 10 years away from death. This rate then accelerates sharply to 45 percent per year in the last two years before death.

[3] Since the data correspond to Medicare expenditures over time for people who died in a single year, one might worry that Figure 3 reveals $\mu + \theta$ rather than just θ. Footnote 3 in Miller (2001) reveals that the expenditures are placed in 1990 dollars by multiplying by a factor corresponding to per capita Medicare expenditures in the two years. This essentially undoes the μ effect. In terms of the model, it introduces an extra term corresponding to the proportion by which life expectancy changed between the two years, but this factor is extremely small relative to the θ effects, so I have chosen to report the original Miller data rather than making the correction.

FIGURE 3. Average Annual Medicare Expenditures per Enrollee



Circles represent average expenditures for all enrollees over age 65, taken from Miller (2001), Table 1.

This suggests a value of θ in the model that is some average of 0.094 and 0.45.[4]

Fortunately, θ can also be calculated using a second piece of data. It is straightforward to show (see the Appendix A) that the fraction of total health expenditures accounted for by people in the last year of life is approximately equal to $1 - e^{-\theta} \approx \theta$.[5] Intuitively, this occurs because the distribution of health expenditures across people is exponential with parameter θ, and this entire distribution shifts in the same way over time. Lubitz and Riley

[4]The model could be augmented to capture this non-exponential pattern by adding a term $\beta y(t)$ to the basic spending relation in equation (4). This term incorporates health expenditures that are largely independent of life expectancy, such as healthy pregnancy and childbirth expenses, vaccinations, athletic injuries, etc.

[5]In the simulations below, this share is indeed very close to $1 - e^{-\theta}$, confirming the validity of the approximation.

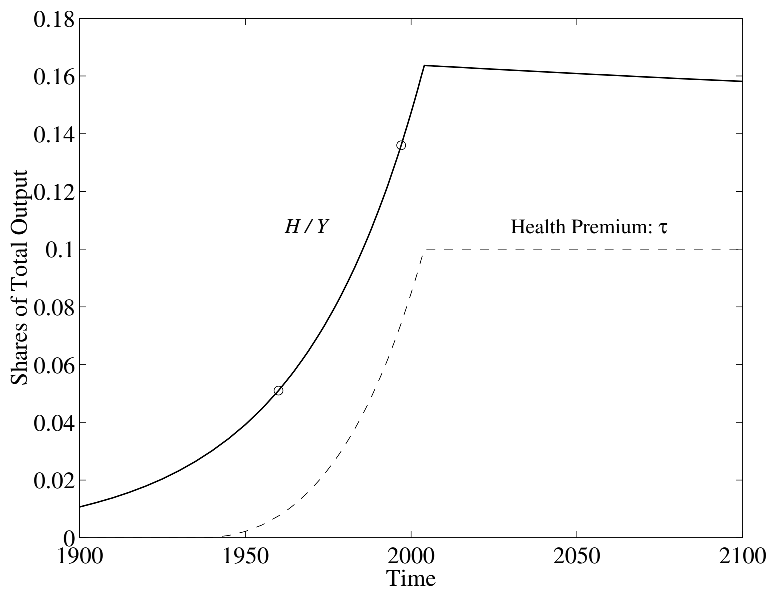
(1993) document that the share of Medicare payments on behalf of people in the last year of life is relatively stable, fluctuating between 27.2 and 30.6 percent over the period 1976 to 1988. This is useful in two ways. First, Lubitz and Riley emphasize that their result is somewhat surprising: many people had suspected that the large rise in Medicare expenditures was disproportionately associated with people in the last year of life. They show that this is not the case, and the model also predicts that this fraction should be stable. Second, we can use these numbers to help calibrate θ. Presumably the fraction of total health expenditures occuring in the last year of life is smaller than the fraction for Medicare.[6] Based on this fraction and on the Miller data discussed above, we choose $\theta = 0.2877$, so that $1 - e^{-\theta} = 0.25$, that is, one quarter of total health expenditures are assumed to go for people in the last year of life. The robustness of the results to changes in this parameter value will be considered below.

The parameter σ captures in a reduced-form fashion the budget constraint limit on health expenditures. We assume, arbitrarily, that $\sigma = 1/2$ so that individuals can spend no more than one half of per capita income on health. Many of the results below do not involve this parameter at all. The exception is in determining the level of transfer payments, where the sensitivity is direct and obvious: for any given aggregate level of spending, a higher value of σ will reduce the amount of transfers required to support that level of spending.

Finally, parameter values for μ and h_0/y_0 have been chosen so that the simulation results match exactly U.S. aggregate health expenditures as a share of GDP in 1960 and 1997. This leads to values of $\mu = .0473$ and $h_0/y_0 = .9771$. Obviously, the value of μ is a critical parameter governing the amount by which H/Y rises, as shown earlier in equation (11). We will

[6]This is not entirely clear, however, since many nursing home expenses are not covered by Medicare.

FIGURE 4. Simulation Results for $\bar{\tau} = .10$



Note: The two circles in the figure indicate data points for the United States: an expenditure share of 5.1 percent in 1960 and 13.6 percent in 1997, when life expectancy was 66.6 and 73.9 years, respectively. See notes to Table 1.

discuss later in Section 5 whether or not this particular value is plausible. For the moment, the simulation should simply be viewed as illustrative.

Figure 4 shows the aggregate health expenditure share and the transfer rate τ over time for a simulated economy. Statistics on H/Y, life expectancy, and the transfer rate τ for this simulated economy are reported in Table 1.

Several results from the simulation deserve mention. First, the aggregate health expenditure share grows rapidly until $\tau(t) = \bar{\tau}$, here set (arbitrarily) to 10 percent. As long as $\ell^* = 0$, that is as long as all technologies are utilized, the exponentially rising cost of the frontier technology causes the aggregate health share to increase. Mathematically, this result is already

TABLE 1.
Simulation Results

Year	H/Y	τ	$\bar{a}$	a^*	ℓ^*	Fraction Subsidized	$h(\ell^*)/y$
				$\bar{\tau} = .10$			
1900	0.011	0.000	54.8	54.8	0.000	0.000	0.168
1920	0.018	0.000	58.7	58.7	0.000	0.000	0.302
1940	0.030	0.000	62.7	62.7	0.000	0.005	0.544
1960	0.051	0.008	66.6	66.6	0.000	0.035	0.977
1970	0.066	0.017	68.6	68.6	0.000	0.050	1.310
1980	0.087	0.032	70.5	70.5	0.000	0.064	1.756
1990	0.113	0.054	72.5	72.5	0.000	0.077	2.355
1997	0.136	0.074	73.9	73.9	0.000	0.086	2.891
2000	0.147	0.085	74.5	74.5	0.000	0.090	3.157
2010	0.163	0.100	76.5	75.9	0.598	0.094	3.564
2050	0.161	0.100	84.4	79.8	4.551	0.087	3.693
2100	0.158	0.100	94.2	84.7	9.499	0.079	3.854
				$\bar{\tau} = .05$			
1900	0.011	0.000	54.8	54.8	0.000	0.000	0.168
1920	0.018	0.000	58.7	58.7	0.000	0.000	0.302
1940	0.030	0.000	62.7	62.7	0.000	0.005	0.544
1960	0.051	0.008	66.6	66.6	0.000	0.035	0.977
1970	0.066	0.017	68.6	68.6	0.000	0.050	1.310
1980	0.087	0.032	70.5	70.5	0.000	0.064	1.756
1990	0.108	0.050	72.5	72.4	0.149	0.075	2.256
1997	0.108	0.050	73.9	73.1	0.843	0.074	2.269
2000	0.108	0.050	74.5	73.4	1.141	0.073	2.274
2010	0.107	0.050	76.5	74.3	2.133	0.072	2.292
2050	0.105	0.050	84.4	78.3	6.104	0.066	2.363
2100	0.102	0.050	94.2	83.1	11.073	0.061	2.450

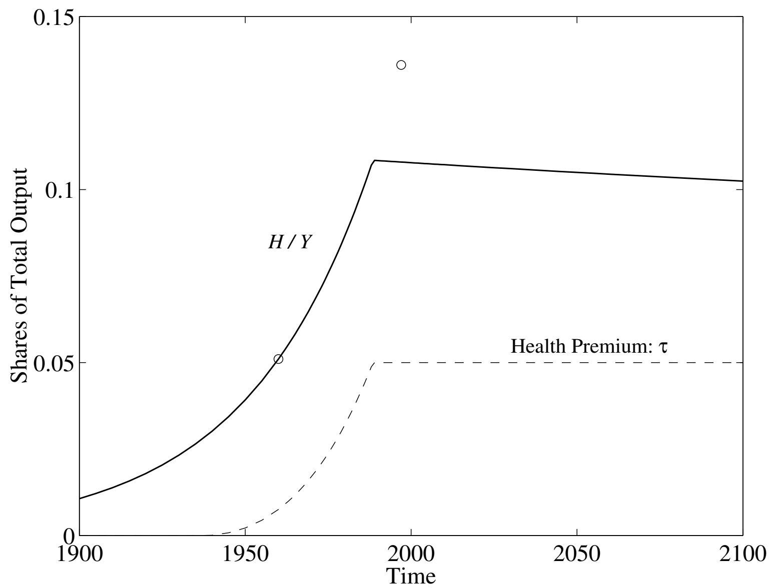
Note: Simulation results with the following parameters: $\alpha = 5.0685, \mu = .0473, \sigma = 1/2$, $g = .018$, $y_0 = 1$, $h_0/y_0 = .9771$. "Fraction Subsidized" denotes the fraction of the population receiving a health expenditure subsidy, $(\bar{\ell} - \ell^*)/a^*$.

familiar from equation (11): the health expenditure share of decedents grows at rate μ while life expectancy grows only linearly. Second, there exist parameter values such that the simulation matches exactly the level of H/Y in 1960 and 1997, and the rate of increase in between.

Third, $\bar{\tau}$ reflects the maximum level of resources that this economy is willing to transfer from healthy people to sick people. In practice, $\bar{\tau}$ is the outcome of a very complicated political economy problem, and it could change over time as the political calculus changes. However, following our basic approach in this model, we treat this political outcome parametrically. What one sees from the simulation is that once this political constraint is binding, the dynamics of H/Y change drastically. Before the constraint binds, the aggregate health expenditure share grows rapidly as society transfers an increasing fraction of resources from the healthy to the sick. Once this source of additional funding is exhausted — which in this simulation happens in the year 2004 — the rise in H/Y comes to a halt and in fact the aggregate health share begins a slow, steady decline. The source of this decline was apparent back in equation (11): if growth in the expenditure share at the end of life is halted (or at least slowed considerably), the dilution effect takes over. Because of technological progress, the rise in life expectancy means that the population consists of an increasing fraction of healthy people. In the simulation in Figure 4, this dilution effect dominates after the $\bar{\tau}$ constraint is hit. H/Y declines, asymptoting to the value given by $\bar{\tau}$ itself, since by construction society is always spending at least this much of its income on health.

But if the individuals near death are always spending $\sigma(1-\tau)y$ of their own resources on health, why does the aggregate share asymptote to $\bar{\tau}$? The answer is suggested by the "Fraction Subsidized" column in Table 1. In particular, notice that the fraction of the population benefitting from the subsidy first rises and then declines after the $\bar{\tau}$ constraint becomes binding.

FIGURE 5. Simulation Results for $\bar{\tau} = .05$



Note: See notes to Figure 4.

That is, the full amount of the subsidy becomes increasingly concentrated within the population after the constraint is met, so that health expenditures financed by the resources of these people shrinks to zero.

The policy parameter $\bar{\tau}$ obviously plays a key role in the simulation results. If it were set higher, the rapid growth of H/Y would continue longer. And if it were set lower, H/Y would peak sooner.

There is a sense in which the "multiplier" associated with a change in $\bar{\tau}$ is large. Figure 5 examines the consequences of cutting $\bar{\tau}$ in half, from .10 to .05, starting from the beginning of the simulation. In this case, the constraint that $\tau(t) < \bar{\tau}$ starts to bind in 1989 rather than in 2004, and the aggregate health expenditure share peaks in that year at 10.8 percent, or 5.8 percent above $\bar{\tau}$. Notice that this is less than the gap when $\bar{\tau} = .10$, which equals

16.3-10=6.3 percent: the transfer rate is associated with a multiplier that is greater than one. The intuition for this result is that transfer payments increase the number of people near death, and these people themselves spend some of their own resources on health. As one would expect, a higher value of σ will increase this multiplier.

The substantial reduction in health expenditures as a share of GDP associated with the lower transfer rate of $\bar{\tau} = .05$ has a relatively small effect on life expectancy, as can be seen in the lower panel of Table 1. In the year 1997, for example, life expectancy in the original simulation matched U.S. life expectancy at 73.9 years. In the simulation with $\bar{\tau} = .05$, life expectancy in 1997 is just under one year less at 73.1 years. About 2.8 percent of GDP goes to pay for the health care of people living the additional 0.8 years, suggesting a ratio of 2.8/0.8=3.5 percent of GDP for each additional year of life expectancy.

This number can be calculated analytically in the following way. Adding one year of life expectancy to the population raises total health expenditures by

$$\int_{\ell^*-1}^{\ell^*} h(\ell, t) d\ell = \frac{1}{\theta} h(\ell^*, t)(e^{\theta} - 1).$$

Dividing this by total GDP, one sees that the additional cost is approximately $H/Y \times (e^{\theta} - 1)$. With $\theta = .2877$ and $H/Y = 10.8$, we get 3.6 percent of GDP.

Alternatively, if we start with the observed $H/Y = .136$ in 1997, this suggests a cost of $.136 \times e^{.2877-1} = .045$, or 4.5 percent of GDP for each additional year of life expectancy at the margin in 1997. Is this a lot or a little? Obviously the answer depends on preferences and other complicated considerations. However, we can shed light on this question in two ways. In 1997, per capita GDP was about \$28,000, and life expectancy was $a^*(1997) = 73.9$ years. Multiplying these two numbers together yields $Y(t) = a^*(t)y(t)$ of about 2 million dollars, and 4.5 percent of this is \$93,000. This number is right in the middle of economists' estimates of the value of one year of life, which range from about \$50,000 to about \$150,000 according to Cutler and McClellan (2001).

Another simple calculation that can shed light on these numbers is the following. Suppose that the utility of a person living T years is given by $\int_0^T e^{-\rho t} u(c_t)dt$, where $u(c) = c^\gamma$, $c_t = c_0 e^{gt}$, and $\gamma \in (0, 1)$. What fraction of her consumption is this person willing to give up each year in order to live an additional year? That is, for what value of x does the following equality hold:

$$\int_0^T e^{-\rho t} u(c_t)dt = \int_0^{T+1} e^{-\rho t} u((1-x)c_t)dt. \tag{16}$$

The answer is approximately given by $x \approx 1/(\gamma T)$, which interestingly doesn't depend on many of the parameters of the problem. With $T = 73.9$, one finds that for values of γ less than 0.30, x is greater than 4.5 percent.[7]

The range of uncertainty surrounding these calculations is large. Still the calculations are useful in that they suggest that spending 13.6 percent of GDP on health care is not obviously crazy. This result is consistent with other calculations in the literature. Cutler and McClellan (2001) calculate that between 1950 and 1990, the present discounted value of the amount an individual could expect to spend on medical care over her entire life rose by \$35,000. During the same period, life expectancy increased by about seven years, which Cutler and McClellan claim is worth a present

[7]This calculation is more problematic that it at first appears. First, there is an important — but not obviously correct — normalization that death is equivalent to zero consumption. Second, empirical work on consumption or finance typically works with preferences of the form $u(c) = c^\gamma/\gamma$. In this case, $1-\gamma$ corresponds to the coefficient of relative risk aversion, which is typically estimated to be greater than one. Second, $1/(1-\gamma)$ corresponds to the intertemporal elasticity of substitution. Estimates of this parameter are typically in the range of 1/3 to 1/2 (e.g. Ogaki and Reinhart (1998)), again suggesting a negative value for γ. It is not clear how to reconcile these empirical estimates with preferences that require $u(c) > 0$.

value of $130,000. They conclude that if health expenditures explain more than about a quarter of the rise in life expectancy, then these benefits from increased health spending exceeded the costs.[8]

A recent report by a panel of experts on the technical merits of the Medicare Trustees' financial projections provides some startling forecasts related to health expenditures as a share of GDP.[9] For example, the middle-range estimate — which assumes health expenditures grow at a rate 1 percent faster than GDP — forecasts an aggregate expenditure share of 25% in 2050 and 38% in 2075, and the report states that "The Panel does not view this [latter] figure as implausible" (p. 39).

The model sheds light on this forecast in two ways. First, the model is potentially consistent with this forecast as long as society is willing to continually transfer more and more resources to people near the end of life, i.e. as long as $\bar{\tau}$ is sufficiently large. On the other hand, the model suggests a reason to be cautious about these kind of forecasts: the dynamics of H/Y look very different before and after the $\bar{\tau}$ constraint becomes binding. If society decides to cap the transfer rate, these forecasts could be far from the mark.

Second, the calculations on the value of a year of life implicitly provide an upper bound on the fraction of GDP that might optimally be spent on medical care. For example, in the first calculation the value of $93,000 was right in the middle of the typical estimates of the value of a year of life, perhaps indicating that the upper bound on H/Y is not too far away. Determining an upper bound from the second calculation — the one that found $x = 1/(\gamma T)$ — is more difficult for reasons noted earlier. However, one should note that the upper bound declines as life expectancy increases,

[8]Other related calculations on the gains from the increase in life expectancy can be found in Murphy and Topel (2002) and Nordhaus (2002).

[9]See Technical Review Panel on the Medicare Trustees Reports (2000).

suggesting that a rising life expectancy makes it more likely that society may wish to impose a cap on the transfer rate.

One of the facts highlighted in the introduction was the extent to which technological progress reduces the cost of treating specific health problems, including heart attacks, cataracts, and depression. How is this aspect of technological progress captured by the model?

To answer this question requires a model in which diseases and health problems enter explicitly. We provide such a model in Appendix B, develop its implications for quality-adjusted treatment prices, and show how this framework leads to the same specification for health expenditures derived in the original model in equation (4). However, the implied rate at which treatment prices decline can be seen, at least to some extent, in the model we have already developed.

Suppose that there is some medical condition (say a heart attack) that in 1997 causes an individual to have a life expectancy of 3 years. According to the model, health expenditures by this person are $h(3, 1997)$. Now consider someone who has a heart attack one year later. How much does it cost to provide the 1997 treatment to this person? In the model, answering this question requires determining the value of ℓ for the heart attack victim in 1998. Clearly, $\ell = 3$ would be the wrong answer. Recall that $h(3, t)$ rises over time at rate μ because of the technological progress that lets sicker people live longer; a person with 3 years life expectancy in 1998 is much more costly to treat than a person with 3 years life expectancy in 1950 (or 1997). Technological progress annually raises life expectancy by $1/\alpha$ years. A person having a heart attack might then expect to live $3+1/\alpha$ years in 1998, not because the heart attack treatment is better, but rather because better technologies will be used throughout the individual's remaining life. Therefore in 1998, health expenditures by someone affected by a heart attack would be $h(3 + 1/\alpha, 1998)$. The expenditures to treat this disease increase by a factor of

$$\frac{h(3+1/\alpha, 1998)}{h(3, 1997)} = e^{\mu-\theta/\alpha}. \tag{17}$$

The percent change in health expenditures for heart attack treatment is then given approximately by $\mu - \theta/\alpha$, a result that is justified formally in Appendix B. Given the parameter values we've assumed, $\mu - \theta/\alpha = -0.0094$. That is, the calibration of the model in the previous section suggests that technological progress reduces the costs of treating a given medical condition by about 1 percent each year. The sensitivity of this rate to changes in μ and θ is straightforward. As one example, if $\theta = .34$ instead of .28, the rate of decline is 2 percent per year instead of 1 percent. These numbers seem roughly consistent with the studies by Cutler et al. (1998), Shapiro et al. (1999), and Berndt et al. (2000) discussed earlier.

5. ROBUSTNESS AND ADDITIONAL EVIDENCE

The survey by Newhouse (1992) suggested that the "march of science" was responsible for at least 50 to 75 percent of the rise in U.S. health expenditures. The calibration in the previous section pushed the model to the extreme in getting it to account for all of the increase. Key to this result was the *ad hoc* assumption that health expenditures at the end of life grow at a rate of 4.7 percent per year as a result of the technological progress that raises life expectancy.

To make a more informed judgment as to how much of the increase in H/Y the model can explain, one would ideally like to obtain the value of the parameter μ without reference to the magnitude of the increase in the aggregate health expenditure share. This is what we do now.

Recall from equation (17) that $\mu - \theta/\alpha$ corresponds in the model to the annual rate at which the quality-adjusted cost of treating a specific medical condition declines over time. As discussed in the introduction, Cutler et al.

TABLE 2.

Robustness Results

	Values of θ			
	0.20	0.25	0.30	0.35
1. Treatment price decline: $\mu - \theta/\alpha = -0.01$				
μ	0.0295	0.0393	0.0492	0.0591
H/Y Factor	1.38	1.98	2.86	4.12
Frac. Expl.	0.23	0.59	1.11	1.87
2. Treatment price decline: $\mu - \theta/\alpha = -0.015$				
μ	0 0.0245	0.0343	0.0442	0.0541
H/Y Factor	1.14	1.65	2.37	3.42
Frac. Expl.	0.09	0.39	0.82	1.45
3. Treatment price decline: $\mu - \theta/\alpha = -0.02$				
μ	0.0195	0.0293	0.0392	0.0491
H/Y Factor	0.95	1.37	1.97	2.84
Frac. Expl.	-0.03	0.22	0.58	1.11

Note: "H/Y Factor" is the factor by which H/Y increases, calculated directly from equation (18). "Frac. Expl." denotes the fraction of the actual increase in H/Y explained by the model and is equal to $(``H/YFactor"{-}1)/(2.67-1)$, where 2.67 is the actual factor increase in the U.S. data between 1960 and 1997.

(1998), Shapiro et al. (1999), and Berndt et al. (2000) found evidence of quality-adjusted price declines for the treatment of heart attacks of about 1 percent, cataracts of about 1.5 percent, and depression of more than 3 percent (though over a much shorter time period of 1991–1996).

Table 2 solves the model for key results using this evidence and a range of values for θ to pin down μ. More specifically, we consider values of θ of 0.20, 0.25, 0.30, and 0.35, and treatment price declines of 1 percent, 1.5 percent, and 2 percent.[10] Table 2 then reports three statistics for each

[10] An important consideration in picking the range for the treatment price declines is that there is also a bias in the total CPI index. The papers considered above attempt to measure

pair of data points. The first is simply the implied value for μ. The second, H/Y Factor, is the factor by which H/Y increases in the model between 1960 and 1997 given this value for μ. Using equation (11), H/Y Factor is given by[11]

$$\frac{H(97)/Y(97)}{H(60)/Y(60)} \approx \frac{e^{(\mu-g)(1997-1960)}}{a^*(97)/a^*(60)}. \tag{18}$$

Finally, "Frac. Expl." reports that fraction of the actual increase in H/Y explained by the model, equal to $(\text{“}H/Y Factor\text{”}-1)/(2.67-1)$, where 2.67 is the actual factor increase in the U.S. data between 1960 and 1997.

The results in Table 2 suggest that it is quite plausible that the "march of science" accounts for a substantial amount of the increase in health expenditures as a share of GDP. Without more accurate estimates of θ and the rate at which treatment prices decline, it is difficult to be precise about the exact explanatory power of the model. Nevertheless, across a wide range of parameter values the fraction is substantial. The main exception seems to be if, simultaneously, θ is small and treatment prices decline very rapidly.

We now turn to other pieces of evidence that can be used to assess the success of the model, the first of which involves health-related transfer payments. The aggregate health expenditure share starts to decline when the economy hits the constraint on transfers, $\bar{\tau}$. Interestingly, U.S. federal and state health expenditures as a share of GDP show a pattern not unlike that of $\tau(t)$ in the simulation. In 1960, this share was 1.3 percent. It rose steadily over time before peaking in 1995 at 6.3 percent and declining slightly to 6.1 percent in 1998.[12] As shown in Figure 1, this peak in the

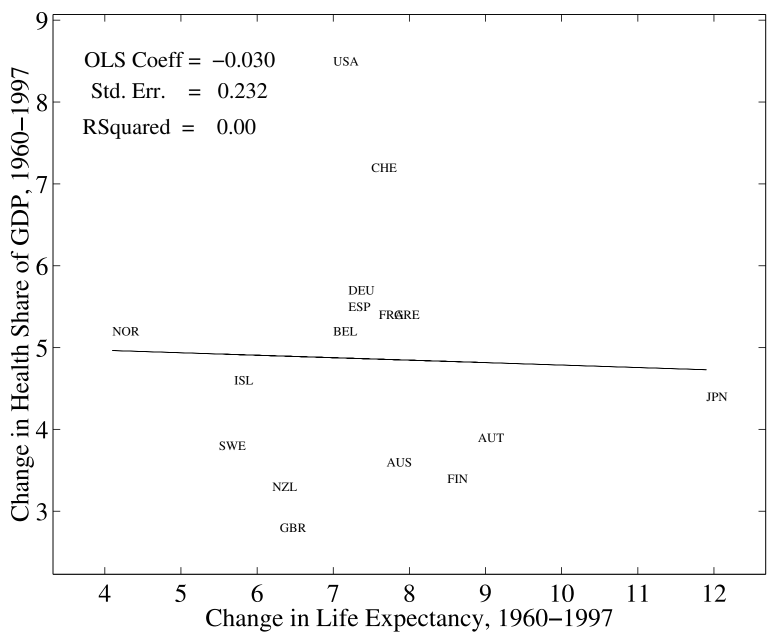
---

the bias in the treatment price index, but to the extent that the total CPI inflation rate is biased upwards, the real price declines implied by those studies may be overstated. This is why we limit the upper end of the range to two percent.

[11] We assume that the $\tau(t) \leq \bar{\tau}$ constraint is not binding so that $\ell^* = 0$.

[12] These numbers are computed from National Center for Health Statistics (2000), Table 115, page 322.

FIGURE 6. Changes in Health Share vs. Changes in Life Expectancy



Notes: Data from OECD (2000), Tables 7 and 16. Life Expectancy is for males only.

public health share corresponds closely to the peak in H/Y, as predicted by the model. Changes in U.S. Medicare policy and the increased emphasis on managed care during the 1990s might be thought of as factors that affected $\bar{\tau}$.

One piece of evidence that is potentially a problem for the model is that there is essentially no correlation between the changes in life expectancy and the changes in the health expenditure share throughout the OECD. The levels are strongly correlated, as shown earlier in Figure 2, but the changes are not, as documented in Figure 6. Countries in which life expectancy increased by a large amount did not on average experience large increases in health expenditures as a share of GDP.

This problem is mitigated upon careful reflection. It is important to appreciate that the model does not predict a monotonic relationship between changes in life expectancy and changes in H/Y. Recall from Figure 4 that H/Y can either rise or fall as life expectancy increases, depending on whether or not the $\tau(t) \leq \bar{\tau}$ constraint is binding. An economy that is unconstrained by $\bar{\tau}$ will see increases in the health share associated with technologically-driven gains in life expectancy but an economy that has reached its $\bar{\tau}$ would experience decreases in the health share associated with such gains. Consider the following comparison. The United States and the United Kingdom had similar changes in life expectancy between 1960 and 1997, which rose by about 7 years in both countries. In contrast, the increases in H/Y were vastly different: H/Y rose by more 8 percentage points in the United States but only by about 3 percentage points in the U.K. A possible explanation of this marked difference is that the U.S. let technological considerations determine its transfer payments while the U.K. limited increases in $\tau(t)$. As discussed earlier, large differences in aggregate health expenditures are typically associated with small differences in life expectancy in the model. These differences are small enough that they could easily be swamped by other considerations such as demography and income distribution.[13]

6. CONCLUSION

Why have health expenditures as a share of GDP been rising in the United States and throughout the OECD? This paper considers an explanation in

[13] More generally, differences in public health, nutrition, diet, and economic growth may help to explain some of the differences in life expectancy gains across countries. For example, both Japan and Norway had similar changes in health expenditure shares, but remarkably different changes in life expectancy. This could be related to the fact that over the period 1960 to 1992, Japan was one of the world's fastest growing countries, and to the nutritional gains that were associated with this growth.

which health expenditures and life expectancy are endogenous variables driven by technological progress. Advances in medicine permit people to spend resources on health care in order to extend life. Starting from initial conditions in which very little is spent on health care because of the absence of such opportunities, this naturally leads to an increase in the health expenditure share. Moreover, advances in life expectancy permit people to live to face more serious medical problems, such as heart attacks and hip replacements, that in turn are eventually cured, but only at a price.

Combining this theory of the effects of technological change with a transfer program that allows health expenditures at the end of life to rise to four times per capita income produces a framework that is broadly consistent with the following stylized facts:

1. Medicare expenditures rise sharply as life expectancy declines in a cross section, at a rate of about 10 percent until life expectancy falls to 3 years and then at a rate of nearly 45 percent in the last couple of years of life.

2. The fraction of Medicare expenditures accounted for by people in the last year of life is about 30 percent, and this number is surprisingly stable over time.

3. While as much as 4 percent of U.S. GDP may go to pay the health expenses of people in the last year of life, it is not at all clear that this reflects an inefficiency in the health care system.

4. Technological progress reduces the quality-adjusted cost of treating specific medical conditions, at a rate that seems to be about one or two percent per year.

5. Life expectancy in the United States since 1960 has risen at a rate of about 2 years every decade.

6. Health expenditures as a share of GDP have risen from 5.1 percent in 1960 to 13.6 percent in 1997. At least half, and most likely three-quarters or more of this change seems to be driven by the "march of science" and medical advances.

7. This trend is present throughout the OECD countries, where the typical increase in health expenditures is from 4 percent of GDP to about 8 percent.

8. Health-related transfer payments as a share of GDP have increased substantially.

9. Across countries, there is very little correlation between changes in life expectancy and changes in health expenditures as a share of GDP.

Simulations of the model suggest that a critical determinant of health expenditures as a share of GDP is the willingness of society to transfer resources to people near the end of life. As long as this willingness is unconstrained, the health share in the model rises rapidly. On the other hand, once a cap on these transfers is reached, the model predicts a halt to the rising health share and even a slow, gradual decline to a long-run level corresponding to the transfer rate. The source of this decline is a "dilution effect" associated with rising life expectancy. There are always five years worth of people with life expectancies of five years or less. As life expectancy rises, the fraction of the population accounted for by these people declines. If the expenditure share of these high-cost people is kept from growing too rapidly, the increasing fraction of low-cost healthy people reduces the aggregate expenditure share.

The model clearly has a number of important limitations. Most obviously, it treats in a reduced-form fashion a number of key parameters, including μ, θ, σ, and $\bar{\tau}$. A productive avenue for future research is to examine the microfoundations that determine these parameter values. Such a study would be forced to confront the important but complicated effects of health insurance, the relation between life expectancy and technological progress, and the political economy of health-related transfer payments.

APPENDIX A

Health expenditures in the last year of life

This appendix calculates the fraction of total health expenditures that are associated with people in the last year of life. Let $H(a, b, t)$ denote health expenditures by people with life expectancies in the range $[a, b]$. Then,

$$\begin{aligned} H(\ell^*, \ell^* + 1, t) &= \int_{\ell^*}^{\ell^*+1} h(\ell, t) d\ell \\ &= \frac{1}{\theta} \left(h(\ell^*, t) - h(\ell^* + 1, t) \right) \\ &= \frac{1}{\theta} h(\ell^*, t)(1 - e^{-\theta}). \end{aligned} \tag{A.1}$$

From equation (11), this relation can be rewritten as

$$\frac{H(\ell^*, \ell^* + 1, t)}{Y(t)} \approx \frac{H(t)}{Y(t)} \times (1 - e^{-\theta}). \tag{A.2}$$

Health expenditures in the last year of life as a share of GDP are proportional to H/Y at rate $1 - e^{-\theta} \approx \theta$. This factor, in turn, is then the fraction of total health expenditures associated with people in the last year of life.

APPENDIX B

Declining quality-adjusted treatment prices

This appendix introduces diseases and health problems explicitly in order to calculate the rate at which quality-adjusted treatment prices decline over time. In addition to deriving this rate, we see that the framework leads to the same specification for health expenditures derived in the original model in equation (4).

Suppose there are a range of health problems that can afflict a person. These problems are indexed by $x \in [0, \infty)$ and are ordered by (increasing) severity. At any point in time, treatments have been discovered only for health problems in the range $[0, \bar{x}(t)]$. Life expectancy $\ell(x)$ for someone who receives a treatment for condition x is given by

$$\ell(x) = \bar{x}(t) - x. \tag{B.1}$$

Absent treatment, the person dies.

Once discovered, the cost of treating condition x is

$$\bar{h}(x,t) = \bar{h}_0 e^{\beta_1 x - \beta_2 (t - \delta(x))}, \tag{B.2}$$

where $\delta(x)$ denotes the date at which a successful treatment for condition x is discovered. The cost of treating a disease when the treatment is first discovered increases at rate β_1 with the severity of the disease. However, after the initial treatment is discovered, technological progress continually reduces the cost of treating the disease at rate β_2. Specified this way, the rate at which the quality-adjusted treatment price declines is given by β_2.

Consistent with the model in the main text, the date at which a successful treatment for condition x is discovered is assumed to be given by

$$\delta(x) = \alpha x. \tag{B.3}$$

That is, to advance the frontier disease by one unit requires α years.

The specification of the augmented model is now complete, and we can now derive health expenditures as a function of life expectancy ℓ rather than as a function of health condition x in order to relate this framework to the basic model in the paper. Some straightforward algebra[1] shows that

$$h(\ell, t) \equiv \bar{h}(x(\ell), t) = \bar{h}_0 e^{\beta_1 t/\alpha - (\beta_1 + \alpha\beta_2)\ell}. \tag{B.4}$$

Over time, the cross section of health expenditures increases at rate β_1/α. Across people with different life expectancies, health expenditures decline

[1]First, notice that $\bar{x}(t) = t/\alpha$, so that $x = t/\alpha - \ell$. Substituting this relation together with the equation describing $\delta(x)$ gives the desired result.

at rate $\beta_1 + \alpha\beta_2$. Comparing this result with equation (4), one sees that the two expressions are identical when μ and θ are given by

$$\mu = \beta_1/\alpha \tag{B.5}$$

$$\theta = \beta_1 + \alpha\beta_2. \tag{B.6}$$

That is, the model given earlier can be viewed as the reduced form of this slightly richer model in which health conditions are specified explicitly.

As noted above, in this richer model, the quality-adjusted price of treating a given disease declines at rate β_2. Solving these last two equations, one sees that

$$\beta_2 = \theta/\alpha - \mu. \tag{B.7}$$

Reversing the sign for convenience, quality adjusted prices *rise* at rate $\mu - \theta/\alpha$.

REFERENCES

Baumol, William J., "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *American Economic Review*, June 1967, *57*, 415–426.

Berndt, Ernst R., Anupa Bir, Susan H. Busch, Richard G. Frank, and Shaon-Lise T. Normand, "The Medical Treatment of Depression, 1991–1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes," 2000. NBER Working Paper No. 7816.

Cutler, David M. and Mark McClellan, "Is Technological Change in Medicine Worth It?," July 2001. Harvard University mimeo.

___ , Mark B. McClellan, Joseph P. Newhouse, and Dahlia Remler, "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics*, November 1998, *113* (4), 991–1024.

Fuchs, Victor, "Though Much is Taken: Reflections on Aging, Health, and Medical Care," *Milbank Memorial Fund Quarterly*, 1984, *62*, 143–166.

Lubitz, James D. and Gerald F. Riley, "Trends in Medicare Payments in the Last Year of Life," *New England Journal of Medicine*, April 1993, *328* (15), 1092–1096.

Lubitz, James, James Beebe, and Colin Baker, "Longevity and Medicare Expenditures," *New England Journal of Medicine*, April 1995, *332* (15), 999–1003.

Miller, Tim, "Increasing Longevity and Medicare Expenditures," *Demography*, May 2001, *38* (2), 215–226.

Murphy, Kevin M. and Robert Topel, "The Economic Value of Medical Research." In *Exceptional Returns* (Murphy and Topel, eds 2002). Forthcoming.

___ and ___ , eds, *Exceptional Returns*, University of Chicago Press, 2002. Forthcoming.

National Center for Health Statistics, *Health, United States 2000* 2000.

Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?," *Journal of Economic Perspectives*, Summer 1992, *6* (3), 3–21.

Nordhaus, William D., "The Health of Nations: The Contribution of Improved Health to Living Standards." In Murphy and Topel, eds (2002). Forthcoming.

OECD, "OECD Health Data 2000 - A Comparative Analysis of 29 Countries," http://www.oecd.org/els/health/software/fad.htm 2000. ISBN 92-64-06876-7.

Ogaki, Masao and Carmen M. Reinhart, "Measuring Intertemporal Substitution: The Role of Durable Goods," *Journal of Political Economy*, October 1998, *106* (5), 1078–1098.

Shapiro, Irving, Matthew D. Shapiro, and David W. Wilcox, "Measuring the Value of Cataract Surgery," 1999. University of Michigan mimeo.

Technical Review Panel on the Medicare Trustees Reports, "Review of Assumptions and Methods of the Medicare Trustees' Financial Projections," December 2000. Mimeo, downloaded from http://www.hcfa.gov/pubforms/actuary/TechnicalPanel/.

Triplett, Jack E., "What's Different about Health: Human Repair and Car Repair in National Accounts and in National Health Accounts," in Ernst Berndt and David Cutler, eds., *Medical Care Output and Producitivity*, University of Chicago Press, forthcoming.

___ and Barry P. Bosworth, "Productivity in the Services Sector," 2000. Brookings Institution mimeo.

White, Kevin M., "Longevity Advances in High-Income Countries, 1955-96," *Population and Development Review*, March 2002, *28* (1), 59–76.