Saving Money or Just Saving Lives? Improving the Productivity of US Health Care Spending

Katherine Baicker,1,4 Amitabh Chandra,2,4 and Jonathan S. Skinner3,4

1Harvard School of Public Health, Harvard University, Boston, Massachusetts 02115; email: kbaicker@hsph.harvard.edu
2Harvard Kennedy School, Harvard University, Cambridge, Massachusetts 02138; email: amitabh.chandra@harvard.edu
3Department of Economics, Dartmouth College, Hanover, New Hampshire 03755; email: jonathan.skinner@dartmouth.edu
4National Bureau of Economic Research, Cambridge, Massachusetts 02138

Abstract

There is growing concern over the rising share of the US economy devoted to health care spending. Fueled in part by demographic transitions, unchecked increases in entitlement spending will necessitate some combination of substantial tax increases, elimination of other public spending, or unsustainable public debt. This massive increase in health spending might be warranted if each dollar devoted to the health care sector yielded real health benefits, but this does not seem to be the case. Although we have seen remarkable gains in life expectancy and functioning over the past several decades, there is substantial variation in the health benefits associated with different types of spending. Some treatments, such as aspirin, beta blockers, and flu shots, produce a large health benefit per dollar spent. Other more expensive treatments, such as stents for cardiovascular disease, are high value for some patients but poor value for others. Finally, a large and expanding set of treatments, such as proton-beam therapy or robotic surgery, contributes to rapid increases in spending despite questionable health benefits. Moving resources toward more productive uses requires encouraging providers to deliver and patients to consume high-value care, a daunting task in the current political landscape. But widespread inefficiency also offers hope: Given the current distribution of resources in the US health care system, there is tremendous potential to improve the productivity of health care spending and the fiscal health of the United States.

Keywords

productivity in health care, Medicare, technology, health care spending
1. INTRODUCTION

Health care spending in the United States now represents one-sixth of GDP and is projected to comprise more than one-quarter of it within 25 years (Congr. Budg. Off. 2010). Spending a large and increasing share of GDP on health care in and of itself is not necessarily troubling: As societies grow richer, it would seem natural that an increasing share of their resources should be devoted to “purchasing” health and longevity. After all, living an extra year in good health may be worth more on the margin than a bigger home, a longer vacation, or a more sumptuous meal. Some studies suggest that, given the relative productivity of health spending, the improvements in health outcomes driven by health care spending outweigh the costs, and we might optimally spend as much as one-third of GDP on health by the mid-century (Hall & Jones 2007, Fogel 2008). If the aggregate health benefits of care truly outweigh the total social costs, there is no a priori reason to be concerned about devoting one-quarter of GDP, or more, to living longer and healthier lives.

Why then is there such distress over rising health expenditures? There are two reasons to be less than sanguine about health care spending in the United States, and indeed in most developed countries (see Table 1). First, public expenditures account for almost half of US health care spending (Cent. Medicare Medicaid Serv. 2010). Most analyses focus on the value of health care spending but neglect the substantial redistributive costs necessary to pay for the health care—the efficiency costs arising from transferring resources through the tax system from rich to poor and from healthy to sick. Federal spending on public programs (including public insurance programs for the poor and elderly and newly enacted subsidies for the poor to purchase private insurance) is anticipated to grow from 5.5% of GDP now to almost 14% or more in 2060 (Congr. Budg. Off. 2010). Unchecked, increases in entitlement spending would double the federal budget as a share of GDP. Although this spending could be financed with the elimination of other public spending or rising debt, higher tax rates would be the most likely outcome, with accompanying economic distortions and diminished GDP (Congr. Budg. Off. 2007, Chernew et al. 2010, Baicker & Skinner 2011). Such costs must factor in to any analysis of efficient levels of health care spending.

Second, there is a growing consensus that health care resources are not being spent efficiently (and may not even be the primary driver of improved outcomes): We are neither allocating resources efficiently between health and other uses nor getting as much health as we could for every dollar spent—making it difficult to evaluate how much we “should” be spending on health care.1 Such inefficiencies are clearly of first-order importance as health care spending encompasses an increasingly larger share of total resources.

We begin by describing the landscape of spending in the United States in the context of the macroeconomy, how it has evolved over time, and how it compares to spending patterns and trends in other developed countries. On first examination, the US health care system may appear to be an impenetrable morass of regulations and acronyms, and it is tempting to view its failures as the sum of disparate inefficiencies. Rather than focusing on each individual inefficiency, we instead outline a simple framework in which one can judge how well a health care system functions. A well-functioning system should exhibit productive efficiency, meaning that health care resources are put to the best use possible and produce as much health as they can, and allocative efficiency, meaning that the right share of resources is being devoted to health care versus other goods in the economy. The

---

1Cutler’s (2004) interpretation is somewhat different. He 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.”
peculiar financing of health care in the United States fosters both allocative and productive inefficiencies. Some have suggested that with system-level improvements, it would be possible to achieve the same level of health with 20%–50% less spending (Fisher et al. 2003a,b; Skinner et al. 2005; Farrell et al. 2008; Buntin & Cutler 2009).

This does not mean, however, that cutting spending on health care will automatically lead to greater efficiency. Cutting spending in an already disorganized hospital may

### Table 1 Health care share of GDP for OECD countries, 1980 and 2008

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1980</td>
<td>2008</td>
<td>Health care spending</td>
</tr>
<tr>
<td>Spain</td>
<td>5.3</td>
<td>9.0</td>
<td>4.5</td>
</tr>
<tr>
<td>Portugal</td>
<td>5.3</td>
<td>9.9</td>
<td>4.4</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5.6</td>
<td>8.7</td>
<td>3.8</td>
</tr>
<tr>
<td>Greece</td>
<td>5.9</td>
<td>9.7</td>
<td>3.1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>5.9</td>
<td>9.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Australia</td>
<td>6.1</td>
<td>8.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Belgium</td>
<td>6.3</td>
<td>10.2</td>
<td>3.3</td>
</tr>
<tr>
<td>Finland</td>
<td>6.3</td>
<td>8.4</td>
<td>3.0</td>
</tr>
<tr>
<td>Iceland</td>
<td>6.3</td>
<td>9.1</td>
<td>2.4</td>
</tr>
<tr>
<td>Japan</td>
<td>6.5</td>
<td>8.1</td>
<td>2.5</td>
</tr>
<tr>
<td>Canada</td>
<td>7.0</td>
<td>10.4</td>
<td>2.9</td>
</tr>
<tr>
<td>France</td>
<td>7.0</td>
<td>11.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Norway</td>
<td>7.0</td>
<td>8.5</td>
<td>4.3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>7.3</td>
<td>10.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Austria</td>
<td>7.4</td>
<td>10.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Netherlands</td>
<td>7.4</td>
<td>9.9</td>
<td>3.1</td>
</tr>
<tr>
<td>Ireland</td>
<td>8.2</td>
<td>8.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Germany</td>
<td>8.4</td>
<td>10.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Denmark</td>
<td>8.9</td>
<td>9.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Sweden</td>
<td>8.9</td>
<td>9.4</td>
<td>1.6</td>
</tr>
<tr>
<td>United States</td>
<td>9.0</td>
<td>16.0</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Data are for 2008, 2007, or 2006(‘). For Germany, the 1980 values are for West Germany. Alternatively, one can calculate growth in health care spending relative to GDP for 1980–1990 and for 1992–2008 (thereby avoiding the transition). This yields an increase of 0.8 percentage points of GDP in health care expenditures. The share of health care spending comprising public dollars does not include tax exclusion of employer-sponsored insurance. Data taken from Chandra & Skinner (2011), drawing on OECD data, 2010.
substantially harm patient outcomes. What the inefficiency estimates highlight is the potential to improve patient outcomes without spending more on health care, for example, by moving elective surgery away from low-volume (and high-mortality) hospitals into high-volume (and low-mortality) regional surgery centers (Birkmeyer et al. 2002). The objective should not be to reduce spending per se, but to increase the health gained for every dollar spent. Indeed, in an efficient system, more spending on health care would be a sign of prosperity and a harbinger of improved health and longevity, not a cause for concern. Achieving this means improving the incentives and infrastructure for providers to deliver—and patients to consume—high-value care, as well as wrestling with the difficult question of whom to cover versus what to cover in public insurance programs.

2. THE US HEALTH CARE–SPENDING LANDSCAPE

To gauge the welfare implications of rising health care spending and evaluate potential policy changes, one must understand the drivers of spending in the current system and how they do or do not align with incentives for efficient use.

2.1. Health Spending in the Economy

Figure 1a (see color insert) shows the growth of health care spending in the United States. Real health care spending (measured in 2009 dollars) has increased from $330 billion in 1966 to an estimated $2.7 trillion in 2011 and from 6% of GDP to 16%. Hospital spending is the largest category, although spending on prescription drugs has been growing more quickly. The usual explanation is that expenditure growth is caused by the growth in health care technology (Newhouse 1993), perhaps coupled with a bad case of “Baumol’s cost disease,” in which labor-intensive sectors exhibit less productivity growth than other industries.2

The problem with these explanations is that they apply equally well to other developed economies with access to all the new devices and treatments available in the United States (Chandra & Skinner 2011), but we see very different trends. Table 1 shows the share of GDP devoted to health care expenditures in 1980 and 2008, along with growth rates in both health care expenditures and GDP. Although the growth of US health care spending is not an outlier, the growth of the share of GDP devoted to US health care spending has outstripped that of other OECD (Organization for Economic Co-operation and Development) countries—rising from 9% in 1980 to 16% in 2008. By contrast, Sweden and Denmark began with nearly identical levels of GDP devoted to health care but rose only 1 percentage point.

Health care is different from other goods in many ways that may interfere with the efficient allocation of resources based on marginal costs and marginal benefits: Patients have limited information about the benefits associated with the care that they purchase (relying on the suppliers of that care for advice) and often need to make decisions in difficult circumstances; both insurers and providers usually operate with limited competition. As noted above, the way we finance health care introduces an important additional set

---

2The classic example of (William) Baumol’s cost disease is a live performance by a string quartet: Four highly skilled people are required to play the performance, and so one would not expect to observe productivity growth in live quartet performances. Yet wages of musicians would presumably keep pace with other skilled workers whose salaries do benefit from productivity growth in other sectors of the economy.
of wedges between the marginal cost of care and the marginal health benefit that care produces. Figure 1b shows how health spending in the United States is financed. Most care in the United States is purchased through private insurance (largely obtained through employment) or public insurance programs (largely Medicare and Medicaid, introduced in 1965). As Table 1 shows, whereas the United States finances a smaller share of spending with public dollars than other OECD countries, the overall level of spending is so much higher that the share of GDP devoted to public health care spending is actually quite similar to other countries such as France and Sweden. Both the public and private financing mechanisms in the United States have particular implications for the economy overall and for the efficiency with which health care resources are deployed.

2.2. Private Health Spending

The majority of private health insurance in the United States today is obtained through employer-sponsored plans, which cover more than 170 million lives (compared with under 20 million nonelderly individuals covered through private nongroup insurance plans) (DeNavas-Walt et al. 2010). The dominance of employer-based health insurance in the private market has historical roots: During a period of wartime price controls, it was determined that such plans were not subject to wage controls and were not taxable, making them a favored form of compensation. At more than $250 billion annually, this tax subsidization of employment-based private health insurance through its exclusion from payroll and income tax bases is as large as federal spending on Medicaid (Gruber 2010) and larger than the mortgage interest deduction. It disproportionately favors those with higher incomes and jobs that offer generous health insurance benefits and comes with the deadweight loss associated with the tax expenditures that finance the subsidy (discussed more below). Employer plans, however, form the primary mechanism for risk-pooling among the privately insured: The tax subsidy drives participation in health insurance by healthy as well as sick at pooled premiums, and risk is pooled to the extent that employees are not choosing jobs based on health insurance benefits. One consequence of this pooling mechanism, however, is that it makes changing jobs costly for those in poor health who would be unable to obtain insurance at favorable rates if they lost their employer plan (Madrian 1994).

Figure 2 (see color insert) shows the increase in premiums for family insurance plans purchased through employers. There is much public discussion of how the rise in health insurance costs affects US businesses’ international competitiveness, but most evidence suggests that in the long run, workers bear the costs of the higher premiums. In the short run, however, there are constraints that affect the ability of firms to shift costs to workers, and thus rising costs can result in fewer people covered by group policies (Cutler &

It is worth noting that much of the debate over health reform muddied the concepts of health care and health insurance. Health insurance is, of course, fundamentally about risk—it is valuable not just because health care is expensive, but because it is expensive and uncertain. Uninsured sick people need health care, not insurance: Once their illness is known, it is no longer insurable (Baicker & Chandra 2008). It is also important to acknowledge that many inputs into long-run health outcomes are independent of the health care system, such as nutrition, exercise, smoking, and environmental hazards.

Additionally, it is important to note that (a) the burden of rising health insurance payments depends also on the extent to which they are valued by workers (Summers 1989), and (b) the share of premiums nominally paid by workers is not indicative of the true burden of insurance payments. What matters is the extent to which worker salaries (or returns to owners of corporate and noncorporate equity) decline in response to the changes in the funding of health care (Ballard & Goddeeris 1989, Currie & Madrian 2000).
Madrian 1998, Currie & Madrian 2000, Baicker & Chandra 2006). Although the vast majority of large employers offer health benefits, over the past 10 years the share of small firms offering health benefits has dropped from 69% to 61%, and rising costs are cited as one of the main factors influencing firms’ decisions to discontinue offering benefits (Kaiser Fam. Found. Health Res. Educ. Trust 2008). This means that rising health care costs can erode the risk-pooling in the private market and exacerbate adverse selection. Although most of the burden of rising private health care spending is borne in slower or stagnant wage growth (as an increasing share of compensation is devoted to health insurance benefits), the consequences of rising premiums can be particularly severe for low-income populations at the greatest risk of losing jobs when the costs of benefits rise (Baicker & Levy 2008).

In addition to the labor market consequences of employer-sponsored health insurance, the tax preference for this insurance relative to out-of-pocket spending promotes first-dollar coverage plans. There is always a trade-off between insurance protection and incentives: Insurance insulates patients against out-of-pocket payments for unexpected and potentially catastrophic health care payments. But it also means that patients have little incentive to economize on care with questionable benefits—why not get the $25,000 treatment even if the $2,000 treatment is just about as good? In a fee-for-service environment, this distortion is exacerbated by the provider’s incentives (the profit margin on the $25,000 treatment is likely to be considerably larger for the physician and hospital), as well as by the patient’s imperfect information about the value of the services.

Why can’t insurance plans in the United States combat insurance-induced overconsumption by offering cheaper insurance policies that disallow expensive and largely unproven treatments? In theory, they could. But in practice there are additional constraints on insurance contracts that limit the availability of lower-cost, higher-value plans. In a world of evolving medical technology and complex care management, it is impossible to fully specify contingent contracts outlining all the care that individuals can receive in every state of the world, and private insurers have limited ability to deny coverage for procedures with unproven benefits when Medicare covers these technologies (Ferguson et al. 1993).

US corporate laws make it difficult for individual insurers and hospitals to reduce the use of technologies with variable payments: Insurers and hospitals are not permitted to interfere with the medical judgment of physicians. State laws also require insurers to pay for any service deemed medically necessary by a physician. Because enrollees eventually leave private insurers (either for other plans or to enroll in Medicare), insurers further face a dulled incentive to invest in care with health benefits that accrue only in the future. Together, these undermine the incentive for insurers to act as a residual claimant, leaving insurers to rely more heavily on negotiating lower prices with providers rather than restraining utilization to control spending. Furthermore, that most private insurance plans are purchased through employers may obscure the costs and benefits for employees (even if employees eventually pay for them in the form of lower wages) and may limit the extent to which plans match the preferences of employees on the margin.

2.3. Public Health Insurance

The landscape of health insurance coverage in the United States has been largely shaped by the 1965 creation of Medicare, the federal program for the elderly and disabled, and Medicaid, the joint federal-state program for the poor and other medically needy groups. Medicare covers 49 million people and costs the federal government over $500 billion annually.
Medicaid covers more than 60 million people (some of whom are also on Medicare, with eligibility criteria that vary across states) and costs approximately $375 billion annually: $250 billion from federal revenues and $125 billion from states and localities. In 1997, the State Children’s Health Insurance Program expanded coverage for low-income children above the poverty level and now covers 8 million with a cost of approximately $11 billion. There are numerous other smaller public programs covering, for example, military personnel and veterans, as well as the direct provision of care through community health centers. As noted in Table 1, public dollars account for nearly half of national health expenditures in the United States, even before accounting for the tax subsidy for private insurance described above. The Affordable Care Act (ACA) expands Medicaid coverage and introduces additional subsidies for low-income adults to purchase private insurance, adding substantially to public spending on insurance (offset at least in part by additional revenue sources).

Whereas Medicaid poses an enormous fiscal challenge for state governments, Medicare poses a greater challenge for the federal government. Medicare expenditures account for 15% of federal spending and 3.6% of GDP (Boards Trustees Fed. Hosp. Insur. Fed. Suppl. Med. Insur. Trust Funds 2011). Medicare spending grew 2.5 percentage points faster than GDP from 1975 to 2008, compared with 1.9 percentage points for health care spending overall (Congr. Budg. Off. 2010). The health care used by Medicare beneficiaries is financed by a combination of dedicated taxes and general revenues (as well as supplemental plans and the 25% of care that beneficiaries pay through premiums or out-of-pocket spending). The payroll taxes dedicated to financing hospital spending are paid into a trust fund, but since 2009 spending has grown more quickly than the tax stream, with the trust fund forecast to be exhausted in 2024. Program costs for physicians, other outpatient care, and drugs are financed mostly through general revenues and beneficiaries’ premiums. Rapid spending growth on these components means that general revenues are an increasingly important component of Medicare financing. Although the aging of the population is clearly one driver of higher spending, the majority of rising expenditures can be attributed to higher spending per beneficiary (Congr. Budg. Off. 2007). Both the rising share of elderly individuals in the population (attributable to an aging Baby Boom generation and greater longevity) and higher spending per beneficiary mean that public health insurance programs will compose a rapidly rising share of both GDP and federal revenues (Figure 3a; see color insert).

Public insurance programs provide benefits that are extremely valuable to beneficiaries. Medicaid redistributes resources toward low-income and sick populations, and Medicare pools risk in a retiree population without access to employer pools. Finkelstein (2007) estimates that in 1963 only 25% of seniors had comprehensive insurance, but shortly after the advent of Medicare virtually all did. This redistribution is a key feature of social insurance that cannot be achieved through private contracts alone, and most countries outside of the United States rely more heavily on publicly financed health care. Of course, public funding of health care does not require that the government actually provides the health care services (as in England) or even act as a large insurance company, paying providers for services (as in Medicare in the United States). A voucher or premium support plan could pay Medicare premiums to private insurance companies or to integrated health care systems, which would then be responsible for health care delivery.5

5An example is in Representative Paul Ryan’s proposal to convert Medicare to a uniform voucher plan [http://www.roadmap.republicans.budget.house.gov/]. Similar features in the Medicare Advantage program are discussed below.
These social benefits must be balanced, however, against the costs associated with the programs’ financing. Raising taxes to pay for public insurance exerts a structural drag on the economy (deadweight loss). To the extent that health sector jobs created by rising spending are not producing substantial improvements in health, they divert resources from higher-value uses such as hiring additional teachers in public schools. Deficit spending on health care also carries an economic cost: Taxes are required to pay interest and principal on any borrowed money, and rising debt-to-GDP ratios eventually have serious adverse effects on the country’s future ability to borrow. Projections of growing federal debt largely reflect anticipated increases in health care spending (see Figure 3b). Even if policy could halve the gap between the growth in health care spending and the growth in GDP, some estimates suggest that our debt-to-GDP ratio would drop only from 300% to 200% by 2050 (Kogan et al. 2008).

There is a substantial body of research exploring the magnitude of the deadweight loss generated for each additional dollar of revenue raised in the US tax structure, with many estimates around 0.3, meaning that every dollar of public spending comes with an additional cost of lowering economic activity by 30 cents (Feldstein 1973, 1999, 2006; Fullerton & Henderson 1989; Gahvari 2006). These tax distortions are not just static in nature (lowering economic activity at a point in time) but can also affect economic growth by distorting investment decisions (Engen & Skinner 1996).

Financing current trends in public program costs would necessitate dramatic increases in tax rates. The Congressional Budget Office estimated (before the ACA) that income tax rates would have to increase by more than 70% to finance health care spending that grew 1 percentage point faster than GDP and would have to increase by more than 160% by 2050 to finance growth at the historical rate of 2.5 percentage points faster than GDP growth. Even with just 1-percentage-point excess growth in health care spending, the Congressional Budget Office estimates that the tax increase would reduce GDP by 3%–14% (Congr. Budg. Off. 2007). This additional cost of public health insurance must be included in any cost-benefit analysis of the program, as rising marginal tax rates inflict rapidly mounting deadweight losses to the economy. Baicker & Skinner (2011), for example, calculate that by the middle of this century, for every $1 in future tax-financed health care spending, nonhealth production would have to fall by more than $2.

Most forecasts of health care–spending growth assume that there must be a break in current trends—health care spending simply cannot continue to grow at historical rates—but it is not clear what will cause that slowdown. Getzen (1992) postulates that health care spending at the country level is driven in large part by the ability to finance health care, and this hypothesis is consistent with evidence presented in Baicker & Skinner (2011) that countries with a higher tax-to-GDP ratio in 1979 experienced significantly slower growth in health care spending between 1980 and 2008. Rising costs of taxation (and debt) may thus ultimately serve as a brake on health care–spending growth, potentially also helping to overcome stakeholder resistance to other policies that improve the efficiency of public programs.

The deadweight loss of financing programs such as Medicare is not the only inefficiency they generate. Medicare performs no evaluation of the benefits associated with new medical

---

6These estimates correspond to a tax with revenues returned through lump-sum or a “negative income tax” scheme. The incremental effects of such a tax could be different from one that is used to provide an in-kind transfer (Currie & Gahvari 2008).
technologies, and in its fee-for-service incarnation does not ask if care could be better managed. This discourages cost-saving innovation and efficient insurance offerings. For example, traditional Medicare does not coordinate hospital, outpatient, ambulatory, and prescription drug care, and it pays for this care without regard to the care’s broader effects or the modes or sites of delivery that work best. If a private insurer providing prescription drug coverage changes program features in a way that reduces prescription drug use but results in patients requiring more hospital care, the Medicare program bears that cost. Chandra et al. (2010a) find evidence of exactly these spillovers.

Other inefficiencies are not restricted to the Medicare program, affecting the style of physician practice in the non-Medicare population. When Medicare covers Provenge for prostate cancer (at a cost of over $90,000 for a few months of survival), private insurers are likely to follow the coverage decision to avoid litigation in which their patients claim that insurers are withholding valuable care (Ferguson et al. 1993). Medicare regulatory boards evaluating new technology focus on whether drugs or procedures provide any benefits and are typically precluded from considering costs.7

There is also a substantial literature at the provider level showing that practice pattern norms drive similar care for all the patients that a provider sees, regardless of individual insurance status. Therefore, changes in the incentives applying to a large share of patients (e.g., Medicare beneficiaries) can drive changes in the care received by all patients (Baker & Corts 1996, Baker 1999, Glied & Zivin 2002, Frank & Zeckhauser 2007).8 Medicare’s administered pricing scheme is frequently out of step with real resource costs and subject to political manipulation (as witnessed by the regular override over the “sustainable growth rate” formula intended to cut physician fees when total Medicare spending rises too quickly).

On the patient side, in an attempt to control the overuse of care, the basic Medicare benefit requires beneficiaries to pay substantial cost-sharing—almost 20% coinsurance for outpatient care. But almost all beneficiaries buy additional Medigap wrap-around plans or receive Medicaid benefits that eliminate most copayments and thus any incentives for consumers to choose lower-cost treatment options. This in turn imposes additional costs on the Medicare program itself. The Medicare Advantage program was intended to promote innovation in insurance coverage and provision of care, but there has been limited evidence of success to date. Current proposals for “premium support” (in essence an insurance voucher) have similar goals.

Thus neither the patient side nor the provider side of the system fosters decision making that allows for a weighing of costs against benefits. Nor is there an evaluation of the benefits of additional spending through one program against the benefits of additional spending through another. The spending on care of questionable marginal benefit fostered by the Medicare program, for example, might come at the expense of covering more people in the Medicaid program who could benefit from any insurance coverage (Baicker & Chandra 2010). Recent research suggests that Medicaid coverage substantially expands

---

7In other countries, boards such England’s National Institute for Health and Clinical Excellence explicitly use costs to determine coverage decisions. Typically these rulings are limited to specific and discrete choices, for example, whether specific drugs can be used for a specific disease. In practice, treatments with heterogeneous benefits are more difficult to classify under “cover/not cover” decision rules.

8However, Franzini et al. (2010) find different patterns of utilization among the under-65 privately insured compared with the over-65 Medicare population in two Texas cities.
the use of preventive care in low-income adult populations, but such populations are not covered by Medicaid in most states (Finkelstein et al. 2011).

3. ASSESSING THE EFFICIENCY OF HEALTH SPENDING

The above discussion focuses on specific inefficiencies in the design and financing of public and private health insurance plans. The tax-preferred nature of financing private plans, the limited use of cost-effectiveness analysis, and the overly generous public reimbursement of care with questionable health benefit might suggest that the United States spends too much on health care relative to other goods. At the same time, we have seen dramatic increases in life expectancy and reductions in disease burden over the past century (Cutler et al. 2006). Some new interventions are cheap (antibiotics, statins, beta blockers, vaccines—many of which are even underused), and others are more expensive but still remain cost-effective [HAART therapy, percutaneous coronary intervention (PCI) after heart attacks], but medical technology has high health returns on average (Cutler et al. 2006). These facts might suggest that US spending on health care has been worth it, and may even be too low in some areas.

How can we reconcile these facts? Drawing on previous work, we develop a framework for understanding the sometimes-conflicting evidence and the drivers of differential productivity of health spending (Chandra & Skinner 2011). We make the key distinction between productive efficiency, meaning that health care resources are put to the best use possible and produce as much health as they can, and allocative efficiency, meaning that the right amount of resources are being devoted to health care versus other goods in the economy. When the evidence on the value of health care spending is evaluated in light of this framework, we see that what might at first appear to suggest that we are spending too little on health care may actually be evidence that we have been spending too much on unproductive care.

3.1. Two Kinds of Efficiency

We begin in a hypothetical world of productive efficiency, in which health care resources are put to the best use possible. Figure 4 illustrates the association between spending on factor inputs (health resources such as physicians, scans, and hospitals) on the horizontal axis and survival/quality of life on the vertical axis. A concave production possibility frontier (PPF) illustrates the aggregate health that is achievable for a given level of inputs for the United States, or any particular delivery system, which could be the United Kingdom or the Mayo Clinic. The production possibility frontier assumes diminishing returns to spending: Higher levels of spending are associated with inputs that generate increasingly smaller health benefits; computed tomography (CT) scanners, proton-beam accelerators, and chemotherapy for metastatic cancers will have lower benefit per dollar spent relative to aspirin for heart-attack patients or antiretroviral therapy for HIV and AIDS patients. Under diminishing returns, physicians are first giving treatments to the patients who benefit most from the treatment, and then moving on to patients with lower benefit. Technological advancements can push the production possibility frontier up and out.

In such a world, different points on the frontier (such as A or B) represent different choices about allocating spending between health and nonhealth consumption (e.g., cars,
vacations, education, and defense). If the system is allocatively efficient, that choice would be made such that the last dollar spent on health generates the same marginal utility as spending an additional dollar on nonhealth consumption; otherwise, society would be made better off by moving resources from the lower-marginal-utility use to the higher one. Because the marginal value of spending on health care is falling, point $B_1$ produces more health than point $A_1$, but it could certainly be the case that the incremental improvement in health outcomes gained by moving from $A_1$ to $B_1$ is not worth the opportunity cost of nonhealth spending forgone. (This gap may be even larger than that shown in Figure 4 once one accounts for the economic burdens of raising tax revenue to finance spending on public programs, as noted above.) If so, we would say that choosing point $B_1$ is allocatively inefficient—we are on the efficient frontier (producing as much health as possible with given health care resources) but have spent too much on health care. Why might a system allocate at $B_1$ rather than $A_1$? The previous section describes reasons that health care markets do not resemble efficient markets for other goods and services. Several of these factors (from the tax preference for employer-sponsored insurance and the moral hazard it generates to the fee-for-service nature of Medicare coverage) push toward a greater allotment of resources to health care.

Evidence suggests, however, that we are likely in a world in which we are not on the productively efficient health care frontier, but in its interior (see the line labeled $PPF_2$ in Figure 4). Productive efficiency involves devoting health care resources to the uses in which they will produce the most health, but we see both the underuse of effective care and the overuse of ineffective care. Providers often fail to do very-low-cost things, such as hand washing, prescribing prophylactic antibiotics before surgery, or using beta blockers for heart-attack patients (Skinner & Staiger 2009). Much (but by no means all) investment in prevention would yield high returns in future health (Cohen et al. 2008). Inadequate health information technology reduces the productivity of all resources used by

Figure 4
Illustration of allocative and productive inefficiency in health care.
increasing fragmentation, errors, and duplication, along with billions of dollars wasted in inconsistent and duplicative billing systems (Emanuel 2011). These sins of omission coexist with sins of commission; estimates suggest that the excess radiation from overuse of CT and magnetic resonance imaging (MRI) scans causes 1.5%–2.0% of total cancers (Brenner & Hall 2007). The forces that interfere with allocative efficiency may also interfere with productive efficiency: Fee-for-service reimbursement and low levels of competition encourage providers to deliver care that they can bill for and underinvest in keeping patients healthy; insurers may pay more generously for intensive technologies than for therapeutic but revenue-shrinking care.

Understanding the degree of productive inefficiency in the health care system begins with characterizing the (variation in) health produced by different types of care.9 Highly effective care—including not just medicine such as aspirin, but also surgical checklists or better care management—has modest costs relative to health benefits (de Vries et al. 2010). Some care, however, is effective for some patients, but less effective (or even nonexistent) for others. For example, stents may have enormous benefits for those who have just had a heart attack but have little demonstrated efficacy for those with stable angina. Finally, there is some care with very high costs but uncertain or very small benefits. A good example of such a technology is cyclotron-based proton-beam therapy, a very expensive treatment with no evidence of better outcomes than conventional and much less expensive approaches (Kagan and Schulz 2010).10 Another example involves aggressive treatments for advanced lung cancer. In one randomized trial, patients receiving “normal” treatments experienced worse quality of life and shorter life expectancy (and higher resource use) than those receiving early palliative care (Temel et al. 2010).11

A key assumption in many productivity analyses is that health care providers adopt the most effective technologies first and apply them to patients with the highest marginal cost-effectiveness first. If this assumption does not hold (as evidence of overuse and underuse suggests), inferences about productive efficiency can be misleading. Consider the illustration of two hospitals adopting technologies in a different order in Figure 5 (see color insert). Each begins with baseline spending and outcomes A. Hospital X then performs highly effective treatment X₁ before adopting moderately effective (or effective

---

9Wennberg and colleagues (2002) delineate care differently, dividing it into effective, preference-sensitive, and supply-sensitive categories. These categories reflect the underlying causes of use (and thus map differently to potential policy solutions). Aspirin for heart attacks would be considered effective under their definition, whereas usual care compared with early palliative care for metastatic lung cancer is more supply sensitive.

10Proton-beam therapy requires an accelerator weighing 150 tons and costing $120 million. Although these accelerators are highly effective for rare forms of tumors, particularly in children, there are not enough of these cases to justify building more than a few facilities in the United States. The business model for the rapid growth in proton-beam therapy is that Medicare pays up to $50,000 for the treatment of prostate cancer, even absent any evidence of greater efficacy than that for alternative treatments (Kagan & Schulz 2010).

11What share of health gains can be attributed to care in each of these categories? One recent study examined factors contributing to the decline in mortality from coronary disease during 1980–2000 (Ford et al. 2007) and found that 35% of the decline could be attributed to highly effective care such as aspirin and 11% to care with variable benefits such as stents. In more detail, over 40% of the decline is attributed to changes in smoking, physical activity, blood pressure, and cholesterol (and offsets from rising obesity and diabetes); 35% to inexpensive but highly effective treatments [aspirin, beta blockers, blood-thinning drugs, antihypertensives, diuretics, and pharmaceuticals such as ACE inhibitors, statins, and thrombolytics (“clot-busters”)]; and 11% to care such as angioplasty (stents), bypass surgery, and cardiopulmonary resuscitation (such as automated defibrillators). The residual health gains, just over 10%, in theory could be attributed to the third category of spending with uncertain or negligible health benefits, but which in turn accounted for the vast majority of costs.
in only some patients) treatment X₂. Hospital Y, however, performs moderately effective treatment Y₁, ineffective treatment Y₂, and only then effective treatment Y₃. An adoption pattern like this could drive some analysts to conclude that more spending was better and others to conclude that it was worse. Comparing Hospital X and Y, as in the line connecting X₂ and Y₃, would suggest a negative correlation between spending and outcomes (even though all technologies shown here do no harm). Comparing point A to X₂ or Y₃ within a hospital would suggest a positive correlation between spending and outcomes (although the improvements in health arise largely from technologies X₁ and Y₃, whereas the increases in spending arise from Y₂). Evidence of the productive efficiency of the US health care system must be evaluated with the potential for this kind of adoption pattern in mind.

3.2. Interpreting the Evidence

What is the evidence on the extent of these inefficiencies? The simplest way to evaluate the presence of allocative inefficiencies would be to measure the spending required to save an additional (quality-adjusted) year of life and to evaluate whether the gain was “worth” the cost. The key idea is that spending on health care may be expensive, but it may produce gains that are more than commensurate with the additional spending. Based on evidence from Cutler and colleagues (Cutler & Madrian 1998, Cutler et al. 2006), Garber & Skinner (2008) report that the average cost per life year gained has risen substantially over the decades, from $64,000 during the 1970s to $247,000 in the 1990s (assuming that half of the gains in life expectancy are attributable to health care spending). This cost is substantially higher than the threshold of $100,000 per quality-adjusted life year often used to evaluate policy, suggesting that we may be spending too much on health relative to other goods. Furthermore, the estimated cost per quality-adjust life year is telling us about average values, but allocative efficiency must be measured on the margin—how much does it cost to save one more life year, and what are the alternative uses of those funds?

An alternative way to measure allocative efficiency would be to compare spending and outcomes in different geographic areas to see the size of the gains in outcomes that accrue to higher spending and then to compare those to productivity in other sectors. Studies taking this approach have mixed findings. Some characterize productivity as high (such as Ong et al. 2009, Bach 2010, Silber et al. 2010, Romley et al. 2011), but similar study designs have also pointed to low productivity (Skinner et al. 2006, Yasaitis et al. 2009, Rothberg et al. 2010). With these conflicting facts, it is unsurprising that some analysts believe that spending more would improve health by enough to warrant the cost, whereas others believe that it would not.

We believe, however, that these methods are flawed. In other work, we have demonstrated that much of the regional variations literature implicitly assumes that all regions

---

12As discussed in Chandra et al. (2011), the willingness to pay for a life year was initially based on a 1984 Canadian study of patients with end-stage kidney disease on dialysis. That study computed these costs at $50,000 per quality-adjusted life year (Winkelmayer et al. 2002). Almost 30 years of inflation have increased it to $100,000 (Lee et al. 2009). Garber & Phelps (1997) offer an alternative rationalization: An annual salary of $30,000 for a 40-h work week would lead to a value of a life year of $100,000 if leisure time were priced at average market wages. Other studies, however, have suggested that values as high as $300,000 per life year should be used (Murphy & Topel 2006).
are on the same production function, whereas in reality they are likely on different production functions (Chandra & Staiger 2007). This would happen if delivery systems specialize in how they deliver care. Some may specialize in technologically intensive treatments, whereas others may specialize in less expensive treatments. Because of this specialization, even though two systems or countries may have similar health outcomes (with one achieved at substantially lower cost), simply cutting spending in the high-cost system will not allow it to achieve the results of a low-cost system, but it could harm patient health in the process. Finding that two systems have similar outcomes at dissimilar spending levels is therefore not informative about whether we should spend more or less on health care: This question cannot be answered in the presence of productive inefficiency; a first step must be to understand why the two systems have different production functions and whether and at what cost the production function might be changed.

What then do we know about productive efficiency—could the same health gains be achieved for less? Chandra & Skinner (2011) argue that the peculiar financing of the health care system results in many high-tech treatments being adopted before lower-tech (but cheaper) ones are. This suggests productive inefficiencies of the type illustrated in Figure 5. Many expensive technologies are used more extensively in the United States than in other countries, such as cardiovascular procedures, at 587 per 100,000 in the United States compared with 207 in Denmark (Peterson & Burton 2007). Simultaneously, there are many cases of underuse of effective treatments in the United States compared with other nations (Cutler & Ly 2011) and for those without insurance. One recent study suggested substantial gains in self-reported health arising from an expansion of the Medicaid program in Oregon (Baicker & Finkelstein 2011, Finkelstein et al. 2011).

We interpret the geographic variations literature as providing further evidence of productive inefficiency. There is evidence that each of these regions is on its own production function, where places that spend less have comparable health outcomes to higher-spending areas because they have adopted all the right low-cost technologies (comparing point X1 to Y3 in Figure 5). Figure 6 (see color insert) shows the variability in spending on Medicare fee-for-service beneficiaries across different states, as well as the fact that higher-spending areas are not those in which beneficiaries receive the highest-quality care. That there is a negative correlation between the use of relatively low-tech treatments and expensive high-tech treatments of dubious value further suggests inefficient adoption (Baicker & Chandra 2004, Yasaitis et al. 2009). These ideas are illustrated in Figure 7 (see color insert), which shows the relationship between hospital-level spending and one-year survival for patients who were diagnosed with heart attacks, hip fracture, or colorectal cancer (Chandra et al. 2010b).13 Regardless of the overall relationship between survival and outcomes (zero in this exhibit), there is tremendous variability in the performance of hospitals around the average relationship (shown by the horizontal line in the figure), highlighting the potential for improving the productive efficiency of the delivery system. Once we move closer to productive efficiency, issues of allocative efficiency can be addressed more effectively.

---

13For each cohort of beneficiaries, we calculated risk-adjusted mortality and costs. The risk adjustment performed used ICD-9 diagnosis codes available on the Medicare Part A claims record. These measures were filtered to adjust for the effect of sampling variability, which introduces noise in the estimates of hospital-specific measures of mortality and costs (a problem that is larger in smaller hospitals). We combined these measures into a single quality dimension and single cost dimension for the 3,804 hospitals in our sample. All spending numbers are reported in 2005 dollars and include both hospital spending and physician spending.
4. POLICY LEVERS FOR IMPROVING THE EFFICIENCY OF HEALTH CARE SPENDING

The distinction between productive and allocative efficiency can help guide analysis of policy levers. Different policy levers operate on different sources of inefficiency, and allocative efficiency cannot be evaluated when productive inefficiency persists. We separate our discussion of policy levers into those that operate on the patient side, those on the provider side, and the environment in which care and insurance are purchased.

4.1. Provider-Side Incentives

Reforming the way public insurance programs pay providers could promote both productive and allocative efficiency. Medicare-administered prices play a particularly large role in creating fragmented patterns of care, and moving away from fee-for-service payments could provide much better incentives for efficient allocation of resources. Current payment structures provide little incentive for cost-saving innovation (Cutler 2010) or for providing lower-cost care. For example, a back-pain clinic loses revenues by steering patients to low-cost rehabilitation rather than equally effective, but costlier diagnostic tests and back surgery (Fuhrmans 2007). A change in payment incentives away from traditional fee for service thus has the potential to improve health and moderate spending growth. Similarly, system efficiency would be improved by creating incentives not only to adopt cost-effective treatments, but also to adopt the most cost-effective treatments first and to apply them to the patients for whom they produce the greatest value first. The imperfect incentives created by current payments reduce productive efficiency by making relatively high payments for care that is not necessarily of the highest value, and they interfere with allocative efficiency by committing (uncapped) resources to care that may have lower value than alternative uses of funds.

Rewarding more efficient delivery is a logical strategy for increasing the productivity of health care spending.14 Medicare has in fact experimented with alternative payment systems. Hospitalizations are now reimbursed by bundled payments based on patients’ diagnoses, rather than on the volume of services delivered. These prospective payments are imperfect but can limit the incentives to provide excess hospital care. Perhaps as a result, since this system was implemented in 1992, there has been limited per-capita real growth in Medicare spending on hospitalizations reimbursed through this system. Weisbrod (1991, p. 537) describes the potential dynamics between reimbursement systems and R&D in the new prospective payment system:

> With a hospital’s revenue being exogenous for a given patient once admitted . . . , the organization’s financial health depends on its ability to control costs of treatment. Thus, under a prospective payment finance mechanism, the health care delivery system sends a vastly different signal to R & D sector, with priorities the reverse of those under retrospective payment. The new signal is as follows: Develop new technologies that reduce costs, provided that quality does not suffer “too much.”

---

14Historically, the United States has not relied on quantity restrictions in health care planning, except for sometimes ineffective certificate-of-need programs (Ho et al. 2009). In theory, regional restrictions to discourage duplication and overbuilding of surgical units, MRI machines, and hospital or intensive-care-unit beds could control unrestricted growth in procedures of questionable value. But this approach to market-level quantity restrictions would require a tectonic shift in the US regulatory and policy environments.
It seems Weisbrod was too optimistic about this system: It created some payment categories defined in part based on procedures (as opposed to illnesses), which unraveled the “prospective” part of the payment and left the incentives for innovation largely unchanged. There are incentives to unbundle outpatient and postacute care, and exempting certain capital costs from the expenditure cap continues to encourage capital spending. Weisbrod’s fundamental insight is still valid: Only when health care is reimbursed based on value will innovations be focused on cost-effective treatments. The greatest saving could arise not so much from new cost-saving devices, but instead from reducing the organizational fragmentation inherent in US health care (Cebul et al. 2007). The large variation in risk- and price-adjusted costs in US academic medical centers (Fisher et al. 2004) is not because some academic centers have access to new technology and others do not. Instead, the variations are explained by how care is provided: the frequency of follow-up visits, referrals to subspecialists, days in the hospital, and the intensity of diagnostic testing and imaging procedures (Chandra & Skinner 2011).

To confront the challenge of inefficient delivery systems, some have embraced the idea of an integrated delivery system with capitated payments. Because such systems retain the savings from better prevention, lower readmission rates, and better medication adherence, they have better incentives to avoid therapies of questionable benefit. The 2010 US ACA focused on accountable care organizations (ACOs), in which shared-saving “bonuses” are provided to health care organizations that are able to provide high-quality care at lower costs. Examples include integrated systems such as Intermountain in Utah and the Geisinger Clinic in Pennsylvania, but also traditional hospital-physician networks (Fisher et al. 2009). These may come closer to the ideal expressed by Weisbrod by providing incentives for cost-saving innovations and gut the incentives that physicians currently have to engage in “financial entrepreneurship,” for example, as observed in the McAllen, Texas, region, where Medicare spending grew from roughly equal to the national average in 1992 to nearly double the national average in 2007 (Gawande 2009).

However, we do not know how well ACOs will sidestep cost-ineffective technologies, particularly if new innovations with limited proven effectiveness, such as robotic surgery, are critical to improving market share. The viability of ACOs will depend on the receptiveness of physicians to capitated payments—some specialists will see their incomes fall and are unlikely to take these cuts quietly. Although their concerns may not resonate with patients, they might if providers claim that their income is falling because valuable care is being withheld. Designers of ACOs are therefore keenly interested in measuring ACO performance and patient satisfaction, but most current quality measures seek to capture only substandard care.

If payment-system reform is particularly promising on the public side, what about the private side? As noted above, private insurance coverage is heavily influenced by the norms driven by Medicare coverage, so these reforms are likely to have system-wide spillover effects. It is natural to ask why private providers have not adopted ACOs or more bundled payments on their own. This remains a puzzle. One explanation is that it is a coordination problem—all insurers may want to adopt larger bundled payments, but no single insurer can make the transition. This is certainly consistent with the historical record on the adoption of prospective payment for hospital care. Once it was introduced in Medicare, private plans were quick to adopt it. Similarly, private hospitals were quick to use the federal government’s efforts to measure quality of care even though nothing stopped them from forming consortiums to measure quality before these federal efforts.
What might be the savings from these reforms? Some estimates would put the number at more than 20% of spending (Fisher et al. 2003a,b; Skinner et al. 2005) but that assumes that low-cost regions are at the frontier. However, even low-cost regions may not be producing efficiently. Medicare is expressly prohibited from selectively contracting with more efficient physicians. Home health care, outpatient visits, office visits, and diagnostic testing (areas of care in which reimbursements are based on volume and the “right rate” is not known) have exhibited very rapid growth in all areas, but at least for home health care, the additional spending appears to have no impact on health outcomes (McKnight 2004). All this suggests that we may save far more if the right incentives are in place; Buntin & Cutler (2009) place the figure at more around 50%. It is worth reiterating, however, that some of the savings from lower quantities may be offset with higher prices to the extent that ACOs are able to exert market power.

System reform such as ACOs may thus improve productive efficiency by reorganizing the delivery system. Capitated reimbursement can affect not just the level, but also the growth of spending by slowing the creation and adoption of new medical technologies. But achieving allocative efficiency also requires carefully calibrated ACO payments and updates over time. Additional tools on the patient side can help promote allocative efficiency.

### 4.2. Patient-Side Incentives and Insurance Design

Although insurance provides highly valuable protection against financial risk, it comes with traditional moral hazard: Patients consume more health care when the cost is lower (Newhouse & Insur. Exp. Group 1993). First-dollar private insurance is promoted by the tax code, as described above, and Medicare’s inconsistent cost-sharing is undermined by the pervasive use of wrap-around supplemental insurance. This promotes greater health care consumption than allocatively efficient and also interacts with provider-side incentives to undermine productive inefficiency by dampening the market discipline that price-sensitive consumers would provide.

Several policy proposals aim to increase the incentives for patients to consume only care that is sufficiently valuable to them, a key condition for allocative efficiency: It is necessary (but far from sufficient) that patients face the right price for the health care that they consume, balancing the increased health care use from moral hazard with the financial protection that insurance provides. The tax exclusion for employer-sponsored health insurance pushes enrollees toward policies with less cost-sharing than optimal. There is widespread support among economists for limiting this tax exclusion, even beyond the limited reform in the ACA (Gruber 2010). Other proposals include expanding the tax preference for high-deductible policies.

How might a demand-side approach affect the adoption or diffusion of new technology? There is some anecdotal evidence that higher out-of-pocket payments would dissuade high-income patients from consuming cost-ineffective treatments, suggesting that cost-sharing could discourage some use of expensive and only marginally effective technologies.15 More

---

15 Some have pointed out the difficulty of judging “true” demand for health care, particularly among those near the end of life (Becker et al. 2007). But at least one former corporate finance officer whose insurance covered the $93,000 prostate cancer drug Provenge was explicit about his willingness to pay: “I would not spend that money; because the benefit doesn’t seem worth it” (quoted in Marchione 2010). Thus high demand for end-of-life care may reflect moral hazard rather than seemingly irrational demand.
sophisticated approaches aim to impose different cost-sharing for different patients and different procedures, based on the health value of the care. In most conventional models, cost-sharing should be based solely on the price elasticity of the demand for health care (which governs the degree of moral hazard). For rational patients with full information who are aware of marginal costs and benefits, this produces efficient use of care, but mounting evidence from the behavioral economics literature suggests that patients may not always make optimal decisions about the use of services when faced with higher copayments and coinsurance (Pauly & Blavin 2008, Baicker et al. 2011).16

“Value-based” insurance plans aim to increase efficiency by imposing higher patient prices for lower-value care. Prices (or the degree of patient cost-sharing) could be set based on evidence from clinical trials and could also vary based on patient characteristics. Although this approach has been applied to lower prices of high-value care with some success (e.g., Chernew et al. 2008), insurance plans that impose significant copayments for lower-value care are rare, perhaps because of the difficulty in determining the idiosyncratic benefit that a patient may attach to these services and because of regulatory or legal barriers. Patient-side approaches thus have the potential to limit low-value spending but require a more nuanced approach than seen thus far (Baicker & Goldman 2011).

The design of coordinated insurance plans such as ACOs may help patients choose higher-value coverage and move toward allocative efficiency. ACOs could, for example, aim to attract patients with greater coverage of marginal medical technologies, and higher-quality and broader provider networks, with the patients responsible for the marginal cost of more generous policies (in which the threshold is based on the value of alternative uses of funds). This is much broader than using cost-sharing to affect the choice of generic drugs versus formulary drugs, or a CT scan in a suburban shopping mall versus one in a hospital outpatient setting. Insights from the behavioral economics literature suggest that patients may be stymied by the cognitive demands of making many price comparisons, but choice may be facilitated by creating larger bundles among which patients can choose.

4.3. Health Care and Health Insurance Markets

There are several features of the way health care and health insurance are purchased that affect both providers and patients—and indeed the types of care and insurance products that are available—with implications for the efficiency of the use of health resources. One popular explanation for inefficiency offered by physicians and provider groups is that the medical malpractice environment provides massive disincentive for efficient use by fostering “defensive medicine,” with providers delivering more care than is socially optimal to avoid potential lawsuits. Such behaviors seem more likely in uncompetitive markets—otherwise if a given doctor provided more care without associated benefits, others could offer higher-value care that was more attractive to patients and insurers. The evidence on the magnitude of defensive medicine, however, suggests that it is not a primary driver of higher health care spending, accounting for less than 3% of health care spending in the United States (Mello et al. 2010). Nor is there much evidence that areas with the greatest malpractice pressure provide the most care (Baicker et al. 2007). Although malpractice

16Most evidence suggests that health is not adversely affected by greater patient cost-sharing on average, but there are important subpopulations, for example, patients with a chronic disease, for which there are health offsets (Newhouse et al. 1993, Chandra et al. 2010a).
reform is needed to ensure that the victims of negligence are compensated and that compensation is not paid to those who are not victims, more stringent damage caps or other reforms are not likely to result in first- or even second-order savings.

A more systemic issue is that provider and insurer market power can result in higher-than-competitive health insurance prices and premiums (Dafny 2008, Dafny et al. 2009). Market power does not affect productive efficiency per se, but it could certainly affect allocative efficiency by increasing prices and reducing quantities relative to a competitive market. Enthusiasm for the potential of ACOs to promote coordinated care must be moderated by their potential to increase market power, making antitrust enforcement even more important.

Perhaps the most important contribution that public policy could make to system-wide efficiency would be to generate more information (for both patients and providers) about what care is in fact high value. There is presently little evidence on the comparative effectiveness for a vast array of treatments: For example, we do not know whether proton-beam therapy offers any advantage over conventional approaches. Moreover, most drug studies compare new drugs to placebos, rather than comparing them with other drugs on the market, failing to generate evidence on which drug works best. One area in which cost-effectiveness analysis may prove to be particularly powerful is in evaluating the relative efficiency of different delivery systems and institutional organization (rather than the relative effectiveness of particular drugs or procedures), such as drop-in clinics rather than emergency room care. Tiered networks, with greater cost-sharing for wider or less efficient networks, provide a way to blend demand-side influences with supply-side reforms.

Because such information is a public good, it is underprovided by the market, which suggests a role for government subsidization of trials. The trials required to ascertain effectiveness will be expensive. At present, the NIH budget is just over $30 billion a year ($100 per person); we are devoting less than 1% of health care resources to learning what works. Improving our knowledge of what works is a prerequisite for increasing productive efficiency and eliminating medical practices that are unsafe at any price or dominated by other treatments.

The challenge with effectiveness studies is not just how to undertake them, but also how to use the results to inform reimbursement or patient cost-sharing. The Patient-Centered Outcomes Research Institute created by the ACA cannot consider costs, as Congress prohibited its use of “a dollars-per-quality adjusted life year [or similar measure] as a threshold to establish what type of health care is cost effective or recommended” (see Chandra et al. 2011 and Garber & Sox 2010 for a discussion). As long as this continues and Medicare is prohibited from using cost-effectiveness analysis, simply providing more information on what works is unlikely to affect coverage and reimbursement decisions by private insurers.

Effectiveness studies could still be used to design patient cost-sharing, but even with better demand-side incentives, government agencies must decide whether (and at what price) Medicaid and Medicare will cover new treatments, or how to adapt coverage and cost-sharing rules as the implicit value of additional health benefits evolves over time. It is not clear how decisions about the size and growth of these large tax-financed government programs will be made. Tying their global budgets to GDP growth would restrain expenditures but might not result in the optimal share of resources being devoted to public insurance versus other responsibilities of government, such as education and defense versus private goods.
5. CONCLUSION

Technological advancement in medicine will continue to produce innovations that offer great benefit to some patients but that can easily be overused in others. It is important that health care systems be designed to foster that innovation and promote its use in patients for whom high health benefits will accrue without incurring massive government debt to cover its use in patients with little or no benefit. “Solving” the fiscal imbalances in current public programs such as Medicare by raising taxes or crudely cutting reimbursement rates will not in and of itself improve the efficiency of the care that is delivered. The need for fundamental reforms is heightened by the demographic pressures of the aging Baby Boom generation but arises more fundamentally from the increasing cost of health benefits per person.

There is no single strategy that is likely to achieve efficient use of health resources, and although the ACA expands health insurance coverage to a group with high marginal benefit of care, it is not clear how much it will improve efficiency. On the provider side, payments through public insurance programs can be bundled to encourage coordination, and providers can share in the financial gains of improving the efficiency with which they deliver care. The cuts to payment for Medicare Advantage (the managed care component of Medicare) in the ACA do not focus on reducing low-value care, but the initiation of ACOs is a step toward better coordination. On the patient side, more nuanced cost-sharing and leveling the playing field for higher-cost-sharing insurance plans can encourage patient involvement in decision making and the balancing of resources costs against health benefits, as well as foster competition. The ACA’s restriction of copayments for preventive care is in the spirit of value-based insurance design, in which different treatments have different levels of cost-sharing. The ACA’s “Cadillac tax” on expensive health plans is a small step toward reforming the inefficient tax treatment of first-dollar employment-based health insurance.

On the system-wide front, better information is needed about which delivery systems—not just which drugs or procedures—are most effective. The ACA creates an innovation center with the authority to “test innovative payment and service delivery models” and finances a number of demonstration projects to improve the malpractice system. These initiatives may uncover promising strategies to reduce productive inefficiency, but their benefit remains unproven at this time.

The United States has yet to wrestle with the question of public policy priorities in a world of scarce resources: Even with perfect productive efficiency, we cannot cover all services for all people, and the ACA fundamentally does not address this trade-off. Care in public programs remains covered largely without regard to effectiveness—and indeed, there is a prohibition against making coverage decisions based on cost-effectiveness—and indeed, there is a prohibition against making coverage decisions based on cost-effectiveness. When public resources come at a cost of lower economic growth, there must be some explicit consideration of the value of redistribution, and the public priority placed on covering different levels of service for different parts of the population. By first ensuring that health care resources are used more productively, we will be in a much better position to move toward spending the “right” amount on health.

DISCLOSURE STATEMENT

K.B. is a director of Eli Lilly, a Medicare Payment Advisory Commissioner, and a member of the Congressional Budget Office Panel of Health Advisers. J.S.S. recently served on the
Congressional Budget Office Panel of Health Advisers. K.B. and J.S.S. have received grant funding from the National Institute on Aging and the Robert Wood Johnson Foundation.

ACKNOWLEDGMENTS

An earlier version of this review was prepared for and presented at the Federal Reserve Bank of Kansas City’s 2011 Economic Policy Symposium in Jackson Hole, Wyoming. This review draws heavily on Baicker & Chandra (2011) and other papers by the authors (while the title borrows from Birkmeyer et al. 2002). The authors thank Jacob Robbins for expert research assistance and gratefully acknowledge financial support from the National Institute on Aging (P01AG19783 and P01AG032952).

LITERATURE CITED

Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance
Trust Funds. Rep., Washington, DC
Buntin M, Cutler D. 2009. The two trillion dollar solution: saving money by modernizing the health
care system. Rep., Cent. Am. Prog., Washington, DC
US health care system. J. Econ. Perspect. 22(4):93–113
Cent. Medicare Medicaid Serv. 2010. National Health Expenditures Accounts: definitions, sources,
Chandra A, Gruber J, McKnight R. 2010a. Patient cost-sharing and hospitalization offsets in the
J. Econ. Perspect. 25(2):27–46
Work. Pap. 16953
Chandra A, Staiger DO. 2007. Productivity spillovers in health care: evidence from treatments of
heart attacks. J. Polit. Econ. 115:103–40
Chandra A, Staiger DO, Skinner JS. 2010b. Saving money and lives. In The Healthcare Imperative:
Chernew ME, Baicker K, Hsu J. 2010. The specter of financial armageddon: health care and federal
on medication adherence within a disease management program. Health Aff: 27:103–12
Washington, DC
New York: Oxford Univ. Press
20(3):97–120
J. Econ. Perspect. 25(2):3–25
Cutler D, Madrian B. 1998. Labor market responses to rising health insurance costs: evidence on
hours worked. Rand J. Econ. 29:509–30
Dafny L, Duggan M, Ramanarayanan S. 2009. Paying a premium on your premium? Consolidation in
the U.S. health insurance industry. NBER Work. Pap. 15434
DeNavas-Walt C, Proctor BD, Smith JC. 2010. Income, poverty, and health insurance coverage in the

Baicker • Chandra • Skinner


Gawande A. 2009. The cost conundrum: what a Texas town can teach us about health care. *New Yorker*, June 1


Figure 2

Figure 3

Figure 5

Productive inefficiency in health care. Comparing outcomes and spending across Hospitals X and Y, as in the line connecting X₂ and Y₃, would suggest a negative correlation between spending and outcomes (even though all technologies shown here do no harm). See the text for further discussion.

Figure 6

Medicare spending and the quality of care based on state-level use of care such as aspirin post-heart attack, blood-level monitoring for diabetics, and flu vaccination for Medicare beneficiaries. Figure copyrighted and published by Project HOPE/Health Affairs in Baicker & Chandra (2004). The published article is archived and available online at http://www.healthaffairs.org.
Figure 7
Association between one-year survival and spending at the hospital level for patients with heart attacks, hip fractures, and colorectal cancer—all conditions with limited discretion in diagnosis. We combined these measures into a single quality dimension and a single cost dimension for the 3,804 hospitals in our sample. All spending numbers are reported in 2005 dollars and include both hospital spending and physician spending. Figure reprinted with permission from Chandra et al. (2010b) by the National Academy of Sciences, courtesy of the National Academies Press, Washington, DC.
Contents

Paul Samuelson’s Legacy
   Avinash Dixit ............................................. 1

Saving Money or Just Saving Lives? Improving the Productivity of US Health Care Spending
   Katherine Baicker, Amitabh Chandra, and Jonathan S. Skinner .............. 33

International Comparisons in Health Economics: Evidence from Aging Studies
   James Banks and James P. Smith ........................................ 57

Rare Macroeconomic Disasters
   Robert J. Barro and José F. Ursúa ......................................... 83

Endogenous Extreme Events and the Dual Role of Prices
   Jon Danielsson, Hyun Song Shin, and Jean-Pierre Zigrand .................... 111

The Distribution of Teacher Quality and Implications for Policy
   Eric A. Hanushek and Steven G. Rivkin ...................................... 131

Economic Modeling and Analysis of Educational Vouchers
   Dennis Epple and Richard Romano ............................................ 159

Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers
   Joseph G. Altonji, Erica Blom, and Costas Meghir ............................. 185

Credit Constraints in Education
   Lance Lochner and Alexander Monge-Naranjo .................................. 225

New Perspectives on Statistical Decisions Under Ambiguity
   Jörg Stoye .............................................................................. 257

The Empirics of Firm Heterogeneity and International Trade
   Andrew B. Bernard, J. Bradford Jensen, Stephen J. Redding,
   and Peter K. Schott ............................................................. 283

Natural Resource Wealth: The Challenge of Managing a Windfall
   Frederick van der Ploeg and Anthony J. Venables ............................... 315
The Economics and Politics of Women’s Rights
Matthias Doepke, Michèle Tertilt, and Alessandra Voena ............ 339

Recent Developments in the Economics of Time Use
Mark Aguiar, Erik Hurst, and Loukas Karabarbounis ............ 373

Life-Cycle Wage Growth and Heterogeneous Human Capital
Carl Sanders and Christopher Taber ................................. 399

Behavioral Economics and Psychology of Incentives
Emir Kamenica ................................................................. 427

The Relationship Between Economic Preferences and Psychological
Personality Measures
Anke Becker, Thomas Deckers, Thomas Dohmen, Armin Falk,
and Fabian Kosse ............................................................. 453

Corruption in Developing Countries
Benjamin A. Olken and Rohini Pande ....................................... 479

A Reduced-Form Approach to Behavioral Public Finance
Sendhil Mullainathan, Joshua Schwartzstein, and William J. Congdon ... 511

Recent Research on the Economics of Patents
Bronwyn H. Hall and Dietmar Harhoff ........................................ 541

Probability and Risk: Foundations and Economic Implications of
Probability-Dependent Risk Preferences
Helga Fehr-Duda and Thomas Epper .......................................... 567

The Theory of Clubs and Competitive Coalitions
Myrna Wooders ................................................................. 595

The Promises and Pitfalls of Genoeconomics
Daniel J. Benjamin, David Cesarini, Christopher F. Chabris,
Edward L. Glaeser, David I. Laibson, Vilmundur Guðnason,
Tamara B. Harris, Lenore J. Launer, Shaun Purcell, Albert Vernon Smith,
Magnus Johannesson, Patrik K.E. Magnusson, Jonathan P. Beauchamp,
Nicholas A. Christakis, Craig S. Atwood, Benjamin Hebert, Jeremy Freese,
Robert M. Hauser, Taissa S. Hauser, Alexander Grankvist,
Christina M. Hultman, and Paul Lichtenstein ................................. 627

Indexes

Cumulative Index of Contributing Authors, Volumes 1–4 ............ 663
Cumulative Index of Chapter Titles, Volumes 1–4 .......................... 665

Errata

An online log of corrections to Annual Review of Economics
articles may be found at http://econ.annualreviews.org