Technology and Expenditure Growth in Health Care*

Amitabh Chandra
Harvard University and NBER

Jonathan S. Skinner
Dartmouth College and NBER

This Draft: February 18, 2008 4:12 PM

Abstract

We examine the parallel trends in technology growth and cost growth in health care. A theoretical model of growth and productivity leads to a typology of medical technology: highly effective and inexpensive innovations (antibiotics, or aspirin and beta blockers for cardiac care), more expensive yet effective treatments for appropriate patients (hip and knee replacements, surgical interventions for heart attack patients), and “gray area” treatments with uncertain clinical value (ICU days among chronically ill patients). We show that the average productivity of treatments depend critically on the heterogeneity of effects across patients, the precise shape of the health production function, as well as the cost structure of procedures such as MRIs with high fixed costs and low marginal costs. Future productivity growth of the current system will be limited by constraints on health care financing because of high tax burdens and the collapse of private health insurance markets. Nonetheless, there are tremendous potential productivity gains from better coordination of care and information technology.

* This research was funded by the National Institute of Aging NIA P01 AG19783-02, the Robert Wood Johnson Foundation, and the Taubman Center at Harvard University. We thank, without implicating, Doug Staiger, Elliott Fisher, and Jack Wennberg for collaborations that have greatly influenced this paper, and to Roger Gordon for illuminating comments. The opinions in this paper are those of the authors and should not be attributed to the NIA, the NBER or any institution that they are affiliated with. Address correspondence to Amitabh.Chandra@Harvard.EDU and Jonathan.Skinner@Dartmouth.EDU.
1. Introduction

The science section of a U.S. newspaper routinely features articles on new surgical and pharmaceutical treatments for cancer, obesity, aging, and cardiovascular diseases, with rosy predictions of expanded longevity and improved health functioning (Gina Kolata, 2006). The business section, on the other hand, features double-digit growth in health care costs, declining health insurance coverage, and gloomy predictions of U.S. Treasury bonds being downgraded to junk-bond status as Medicare threatens implosion.\textsuperscript{2} It is perhaps not surprising that there is some ambiguity as to whether these two trends taken together are a cause for celebration or concern.

We can certainly identify the source of both trends: technology growth. Newhouse (1992) considered a wide variety of factors that could reasonably have caused the rapid growth in health care expenditures, such as an aging population or the expansion of health insurance, and concluded that only technology growth could explain the vast majority of expenditure growth. Similarly, Cutler (2004) has argued that technological advances have generated both the far-reaching advances in longevity, and the rapid growth in costs.\textsuperscript{3} Even Finkelstein (2007) argued that the introduction of Medicare in the U.S. fueled growth rates in health care by boosting diffusion rates of then-new cardiac catheterization and intensive care units across hospitals.

Unfortunately, simply attaching the label of “technological growth” as a cause of these increases in health care expenditures doesn’t get us much closer to understanding either the pathology of the disease – if it is a disease – or a meaningful way to address the problem – if it is a problem. A better understanding of technological growth in healthcare is the first goal of this paper.

In a classic article, Weisbrod (1991) provided a sweeping overview of technological change in health care, and argued persuasively that the structure of insurance coverage was an integral part of this rapid growth. The intuition is straightforward: if you pay for it, they will build it. He concluded with a relatively optimistic view that the then-new prospective payment system (PPS) in the Medicare program would provide greater incentives for technological innovation that provided better care at lower cost, and would thus be profitable for

\begin{footnotesize}
\begin{itemize}
\item[2] See Stein (2006). Currently, the Medicare Hospital Insurance Trust Fund, which pays for hospital benefits, is expected to be exhausted in 2020 (Centers for Medicare and Medicaid, 2005).
\item[3] Cutler (2004) writes “Money matters in health care as it does in few other industries. Where we have spent a lot, we have received a lot in return.”
\end{itemize}
\end{footnotesize}
health care providers under the PPS program. Regrettably, this optimism has not been borne out. Health care costs have continued to rise rapidly, and there is little evidence that technological innovation (and diffusion) have slowed in response to the PPS system.4

Some of the concerns about health care spending growth are not unique to the US experience, given that other industrialized countries also grapple with this dilemma (Cutler, 2002). Figure 1 illustrating the striking increase in the share of GDP spent on healthcare in six OECD countries.5 Thus to find explanations for health care growth overall, we need to consider global trends, but again one where technology growth is a primary suspect. On the other hand, as Figure 1 emphasizes, health care expenditures in the U.S. have sprinted far ahead of any other country since the late 1960s, and so our article will also consider ways in which the US experience is unique.6

In this paper, we first develop a parsimonious model of technology growth and cost growth that is anchored by a simple assumption – that physicians will do everything in their (perceived) power to cure their patient, given financial or resource constraints. This in turn leads to strong implications for both the level and growth of technology in the presence of heterogeneity in economic incentives, physician beliefs, and the distribution of treatment benefits across patients.

The model can also be used to distinguish among general categories of innovations, ranked in order of their contribution to health care productivity (health outcomes per dollar spent).7 Not surprisingly, we find that the greatest contribution occurs for low cost but highly effective treatments: antibiotics for bacterial infection, or aspirin and beta blockers for heart attack patients. But the

---

4 Newhouse (2002) noted that the PPS program in fact exempted certain capital costs from the expenditure cap; more recently Acemoglu and Finkelstein (2006) found a significant impact of PPS on the ratio of labor to capital costs; labor costs declined while capital costs held steady through the introduction of the PPS program. More importantly, McClellan (1996) demonstrated that the PPS program was not always specific to the disease (e.g., heart attack) but was often specific to the treatment (e.g., bypass surgery).

5 Between 1970 and 2002, Lawrence Kotlikoff and Christian Hagist (2006) find that real per capita benefit levels grew at a 4.1% annual rate on average in Australia, Austria, Canada, Germany, Japan, Norway, Spain, Sweden, the UK, and the U.S., a full two percentage points higher than the average growth rate of real per capita GDP.

6 See also Garber and Skinner (2008).

7 Here we follow Wennberg, Fisher, and Skinner (2002) although our second category is for hi-tech care rather than preference-sensitive care, a category we consider below.
incentive structure also matters; even when treatments are highly efficacious, diffusion can be slow when there are few incentives to encourage adoption (Berwick, 2003; McGlynn et. al., 2003).

A second broad category of treatment or procedure are those where benefits are substantial for at least some patients, but where the costs are also large; examples include “hi-tech” treatments such as hip and knee replacements, interventional cardiac procedures, and pharmaceutical treatments such as anti-retrovirals for HIV or treatments for depression. For anti-retroviral drugs, there are clear and well-defined limits for which patients will benefit – those with diagnosed HIV – and who will not benefit – everyone else. In other words, the second derivative of the production function is strongly negative, in that benefits quickly turn negative for non-HIV-infected patients. For these treatments, we expect very high average productivity returns.

The model predicts more modest productivity gains from hi-tech treatments such as angioplasty or diagnostic testing. Angioplasty is a procedure in which a catheter is used free blockage in the heart, typically with a wire stent left to keep the artery open. For heart attack patients within the first 12 hours following the heart attack, the benefits of angioplasty (and stents) are substantial. But there are many more patients where the value of angioplasty is less clear, for example among those with stable angina, or several days after the heart attack, yet rates for these procedures continue to rise. Because there are far more people in this latter category than in the former category – the second derivative of the production function is small in magnitude – average productivity is diminished.

In the United States, the insurance system provides a particularly strong incentive for the growth of high-tech treatments with high fixed cost and low marginal cost. For example, staffing a 24-hour cardiac catheterization laboratory is expensive, but once the facility is in place, there are financial incentives to provide cardiac screening and subsequent interventions given the minimal marginal cost and substantial reimbursement for each procedure. This in turn encourages further diffusion and growth in both expenditures and rates of screening. This distinction between average and marginal costs is also central to understanding the double-digit growth in imaging procedures as physician groups buy their own MRIs and collecting both the facility charge (for the MRI) and the

---

8 The classic example of a high fixed cost but low marginal cost procedure is the use of nuclear particle accelerators – huge machines costing more than $200 million to build – to treat rare forms of cancer. Once built, they are used primarily to treat cases of far more common prostate cancer, despite the absence of evidence on clinical benefits (Pollack, 2007).
diagnostic charge (for reading the MRI).

The third general treatment category can be best characterized as “care management,” and relates to treatments for patients who have serious diseases such as pulmonary disorders, congestive heart failure, diabetes, or the metabolic syndrome, where the primary goal of the physician is not so much to cure the disease, but instead to keep the patient alive and to maintain quality of life for as long as possible. But how to manage disease is often uncharted territory; there are no randomized trials for when pneumonia patients should be admitted to the hospital, how many specialists should be treating a patient (Becker and Murphy, 1992), or whether to institute mechanical ventilation for a severely ill patient with dementia.

We can use these three categories to consider several critical questions in health policy. The first is: have health care expenditures been worth the cost? Many have found the answer to be a resounding yes (Cutler, et. al., 2006; Cutler, 2004; Murphy and Topel, 2006), with Hall and Jones (2007) suggest that health care spending should reach 30 percent of GDP by 2050. But more recent estimates of discounted productivity growth suggest a more worrisome trend – a sharply diminishing value of health care dollars, with average discounted cost effectiveness ratios of more than one-quarter of a million dollars per life-year saved during 1990-2000 (Rosen, personal communication).

The worsening average value of health care expenditures suggests that prospects for a “soft landing” in expenditure growth are dim; few new innovations both save money and lives. More likely, a slowdown will occur because of the inability or unwillingness of government spending to finance the current Medicare and Medicaid programs, coupled with growing inefficiency from the higher taxes or insurance premiums necessary to pay for the health care (Baicker and Skinner, 2008). One guess for the future is that the U.S. system will resemble the U.K. system; minimal universal coverage coupled with an active private market allowing high income subscribers to jump ahead of the queue or for “luxury” health care services that may prove too expensive for government coverage.

We further speculate that the next frontier for productivity improvement lies precisely in more efficient management of “care management” through the use of computerized information technologies and standards of care based on practice patterns at highly efficient clinics such as Mayo and Intermountain, where costs for treating similar patients are up to one-third lower than high-cost hospitals (Wennberg, et. al., et., 2007), as well as providing demand-side incentives to use
cost-saving drugs (Dor, 2002; Chandra, Gruber, and McKnight, 2008).

2. A Simple Model of Technology Growth and Expenditure Growth

The traditional introduction to the economics of health care begins with the question of why is health care so different from other goods. A reasonable place to begin is what appears to be a fundamental human concern with the pain and suffering of others, which in turn has been reflected in a collectivist view of health care. As health care has become more expensive, insurance markets appeared, both to ensure that providers actually got paid (as in the original Blue Cross plans during the 1920s) and to provide health care to the widest range of citizens (as in Bismark’s Germany).

Widespread insurance markets lead naturally to the twin problems of moral hazard and adverse selection. Moral hazard occurs because a third-party, the insurance company, is paying most of the cost of the transaction between the first party (the physician or health provider) and the second party (the patient). This can lead to individuals not taking sufficient precaution against illness, or more commonly, patients and providers will end up consuming and providing “too much” health care. The degree to which this happens will depend on the price elasticity of demand for health care and the structure of the particular insurance contract (e.g., Newhouse, 1993).

Adverse selection can occur on both sides of the market. Insurance companies structure contracts in a way to attract low-risk enrollees away from other firms, for example by offering managed care options that are most appealing to healthy enrollees. Adverse selection can also occur on the consumer side of the market, for example when people with higher expectations of poor health sign up for long-term care insurance (e.g., Finkelstein and McGarry, 2005).

A third issue that arises in health care is the principal-agent problem; that because of asymmetric information, the physician is making decisions on behalf of the patient, leading to the concern that the decisions that do occur favor the interests of the physician. For example, a physician may schedule a sequence of short visits which may prove to be more financially lucrative than a single visit of longer duration. This is the basis for a long-standing debate over the existence of “supplier-induced demand,” a literature extending back several decades that seeks to find evidence of physicians acting in their interest – to shift out demand for their services – even when it may harm, at least financially, the patient.
While each of these three issues is clearly important for health care policy, we argue that they are not equally important in explaining the twin growth rates in technological progress and health care costs. Adverse selection is clearly of importance in the design and long-term equilibrium of private health insurance plans, but the patterns we observe in Section 3 are true for the elderly Medicare population, and in other countries with national health insurance, where adverse selection issues are less important.

Instead, we argue that the combination of moral hazard and the principal agent problem has provided fertile ground for the particularly rapid growth in health care costs. We are not thinking as much of the traditional static moral hazard problem. The rising fraction of uninsured has not exerted a substantial moderating influence on health care costs, as one would expect if the loss in coverage would have had a large impact on utilization. Newhouse (1993) reiterates the message in Schwartz (1987) that it is important to distinguish between factors that contribute to the level versus the growth of healthcare expenditures. By examining the time-series evidence, Newhouse (1992, 1993) examines the time-series evidence and concludes that factors such as the aging of the population, the spread of health insurance, increased incomes, physician supply, greater defensive medicine, administrative costs, and the share of expenditures on the terminally, have all contributed to the growth of US healthcare spending, but these factors do not explain the evidence in a substantive manner (Kotlikoff and Hagist, 2006; Getzen, 1991).

Nor can typical estimates of price elasticity explain more than a tiny fraction of the overall growth in health care costs (Newhouse, 1992; Manning et. al., 1987). The more important role of insurance is in financing the dramatic growth in health care expenditures through innovation and rapid diffusion of technology (Finkelstein, 2005). The real question is whether implementing market-based incentives systems, such as capping the deductibility of employer-provided health insurance, or health savings accounts (HSAs) can have a real impact on growth rates, a question to which we return below.

2.1 The Demand for Health Care and Social Willingness to Pay

We begin with a simple model of consumer demand. Suppose that utility for patient $i$ is a function of quality-adjusted life years, $s_i$, and non-medical goods $C_i$:

$$U_i = U(s_i(x_i), C_i)$$  \hspace{1cm} (1)
where $x_{it} = x_{i1}, x_{i2}, \ldots, x_{in}$ represents the vector of treatments received by individual $i$, and $n$ is the total possible set of treatments available. We discuss below what happens when $n$ grows over time because of technological improvements in treatments. Note that $s_i$ depends on the initial health status of the individual; so that the marginal productivity of a treatment (such as mechanical ventilation) is clearly different for a healthy person compared to a sick person. Ignoring intertemporal issues, the one-period budget constraint is $Y_i \geq x_i p_i + C_i$, where $p_i$ is the vector of prices faced by individual $i$ (which may differ across individuals depending on the generosity of their insurance), and $Y_i$ is income.9

Key to understanding the demand for healthcare is the marginal value of increasing spending that generates quality-adjusted life years relative to spending on the consumption good (Murphy and Topel, 2006):

$$\varphi_i = \frac{\partial U}{\partial s_i} \bigg/ \frac{\partial U}{\partial C_i} \quad (2)$$

The demand condition is that the marginal benefits in dollar terms from spending more on treatment $k$ exceeds or is equal to the price faced by the patient; $\varphi_i s_{ik} \geq p_{ik}$ where $s_{ik} = \frac{\partial s_i}{\partial x_{ik}}$.

Within this framework, it is instructive to see the derivation of the marginal benefit curve which ranks patients from highest to lowest marginal benefit for procedure $k$. Figure 2 shows two demand curves for treatment. The solid line JJ’ represents the marginal value of each treatment where $\varphi_i$ is fixed for everyone at the “social” value of extra quality-adjusted life years; e.g., $50,000$ per life year, a common hurdle used in cost-effectiveness analysis. The social optimum in this case is at point $x^*$, where the social cost of treatment $q$ (discussed in more detail below) is equal to the incremental value for the marginal patient.

In contrast this ranking, consider one where there are no insurance markets, and patients are ranked on the basis of their valuation for healthcare, as shown by ZZ’, so that $\varphi_i$ is larger for higher income households, but far smaller for low income households. The variation in the marginal valuation may be higher for some people (because they both benefit from the procedure and their individual $\varphi_i$ is large) but for others, the lower value of $\varphi_i$ leads to fewer people showing up in the physician’s office or emergency room for income considerations. The

---

9 For more complex demand-side models, see also Grossman (1976), Murphy and Topel (2003, 2006) and Hall and Jones (2006).
optimum allocations are quite different; in the first case (JJ’) where survival benefits across patients are valued similarly, productivity as measured by quality-adjusted life years will be maximized at the optimum x*. In practice insurance reduces the variance of φ across individuals by providing financial support to people with low income who would otherwise not be able to purchase health care (Nyman, 1999), thus moving the outcome away from x toward x*. At the same time, of course, insurance reduces the price faced by consumers (through copayments and coinsurance), increasing utilization to a point that is likely beyond x2.

2.2 A Static Model of Provider Behavior

We use the most parsimonious assumption to characterize physician (or provider) behavior, one which accords with what physicians say and how they behave:

*Health care providers maximize the perceived health of their patients, but are sometimes constrained from doing all they want by capacity constraints and/or financial incentives.*

That is, every physician and health care provider wakes up in the morning to save as many lives as they can. There are always exceptions, rogue physicians who game the system at the expense of patients, but most physicians believe that what they are doing is best for their patient. We are therefore trying to avoid the standard models of “supplier-induced demand” in which physicians harm their patients if they are paid enough to do so.10 While there is remarkable heterogeneity in physician beliefs, as we document below, for the moment assume that there is a single known “production function” at time t.

In practice, there are both explicit and implicit (or hidden) constraints on health care. In the U.S. explicit restrictions are rare, for example cost-effectiveness rules attempted in the Oregon Medicaid program, or in gatekeeper models of health maintenance organizations (HMOs) which required a second opinion or a sign-off for common medical procedures. Implicit restrictions are more common; the limitations of physicians who only have 24 hours in the day, a lack of diagnostic or surgical facilities (or catheterization labs not staffed at night), or financial constraints, for example when patients are uninsured or otherwise unable to pay. Physicians often do not chafe against these implicit

10 Chandra and Staiger (2007) reject models of supplier induced demand.
restrictions, simply adjusting their “best practice” to the resources they are accustomed to using.\textsuperscript{11}

Suppose that there $i = 1, \ldots, m_j$ patients with provider $j$ as their primary physician or hospital. As before, survival for individual $i$, $s_i$, is assumed to be a function of the vector of treatments $x_i$. We further assume a Leontief production function for the physician’s time; so that each procedure or treatment (whether reading an X-ray, an office visit, or surgery) is a fixed fraction of the physician’s total time, $\omega_k$, so that the total fraction of hours for the physician is $L_j = \sum_i \omega_j x_{ij} \leq 1$ where 1 is the maximum hours worked.

Physicians typically face a resource capacity constraint, for example a target income or a minimum requirement that the hospital must not lose revenue (see Rizzo and Zeckhauser, 2005). We focus on the individual physician’s incentives for the moment, taking as given the institutional structure of his or her practice. Revenue, for the $j$th provider, depends on the net cost to the provider of the medical treatments, subject to the assumed revenue constraint $R_j$ which may include fixed costs:

$$m_j B + \sum_i (r_i x_i + wL_i) \geq R_j$$

(3)

where $B$ is the average capitated payment per patient and $m_j$ is the number of patients seen by the physician (as in the U.K.). Alternatively, the physician may be paid purely on a salary basis (in which case $L_i$ is set to one, $B = 0$, and $w$ is the salary rate), or in the more common case in the U.S., where the physician shares in reimbursements depending on the volume of procedures; $r_i$ is the vector of net payments received by the physician (or received by the provider and implicitly shared with the physician by salary adjustment or other compensation). If patient $i$ is uninsured, the resource cost will be borne by the physician or hospital, so that $r_{ik} < 0$. If instead a specific procedure yields marginal profits because of Medicare setting high reimbursement rates based on average costs, then $r_{ik} \gg 0$, as is the case for coronary artery bypass grafts (CABG) or stents (Hayes, et. al., 2007).

Note that the price paid by the patient for procedure $k$, $p_{ik}$, will generally be quite different from the amount received by the provider, $r_{ik}$, because of third-party insurance programs and different incentive systems designed to pay

\textsuperscript{11} For example, in the late 1980s New Haven had roughly half of the bed per capita ratio of Boston (Fisher et al, 1994) largely because the dominant hospitals in New Haven made a conscious effort to restrict expansion. Yet physicians in New Haven did not view their hospital capacity as being constrained.
providers. The social cost of the $k$th procedure will generally be equal to something quite different. If we assume competitive wages for physicians and costs for procedures, and simplify by ignoring the inputs of other health professionals, then the social resource cost will equal the average cost $C(.)$ of the treatment plus the labor input of the physician $q_k = C(x_k) + \omega_k w$. Why average and not marginal cost? For procedures such as stents, MRIs, and other “high-tech” treatments, fixed costs are large and compounded by 24 hour/day staffing, but where the incremental marginal cost is low, at least up to the capacity of the machine.

Capacity constraints are more likely in countries such as the U.K. or other countries with centralized financing. In this case, the central health agency has a global budget which they use to provide inputs to individual hospitals and clinics, for example a certain number of artificial knee or hip joints devices (Aaron and Schwartz, 2005). Similarly, some regions may not be allocated as many magnetic resonance imaging (MRI) machines, or cardiac catheterization units. In many cases, physicians adjust implicitly to these constraints, although waiting lists for surgery are often the subject of public demonstrations and political controversy (Aaron and Schwartz, 2005). Whether implicit or explicit, physicians and hospitals in these cases are operating under quantity constraints, which can be expressed as $\sum_i x_i \leq X_k$.

Let the social marginal value of an extra year of life be given by $\varphi^*$ at time $t$. If providers solve the social planners allocation problem they should maximize over the patients who appear in their waiting room (i.e., those patients for whom the perceived marginal benefit of the treatment is greater than the price, $\varphi_i s_{ik} \geq p_{ik}$):

$$L = \sum_{\varphi^*_i \geq p_{ik}} \varphi^*_i s_i(x_i) - \lambda_0 \left( \sum_i \omega x_i - 1 \right) - \gamma \left( \sum_i (-r_i x_i + wL) + mB - R \right) - \sum_k \lambda_k \left( x_k - X_k \right)$$

(4)

The first-order conditions imply that:

$$\varphi^* \frac{\partial s_i}{\partial x_{ik}} = \lambda_0 \omega_k - \gamma r_{ik} + \lambda_k$$

(5)

This set of first-order conditions imply that, in the absence of constraints on their

---

12 We are grateful to Christopher Snyder for helpful discussions on this question.
time ($\lambda_0$), constraints on target income ($\gamma$), or capacity constraints ($\lambda_k$), physicians or health care providers should treat patients to the point where there are no more perceived benefits of treatment $x_k$ (that is, treat to the point that $\partial s_i / \partial x_{ik}$ is zero). Of course, most physicians will be subject to at least one of these constraints—one might expect the time constraint and the capacity constraint to restrain utilization so that, at the margin, physicians allocate treatments to those who would benefit most.\(^{13}\)

Consider Figure 2a again, where the assumption of a common $\phi^*$ implies that the social planner is maximizing social benefits along the curve $JJ'$, at point $x^*$. However, financial incentives on both the demand and supply side affect the quantity of procedures. Suppose we make a simplifying assumption that the marginal value of quality-adjusted life years is the same across individuals ($\phi^*$), as is the price faced by patients ($p_k$). Then in order to be considered for treatment by the physician, $\phi^* s_{ik} \geq p_k$, and, absent other constraints, the physician would need to be sufficiently compensated to provide the service, $\phi^* s_{ik} \geq -\gamma r_k$. Thus the market outcome in this hypothetical market would be one where physicians provide treatment $k$ for all patients for whom the marginal benefit $\phi^* s_{ik} \geq \max(p_k, -\gamma r_k)$. In the presence of third-party insurance, the outcome $x_1$ will be to the right of the optimum $x^*$. So far, this is the conventional analysis of excess burden with a subsidized good, and static moral hazard is demonstrated in a straightforward way; it’s the difference in input use and treatment behavior between points $x^*$ and $x_1$, given by the area denoted $C$.

More complicated is where net reimbursement ($r_k$) is positive, so that a procedure brings money to the provider, but $\gamma > 0$, one is still left with the uncomfortable implication that the provider will attempt to convince the patient of beneficial effects, even when the marginal value of the procedure leads to harm. This occurs in isolated cases, for example the cardiac surgeons in Redding, CA who, because of very high reimbursement rates, pulled a healthy patient off the golf course for emergency cardiac surgery (Enkoji, 2002). More plausible is that there is a sufficiently large "reservoir of disease" in the general population that physicians can reach their time capacity ($\lambda_0 > 0$) by treating more patients for whom at least perceived benefits are non-negative.

\(^{13}\) A Roy model of treatment allocation is developed by Chandra and Staiger (2007), who find evidence that providers rank patients on the basis of diminishing therapeutic benefit and work down this distribution.
2.3 A Dynamic Model of Provider Behavior

To this point, the analysis has been in a static framework. We next turn to one that attempts to model both the dynamics of change. How does a static model translate to a dynamic framework for technology and expenditures? Here we borrow from the literature on diffusion in economics (e.g., Jovanovic and Nyarko, 1995, 1996; Hall, 2004) and sociology (e.g., Rogers, 2005). In a world where the binding constraints are financial (i.e., those for which \( \gamma \) represents the shadow price), individual health providers each make decisions about how much to adopt or not adopt the new innovation. Incentives to adopt the innovation depend critically on the nature of their own perceived production function – which in turn reflects how rapidly they themselves absorb new clinical evidence on treatments – and the financial structure and rewards of the innovation. When the innovation is both financially remunerative, and is perceived to be highly effective, the speed of adoption will be rapid, as for example the rapid diffusion of tetracycline in the 1950s documented in the classic Coleman study (Coleman, et. al., 1966).

If there are few economic incentives to adopt, as in the case of beta blockers for the treatment of AMI, then diffusion will depend almost entirely on the extent to which individual perceptions of physicians change over time (Skinner and Staiger, 2008). In other words, we need to consider the convergence of knowledge and understanding of treatment effects, so that physician-specific perceived survival “production functions” may differ from the true frontier of knowledge \( s_{it}\ast(x_{it}) \) at time \( t \). Note that we now must subscript the survival function by time \( t \) as new innovations are developed. Diffusion in the sociological literature (Rogers 2005) emphasizes not economic or profit-maximizing factors, but instead on a cultural or intrinsic inability to perceive the benefits of new technologies despite evidence of their effectiveness.\(^{14}\)

2.3.1 Diffusion of Innovations

Over time, elements of the vector \( x \) may be added as new treatments in the health care provider armamentarium (Skinner and Staiger, 2008). Thus technology growth can be represented by both the introduction of new inputs \( x_k \)

\(^{14}\) Parente and Prescott (1994) have focused on the importance of relatively minor differences in barriers to technology adoption in explaining differences in income growth across countries. This emphasizes the importance of cross-country differences in productivity arising from variation in technology adoption (Comin and Hobijn, 2004).
and the diffusion of existing treatments $x_j$ when $j < k$. These will have dynamic
effects both on health outcomes as well as costs. Returning to the model
described in equation (4), we characterize how innovations affect both outcomes
and costs for different types of health care systems, in particular those like the
U.S. where there are few quantity restrictions, and those like the U.K. and Canada
where central planning is far more common.

We integrate both dynamics and heterogeneity with the use of an adjustment
function approach, in which health care providers evolve towards the time-t
frontier, but at differing rates depending on how far they are from the optimum,
and their ability to integrate this new information (Skinner and Staiger, 2008).
This approach has obvious parallels in the productivity literature, as in Nelson and
Phelps (1965) and Parente and Prescott (1994). At a point in time, some health
care providers may be further from the current production possibility frontier,
whether because of barriers to acquiring information, or because of intrinsic
differences in skills and abilities.  

We assume a standard partial-adjustment framework where the growth in inputs
depends both on how far away last year’s procedure was from the time-t optimal
solution to the Lagrangian in Equation 4, $x_{kt}^*$, and the individual health care
provider’s speed of diffusion or adoption, which we denote $\pi_j$. Recall that the
optimal solution reflects both best-practice health care known at time $t$, but also
the incentive structure and capacity constraints for the individual provider at time
$t$. That is,

$$\Delta x_{kt}^* = \pi_j (x_{kt}^* - x_{kt-1})$$  \hspace{1cm} (6)$$

For example, in Figure 2b the time-t solution to the Lagrangian is given by $x_t^*$.
During the last period, input use was $x_{t-1}$, and the new input choice, $x_t$, is a
function of the distance to $x_t^*$ times the provider-$j$ adjustment parameter $\pi_j$. We

---

15 There are other sources of productivity growth; the incremental improvement of existing
technologies, and closely related, learning by doing. In our framework, redefining a new $x$
to reflect the improved procedure for a hip replacement, for example, captures such improvements.

16 Medieval physicians engaged in “best practice” care by the expeditious use of leeches and other
types of bloodletting, confident in the knowledge that they were providing the high quality
medical care possible care to their patients. One question is why don’t these suboptimal physicians
learn? Unfortunately, most simple learning models (e.g., Jovanovic et al, 1995) reveal the “truth”
at the end of each play, making it easy for individuals to apply Bayesian learning rules to the
sequence of patients. In real life, there is rarely such a revelation; patients may get better on their
own, despite the use of leeches, and the lack of variation in treatment techniques and the noisiness
of the individual data precludes widespread and rapid learning by the median physician.
can further approximate the change in the $k$th input by taking a first-order Taylor-Series approximation around the difference in Equation (5);

$$\Delta x_{jk} = \frac{\pi_j (\Phi_k (x_{jk-1}) - \gamma_k r_{kij})}{s_{kk}}$$

(7)

where $s_{kk}$ or the second derivative of the production function for treatment $k$, is approximated as a constant over the range of integration. That is, the growth in the use of input $k$ is (1) positively related to the speed of diffusion parameter $\pi$, (2) positively related to the social value placed on health, so that high-income countries may be expected to increase expenditures on health more rapidly, (3) positively associated with the marginal survival value of the input, evaluated at $x_{jk-1}$, (4) negatively associated with the financial constraint, for example if the physician faces binding financial constraints and loses money on the procedure, or conversely, if the provider is making money on net, and (5) negatively associated with the magnitude of the slope of the marginal survival curve $-s_{kk}$. In other words, when there are many patients who could potentially gain even minor benefits from this procedure ($-s_{kk}$ will be close to zero), the diffusion of the procedure will tend to be rapid, for example in medical imaging of patients where a large fraction of the US population may be asymptomatic but potential candidates for a given disease.

2.3.2 Quantity constraints versus financial constraints

The preceding discussion assumed that there were no capacity constraints. In the case of binding quantity constraints (for purposes of exposition, assume that $\gamma \lambda_k = 0$, so that just one type of constraint is binding), the problem is much easier; assuming global decisions on the overall quantity of services (for example, knee prostheses or the number of MRI machines), one finds an expression that restricts the growth of a given procedure across all providers:

$$\sum_j \Delta x_{jk} \leq \Delta X_k \quad \text{when } \gamma = 0 \text{ and } \lambda_k \geq 0$$

(8)

The key insights from this simple model are that the level of care is affected by the desire of providers to maximize the perceived benefits of their patients,

---

17 The derivation is straightforward: the value of the survival gain from increasing utilization of procedure $k$ from $x_{t-1}$ to $x^*$ is $s(x_{t-1}) - s(x^*)$, where $s(x^*) = \eta y$ in the presence of a revenue constraint; in its absence $s(x^*) = 0$. If the slope of the marginal survival curve is approximately linear, this value will be close to $-s_{kk} \Delta x_{jk}$. 

p. 14
tempered by informational barriers to adoption, and financial incentives arising from capitation, generous or less-generous reimbursement and revenue constraints. The dynamic implications of (7) and (8) illustrate the contrast between the U.S. system where the primary constraint relates to financial decisions of the health care provider ($\gamma > 0$), and a system like that in the UK or Canada, where quantity restrictions based on central decisions of the government are the primary determinants of changes in both outcomes and costs ($\lambda > 0$). Of course, there still may be slow diffusion when physicians are not convinced that the treatment is effective, but in that case the quantity restriction is no longer binding. Clearly, any change induced by the introduction of $x_k$ is going to be as a result of either central planners providing a new supply, or of local health agencies diverting a part of their budget towards the new innovation.

While the control over growth is quite explicit in the context of quantity restrictions, the primary control over U.S. health care expenditure growth is through the setting of prices, for example reimbursement rates for physician or inpatient services by Medicare and Medicaid, or through the negotiation of discounts by large managed care organizations and insurance firms. Newhouse (2003) has described the efforts of U.S. government rate setters adjusting reimbursement rates (changing $r$) in an effort to restrain quantity, with decidedly mixed results. As the literature on physician behavior has emphasized, cutting reimbursement rates can have ambiguous effects on costs and quantities once one allows for both income and substitution effects. This can be seen in the context of our model, where reductions in the price vector $r$ will either have no impact on utilization when $\gamma = 0$, or negative substitution effects and positive income effects when $\gamma > 0$.

Many observers have called attention to the phenomenon that health care costs are rising in all countries, regardless of whether the provision was nationalized or without. This observation is certainly not inconsistent with the model; after all, many new innovations, even costly ones, have yielded great benefits, and thus represents money well spent for countries such as the UK and Canada where health care accounts for a much smaller fraction of GDP. But at the same time, we would certainly expect from Equations (7) and (8) that remuneration for procedures with uncertain value could account for more of the increase in expenditures in the U.S. To the extent that British or European countries are better able to order priorities by enforcing capacity constraints of therapies of dubious value, one might expect better incremental value per incremental dollar or Euro spent in European countries, a result shown in Garber and Skinner (2008).
Aaron and Schwartz (2005) suggest that in the U.S., ultimately only a “quantity constraint” program like that in the U.K. is sustainable in the long run.

2.4 Insurance and induced innovation

To this point, we have not specified how the U.S. insurance system affects the rate and nature of innovation. A previous essay by Weisbrod (1991) convincingly argued that the willingness of the U.S. insurance programs to reimburse for new innovations had led to dramatic growths in expensive technological innovations. Amy Finkelstein (2007) used evidence from the introduction of Medicare to estimate that the aggregate (general-equilibrium) effects Medicare on healthcare spending are substantially larger than individual estimates from the RAND Health Insurance Experiment. Of the total increase in spending on healthcare between 1950 and 1990, Finkelstein attributes 40 percent to the growth of health insurance. Figure 2a illustrates how widespread insurance coverage can affect the incentive to innovate or diffuse existing technology, as in Finkelstein (2007). At point $a$, the use of $x_1$ is substantially higher than what one would observe in the absence of any insurance ($x_1$ versus $x_3$). With insurance, the larger extent of the market means that drug and device manufacturers, as well as individual hospitals or physician offices, can recover the initial fixed costs of development and instillation of expensive machinery such as PET scanners and MRIs.

While the evidence is less clear on whether innovation occurs differentially because of specific incentives from insurance markets (see Acemoglu and Linn, 2004; Acemoglu, Cutler, Finkelstein and Linn, 2006), we would still expect a far more rapid rate of innovation when there is an implicit promise that large insurance companies will pay for those innovations. On the other hand, we would also expect that those marginal innovations would provide less value per dollar. Even in the absence of widespread insurance coverage, low-cost effective innovations such as penicillin and beta blockers, or even high-cost but highly effective treatments such as anti-retroviral pharmaceutical treatments, are still worth developing given the high willingness to pay among nearly anyone affected by the disease.

2.5 Aggregate Productivity Improvements

With this framework, what can we say about aggregate productivity? Summing the change in utilization over all providers and patients yields:

$$\Delta x_{kt} = \sum \sum \Delta x_{jikt}$$  \hspace{1cm} (9)
and the accompanying change in health care expenditures:

\[ \Delta M_{kt} = q_{kt-1} \Delta x_{kt} + x_{kt-1} \Delta q_{kt} \]  \hspace{1cm} (10)

where the change in the social cost \( q \) reflects changes in real costs of production (some of which will be reflected in reimbursements \( r_k \) for procedure \( k \)). The change in the value of the quality-adjusted survival is approximated by:

\[ \phi_t \Delta S_{kt} = \phi_t \sum_j s_k(x_{jkt-1}) \Delta x_{jkt} \]  \hspace{1cm} (11)

Aggregating over all \( k \) procedures:

\[ \Delta M_t = \sum_k \Delta M_{kt} + w_{t-1} \Delta L_t + L_t \Delta w_t \]  \hspace{1cm} (12)

and the aggregate improvement social value of quality-adjusted survival as:

\[ \Delta(\phi_t S_t) = \Delta \phi_t S_t + \phi_t \Delta S_t \]  \hspace{1cm} (13)

Studies characterizing average benefits of health care spending, such as Murphy and Topel (2006) and Cutler, et. al. (2006) focus on the ratio of \( \Delta M_t / \Delta(\phi_t S_t) \). Because sources of increased spending may be quite different from the sources of increased survival (as in Equations 10-13) we consider the empirical literature for guidance on which factors may have contributed most to increased healthcare costs and which to improvements in quality-adjusted life years.

To motivate our empirical review, we rank these innovations in three groups in terms of their relative ranking in \( \phi_t \Delta S_k / \Delta M_k \), a convenient measure of the contributions to productivity. This measure is simply the inverse of the cost-effectiveness ratio, which, because of its wide prevalence in the clinical and economic literature, we follow here.\(^{18}\) A common threshold in the cost-effectiveness literature is that, for the treatment to be considered cost-effective, that the ratio be $50,000 per quality-adjusted life year, or lower. Some have argued that, since this has been the nominal standard for decades, threshold values in excess of $100,000 would not be unreasonable (Hirth, et. al., 2002). However, the medical profession, and technology assessment boards in other countries, generally look askance at cost-effectiveness ratios in excess of $100,000.

We consider three general categories: (1) low-cost highly effective care; “home run” treatments with cost effectiveness ratios of as little as $50 per life-year (e.g., Cutler, 2004); these we argue are the critical fuel for productivity

\(^{18}\) Another approach is to use the “value of a life” approach. The two differ with regard to how people of different ages are treated; clearly valuing a life presumes a fixed value of a life for any age, while life-years are highly sensitive to age – so that saving a baby’s life yields more in social value than an elderly person’s life.
growth; (2) (b) high-cost effective care, with cost-effectiveness ratios from $5,000 to $50,000, and which ranges – for successful treatments – at $25,000 per life year, at least among a defined subset of patients who benefit most from such treatment. However, for this group, “off-label” use may be extensive, with marginal benefits being driven considerably lower. For example, digital mammography exhibits a cost-effectiveness ratio of $26,500 for women under age 50, but no incremental value over X-rays for women over age 50, so on average, its cost-effectiveness is $331,000 (Tosteson, et. al., 2008). The final category, (3) is care where benefits are simply not known but where practice patterns arise from tradition or rules of thumb, and the cost-effectiveness ratio is either not defined (because the ratio is divided by a number approaching zero or even negative) or is simply not known.

3. A Micro-level Look at Technological Innovation in Health Care

3.1 Low Cost Effective Care

The first category of innovation to consider is “low-cost effective care”, that which is cost-effective for nearly any hurdle, and useful for nearly everyone in the relevant population. They would be characterized by high-value low-cost ($q_0 = q_0^*$ in Figure 2b) situation where the net benefits to society of their greater use (from $x_{t-1}$ to $x_t$) would be represented by the area $A + B$ underneath the marginal productivity curve; these are often characterized by a sharply negative $s_{kk}$, but they need not be, for example the use of antibiotics.

Perhaps the most famous example of such an intervention is Joseph Lister’s recommendation that surgeons wash their hands, use gloves, and swab wounds with carbonic acid (Lister, 1867).\(^19\) The formal development of the ‘germ theory of disease’ at the end of the 20th century greatly facilitated public health initiatives to combat leading killers like malaria. The ability to mass-produce penicillin boosted the ability to fight a number of bacterial infections (National Academy of Sciences, 1970). Cutler, Deaton, Lleras-Murray (2006) argue that

---

\(^19\) Lister acknowledged that his research was influenced by the Austrian physician, Ignaz Semmelweis. Semmelweis demonstrated that maternal mortality from puerperal fever (an infection of the genital tract after giving birth) could be reduced from 12.2 percent to 2.4 percent by making physicians wash their hands with chlorinated lime between autopsy and obstetrical rotations. Semmelweis did not “have a model,” which in medical parlance would correspond to a germ theory of disease to explain his results, and that is thought to have contributed to the medical establishment’s reluctance to embrace his findings.
new drugs, primarily antibiotics such as sulphonamide drugs, played an important role in reducing US mortality between the 1930s and 1960. In particular, the development of these drugs allowed for progress against a range of bacterial killers including pneumonia, tuberculosis, dysentery and venereal disease. But these discoveries, and accompanying surges in survival, were discontinuous events.

To illustrate, in Figure 3, we graph the historical trends of mortality associated with community outbreaks of meningitis, along with the innovations developed to treat generally younger patients who had been infected with different strains of bacterial meningitis. There were some early technological developments in anti-meningitis treatments, but it was the development of low-cost antibiotic “sulfa” drugs that lead to the remarkable improvements in outcomes following outbreaks of bacterial meningitis. After the 1950s, there were incremental gains in mortality, but a large portion of the new developments, such as the third-generation cephalosporins, were designed to treat infections that had developed resistance to the first- and second-generation antibiotics.

A similar pattern was demonstrated for survival following heart attacks, with a brief period of stasis in the late 1970s and early 1980s, followed by a very rapid decline in mortality for about a decade (1986-96), a period of rapid improvement for which the primary causes were (in rough order of importance) the widespread diffusion of aspirin, other drugs, and surgical interventions such as bypass surgery (Heidenrich and McClellan, 2001). In the modern era, many quality measures developed by the Medicare Quality Improvement Organization (QIO), such as the prescription of warfarin for atrial fibrillation, influenza immunizations, and biennial eye examination for diabetics fit this category of care (Jencks (2002, 2003)).

At the beginning of life, an example of highly cost-effective interventions comes from the introduction of surfactants to treat neonatal acute respiratory distress. In the last quarter of 1989 the Food and Drug Administration (FDA) approved the use of two new formulations of this class of drugs. Schwartz et. al (1994) estimate that this drug reduced the odds of in-hospital mortality by 30 percent, and attribute 80 percent of mortality decline between 1989 and 1990 to this drug. In addition to being a miracle drug for thousands of low infant babies, surfactants also proved to be protective of costs—which declined by 10 percent for survivors, and 30 percent for decedents.

A final factor that must belong to the category of “low cost high value”
innovations is improved health behaviors. If one is judging the cost-effectiveness of medical innovations, then perhaps the most cost-effective innovation was the Surgeon General’s Report on the risks of smoking in 1964. More recently, Ford et. al. (2007) estimated that half of the decline in cardiovascular disease between 1980 and 2000 was because of changes in dietary, smoking, and other behavioral factors that had less to do with actual medical treatments and more to do with medical research (e.g., Murphy and Topel, 2006) and cultural tastes for red meat, cigarettes, and exercise.20

There are two observations that we can draw from the growth of low-cost effective care. First, medical innovation in this category is more likely to be “episodic” than continuous. As Figure 4 shows, following stagnation in the early 1980s, one-year survival following a heart attack rose by one percentage point annually between 1986-1995 (Cutler, et. al.,1998). From 1995 to 2004, however, the rate improvement in survival following a heart attack has been just 0.2 percentage points (Skinner and Staiger, 2008), while costs have continued to climb. These examples suggest a process that, rather than resembling Darwinian evolution, is closer to “punctuated equilibriums” in the sense of Steven Jay Gould; periods of stagnation followed by rapid improvements in outcomes.21 The second observation is that, by the logic of Equations (12) and (13), these low-cost effective innovations are the driving force behind the remarkable improvement in survival and health care functioning across the world. These include preventive behavior; better diets and exercise, and lower smoking rates (Ford, et. al., 2007; Cutler and Kadiyala, 2003), while examples in developing countries include low-cost interventions such as treated mosquito netting and immunization programs (Soares, et al, 2006).

3.2 High Cost Effective Care

The second classification is “high-cost effective care”, which, while clearly beneficial for most patients, also entails a substantial cost. Figure 2b provides an example in which the social price of the procedure is given by \(q_1^*\) but the resource cost to the health care provider is again very low, for convenience we set

20 Thus Richard Simmons, the exercise guru of the 1970s, likely saved more lives than Michael DeBakey, the pioneering heart surgeon.

21 We do not mean to suggest that these “episodic” changes are the necessarily consequence of supply-side factors such as serendipity or exogenous technological breakthroughs that occur ever so often. Finkelstein (2004) demonstrates that the vaccine industry responds to demand-side incentives. However, she demonstrates that this response operates through the commercialization of existing technologies, not the initiation of fundamentally new ones.
it equal to $q_0$. Thus providers continue to iterate towards the point at which marginal value is set equal to $q_0$ which is clearly below $q_1^*$. Nonetheless, at this stage there are still gains in consumer surplus which are not as large as previous – area $A$ instead of $A+B$ – but which can be highly cost-effective and contribute on net to productivity growth in health care.

Within this class, we distinguish between expensive treatments with large benefits for a clear subset of patients, but with rapidly declining or even negative benefits outside of this group (that is, $-s_{kk}$ is large in magnitude). One example of this type of treatment is Ceradase, a drug developed by Genzyme that provides a critical missing enzyme to patients with Gaucher’s disease, a rare genetic disorder. While expensive (initially, it cost upwards of $60,000 per year) it offered a complete cessation of symptoms to children who begin treatment sufficiently early before damage to bones can occur. Because no individual without Gaucher’s disease would be administered this drug, the average productivity is extremely high.

Another example is the development of anti-retroviral drugs for the treatment of HIV patients to prevent the development of full-blowns AIDS, a treatment which, despite its high costs of both development and treatment, exhibits favorable cost-effectiveness ratios under $25,000 per life year (Duggan and Evans, 2005; Lichtenberg, 2006). Indeed, one study has found that the marginal dollar value to patients in terms of future life expectancy was roughly 20 times the costs of development (Philipson and Jena, 2006).

Unfortunately, these two “blockbuster” pharmaceutical advances are the exception rather than the norm in medical care. A more interesting case, and one with much larger cost implications, is primary angioplasty, an invasive procedure where a cardiologist inserts a thin wire into the coronary arteries, and inflates a balloon at the tip of the wire to restore blood flow in the minutes after a heart-attack; the costs depend on who pays for it, but it generally runs above $15,000. It is a good example of a medical procedure with differing rates of productivity depending on characteristics of the patient, so that $-s_{kk}$ is small. Angioplasty has been shown to dramatically improve survival following a heart attack if performed within the first 12 or 24 hours following its onset, and is highly cost-
effective for this use (Hartwell, et. al, 2005). This “primary” angioplasty accounts for approximately 20 percent of heart attack patients. The other half (about 20 percent of heart attacks) occurs more than 24 hours after the heart attack, and for many of these patients with arterial occlusions, surgery has an adverse (but not statistically significant) impact on survival (David Bates, et. al., 1997), with modest benefits arising from reduced chest pain. Its value is also higher for relatively younger patients, and is highly cost effective for this group, but it is of lesser value in older populations.

During the past several decades, the fraction of patients treated surgically, whether with angioplasty or conventional cardiac bypass surgery, has diffused to about 40 percent of the population of patients with heart attacks. Given relatively modest growth (or even a potential reduction) in the number of heart attacks every year, further growth in the use of survival-producing angioplasty for this well-defined group is likely to be modest. Most of the growth arises in the general population of people with cardiac conditions, and where substantial differences in opinion exist among cardiologists as to what constitutes blockages amenable to surgical intervention. For example, per capita angioplasty and stent rates in Elyria Ohio, were four times the U.S. average (New York Times, 2006), which is itself more than twice as large as the average rate in Canada (Lucas, et. al., 2006; Alter, et. al., 2006). Thus one might expect average productivity improvements for angioplasty per procedure to be lower in the U.S. to the extent that physicians

---

22 A review of over 23 trials by Keeley, Boura and Grines (2003) noted the superiority of the intensive intervention over fibrinonic (medical) therapy in the hours immediately after a heart-attack. But this benefit is concentrated in primary angioplasty: a recent randomized clinical trial showed no better survival for angioplasty in order to open occluded arteries when performed more than a day or so after the heart attack (Hochman 2007).

23 Chandra and Staiger (2007) note that the causal effect of intensive treatments for heart-attacks improves one year survival by 18.5 percent in patients less than the age of 80, but increases one year survival by only 3.5 percent in those over the age of 80 (the latter result was not statistically different from zero).

24 Chandra and Staiger (2007) offer a dissenting interpretation of these facts. Models invoking flat-of-the-curve medicine predict that the marginal return to more intensive care is lower is more aggressive areas. Chandra and Staiger offer evidence to the contrary: at least for heart-attack treatments, more aggressive area have substantially higher returns to intensive treatments. They appeal to the notion of “productivity spillovers,” where specialization in one type of care (for example, more intensive care) crowds out the alternative form of care (the use of high-quality medical management) to rationalize their result with the more fundamental fact that overall intensity is not associated with better outcomes. In other words, areas that are practice a lot of “high-tech” medicine are good at such care, but this specialization simultaneously results in reducing the quality of alternative less-intensive therapies. The net result is that despite being better at more intensive care, more intensive areas are not more productive per se.
“work down” the appropriateness curve, a result found by Chandra and Staiger (2008).

It’s difficult to quantify the marginal value of angioplasty among the majority of people having stents inserted. A Wall Street Journal article quoted Dr. Stephen Nissen, chair of cardiovascular medicine at the Cleveland Clinic, on the classic “marginal” patient:

A common scenario, says Dr. Nissen, is a patient without symptoms whose treadmill test during a routine physical reveals potential trouble. That leads to an angiogram, where a doctor sees some narrowing and puts a stent in. ‘What is the scientific evidence that we’ve done anything positive for that patient?’ Dr. Nissen asks. (Winslow, 2007; A16).

Often, rates of diffusion occur for “off-label” use, a term for when a pharmaceutical treatment is used among patients which are not specified either by FDA approval. One example in which “off-label” use presumably attenuated average productivity was for nesiritide, which was approved by the FDA in 2001 for the treatment of acute congestive heart failure. It was quickly adopted by doctors, who were able to charge $500 for an infusion, 50 times the cost of the competing drug, nitroglycerin. By 2005, almost 600,000 people had been treated with this drug, most of whom were receiving it for off-label “tune-up” therapy, or prophylactic use against the onset of heart failure, for which there was no trial support. Subsequent research, however, demonstrated that nesiritide had an adverse impact on 30-day mortality and renal function for this off-label patient population (Topol, 2005). Similarly, roughly half of all stents have been described as “off-label” (Marroquin, et. al., 2008), although the marginal value of these “off-label” stents may not be much different from “on-label” stents.

Another category of treatment with very high incremental value for a subset of patients, but whose use is increasingly being driven by our primary assumption – that physicians will, if they can, utilize to the point where the marginal value of additional utilization is zero -- is screening and imaging.\textsuperscript{25} Figure 2c illustrates a scenario where utilization is growing rapidly (because $x^*$ is so far from current usage), where the compensation is very generous (because the compensation

\textsuperscript{25} It is tempting to attribute the growth of imaging services to malpractice fears. Baicker, Fisher and Chandra (2006) find support for this mechanism; a 10 percent increase in malpractice liability increases the use of these services by 2 percent. But variation in malpractice pressure only explains 10 percent of state-level variation in the growth of imaging services, suggesting that other factors are more important in explaining the diffusion of this service.
covers average cost, but the physician/provider faces marginal costs; \( \gamma \) is negative), but where the marginal net value is (under the assumption of our model) driven to zero under the U.S. reimbursement system.

Iglehart (2006) notes that physicians are now using computed tomography (CT), MRIs, and even positron-emission tomography (PET) for an ever-expanding list of diseases. Figure 5 notes the growth of physician services per Medicare beneficiary between 1999-2004, by category of spending. Reimbursements for major and minor procedures are small relative to the growth in imaging services, which grew by 45 percent over this time-period. Yet at least report to Congress suggests low marginal effectiveness of imaging services (MedPAC, 2003; Miller, 2005). The use of digital mammography provides a particularly good example of this combination of high average value for one group of patients – women under age 50, for whom the cost-effectiveness is a very favorable $26,500 per life-year – but quite poor average cost-effectiveness for all women, $331,000 per life-year (Tosteson, et. al., 2008).26 Pharmaceutical treatments for depression are another example where the first-generation selective serotonin reuptake inhibitors (SSRIs) provided dramatic benefits for patients with severe depression (e.g., Berndt, et. al., 1997), but where the more expensive second-generation treatments have not proven to exhibit large incremental effectiveness, particularly among the wider group of patients now being treated.

Aaron and Schwartz (1987) offer a related, but different, view of technological diffusion and cost growth. They note that new technologies involve fewer risks, for example the use of laparoscopy rather than an open surgery to remove inflamed gallbladders, and therefore increase costs not because each procedure is more expensive, but because so many more patients may now experience relative gains as a consequence; per patient costs fall, but total costs increase as the new procedure diffuses.

### 3.3 Patient preferences and choices

A final issue relates to differences in preferences across patients. This is most likely to occur where treatments entail both benefits and side-effects entailing loss of quality-of-life. In the context of our model, one might think of the true benefit from treatment \( k \) being \( s_{ik} - e_{ik} \) where \( e_{ik} \) measures individual \( i \)'s side-effects of

---

26 This is because there is no incremental value of digital mammography over standard X-rays for women over age 50.
treatment. In this case, by only focusing on \( s_{ik} \) physicians may provide the patient with a positive survival benefit (in average quality-adjusted life years) but reduce patient welfare.

For example, testing for prostate cancer, current science suggests very small survival gains for younger (pre-65) men in terms of survival from the early detection of prostate cancer, but there are also potential side effects from loss of sexual functioning and incontinence, even when the patient would have died from some other disease.\(^{27}\) Similarly, there are two options for the treatment of early stage breast cancer: mastectomy (complete breast removal) versus lumpectomy (partial breast removal which involves less invasive surgery of the lymph nodes). Both options have been shown to provide similar survival outcomes, but with different complications and side-effects. In these cases, there is no “best” option for all patients; care should be tailored to patient preferences. Despite the similarity of the two options in terms of effectiveness, a Pareto improvement can occur when people who prefer one option over the other are able to choose their preferred option (e.g., Meltzer, et. al., 2007).

While the internet and commercial decision aids are increasingly helping patients to become better informed about costs and benefits of medical options, there is still strong evidence that physicians develop rules of thumb for choosing one type of option over another. For example, 48 percent of women with breast cancer in Elyria; in Columbus, Ohio the corresponding rate was 12 percent. It is difficult to believe that patient preferences exhibit such pronounced discontinuities across adjacent geographic areas. Generally, fundamental beliefs about treatment intensity – for example, whether one wants to die at home or in the hospital – are uncorrelated with the actual health care intensity (Barnato et.al., 2007; Pritchard, et. al., 2001) suggesting large potential efficiency gains from a better understanding of patient preferences.

### 3.4 Treatments lacking strong clinical evidence of effectiveness

This final category attempts to capture the notion that for many treatments, there is remarkably little evidence on effectiveness for its typical use. Often these treatments occur because neither physicians nor the clinical literature knows about the shape of the survival function \( s(x) \). The classic example was a trial published

\(^{27}\) Nearly one-third of men over age 80 have some form of prostate cancer when they die, but only 3 percent actually die from prostate cancer.
in the *New England Journal of Medicine* on arthroscopic surgery for osteoarthritis of the knee, in which surgeons would enter the knee and clean out particles from the joint (Moseley, et. al., 2002). Over 650,000 such surgeries were performed each year at a cost exceeding $5,000 per surgery. Those in the control group received “placebo surgery”— skin incisions and simulated surgery. The trial found no gain either from arthroscopic surgery relative to sham surgery.

For ethical and logistical considerations, it is extremely difficult to conduct trials of this type to establish the efficacy of every medical and surgical therapy. Nor does the Food and Drug Administration (FDA) in the U.S. require that specific surgical devices be more effective than some alternative, only that they cause no harm. Thus the requirements for new surgical devices are far less stringent than for drugs. These problems are exacerbated when the “treatment” does not have a well-defined end-point (such as death or hospitalization) associated with measuring the intervention’s efficacy. For example, there are a variety of new (and more expensive) artificial hips, but their efficacy is not well understood since even the standard low-cost hips last for several decades, and we won’t know for many years whether these new hips are superior to the less expensive versions.

We include other services such as the frequency of office visits, and referrals to specialists along with imaging services in defining this category of care. Growth in this class of technological innovations will not necessarily be stymied by a greater reliance on cost-effectiveness analysis, because these innovations aren’t designed to treat specific abnormalities whose progress can be monitored in a trial. The remarkable variations in per-patient spending observed across academic medical centers with similar outcomes are largely due to differences in use of largely discretionary services such as the frequency of physician office visits or specialist consultation, differences in the relative intensity of imaging services, and how much time similar patients spend in institutional settings (Fisher et al., 2004). There is some evidence that suggests the growth of these services, as opposed to treatments that are administered in an inpatient setting (and amenable to evaluation by CEA), account for the lion’s share of cost growth in U.S. healthcare. For example, whether patient safety is compromised by 6-month revisit rates (for office visits) instead of 3-month revisit rates is not known. Yet there is enormous disagreement across physicians in the appropriate rates of office visits (Brenda Sirovich, et. al., 2006).

Looking at the brighter side, there is also the potential for tremendous gains by more efficient use of existing technology and resources already in place. As
noted earlier, physicians are often slow to adopt treatments with well-established benefits, for example the use of β Blockers for heart attacks and congestive heart failure. One study suggested that raising rates of β blocker use to those suggested by scientific evidence would, at negligible cost, lead to an extra 447,000 life years over the next 20 years (Philips, et. al., 2000). Regions that adopted quickly the low-cost high value treatment strategies for heart attacks experienced below-average growth in expenditures, while regions with initial levels of complex patient management experienced more rapid growth in expenditures (Skinner, et. al., 2006).

4. Is Health Care Technology Worth It?

One of the most remarkable trends of the 20th Century has been the dramatic decline in mortality and improvement in functioning, with life expectancy rising from 47.3 years at birth in 1900 to 77.8 in 2004. Nordhaus (2003) estimated that while market expenditures per capita have grown by 2 percent annually since 1900, the imputed value of health capital has grown at an even faster rate of 2.4 percent. Thus the improvement in health, properly valued, has added as much to full “consumption” as traditionally measured consumption.

Murphy and Topel (2006) and Hall and Jones (2007) adopt a life cycle model of consumption, health, and longevity to “price” the improved health benefits. In the life cycle model, lifetime well-being is the sum over one’s lifespan of felicity or utility drawn from leisure and the consumption of goods and services. More income is better, of course, but diminishing marginal utility implies that the incremental value of the last dollar spent on non-medical goods, for the third TV or additional vacation house, will decline as affluence increases. By contrast, a health innovation that adds an extra life-year will increase the individual’s well-being by “total” utility, or the entire flow of utility from goods and services consumed during that extra year, and thus become more valuable as affluence rises. Murphy and Topel (2006) find that the value of improved health through longevity has risen by 95 trillion dollars between 1970 and 2000, roughly three times medical spending during this period. Hall and Jones (2007) suggest that, as GDP per capita continues to rise (and marginal utility continues to fall), health care should account for roughly 30 percent of GDP by 2050.

One might be tempted to interpret the improved value of health capital documented by Murphy and Topel (2006) as being somehow caused by the rise in
health care spending. However, Ford, et. al. (2007) documented the causes of cardiovascular mortality decline between 1980-2000 – the vast majority of overall mortality decline during this period – and found that 56 percent of the reduction was the consequence of behavioral factors such as smoking, drinking, and lower cholesterol and hypertension prevalence without any medical intervention. (Diabetes and obesity, on the other hand, were responsible for a 17% increase in mortality during the same period.) Another 20 percent of the decline in mortality was the consequence of our first category of treatments, with remarkably low cost-effectiveness ratios: off-patent aspirin, beta blockers, and anti-hypertensives. Thirteen percent of the gains in survival occurred through effective but higher cost pharmaceuticals: ACE inhibitors, anti-cholesterol “statins” and thrombolytics, while just 7 percent derived from “high-tech” surgical interventions such as angioplasty, stents, and bypass surgery (Ford, et. al., 2007). This suggests caution in interpreting ratios involving aggregate survival gains and aggregate health care expenditures (Weinstein, 2005).

Cutler, Rosen, and Vijan (2006a) provide a better measure of average returns to health care expenditures by distinguishing between behavioral changes including smoking-related deaths and accidental deaths, and health care changes. Their published estimates suggest highly favorable average cost-effectiveness measures of $19,900 per life year from 1960 to 2000. But even these careful estimates are subject to methodological criticism, because the authors discount future expenditures, they do not discount future life years. As Braithwaite and Roberts (2006) note, standard cost effectiveness guidelines discount both future life years and dollars (although this view is not universally shared; see also Cutler, et. al. (2006b) and Eisenberg and Freed (2007)). Allison Rosen, MD, has kindly provided corresponding measures of average cost-effectiveness with both costs and lives discounted for a representative individual age 45. By this measure, average cost-effectiveness (in $2000 dollars) was most favorable in the decade of the 1970s ($64,000), but since that time, average cost-effectiveness has climbed steadily, to $159,000 in 1980-90, and $247,000 during 1990-2000. To the extent that this worrisome trend continues, it suggests that individuals may not be entirely pleased with the possibility that any future growth in disposable income.

Murphy and Topel (2003, 2006) avoid statements about causality; their primary point is that even if biomedical research accounted for just one percent of the overall improvement, it must be worthwhile. However, this point is independent of whether average costs are less than or greater than average benefits.

These calculations are based on the extensive tables in Ford, et. al. (2007) and available upon request from the authors.
will be absorbed entirely by rising health care costs (Chernew, et. al., 2003).

On the other hand, restricting attention to longevity alone will tend to understimate the benefits from medical technology (e.g., Fogel, 2004). Hip replacements, knee replacements, and back surgery for disk herniation, are all procedures not expected to improve survival, but instead are designed to improve the quality of life. Furthermore, new technology such as treatments for cardiac disease may have as much impact on functioning and disability as on longevity (Cutler, 2005).

In sum, health care is characterized by wide differences in marginal valuation, where some treatments account for very little in costs but a disproportionate fraction of the overall gains (e.g., aspirin for AMI) while other treatments account for a larger fraction of costs, but do not apparently result in large differences in health outcomes. Average benefits are driven by the former sets of treatments, while average costs are driven largely by the latter categories. On average, we may have gotten good value from health care in the past, but the trend for average value in the future is perhaps more tenuous.

5. Looking ahead: Policy options to pay for health care in the future

There is strong evidence that the gradual decline in private health insurance is fueled by growth in health care costs, particularly among workers with the lowest wage rates where health benefits comprise a large and rapidly growing fraction of wages, and for whom a wage offset is difficult (Baicker and Chandra, 2006). There is also evidence that recent increases in the price of employer provided health insurance are not valued by workers—and employers have responded by shifting workers from full-time jobs with health insurance to part time jobs without. If healthier workers, who value health insurance the least, are the first to refuse employer provided health insurance, the integrity of insurance markets to pool risks could be severely compromised. Similar trends affect the provision of supplemental retirement health insurance—by 2005, fewer than 15 percent of

---

30 In Summers’ (1989), there are no employment effects if workers valued increases in the price of benefits. Baicker and Chandra (2006) examine the effect of rising health insurance premiums (on a margin that is not valued by employees) and find that a 10 percent increase in health insurance premiums reduces the aggregate probability of being employed by 1.6 percent, increases in prevalence of part-time employment by 1.9 percent, decreases wages for workers with employer health insurance by 2.2 percent, and results in reduced hours for workers moved from full time jobs with benefits to part time jobs without.
firms offered such insurance, with a strong and continued trend downward. One response to this trend is the imposition of mandated coverage, as in Massachusetts. While not explicitly a government-run program, it has similar effects in that coverage is guaranteed for everyone, thus making any premiums paid an implicit “tax” rather than a benefit (e.g., Summers, 1989).

One possible solution to rising health care costs is the introduction of Health Savings Accounts (HSAs) which allow individuals to contribute (in pre-tax dollars) a certain amount to their account and then draw down that amount for health care spending. In the context of Figure 2a, this would ideally push utilization back to point x* and at least ensure that patients will be judicious with their spending decisions—especially in ways that reduce the use of therapies with uncertain clinical value. Some, particularly the educated, may even seek out low-cost effective therapies, and discourage the growth of our third category of expensive treatments with uncertain value.

There are two aspects of HSAs which may limit their ability to reduce growth rates. First, while patients are quite price-sensitive in their health care choices, they tend to be price-sensitive across both the highly effective (type 1) treatments and drugs as well as those of questionable effectiveness (type 3). Indeed, Chandra, Gruber and McKnight (2007) found that patients decrease their use of effective drugs in response to increased cost sharing, and as a consequence some became sicker and need additional hospital care, offsetting any of the original savings. Similarly, Hsu et.al. (2006) demonstrate that a $1000 cap on drug benefits was associated with lower drug consumption and deleterious health outcomes, especially in the sickest patients. Second, a well-designed HSA includes an upper limit on spending from the HSA with full insurance coverage above several thousand dollars. Thus many of the large-ticket items are entirely covered by insurance, attenuating any cost-saving effects.31

Thus we expect that the fundamental tension in the future will be the rising marginal efficiency costs of funding health insurance. This might appear to contrast with the results in Hall and Jones (2007) who suggest a rising fraction of health care expenditures as the marginal utility of consumption declines and the average value of an additional life-year increases. While they assume a representative agent model, the model becomes more complex in a national health insurance program, or where the government pays for the majority of health care,

31 Because HSAs are attractive to the healthiest risks, there remains that the growth of these accounts will contribute to insurance markets unraveling as the healthiest risks select out of insurance pools.

p.30
as it does currently (Himmelstein and Woolhandler, 2003), but private insurance plays an important role. As preliminary results from Baicker and Skinner (2008) suggest, the marginal value of future health care innovations may be attenuated by the necessity to raise a very large fraction of GDP through premiums or (more likely) through higher federal taxes, with the accompanying marginal efficiency losses (e.g., Gruber and Saez, 2002).

6. Discussion and Conclusion

This has been a somewhat extended journey through both the clinical and economic literature. In organizing the discussion in this way, we have sought to bring some structure to a better understanding of technological progress and cost growth, in health care. There are no easy answers to the problem of rising costs in health care; Douglas Holtz-Eakin, then Director of the Congressional Budget Office (CBO) put it best: “Social-Security is Grenada. Medicare is Vietnam.” (Wolf, 2005) Still, we suggest five broad conclusions arising from our survey of the economics and medical literature.

First, a fundamental problem facing technological developments in health care is that while there are occasional highly effective wonder treatments – aspirin for heart attacks, synthetic erythropoietin for anemia, surfactants for premature births, and so forth – the vast majority of innovations are useful for some and not as useful for others. Trials are not always designed to identify the optimal stopping point, and many therapeutic technologies are also characterized by overuse at the margin. Thus it becomes difficult to disentangle technological progress from inefficiency. This problem becomes particularly clear in light of the rapid growth in imaging, which creates both great opportunity, but also the potential for much greater costs and potential harm from the inappropriate identification of disease, including the identification of ‘pseudo-disease,’ or tumors and pre-cancerous nodules that will not actually harm or kill the patient.

Second, there are strong economic incentives to overuse technology exhibiting a high fixed cost but low marginal cost. This point has been made with regard to hospital beds – Roemer’s law, that a hospital bed tends to be filled – but applies more generally, even in the realm of prospective payment – to diagnostic and surgical treatments such as MRIs, stents, bypass surgery, and most tellingly, the $200 million fixed cost for a nuclear accelerators with unclear incremental value.

Third, health savings accounts (HSAs) have the potential to ensure that new
innovations are of real value to patients. The downside is that even HSAs must ultimately provide “catastrophic” coverage, which is increasingly being met for modern hi-technology treatments, and thus the marginal incentives are no longer present. The best hope for competition and private markets to reign in growth is with regard to choosing health care insurance plans that are genuinely less expensive—a health care plan like Kaiser that uses quantity restrictions and a philosophy of conservative care that allows for lower cost but similar quality care compared to the fee-for-service sector. Some recent proposals (for example by Victor Fuchs and Ezekiel Emmanuel) have argued for the use of government-provided vouchers so that individuals can shop for low-cost and efficient insurance policies. Demand-side design is also of particular importance to ensure that high-value drugs are actually used, since patients have a tendency to scale back on consumption of all health care (both high-value and lower value) when prices rise (Chandra, Gruber, and McKnight, 2007).

Fourth, continued pressure on private US insurance markets will make it increasingly likely that the government will occupy a central role in the future health care system, whether through mandates, subsidies, or by direct financing. Such a model would have several advantages over the present system (not least by undoing large inefficiencies in the labor market by untying health insurance from employment, the promise of reduced administrative costs, and providing, in principle, the opportunity to only reimburse productive therapies). But there are also fundamental concerns about whether centralization would dramatically improve the productivity of healthcare. As well, the literature on the shadow or marginal cost of raising tax revenue cautions that raising additional revenues may be difficult, as noted in Section 5. Other countries are straining to balance budgets in the face of increasing health care costs, but in nearly every other country, expenditures are 11 percent of GDP or less, as compared to 16 percent for the U.S. To the extent that a private market arises in the presence of government guarantees (as in the U.K.), one might expect to find greater value in “demand-side” policies designed to scale back inefficient use in these markets.

Finally, the greatest potential for productivity improvements will come from the management of chronically ill patients. The nearly two-fold differences in longitudinal costs observed across academic medical centers in the care for patients with AMI are largely due to discretionary decision-making about how frequently patients should be seen, how often similar patients are referred to subspecialists, whether patients are cared for in the hospital, and the intensity of diagnostic testing and imaging procedures (Fisher, et. al., 2004). Differences in
spending aren’t due to “what” is provided (nearly all academic medical centers have access to the latest technology), but how it is provided; that is, the organization of hospitals and health care, which – at least according to evidence from England – lags well behind other industries (Bloom, et. al., 2007).

Information technology and competition among insurance plans may provide the best opportunity to ensure efficient spending for health (Walker, et. al., 2005). One recent randomized trial found that physician prompting programs reduced both medical “missteps” and scaled back costs by 6 percent (Javitt, Rebitzer, and Reisman, 2007). That one-half of women without a cervix are tested for cervical cancer (Sirovich and Welch, 2004) suggests a failure of organization rather than supplier-induced demand. But the current economic incentives work against this type of quality improvement, since providers lose revenue, as in the case of a back pain clinic steering patients to low-cost rehabilitation rather than straight to hospital-based diagnostic tests and potential back surgery (Fuhrmans, 2007). One exception is Medicare’s Physician Group Practice demonstration (CMS, 2005) in which CMS offers physician groups a share of any savings below a projected target growth rate if they also meet quality targets. Any future productivity gains in U.S. health care suggests a shift away from a “craftsman” model of physician care in a cottage industry of individual hospitals, to one where there are more common computer-assisted informational templates. Of course, physician practices could coexist with this type of system, much as highly skilled woodworkers coexist with the Home Depot.

Determining which treatments are beneficial to patients – not simply with regard to survival, but also with regard to whether they actually want the treatment – is critical to ensuring that the marginal dollar spent in the U.S. health care system yields real value. Ensuring this standard for efficiency also would make it that much easier to justify and allow for growth in health care expenditures – when the dollars spent are worth it, growth in spending is no longer really a problem. Our major concern is the ever-larger fraction of GDP drawing resources from other sectors of the economy, but with poor marginal benefits provided to patients will, in the absence of major renovations in the current health care system, lead to significant stagnation in overall U.S. productivity growth.
References


Baicker, Katherine and Amitabh Chandra, “Medicare Spending, The Physician Workforce, and The Quality Of Health Care Received By Medicare Beneficiaries.” Health Affairs, April 2004: 184-97.


Productive Inefficiency, Expected Outcome Variations and Price Indexes,
*Journal of Health Economics*, 21(3): 373-396,


Enkoji, M.S., "FBI targets 2 heart doctors: Redding MDs are suspected of possible health-care fraud," *Sacramento Bee*, November 1, 2002.


p.36


Hartwell, D, J Colquitt, E Loveman, AJ Clegg, H Brodin, N Waugh, et. al.,


National Academy of Sciences (1970) The Life Sciences, Committee on Research
in Life Sciences of the Committee on Science and Public Policy, Washington DC.


Sirovich, Brenda and Welch, H. Gilbert. 2004 “Cervical Cancer Screening Among Women Without a Cervix,” *JAMA* 291: 2990-2993


Figure 1. Percent of GDP spent on healthcare. Data for Germany refer to West Germany. Data for 2005 are estimates based on actual expenditures through 2004. See OECD (2005) for other countries.
Figure 2a: Marginal Productivity of a Specific Health Care Treatment.

Note: The solid line (JJ’) represents patients ranked in the order of decreasing social benefit \( \varphi_{sk} \) from the use of intervention \( k \). \( q \) represents the social cost of the input \( k \), \( x^* \) is the corresponding socially optimal allocation. The curve ZZ’ denotes the demand curve for healthcare in the absence of insurance; patients are ranked by \( \varphi_{isk}(x) \). \( X_1 \) represents the use of procedure \( k \), when reimbursement is lucrative, and the revenue or price constraint is binding. Area C denotes inefficiency from this allocation relative to the social optimum.
Figure 2b: Schematic of a High Value Technology with sharply declining marginal value

Note: This graph illustrates both high-value low-cost technologies, such as aspirin for heart attacks or antibiotics for infection, where the social cost of the treatment is very low ($q_0$) and hence the social surplus is exceedingly large. Here, diffusion from $x_{t-1}$ to $x_t$ generates a consumer surplus of area $A + B$. It also illustrates the case where the costs may be quite large ($q_1$) but where the sharply diminishing return for additional patients ($-s_k$ is large) limits its diffusion. Examples include antiretroviral HIV treatments or Ceradase for the treatment of Gaucher’s disease. Diffusion in this case would be marked by $A$, but even when the social optimum of $x_t$ is exceeded, the extent of the losses are limited given the large magnitude of $-s_k$. 

p.44
Figure 2c: Schematic of Technology Growth with Low Social Value

Note: Marginal net value is minimal when there are many people who qualify for the treatment, and when compensation is very generous. In this graph, there is no net marginal benefit over costs when utilization increases from $x_{t-1}$ to $x_t$; the costs of the procedure are equal to the social benefits gained.
Figure 3. Mortality Rates Associated with Community Acquired Bacterial Meningitis over the past 90 years. Source: Swartz, MN, *NEJM* Oct 28, 2004
Figure 4. Adjusted One-Year Survival For Elderly Medicare Enrollees with an AMI: 1986-2002. The vertical axis measures the percentage who survive the index acute myocardial infarction (AMI). This is equal to 100 minus the percentage one-year mortality. The left dashed line (1984-94) is from Cutler and McClellan, right solid line (1986-2002) is from the authors’ calculations.
Figure 5. Growth in the Expenditures on Physician Services per Beneficiary, 1999–2004. Evaluation and management service includes office visits and hospital visits. The category "Tests" excludes imaging. Source: Iglehart (NEJM 2006)