This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Analyses in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90286-2

Volume URL: http://www.nber.org/books/wise05-1

Publication Date: August 2005

Title: The Efficiency of Medicare

Author: Jonathan S. Skinner, Elliott S. Fisher, John Wennberg

URL: http://www.nber.org/chapters/c10359

4 The Efficiency of Medicare

Jonathan Skinner, Elliott S. Fisher, and John E. Wennberg

4.1 Introduction

The United States spends more on health care in per capita terms and as a percentage of gross domestic product (GDP) than any other developed country (Reinhardt, Hussey, and Anderson 2004). This can be interpreted in two ways. One is that the elevated spending is symptomatic of failure in the health care system. Money is wasted through administrative overhead, the overuse of fully insured health care, or the provision of expensive tertiary but only marginally useful technology.[1] A different view is that U.S. citizens demand, and get, a higher quality level of health care than anywhere else in the world. It may be expensive, but the technological advances provided in the United States have led to dramatic improvements in functioning and life expectancy.[2]

Jonathan Skinner is the John French Professor of Economics and professor of community and family medicine at Dartmouth College, and a research associate of the National Bureau of Economic Research (NBER). Elliott S. Fisher is a professor of medicine and of community and family medicine at Dartmouth Medical School. John E. Wennberg is the Peggy Y. Thomson Professor for the Evaluative Clinical Sciences and a professor of community and family medicine (epidemiology) and of medicine at Dartmouth Medical School.

Support by the Robert Wood Johnson Foundation and the National Institute on Aging (PO1 AGI9783) is gratefully acknowledged. The views expressed here are those of the authors alone and not of the Department of Veterans' Affairs. We are grateful for very helpful comments from Esther Duflo, Alan Garber, David Laibson, Douglas Staiger, Duncan Thomas, Frank Vella, and seminar participants at the NBER Conference on Aging, Boulder, Arizona; at the University of California, Los Angeles; Rand; Georgia State University; and the Universities of Virginia, Maryland, and North Carolina. Dan Gottlieb provided excellent data analysis.

1. For a good presentation of this view, see Evans and Stoddart (1994).

2. For a general exposition of the view that current high spending levels for medical technology will yield benefits that could even lower costs in the future, see Pardes and others (1999). For specific measures of improvements in outcomes following the use of more intensive technology, see Cutler and others (1998) and Cutler and Meara (1999).

Knowing which story holds true is crucial for any kind of health care reform, and particularly for Medicare reform. Unfortunately, the answer is elusive. While the evidence strongly suggests substantial technological gains in the treatment of specific diseases such as heart attacks or specific groups such as low-birth-weight infants (e.g., Cutler et al. 1998; Cutler and Meara 1999), it is not clear how well these specific paradigms generalize to the entire health care system where medical progress has not been nearly so robust. More importantly, the secular improvements in mortality are "average" effects of technology rather than the marginal impact of greater health care intensity on health outcomes (see Cutler 2000).

Researchers have attempted to exploit "natural randomization" in outcomes data to estimate the marginal effectiveness of specific medical technologies on outcomes such as mortality for people with heart attacks (McClellan, McNeil, and Newhouse 1994) or for infants (Currie and Gruber 1996). For example, in McClellan, McNeil, and Newhouse the "treatment" group was people experiencing heart attacks who lived relatively near a hospital equipped with diagnostic laboratory facilities that helped physicians decide whether to proceed with surgery. The "control" group was those living relatively far away. They found minimal benefits among the population treated most intensively for heart attacks. It is not known how well this finding for heart attack patients, where the quality of the scientific evidence is high, generalizes to daily decisions about treatment for chronically ill patients, where the scientific evidence is quite modest.

In this paper, we use the idea of natural randomization to evaluate the efficiency of the Medicare program more generally. The macro-level equivalent of living near a hospital with advanced diagnostic laboratories for heart attack patients is whether the health care *system* provides a higher-than-average intensity of health care, ranging from flu shots in the outpatient setting, to hospitalization instead of outpatient care, surgery instead of watchful waiting, and three-month waiting periods for physician appointments instead of six-month waiting periods for appointments.

Simply comparing outcomes between regions with higher-than-average and lower-than-average Medicare expenditures risks the reverse causality problem; the sickest regions tend to experience more spending on health care.[3] A deeper issue is the difficulty in comparing multidimensional health care systems across regions. One region may spend more to provide "effective" care of proven clinical value (such as flu shots, mammograms among women over age fifty, and eye examinations for diabetics), another region

3. In an earlier study, Hadley (1988) used as instruments for Medicare expenditures factors such as nursing home residence rates; he found a positive impact of Medicare spending on survival. However, nursing home residence rates are probably correlated with unmeasured regional health status.

may spend more because it provides surgery where nonsurgical alternatives often exist (such as back surgery or knee replacements), while physicians in a third region could be more likely to admit chronically ill patients to a hospital or refer them to a specialist.

In the empirical section, we focus on several "markers" for the different dimensions of intensity. With regard to effective care, we need not instrument for reverse causality, since there is no reason why annual eye examinations for diabetics (for example) should be lower or higher in regions with greater incidence of disease. In every region, the "right rate" among the relevant population is nearly 100 percent, and so we can use this marker for quality without adjusting for potential reverse causality. However, finding a marker for intensity of care among chronically ill patients is more difficult, since patients in regions with greater prevalence of disease might be expected to account for more Medicare spending. We use two types of instruments that we believe on clinical grounds should not be biased by differences across regions in underlying health status. The first focuses on utilization (physician visits and intensive care unit [ICU] days) in the last six months of life. By definition, these patients are similar with regard to health status across regions for one important dimension of health: their six-month survival rate. Obviously, one cannot use treatment patterns in the last six months to make inferences about Medicare efficiency in this group, since they are all dead by the end of the period. Instead, the intensity of care in the last six months is used as an instrument to measure whether Medicare enrollees who are alive at the beginning of the year gain survival benefits from higher levels of Medicare spending for the type of care typically provided to those with chronic illnesses.

The second instrument takes a prospective approach for a disease with "high-tech" treatments: regional price-adjusted averages of one-year Medicare part A (hospital) expenditures for a cohort of patients following acute myocardial infarction (AMI), or heart attacks. Again, expenditures for AMI are likely to be good measures of the intensity of health care in the region, although in this case the dimension of intensity may differ since regions with more aggressive surgical treatment of AMI will tend to experience higher costs for the treatment of AMI patients. Like end-of-life Medicare enrollees, patients who have experienced an AMI tend to be similar across regions with regard to their initial health status.

Using data from the *Dartmouth Atlas of Health Care* as well as supplemental information from the U.S. Census and the Centers for Disease Control, we estimate the incremental effects of these different types of spending, both in a linear model and, where appropriate, in a nonlinear framework (Newey, Powell, and Vella 1999). We find that higher measures of "efficient" care are associated with better survival for the general population but are not associated with higher costs. However, we also find that higher levels of spending associated with more intensive end-of-life physician visits, and with more intensive treatment of AMI patients, are not associated with greater overall rates of survival in the Medicare population; in other words, we estimate a "flat of the curve" segment in which regions spend more but gain no benefit in higher survival.

These results are consistent with two recent studies focusing on outcomes for cohorts of AMI, colon cancer, and hip fracture patients (Fisher et al. 2003a,b), rather than the entire Medicare population, as we do here. They found evidence of worse outcomes—that is, higher mortality rates—associated with higher levels of health care intensity. In short, our estimates imply that physicians and hospitals participating in the Medicare program provide too little in the way of inexpensive and effective care, while at the same time spending $26 billion annually, or nearly 20 percent of the Medicare program's budget for health care that appears to be of questionable value with regard to survival benefits.

4.2 The Nature of the Problem: Per Capita Medicare Expenditures in the United States

We first consider the magnitude and extent of regional differences in Medicare expenditures in the United States. Primary data sources are samples of either 20 percent (outpatient) or 100 percent (inpatient) of the Medicare claims data during 1995 and 1996. The basic unit of analysis is the hospital referral region (HRR), of which there are 306 in the United States. The HRR was constructed in the *Dartmouth Atlas of Health Care* as a unit of analysis that reflected the actual hospital migration patterns of Medicare patients for tertiary care. An HRR must include at least one hospital that performs cardiac surgery and neurosurgery. Each zip code in the United States is assigned to an HRR depending on what hospital the majority (or, in some cases, the plurality) of Medicare enrollees seek their hospital care; see Wennberg and Cooper (1999) for details. Thus, the HRR may cross county or state boundaries or, in some cases, follow interstate highways.

The important thing to note about HRRs for this study is that all rates are based upon the zip code of residence and not where the person actually sought care. Thus, if an individual lives in Lebanon, New Hampshire, and is admitted to a hospital in Boston, all utilization is assigned to the Lebanon HRR. Most care is delivered locally; 80 percent of the U.S. population lives in HRRs in which over 85 percent of care is delivered by providers within the HRR. In the analysis that follows, utilization rates have been adjusted for differences across HRRs in the age, sex, and racial composition of the population, and (where necessary) differences in the price level. We restrict our attention to the fee-for-service Medicare population

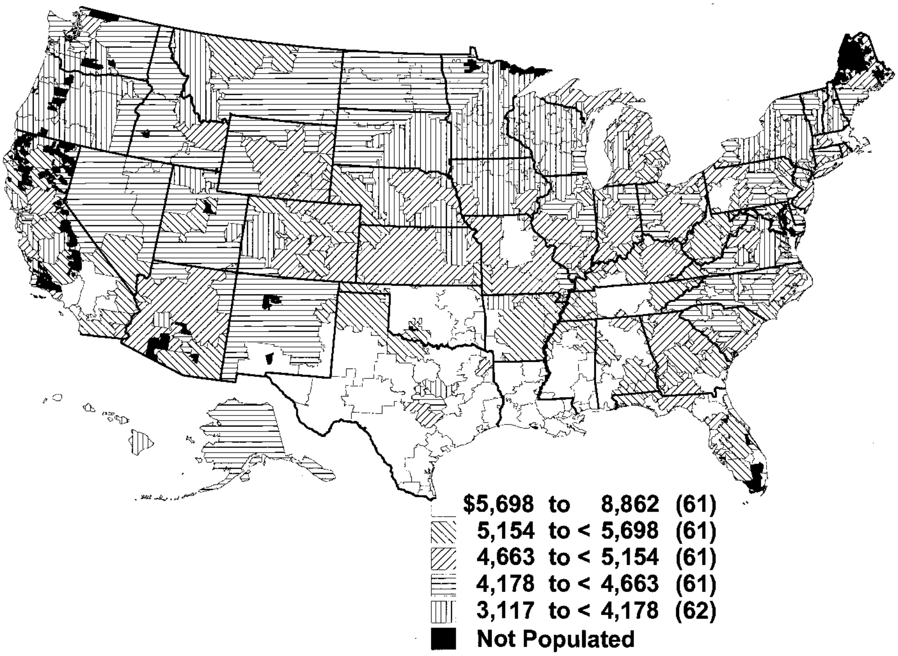


Fig. 4.1 Noncapitated Medicare expenditures per enrollee, 1996

that during the study period accounted for more than 85 percent of the total Medicare population.[4]

Figure 4.1 uses these data from Wennberg and Cooper (1999) to construct a map showing the distribution of per capita Medicare expenditures across the United States in 1996; these are adjusted for differences across regions in age, sex, and race (black and nonblack). There are clearly wide variations in the extent of spending, with per capita expenditures ranging from $3,341 in Minneapolis, for example, to $8,414 in Miami. There are clusters of high-expenditure regions largely concentrated in Florida, the deep South, and urban areas on the East and West coasts. There are exceptions as well: inexpensive regions in Florida and low-cost cities on the West Coast (e.g., Portland, Oregon) and the East Coast (Richmond, Virginia).

This figure raises some immediate policy issues. If the higher expendi-

4. One concern with using the fee-for-service population is the selection problem; healthier patients tend to enroll in health maintenance organizations (HMOs). We control for this in part by measuring health status (discussed presently) for the same fee-for-service population. Thus, if a healthy region experiences a high rate of HMO enrollment, leaving an unhealthy fee-for-service population, this will be reflected in both health status measures and in per capita spending. In practice, either including the percentage of the HMO population in the HRR, or excluding regions with more than 10 percent HMO enrollment, has little impact on the results.

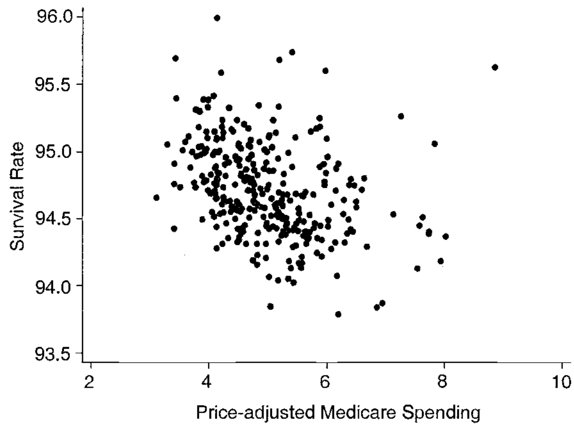


Fig. 4.2 Per capita Medicare expenditures and mortality rates in the Medicare population, 1996

Note: Medicare expenditures and mortality rates are age, sex, and race adjusted; both apply to the fee-for-service population.

tures in some regions actually lead to better health, then the Medicare program may be inequitable to the extent that taxpayers in the low-expenditure regions are paying for the better health of those in the high-expenditure regions (Feenberg and Skinner 2000). Conversely, if the higher expenditures yield nothing in health benefits, then there is tremendous waste in the program; reducing spending in the high-expenditure areas can save enough money to preserve the solvency of the Medicare trust fund by a decade (Skinner and Wennberg 2000). It could also be the case that people in the high-expenditure areas simply prefer more intensive care. One might then ask why other regions should be subsidizing their preferences.

To consider the efficiency of Medicare, we provide a simple graph in figure 4.2 that shows per capita Medicare expenditures (age, sex, race, and price-adjusted) on the horizontal axis and survival, which we define as the (1 – age-sex-race-adjusted mortality) $\times$ 100, on the vertical axis. There is a clear negative correlation between expenditures and survival. This in itself is not too surprising; spending should be higher in regions with poorer levels of health, so we might expect to observe Mobile, Alabama, spending more than Grand Junction, Colorado. In the next section, we consider a simple model that formalizes how to evaluate the efficiency of Medicare given that Medicare spending is likely to be higher in regions with poor health.

4.3 Medicare Expenditures and Outcomes: Theoretical and Measurement Issues

In this section, we first develop a theoretical model of Medicare expenditures and outcomes and show that the observed negative correlation between Medicare expenditures and survival in figure 4.2 is at least consistent with the Medicare program meeting stringent efficiency conditions. We then develop criteria for determining what one *should* expect in terms of better survival.

4.3.1 A Theoretical Model of Regional Differences in Medicare Expenditures

We next turn to evaluating the efficiency of the Medicare program with regard to differences across regions in health care spending. It is important to emphasize at the outset that there are many inputs into the production of health care, some of which may yield substantial health benefits at little costs, and others of which may provide little in benefits but exert a larger impact on costs. In this study, it is impossible to characterize all of these inputs, but we can try to provide a rough categorization of regional differences in the intensity of three dimensions of inputs to health care.[5]

The first is "effective" care, treatments where there is strong clinical evidence as to their efficacy and where, among the relevant population of patients, nearly 100 percent should be receiving the treatment. Examples include eye examinations for diabetics, flu shots for the elderly, mammograms for women over age fifty, and the use of beta blockers and aspirin among appropriate heart attack candidates. The second categorization is "preference-sensitive" care, or treatments where there are valid options for how to treat the specific disease, and choices often involve trade-offs between risks and benefits of treatment. For example, knee replacement surgery can provide long-term benefits in functioning, but there is a nontrivial risk of surgical complications and a lengthy period of recovery. Similarly, angioplasty for the treatment of ischemic heart disease carries with it the risk of stroke or operative death, but it can provide relief of symptoms.[6] These treatment options typically do not have a large impact on long-term survival but can potentially affect functioning and quality of life. Unfortunately, we cannot measure quality of life, only quantity in terms of survival rates. Thus, there is unlikely to be any influence on mortality of greater intensity of preference-sensitive care.

Finally, the third category is "supply-sensitive" care, where the quality of scientific evidence regarding use of care is poor and the quantity of care tends to be associated with the supply of resources such as specialists and hospital beds. Measures of the prevalence of supply-sensitive care include the frequency of physician visits and the use of ICU days among chronically ill patients. Note that we do not presume causality from supply to the

5. This categorization is developed in Wennberg, Fisher, and Skinner (2002).

6. Angioplasty is proven to be effective for reperfusion within twenty-four hours of the onset of an AMI, and for this use it is less likely considered preference sensitive. However, the vast majority of angioplasties (more than 85 percent) are performed on *non-AMI* patients.

use of supply-sensitive care, simply a correlation between factors of supply and the quantity of supply-sensitive care.

There are separate efficiency conditions for each of these inputs and in each region. Suppose that the value function of the "social planner" is written in terms of

(1) $$\mathrm{V} = \mathrm{V}[\mathbf{S}(\mathbf{R}, \mathbf{H}), \mathbf{Q}(\mathbf{R}, \mathbf{H}), \mathbf{Y}(1 - \tau) - (1 - \theta)\mathbf{M}^*, \mathbf{P}],$$

where V is a concave value function, the bold-faced **S** denotes the vector of (regional) per capita survival measures for regions $i = 1, \ldots N$, **Q** the vector measuring quality of life (conditional on survival), and **R** is the $K \times N$ vector of inputs (for example, inputs that are efficient, preference sensitive, and supply sensitive) for each of the N regions. With regard to the third argument, after-tax income (assumed to be equal to consumption), $\mathbf{M}^*$ is the level of real per capita total health care expenditures, and within each region total expenditures are a function of the intensity of inputs conditional on health status and the health of the population **H** so that $\mathbf{M}^* = H(\mathbf{R}, \mathbf{H})$.

Nonmedical consumption is given by $\mathbf{Y}(1 - \tau) - (1 - \theta)\mathbf{M}^*$: gross income Y after the Medicare tax τ has been paid and Medicare's share of (out-of-pocket) medical expenses θ has been paid. Note that total Medicare expenditures are therefore $\mathbf{M} = \theta\mathbf{M}^*$. Medicare taxes are assumed to be proportional to income.[7] Finally, the population of each region is given by **P**; this is to allow for larger regions to receive a larger weight in the social welfare function. While the Medicare program is a complex intergenerational transfer mechanism in which younger workers pay most of the taxes ultimately consumed by the elderly, we assume for analytic simplicity that the people paying the taxes in region i are the same ones experiencing the benefits in region i.[8]

Increasing Medicare spending in just one region i is assumed to result in an increase in the overall Medicare tax rate τ: $\Delta M_i \cong \Delta\tau\Sigma_j[(P_j Y_j)/P_i]$. Thus, the balanced budget change in the tax rate necessary to fund an extra (per capita) Medicare dollar spent in region i is

(2) $$\frac{d\tau}{dM_i} = \left[\sum_{j=1}^{k} \frac{P_j}{P_i} Y_j\right]^{-1},$$

and the first-order condition for Medicare expenditures across regions is, for each i,

(3) $$\frac{dV}{dR_{ki}} = \mathrm{V}_{1i}\frac{dS_i}{dR_{ki}} + \mathrm{V}_{2i}\frac{dQ_i}{dR_{ki}} - \sum_{j=1}^{k} \mathrm{V}_{3j} Y_j \frac{d\tau}{dM_i}\frac{dM_i}{dR_{ki}},$$

7. The Medicare payroll tax that funds part A, or the hospital component, is proportional to earnings. While the part B (physician) premium is regressive, the larger proportion funded by general tax revenue is progressive; overall, the tax is not far from proportional. See McClellan and Skinner (1999).

8. See, for example, Feenberg and Skinner (2000).

where V_{1i} is the contribution of an incremental increase of survival in region i to the social welfare function, V_{2i} the contribution of quality of life (conditional on survival), and V_{3i} the impact of after-tax nonmedical income on social welfare of the entire country. (Note that the constant fraction $\theta = dM_i/dM_i^*$ drops out of equation [3].)

The first-order condition can be written

$$(4) \qquad \frac{dS_i}{dR_{ki}} + \frac{V_{2i}}{V_{1i}}\frac{dQ_i}{dR_{ki}} - \left(\frac{\sum_{j=1}^{k} V_{3j} Y_j \frac{d\tau}{dM_i}\frac{dM_i}{dR_{ki}}}{V_{1i}} \right) = 0$$

The first term on the left-hand side measures the marginal value of Medicare expenditures on survival in region i, while the second term is the improvement in functioning and quality of life, expressed in terms relative to the incremental social value of increasing survival (V_2/V_1). Unfortunately, we do not have good data on Q_i, the level of functioning in a region, and so we will not be able to estimate dQ_i/dR_{ik} directly, although evidence on this question is discussed presently.

Suppose that the value function V is linear, so there is a uniform social trade-off between increasing survival by one unit and reducing after-tax income by β. Thus, β is the conventional cost-effectiveness "hurdle," or how much is society willing to spend to increase survival rates by a given amount under the assumption that the health inputs have no impact on quality of life, hence $dQ_i/dR_{ik} = 0$. In this special case, $\beta = V_3/V_1$, $\forall\ i, j$, which allows us to simplify equation (4) to

$$(5) \qquad \frac{dS_i^k}{dM_i^k} = \frac{\frac{dS_i}{dR_{ki}}}{\frac{dM_i}{dR_{ki}}} = \beta^{-1}.$$

In other words, all regions should devote expenditures up to the point where the marginal gains are equal, both across types of care, and across regions. In the case of a specific factor that raises spending, we can consider this scenario in figure 4.3, with the same dimensions as those shown in figure 4.2: expenditures (or intensity) on the horizontal axis and survival on the vertical axis. Combinations of expenditures and survival rates are shown for three regions, A, B, and C, as well as each of their concave health "production functions." The slopes of each of the straight tangential lines are equal to β so that $dS_i/dM_i = \beta$ across regions. Furthermore, this graph replicates the general pattern of spending and survival shown in the empirical data in figure 4.2. Accounting for the concavity of the value function V would imply efficiency conditions that would move region A further along its production function to A′, and would move C to C′, resulting in a reduction in health disparities across regions.

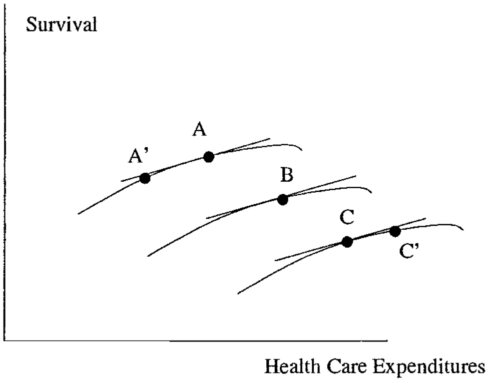


Fig. 4.3 Efficiency in health care

Note: A, B, and C represent regions. At each point, the slope of the health care productivity curve is equal (and shown by the straight line passing through points A, B, and C).

To capture the implicit function dS_i/dM_i as "traced out" by the markers (or instruments) for the intensity of care, we specify a model where survival S_i is a general nonlinear function of Medicare expenditures in that region, $S(M_i)$, that is,

$$S_i = S(M_i, \mathbf{Z}_i) + u_i, \quad (6)$$

where $\mathbf{Z}_i$ is a vector of underlying observable health characteristics. A simplified version of this equation is

$$S_i = S(M_i) + \varphi \mathbf{Z}_i + u_i, \quad (6')$$

where φ is the corresponding vector of coefficients; thus, $\varphi \mathbf{Z}_i$ shifts the productivity curves vertically with respect to observable differences in health status. To the extent that *unobservable* health status is reflected in the error term u_i, it will be correlated with Medicare spending, leading to inconsistent estimates in a single-equation model. We therefore model Medicare expenditures as a nonlinear function of the (uncontaminated) measures of intensity R_{ik} so that as regions increase R_{ik}, spending on this input (and associated treatments) also would tend to increase in a potentially nonlinear way:

$$M_i = M(R_{ik}) + \mathbf{Z}_i \prod + e_i, \quad (7)$$

where Π is a vector of coefficients and e_i the error term. This block-diagonal structure is well suited to estimation using the methods developed in Newey, Powell, and Vella (1999); we return to estimation issues in section 4.4.

4.3.2 How much should survival rates differ across regions?

The theoretical model suggested that it is important in assessing efficiency to measure the impact of the marginal dollar of Medicare spending (for each dimension) on health outcomes. As a first step, we would like to know how much difference in survival rates we would expect to see under the null hypothesis that incremental Medicare expenditures yield first-order health benefits. In the short term, we would expect to see a jump in survival. If we viewed the social β to be $100,000 per additional year of life, then spending an extra $1,000 per capita in Medicare spending should, in the short term, yield a drop in mortality rates (or increase in survival rates) of 1.0 percentage points. Over the long term, the mortality rate would climb back up as those patients saved early on ultimately die.

To quantify the change in survival and mortality rates at the level of the population that is consistent with the micro-level cost-effectiveness benchmarks, we develop a simple model using the life tables for 1991 from Wilmoth (2001). Figure 4.4 shows the benchmark survival pattern for the U.S. population in 1991 (the leftmost curve); the average mortality rate is 5.2 percent in this population, which is consistent with mortality in the Medicare population. Next, suppose an innovation is introduced that reduces annual mortality rates by 25 percent, for example, leading to the

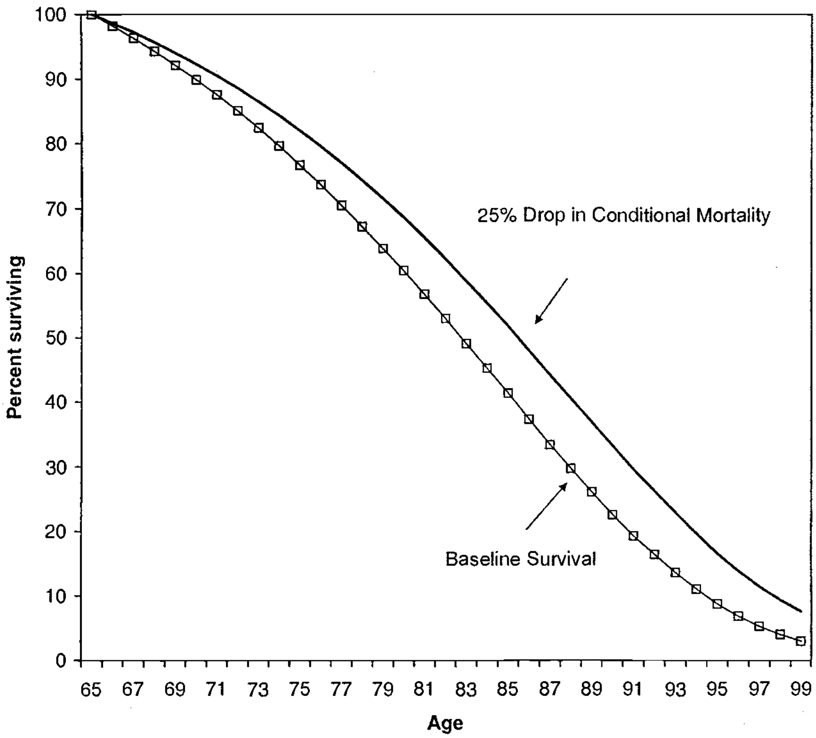


Fig. 4.4 Baseline survival curve and counterfactual survival curve after 25 percent decline in mortality rate

rightward shift in the survival curve (figure 4.4). The population-weighted decline in the mortality rates is 1.3 percentage points, down to 3.9 percent. In the steady state (assuming no population growth), the population is larger by 11 percent, the area between the two curves. In other words, the implicit numerical derivative of the percentage change in steady-state life years per 1 percentage point change in the mortality rate is equal to 11/(5.2 – 3.9) or 9.0.[9] Conversely, a "hurdle" rate of $100,000 per life year implies that every increase of $1,000 in per capita Medicare expenditures should increase survival rates by 1/9, or 0.11 percentage points.

This hurdle rate may be too low. As noted previously, short-run effects are larger than long-run effects, and to the extent that Medicare spending has risen dramatically in the last decade, we might expect to observe a larger impact on survival. Second, Medicare does not cover all expenditures for the care measured in the claims data, so we are underestimating the true cost of the medical care, which will tend to bias our results toward finding that Medicare spending at the margin is efficient. Finally, the $100,000 hurdle is often given for a year of life in perfect health, and one is rarely extending life years in perfect health; indeed, the increased survival may be at the expense of quite poor health functioning. On the other hand, if functioning is improved as well as survival, one may require less improvement in survival to justify the incremental Medicare expenditures, an issue to which we return presently.

4.4 Measuring Health Care Intensity

First, with regard to the intensity of "efficient" care, Wennberg and Cooper (1999) measured the frequency of use where appropriate for eleven treatments or screening methods that are generally agreed upon to be effective in medical care; this includes the fraction of women in the Medicare population who were screened for breast cancer (mammography), the percentage of diabetics administered blood tests and eye exams, and the percentage receiving a pneumonia vaccination. Averaged in with these measures are quality measures for the treatment of heart attacks; the percentage of appropriate patients in each HRR who received effective drug treatment such as aspirin, beta blockers, and angiotensin-converting enzyme (ACE) inhibitors, for example.[10] One would not expect 100 percent compliance for a variety of reasons, but in general the quality indexes were quite low, with a (weighted) mean of 48 percent and a range from 32 to 57 percent (see table 4.1), with a higher index indicating better quality care.

9. We chose the 25 percent mortality decline to make the differences apparent in figure 4.4. Smaller changes in mortality yield the same numerical derivatives, however.

10. These latter indicators are drawn from the Cooperative Cardiovascular Project, or CCP, a detailed study of more than 200,000 heart attacks in the United States. See Jencks, Huff, and Cuerdon (2003) for a more comprehensive measure of quality by state.

Table 4.1 Summary statistics

Dependent variable	Mean	Standard deviation	Minimum	Maximum
Medicare expenditures (in $1,000)	5.08	0.86	3.12	8.86
Heart attack (AMI) rate (per 1,000)	19.45	2.91	11.44	29.44
Stroke (CVA) rate (per 1,000)	22.95	2.84	15.24	32.47
Gastrointestinal bleeding rate (per 1,000)	15.16	1.64	10.54	20.43
Colon cancer rate (per 1,000)	4.74	0.54	2.83	6.34
Lung cancer rate (per 1,000)	1.42	0.28	0.50	2.28
Hip fracture rate (per 1,000)	15.55	1.53	9.20	19.62
Physician visits in the last 6 months (per decedent)	24.20	7.12	8.5	47.9
Decedents admitted to ICU in last 6 months (%)	31.26	5.71	14.2	49.3
Avg. AMI Medicare Part A expenditures (in $1,000)	15.63	1.69	11.66	21.92
Fraction living in poverty	13.10	5.59	4.66	32.73
Avg. Social Security income (in $1,000)	7.760	0.61	6.056	9.532
Elderly living alone (%)	35.02	3.62	20.8	41.9
Cigarette smokers (%)	23.47	2.47	13.7	30.8
Obese (%)	17.09	1.84	11.9	22.0
Seat belt users (%)	28.39	8.23	12.8	59.6
Binge drinkers (%)	13.41	3.61	6.3	23.2
High school graduates (%)	49.37	4.68	31.91	58.89
College graduates (%)	25.55	6.23	11.49	52.24
Effectiveness index	47.08	4.06	32.21	56.74
Survival rate (%)	94.70	0.34	93.78	95.99

We also include a measure for preference-sensitive care that averages age-sex-race-adjusted rates of ten surgical procedures by HRR that include back surgery, angioplasty, bypass surgery, prostatectomy, knee replacements, and hip replacements. We note, however, that because our outcomes measures do not reflect functioning and quality of life—the factors most likely to be affected by these procedures—we would *not* expect that survival should be affected strongly by the intensity of surgical rates. There may even be trade-offs where more intensive surgical rates lead to greater operative mortality, but better functioning, for those who survive. Thus, while we do not consider whether preference-sensitive care is associated with better survival, we can ask whether regions that are more aggressive with regard to surgery also experience higher Medicare costs.

We also consider instruments for supply-sensitive care where there are few guidelines for the appropriate treatment of chronically ill patients. As noted earlier, we consider two approaches to providing a marker for this type of care. The first is physician visits per decedent in the last six months of life. We focus on this group because by looking at those near death, we are at least comparing health care utilization across a group of largely very

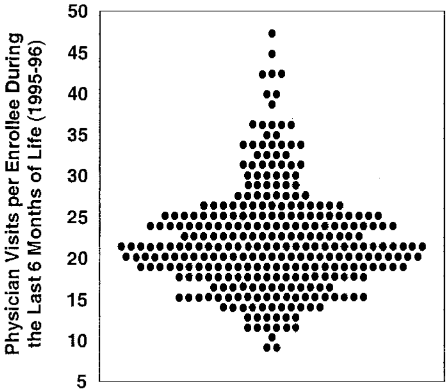


Fig. 4.5 Average number of physician visits per decedent during the last six months of life (1995–96)

Source: www.dartmouthatlas.org, Wennberg and Cooper (1999).

sick patients. Figure 4.5 shows a turnip graph of the distribution of physician visits in the last six months of life; each dot represents average rates for a hospital referral region (HRR).[11] The rates vary widely, from 8.5 in Grand Junction, Colorado, to 43 in Ridgewood, New Jersey, and nearly 48 in Miami. As well, we also consider the percentage of Medicare enrollees who in their last six months are admitted to an ICU. This value again varies widely, from 14 percent in Sun City, Arizona, to 49 percent in Miami, Florida, and Munster, Indiana.

In addition to end-of-life care, which might be expected to reflect that fraction of health care spending devoted toward ineffective care, we also take a prospective approach and consider twelve-month part A Medicare expenditures for a 100 percent sample of fee-for-service Medicare patients aged sixty-five and over who were admitted during 1993–94 for AMI. (Different years are used to ensure that AMI expenditures are not themselves components of overall Medicare expenditures in 1995–96.) Costs have been adjusted for the underlying demographics (age, sex, race) and comorbidities of each patient (at the individual level); as well, price differences across regions have been adjusted using the *Dartmouth Atlas* price index, which is in turn based on Zuckerman, Welch, and Pope (1990). This measure of intensity reflects both the propensity of hospitals to treat heart attacks using high-tech revascularization and the likelihood of follow-up hospitalization and other inpatient treatments during the subsequent year, and thus reflects a broader dimension of care than just treatments for chronically ill patients. The mean level of expenditures is $15,632, with a range from $11,664 to $21,917.

11. Think of this as a vertical histogram with the rows centered.

4.5 Estimating the Efficiency of Medicare Expenditures

The basic set of covariates *Z* includes age-sex-race-adjusted population-based measures at the HRR level of hospital admissions (per 1,000 elderly population) for AMI, stroke (CVA), gastrointestinal bleeding, hip fractures, colon cancer, and lung cancer (Wennberg and Cooper 1999). Hospital admission rates for these diseases are accurate measures of the population incidence since nearly every person with these diseases is admitted to the hospital. In addition, we include rates of poverty in the elderly population and average Social Security income among households receiving this source of income, measured using the CensusCD data (based on primary census data) at the zip code level in 1990. Since Social Security benefits are based on lifetime earnings, these provide a good (albeit nonlinear) index of lifetime earnings. We also use the census data to measure the fraction of elderly people living alone, since such patients may have fewer potential caretakers and thus would be more likely to spend their last six months receiving inpatient care.[12] We consider measures of behavioral health status from the general population such as obesity, binge drinking, cigarette smoking, and seat belt use; these are derived from Centers for Disease Control data (Bolen et al. 2000) at the state level and assigned to the HRR level according to the relative state population weights in each HRR. These variables are summarized in table 4.1.

Table 4.2 demonstrates that these variables combined are strongly associated with survival rates, with an R^2 of 0.56 for the full model. Interestingly, the poverty rate enters with a positive coefficient (that is, a higher poverty rate implies a higher survival rate, holding constant income). In part, this is the consequence of including comorbidities such as heart attacks and hip fractures; poverty works primarily through its impact on disease categories.[13]

We first consider factors that can affect Medicare expenditures in the context of a first-stage regression. Here we use expenditures that have been adjusted for age, sex, race, and differences in cost-of-living reimbursement rates for Medicare services (Wennberg and Cooper 1999). To show the extent to which variations in Medicare expenditures across regions are correlated with covariates *Z*, we first consider a parsimonious model that includes just the six measures of health status (heart attack rates, stroke rates, etc.), with results in column (C) of table 4.2. The set of underlying health measures explain 27 percent of the total variation in Medicare spending. Column (D) of table 4.2 includes additional socioeconomic and behavioral

12. This is defined as the ratio of people over age sixty-five living alone divided by the total number of people over sixty-five not in institutions or living with unrelated people.

13. Holding income constant, a larger poverty rate is consistent with a widening of the income distribution—for example, more income inequality. However, the hypothesis that income inequality is bad for health (e.g., Wilkinson 1997) would predict a negative coefficient on poverty rates, not a positive coefficient. See also Deaton and Paxson (1999).

Table 4.2 **Regression models of survival and Medicare expenditures**

	Survival		Medicare expenditure ($1,000)		
Dependent variable	(A)	(B)	(C)	(D)	(E)
AMI rate (per 1,000)	–0.010	–.007	–0.061	–0.004	0.028
	(1.2)	(0.8)	(2.3)	(0.2)	(1.3)
Stroke rate (per 1,000)	–0.038	–0.043	0.048	–0.022	–0.030
	(4.5)	(4.4)	(1.8)	(0.7)	(1.1)
Gastrointestinal bleeding rate (per 1,000)	–0.051	–0.054	0.190	0.168	0.050
	(2.6)	(3.4)	(3.6)	(3.8)	(1.2)
Colon cancer rate (per 1,000)	–0.011	–0.006	–0.062	0.152	–0.172
	(0.4)	(0.2)	(0.6)	(1.4)	(1.7)
Lung cancer rate (per 1,000)	0.133	0.140	0.467	0.590	0.360
	(1.4)	(1.9)	(2.3)	(3.2)	(2.3)
Hip fracture rate (per 1,000)	–0.047	–0.031	0.056	0.123	0.172
	(3.8)	(2.0)	(1.4)	(3.1)	(5.8)
Fraction living in poverty		.026		0.068	0.058
		(3.0)		(2.8)	(3.4)
Avg. Social Security income (in $1,000)		0.205		0.440	0.040
		(2.8)		(2.4)	(0.3)
High school graduate (%)		–0.008		–0.045	0.018
		(1.6)		(2.5)	(1.3)
College graduate (%)		–0.003		–0.016	–0.005
		(0.6)		(1.3)	(0.5)
Binge drinking (%)		0.003		0.060	0.064
		(0.6)		(3.7)	(4.8)
Cigarette smokers (%)		–0.027		–0.073	–0.067
		(2.8)		(3.0)	(3.1)
Obesity (%)		–0.002		0.003	0.021
		(0.1)		(0.1)	(0.8)
Seat belt use (%)		0.011		–0.003	–0.000
		(4.6)		(0.3)	(0.1)
Elderly living alone (%)		–0.017		–0.051	–0.033
		(.2)		(2.3)	(2.3)
Index of effective care				–0.037	–0.023
				(2.3)	(1.8)
Index of preference-sensitive care				–0.001	0.010
				(0.0)	(1.0)
Physician visits in the last 6 months (avg.)					0.063
					(6.5)
One-year expenditures for AMI					0.112
					(3.8)
Constant	97.124	96.372	1.052	2.474	–0.792
R^2	0.46	0.56	0.27	0.46	0.64

Notes: $N = 306$. Robust standard errors; absolute value of *t*-statistics in parentheses. All sample sizes weighted by Medicare population. Low variation conditions (e.g., AMI, stroke) and effectiveness index are from 1995–96, Medicare expenditures data are for 1996, poverty and Social Security data from 1990 Census, and CDC behavioral data from 1997.

explanatory variables, as well as measures of effective and preference-sensitive care; these combined raise the R^2 to 0.46. The counterintuitive signs of some coefficients (such as for smokers and elderly living alone), as well as the robust correlation between income and Medicare spending, suggest that the coefficients may reflect factors other than health status (McClellan and Skinner 1999). As well, note that the quality index is *negatively* associated with Medicare expenditures; in other words, Medicare spending does not appear to buy better quality of care. Finally, the preference-sensitive index is not correlated with overall spending. While certain components of surgery may be correlated (for example, cardiac bypass surgery is positively associated with per capita Medicare expenditures), on the whole, high-tech surgery does not appear to be the primary (or even secondary) cause of geographical variations in spending (see Wennberg, Fisher, and Skinner 2002).

Finally, physician visits in the last six months and expenditures for AMI patients are highly significant (table 4.2, column [E]) and raise the R^2 to 0.64. To give some sense of how expenditures differ by physician visits in the last six months, we also report predicted Medicare spending (controlling for Z) from a regression with dummy variables reflecting the regional decile ranking for physician visits in the last six months. These coefficient estimates are shown visually in figure 4.6, where decile 1 (the lowest decile) is anchored at the average Medicare spending in that decile. We can use these estimates to calculate overall Medicare expenditures that are explained by end-of-life physician visits; noting that each decile contains 2.77 million elderly people in 1996, we find net spending relative to the bottom decile is $26 billion larger. Since overall expenditures during the same year were about $138 billion, the variation in expenditures attributable to end-of-life care accounts for nearly 20 percent of overall spending.

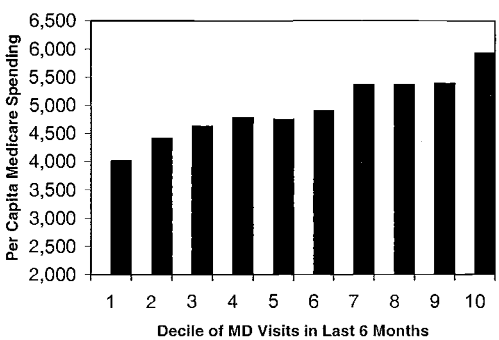


Fig. 4.6 Average per capita Medicare spending 1995–96, by decile of physician visits in the last six months

4.5.1 The Association between Health Care Intensity Measures and Regional Health Status

To what extent are the markers for intensity associated with the measures of health Z? Presumably the instruments or measures of intensity are chosen because on a theoretical basis they are deemed exogenous. Further findings that the health care intensity measures are uncorrelated with measured health status Z would be supportive of the exogeneity presumption for the following reason. If the health intensity measures R are uncorrelated with Zs, which are themselves randomly measured from among the wide variety of health status measures, it would be difficult to argue that the health intensity measures are in addition correlated with unmeasured health status (see Altonji, Elder, and Taber 2005). To examine this correlation, we create terciles of HRRs based on three measures of intensity: the efficiency index, physician visits in the last six months, and AMI expenditures, with results in table 4.3. To summarize the correlation of the health care intensity measure with regard to underlying health status in the region, we create a predicted survival measure based on coefficient estimates in column (C) of table 4.2. As noted above, this predicted measure explains 56 percent of the variation in mortality.[14]

For the index of effective care, there is a positive and significant correlation with predicted survival, suggesting that regions with better initial health status (as measured by the incidence of heart attacks, stroke, colon cancer, obesity, binge drinking, poverty, and other factors) also have higher quality physicians as well. Here we do have a strong theoretical presumption that there is no reverse causation; both sick and healthy patients should be receiving effective care. As noted previously, effective care is negatively correlated with Medicare expenditures ($p < 0.01$). By contrast, for physician visits in the last six months of life, and one-year expenditures for AMI, predicted survival is nearly identical across the terciles. For example, the correlation coefficient between end-of-life physician visits and predicted survival is just 0.01 with a p-value of 0.88, suggesting that they are unlikely to be associated with unmeasured health status. As expected, these latter two factors are positively associated with overall Medicare per capita expenditures.[15]

4.5.2 Linear Two-Stage Regressions

We next consider the linear two-stage regression, presented in table 4.4. The first two columns present regression results for the limited set of covariates (the six health conditions, Social Security income, and poverty

14. Note that this predicted measure of mortality is a linear function only of Z, not of S or u.
15. The final measure of intensity, ICU days in the last six months, is negatively associated with survival, with a correlation coefficient of –0.27 ($p < 0.01$).

Table 4.3 **Terciles of Medicare expenditures and predicted survival by effective care, average physician visits in the last six months, and one-year inpatient expenditures for a cohort of AMI patients**

Dependent variable	1st (lowest) tercile: Effective care	2nd (middle) tercile: Effective care	3rd (highest) tercile: Effective care	Correlation coeff. (p-value) (effective care)
Effective care (%)	43.76	47.93	51.48	1.00
Medicare expenditures (in $1,000)	5.375	4.983	4.906	–0.263 (<0.01)
Predicted survival (from table 4.2, column [B])	94.58	94.67	94.83	0.39 (<0.01)
	1st tercile: Physician visits	2nd tercile: Physician visits	3rd tercile: Physician visits	Correlation coeff. (p-value) (Physician visits)
Avg. physician visits in last 6 months	17.49	22.90	32.15	1.00
Medicare expenditures (in $1,000)	4.486	5.187	5.576	0.53 (<0.01)
Predicted survival (from table 4.2, column [B])	94.74	94.60	94.74	0.01 (0.88)
	1st tercile: AMI expenditures	2nd tercile: AMI expenditures	3rd tercile: AMI expenditures	Correlation coeff. (p-value) (AMI expenditures)
Avg. 1 year Part A AMI expenditures (in $1,000)	14.06	15.40	17.42	1.00
Medicare expenditures (in $1,000)	4.82	4.98	5.46	0.36 (<0.01)
Predicted survival (from table 4.2, column [B])	94.70	94.68	94.71	0.01 (0.89)

rates), with column (A) weighted by the Medicare population and column (B) unweighted and where the model is just identified using physician visits in the last six months as an instrument. The coefficient estimates are stable with regard to whether they are weighted or not. The estimated linear coefficient estimates in columns (A) and (B) of table 4.4 of Medicare expenditures on survival are 0.009 and –0.047, respectively. Neither is significantly different from zero. In other words, these results imply that the roughly $26 billion in Medicare expenditures explained by the instrument generate no clinically important (or economically important) influence on survival. Using the Altonji, Elder, and Taber (2005) approach, we also compare the regression coefficient when all covariates Z are excluded. The

Table 4.4 **Instrumental variable regression estimates of factors affecting survival (dependent variable is the one-year survival rate)**

	Physician visits, last 6 months		Medicare spending, AMI 1993–94		All instruments[a]
	(A)	(B)	(C)	(D)	(E)
Medicare expenditures (in $1,000)	0.009	–0.047	–0.011	–0.002	0.003
	(0.2)	(0.9)	(0.2)	(0.0)	(0.1)
AMI rate (per 1,000)	–.014	–.015	–.014	–.010	–.014
	(1.7)	(2.2)	(1.7)	(1.3)	(1.7)
Stroke rate (per 1,000)	–0.039	–0.037	–0.035	–0.028	–0.035
	(4.3)	(5.0)	(3.8)	(3.6)	(3.7)
Gastrointestinal bleed rate (per 1,000)	–0.048	–0.044	–0.047	–0.057	–0.049
	(2.5)	(2.7)	(2.5)	(3.1)	(2.7)
Colon cancer rate (per 1,000)	0.002	0.003	–0.011	–0.019	–0.013
	(0.1)	(0.1)	(0.3)	(0.6)	(0.3)
Lung cancer rate (per 1,000)	0.051	0.019	0.100	0.051	0.089
	(0.7)	(0.4)	(1.3)	(1.0)	(1.3)
Hip fracture rate (per 1,000)	–0.052	–0.056	–0.032	–0.040	–0.034
	(3.8)	(4.9)	(2.0)	(2.9)	(2.3)
Fraction living in poverty	0.030	0.037	0.030	0.032	0.029
	(3.8)	(5.2)	(3.5)	(4.0)	(3.4)
Avg. Social Security income (in $1,000)	0.205	0.215	0.205	0.200	0.199
	(2.9)	(3.3)	(3.0)	(4.0)	(2.9)
Effectiveness index	0.019	0.013	0.020	0.018	0.021
	(3.4)	(2.7)	(4.0)	(3.9)	(4.1)
Constant	94.51	94.83	95.38	95.14	95.38
Regression weighted by Medicare population?	Yes	No	Yes	No	Yes
R^2	0.52	0.52	0.58	0.58	0.58

Notes: $N = 306$. Additional health measures include percent high school graduates, percent college graduates, percent of elderly population living alone, and percentage of population who are binge drinkers, cigarette smokers, obese, and seat belt users. Robust standard errors; absolute value of *t*-statistics in parentheses. All sample sizes weighted by Medicare population. Low variation conditions (e.g., AMI, stroke) and effectiveness index are from 1995–96, Medicare expenditures data are for 1996, poverty and Social Security data from 1990 Census, and CDC behavioral data from 1997.

[a]Instruments are physician visits in the last six months, percentage of decedents admitted to the ICU in their last six months, and Medicare expenditures for a cohort of AMI patients in 1993–94.

results are similar (the coefficient for Medicare expenditures is –0.015, with a *t*-statistic of 0.25), providing further support for the view that the results are not systematically biased by unobservable health status.

Recall from our earlier discussion that one should expect to find in the steady state an improvement in survival rates of 0.11 from an incremental $1,000 in Medicare expenditures per capita. However, we can reject the null hypothesis that the coefficient is 0.11 at the 0.05 significance level for both columns (A) and (B). Similar results obtain when AMI expenditures are used as the instrument, as shown in columns (C) and (D); once again the coefficients on Medicare expenditures are not significantly different from zero, and we can reject the null hypothesis of 0.11 at the 0.05 level, while ruling out statistically the hypothesis that regional variations in Medicare spending satisfy conditions for efficiency.[16]

Note also that in these four columns of table 4.4 the coefficient on the index of effective care is positive and significant; a 10 percent increase in the index is associated with survival rates roughly 0.2 percentage points higher. This is large in both a clinical sense and in an economic sense; it implies that increasing the index by 4 percentage points, or 1 standard deviation (table 4.1), would increase survival rates by 0.8 percentage points and, using the benchmark of $100,000 per life-year, yield benefits equivalent to about $7,000 per person. This estimate should be interpreted cautiously, since it may be the case that the effectiveness index is associated with a variety of other characteristics of the health care system that we have not controlled for in the regression analysis.

We include all instruments (percent with ICU days in the last six months, as well as physician visits in the last six months, and AMI expenditures) in column (E) of table 4.4. Once again, the coefficient on Medicare expenditures is not significant and is near zero in magnitude. Furthermore, the model does not reject an overidentification test for the additional two instruments ($p = 0.92$).

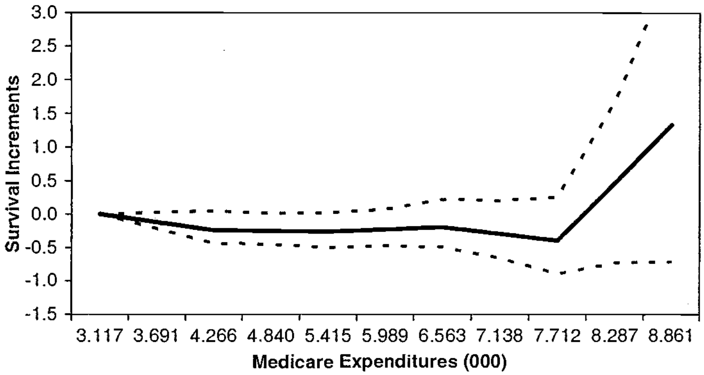
4.5.3 Semiparametric Instrumental Variable Estimates

To this point, we have not allowed for nonlinearities in either $S(M, Z)$ or in $M(R)$ as in equations (6) and (7). We adopt the general model developed in Newey, Powell, and Vella (1999; hereafter NPV), but because of the limitations on the size of the data set, we adopt a simple splined function $S(M)$ and $M(X)$ where the five intervals of the spline are evenly distributed along the ranges of M and X. The NPV method estimates the first-stage regression as in equation (7). Rather than using the fitted value of M in the second stage (as one would normally do in two-stage least squares), one uses the nonlinear transformations of M and the fitted value of the error term from the first-stage regression.

We estimated the simplified $S(M)$ function with results shown in panels A (weighted) and B (unweighted) of figure 4.7 using physician visits in the last six months as the instrument, with the incremental gain in survival set

16. We also experimented with a variety of other variables, such as the percentage in HMOs and the percentage urban population. These variables tended to have similar effects on the coefficient estimates (since they were quite closely correlated) and tended to increase the coefficient estimate to between 0.04 and 0.06, but not significantly different from zero, or 0.11. (Using the urban variable resulted in the loss of one HRR; Ocala, Florida, revamped its zip codes substantially between the 1990 census and the 1996 HRR crosswalk.) Excluding HRRs with more than 10 percentage point HMO enrollment resulted in a *negative* estimated effect of expenditures on survival (–0.147, significant at the 10 percent level).

a. Weighted Data



b. Unweighted Data

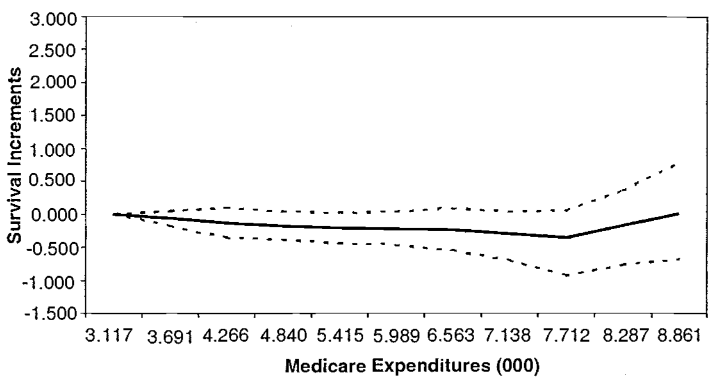


Fig. 4.7 Semiparametric estimates of the marginal effectiveness of Medicare expenditures: *A*, weighted data; *B*, unweighted data

at zero for the minimum level of Medicare expenditures. The 95 percent confidence intervals are also shown in figure 4.7; these are estimated by bootstrapping the combined first- and second-stage regressions.[17] Rather than having a concave shape, as is suggested by diminishing returns (as in figure 4.3), the predicted function is convex in panel A and nearly flat in panel B. The jump up in survival in the weighted regression can be explained almost entirely by the presence of Miami, which, conditional on other factors, is a surprisingly healthy region (e.g., Fuchs, McClellan, and

17. This is done by predicting the entire $S(M)$ curve for each of the bootstrapped iterations, and then graphing the 2.5th and 97.5th percentiles of the bootstrapped values for $S(M)$ along the entire range of M. For computational convenience, the true distribution is evaluated for ten evenly spaced values of M, and interpolated linearly.

Skinner 2004). However, neither estimated model, nor the models estimated using AMI expenditures as the instrument, rejects the hypothesis that Medicare expenditures have zero incremental marginal effectiveness.

4.5.4 Does supply-sensitive care improve functioning and health status rather than survival?

One might object to the statistical models because the higher rates of health care spending may improve patient functioning rather than just survival. We have already shown that higher surgery rates in a region do not appear correlated with higher Medicare spending, suggesting that higher-cost regions are not simply the result of elderly patients' demanding (and getting) more invasive high-tech surgery. But perhaps the end-of-life spending also provides benefits with regard to a better process of care for chronically ill patients. We provide a heuristic impression of differences in health care intensity by considering utilization rates for two groups of regions in the United States. The first sample for the top decile of regions as ranked by physician visits in the last six months of life (weighted by the size of the Medicare population) during 1995–96. These regions include much of the New York and New Jersey metropolitan areas; Takoma Park, Maryland; Philadelphia; and McAllen, Texas; average visits per decedent range from thirty-six to forty-eight. The second sample is for HRRs that are classified in the lowest decile of regions (decile 1); these are all regions with fewer than sixteen visits per decedent. These include Lynchburg, Virginia; Minneapolis, Minnesota; Portland, Oregon; and Salt Lake City, Utah.

In table 4.5, we consider rates of physician procedures (from part B Medicare claims data), using a sample of about 32,000 decedents who died during 1995–96 between the ages of seventy and ninety. The two groupings of HRRs were nearly identical with regard to the age distribution of death and broadly similar with regard to (state-level) causes of death.[18] Rates are expressed as average counts per person in the sample; thus a higher number may reflect more people receiving the treatment or a larger number of treatments per person.

It is perhaps not surprising that physician visits per decedent are higher in decile 10 than decile 1, since that is how the categories were chosen. However, the types of visits show quite different patterns. Outpatient office visits were 46 percent higher in decile 10 regions, relative to our benchmark of the lowest (decile 1) regions, and 79 percent higher for the initial visit by

18. Cause-of-death data are from the Centers for Disease Control web site (www.cdc.gov). Rates are adjusted for age and sex for the population over age sixty-five and are drawn from state measures and merged (by zip code) to the relevant hospital referral regions (HRRs). The age-adjusted mortality rates were roughly identical in the two deciles, 4.97 percent in decile 1 and 4.96 in decile 10. While the percentage of deaths due to cancer in the two groups was similar (22.7 percent in decile 10, 22.1 percent in decile 1), as was diabetes (2.6 and 2.7 percent), cardiovascular diseases were higher in decile 10 (49.1 percent versus 44.1 percent), and chronic obstructive pulmonary diseases were lower (4.9 versus 5.8 percent).

Table 4.5 **Rates of specific procedures per 1,000 decedents, by regional frequency of physician visits in last six months**

Physician service code	Highest decile of physician visits during last six months	Lowest decile of physician visits during last six months	Ratio
	Physician visits		
Physician office outpatient visits	4,453	3,051	1.46
Physician visit for initial hospital care	1,442	804	1.79
Specialist initial inpatient consult	3,087	628	4.92
	Diagnostic testing		
Upper gastrointestinal endoscopy	132	64	2.06
Cat scan of head/skull/brain	492	236	2.08
Chest x-ray	6,631	2,700	2.46
Doppler/echocardiagram of heart	799	268	2.98
Electrocardiogram and report	3,888	1,161	3.35
	Treatments for serious chronic illnesses		
Insertion of emergency airway	140	42	3.33
Hemodialysis (related to kidney failure)	384	87	4.41
Gastrostomy tube placement/change (feeding tube)	136	25	5.44
Continuous ventilator management (mechanical breathing apparatus)	387	50	7.74

Notes: Age 75–90; N = 15,097 (Decile 1); N = 17,225 (Decile 10).

the physician when the patient was admitted to the hospital. The real differences occur for the initial visit by a specialist newly brought on to the case (392 percent more in decile 10 regions). In other words, there is greater *scope intensity*—more specialists treating separate organs or systems—in regions with a large number of physician visits in the last six months of life.

Regions in the top decile are also characterized by their greater use of diagnostic techniques such as endoscopies, X-rays, Doppler echocardiograms, and electrocardiograms; their use is between 106 and 235 percent greater in decile 10 regions. Finally, the greatest divergence in specific medical procedures comes in those that are used to maintain survival among seriously and chronically ill patients: insertions of emergency airways, dialysis for failing kidneys (hemodyalisis), feeding tubes inserted into the stomach (gastrostomy tube placement), and mechanical breathing assistance (continuous ventilator management). These rates are consistently higher, with rates in decile 10 ranging from 233 percent to 674 percent above those in decile 1. Note that these procedures are not designed to improve quality of life but instead are directed toward maintaining short-term survival. In short, it seems unlikely that this supply-sensitive care could be justified on the basis of improving quality of life for Medicare patients.

4.6 Discussion and Conclusion

In this paper, we have attempted to test whether the Medicare program is broadly consistent with the efficiency criterion commonly used in public economics where the marginal social value of the last dollar spent on specific types of health care (in each region) is equal to the marginal social benefits of the dollar that could have been spent for other worthy causes. We used data on survival rates, Medicare expenditures, and health status measures across 306 hospital referral regions in the United States to test these hypotheses. Our best estimates of the incremental value of Medicare spending with regard to *effective* care suggests that spending for these types of services is too low, especially considering how this type of care is associated with overall Medicare expenditures. On the other hand, the *supply-sensitive* dimension of care is a major factor in explaining overall Medicare expenditures—roughly 20 percent annually—but does not show any impact in terms of improving survival rates across regions. These results suggest that the inefficiency inherent in the Medicare program is as much as 20 percent of total Medicare expenditures.[19]

One explanation for these results is that while survival may not be improved, patients either enjoy better health functioning or they may simply prefer preference-sensitive care. However, neither the results presented here nor other studies provide much support for this view (see, e.g., Guadagnoli et al. 1995). In the SUPPORT study, seriously ill patients were asked about their preferences regarding dying in the hospital and intensive life-saving care, and efforts were made to ensure that they got what they wanted (Lynn et al. 2000). However, there was no correlation between expressed preferences and what they actually got; Pritchard and others (1998) found the only predictor of whether patients died in the hospital was the supply of hospital beds in the area. And when asked directly, patients in high-intensity regions (based on overall expenditures in the last six months of life) did not respond that they enjoyed better access to care, nor were they more satisfied with their care (Fisher et al. 2003a,b).

One potential shortcoming with this study is the difficulty in interpreting "the" effect of an instrumental variable on health care across regions. The problem is the highly multidimensional nature of a health care system. For example, when we observe that regions with higher values of physician visits in the last six months of life also experience higher Medicare expenditures, we cannot specify exactly along what dimensions the quality and quantity of care differ. Physician visits for chronically ill patients (many of whom are in their last six months), or AMI inpatient care, could be a pro-

19. Fisher and others (2003a,b) suggest an even larger percentage based on their cohort analysis.

ductive use of Medicare funds, but the other unobserved treatment characteristics with which these are correlated may not be. Or it could be the case that incremental physician visits in the last six months of life are not particularly efficacious, and by choosing a marker most likely associated with inefficient care, we are predisposed toward finding that Medicare expenditures don't have a large impact on outcomes. In this latter case, the surprising result is therefore not that these factors have an indifferent impact on outcomes but that these factors are so highly correlated with overall Medicare expenditures, even after controlling for obvious covariates.

There are three other potential objections to the methodology of the study. The first is that using end-of-life measures of treatment can bias results, given that the effectiveness of the regional health care system will tend to affect the denominator of who is in the end-of-life sample. However, in this case the bias will tend to go toward finding that Medicare expenditures improve survival. Suppose that region A is more aggressive and more effective in treating patients than region B. As a consequence, some of the patients receiving the most aggressive (and expensive) care in region A survive, and hence they do not appear in the end-of-life sample. Hence, region A's end-of-life spending would be biased downward, making it appear relatively more cost effective.

A second objection is that if regions are already optimizing with regard to their Medicare expenditures (i.e., they are at points A, B, and C in figure 4.3), one would not expect to find any difference in outcomes conditional on expenditures. Any variation in Medicare expenditures (and in outcomes) observed in the data is for other factors not adequately captured by the regression model—for example, because of unobservable health needs. However, it seems difficult to reconcile how these unmeasured factors would be so highly correlated with the measures of health care intensity, given that our health care intensity measures for supply-sensitive care are not themselves correlated with observable health needs.

A final concern with the paper is that we are not learning about the efficiency of Medicare per se, because Medicare is an insurance program that is designed to protect the financial health of enrollees, rather than a direct provider of health care such as the Veterans Affairs hospital system. Strictly speaking, to evaluate the efficiency of the Medicare program, one would need to compare the current system to the counterfactual structure of health and health care that would have occurred had Barry Goldwater won the election in 1964 and chosen not to push forward universal health care for the elderly. Absent such an exercise, however, we can ask whether the Medicare program, in its developing role as health regulator rather than passive insurance agency, is getting its money's worth by allowing providers in some states to provide a much higher level of health care intensity (or allowing consumers to receive a much higher level of health care intensity) than in other states.

This macro-level study does not provide an easy prescription for how to fix the Medicare program or how to encourage the greater use of effective care. The Medicare program is federal, with roughly uniform prices paid for procedures (apart from cost-of-living adjustments), so that one cannot appeal to differences in relative nominal prices to explain why Miami spends so much more than other regions in the United States. Nor can reimbursement rates be lowered in high-cost regions without risking a response by physicians of increasing the number of procedures. (As well, lower reimbursement rates in high-cost regions punish those physicians who do practice conservative care.) Still, the potential benefits of reducing regional variation—a saving of nearly 20 percent of current Medicare expenditures—suggests that a central focus of Medicare research should be to better understand why and how some regions are able to provide effective care at relatively low cost.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy,* forthcoming. Northwestern University, Department of Mimeograph.

Angrist, Joshua D. 1990. Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records. *American Economic Review* 80 (3): 313–36.

Bolen, Julie C., Luann Rhodes, Eve E. Powell-Griner, Shayne D. Bland, and Deborah Holtzman. 2000. State-specific prevalence of selected health behaviors, by race and ethnicity—behavioral risk factor surveillance system, 1997. *MMWR* 49 (SS02): 1–60.

Cutler, David M. 1995. The incidence of adverse medical outcomes under prospective payments. *Econometrica* 63 (1): 29–50.

Cutler, David M. 2000. Walking the tightrope on Medicare reform. *Journal of Economic Perspectives* 14 (2): 45–56.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler. 1998. Pricing heart attack treatments. *Quarterly Journal of Economics* 113 (4): 991–1024.

Cutler, David, and Ellen Meara. 1999. The technology of birth: Is it worth it? NBER Working Paper no. 7390. Cambridge, MA: National Bureau of Economic Research, October.

Currie, Janet, and Jonathan Gruber. 1996. Saving babies: The efficacy and cost of recent changes in the Medicaid eligibility of pregnant women. *Journal of Political Economy* 104 (6): 1263–96.

Deaton, Angus, and Christina Paxson. 1999. Mortality, education, income, and inequality among American cohorts. NBER Working Paper no. 7140. Cambridge, MA: National Bureau of Economic Research, May.

Evans, Robert G., and G. S. Stoddart. 1994. Producing health, consuming health care. In *Why are some people healthy and others not?* ed. Robert G. Evans, Morris L. Barer, and Theodore R. Marmor, 27–64. Hawthorne, NY: Aldine de Gruyter.

Feenberg, Daniel, and Jonathan Skinner. 2000. Medicare transfers across states: Winners and losers. *National Tax Journal* 53 (September): 713–32.

Fisher, Elliott S., David Wennberg, Therese Stukel, Daniel Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003a. The implications of regional variations in Medicare spending. Part 1: The content, quality, and accessibility of care. *Annals of Internal Medicine* 138 (4): 283–87.

Fisher, Elliott S., David Wennberg, Therese Stukel, Daniel Gottlieb, F. L. Lucas, and Etoile L. Pinder. 2003b. The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine* 138 (4): 288–99.

Fuchs, Victor, Mark McClellan, and Jonathan Skinner. 2004. Area differences in utilization of medical care and mortality among U.S. elderly. In *Perspectives on the economics of aging,* ed. D. Wise, 367–406. Chicago: University of Chicago Press.

Guadagnoli, Edward, Paul J. Hauptman, John Z. Ayanian, Chris L. Pashos, Barbara J. McNeil, and Paul D. Cleary. 1995. Variation in the use of cardiac procedures after acute myocardial infarction. *New England Journal of Medicine* 333: 573–78.

Hadley, Jack. 1988. Medicare spending and mortality rates of the elderly. *Inquiry* 25:485–93.

Hearst, Norman, Thomas B. Newman, and Stephen B. Hulley. 1986. Delayed effects of the military draft on mortality. *New England Journal of Medicine* 314 (10): 620–24.

Jencks, Stephen F., Edwin D. Huff, and Timothy Cuerdon. 2003. Change in the quality of care delivered to Medicare beneficiaries, 1998–1999 to 2000–2001. *Journal of the American Medical Association* 289 (3): 305–12.

Lynn, Joanne, Hal R. Arkes, Marguerite Stevens, Felicia Cohn, Barbara Koenig, Ellen Fox, Neal V. Dawson, Russell S. Phillips, Mary Beth Hamel, and Joel Tsevat. 2000. Rethinking fundamental assumptions: SUPPORT's implications for future reform. *Journal of the American Geriatric Society* 48:S214–S221.

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. 1994. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association* 272 (September): 859–66.

McClellan, Mark, and Jonathan Skinner. 1997. The incidence of Medicare. NBER Working Paper no. W6013. Cambridge, MA: National Bureau of Economic Research, April.

McClellan, Mark, and Jonathan Skinner. 1999. Medicare reform: Who pays and who benefits? *Health Affairs* 18 (1): 48–62.

Newey, Whitney K., James L. Powell, and Francis Vella. 1999. Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67 (3): 565–603.

Pardes, Herbert, Kenneth Manton, Eric Lander, Dennis Tolley, Arthur Ullian, and Hans Palmer. 1999. Effects of Medicare research on health care and economy. *Science* 283:36–37.

Pritchard, R. S., E. S. Fisher, J. M. Teno, S. M. Sharp, D. J. Reding, W. A. Knaus, J. E. Wennberg, J. Lynne, and SUPPORT investigators. 1998. Influence of patient preferences and local health system characteristics on the place of death: SUPPORT investigators study to understand prognoses and preferences for risks and outcomes of treatment. *Journal of the American Geriatric Society* 46: 1242–50.

Reinhardt, Uwe E., Peter S. Hussey, and Gerard F. Anderson. 2004. U.S. Health care spending in an international context. *Health Affairs* 23 (3): 10–25.

Skinner, Jonathan, and John E. Wennberg. 2000. Regional inequality in Medicare spending: The key to Medicare reform? In *Frontiers of health economics 3,* ed. A. Garber, 69–90. Cambridge: MIT Press.
Wennberg, John E., and Megan M. Cooper. 1999. *The Dartmouth atlas of health care in the United States: The quality of medical care in the United States; A report on the Medicare program.* Chicago: American Health Association Press.
Wennberg, John E., Elliott Fisher, and Jonathan Skinner. 2002. Geography and the debate over Medicare reform. *Health Affairs* 21:W96–W114.
Wilkinson, Richard. 1997. *Unhealthy societies: The affliction of inequality.* London: Routledge.
Wilmoth, John R. 2001. Human mortality database. http://www.mortality.org.
Zuckerman, Stephen, W. P. Welch, and Gregory C. Pope. 1990. A geographic index of physician practice costs. *Journal of Health Economics* 9 (1): 39–69.

Comment Alan M. Garber

This paper is part of a rich tradition of research exploring geographic variation in the utilization of medical services. John Wennberg, one of the coauthors, is a pioneer in this field of research, which has demonstrated the pervasiveness of variation (Wennberg and Gittelsohn 1973). Variation has been documented in multiple settings and at different periods of time, and it is not fully explained by differences in the prevalence of disease or in other observed health characteristics of the populations (Phelps and Mooney 1993). The literature suggests that supply factors, though they are incompletely measured, may play an important role (Wennberg, Barnes, and Zubkoff 1982). This literature also documents that variation tends to be greater when there is greater uncertainty about the effectiveness of care. Thus, the propensity to hospitalize patients with heart attacks, for example, is much less variable than the use of surgery as a treatment for back pain, for which high-quality evidence of effectiveness is scant.

If variation in utilization is not attributable to unmeasured health characteristics, there is at least a presumption that someone, somewhere, is not providing care efficiently. Then it might be possible to reallocate health resources to improve health without increasing resource consumption. This is the issue that Skinner and colleagues explore, looking specifically at the Medicare program.

The Medicare population might appear at first to be an odd choice for a study of practice variation. One would expect to see extensive variation in utilization among the nonelderly population, if only because many lack in-

Alan M. Garber is a staff physician with the Veterans Affairs Palo Alto Health Care System; Henry J. Kaiser, Jr., Professor and professor of medicine at Stanford University; and a research associate and director of the Health Care Program at the National Bureau of Economic Research.

surance and cost-sharing provisions of commercial health insurance policies vary greatly. Among the elderly, hospital insurance coverage under Medicare part A is effectively universal, and the overwhelming majority of elderly have Medicare part B coverage. A large majority also have supplemental (Medigap) private insurance. Thus, health insurance coverage among the elderly would seem to be far more uniform than among Americans younger than sixty-five years of age. Nevertheless, as Skinner and colleagues show, the elderly in different parts of the country receive different amounts and forms of care.

What does "efficiency" mean? According to Skinner and colleagues, efficiency requires that benefits from Medicare expenditures be equated at the margin across different geographic units. One can readily imagine that if beneficiaries in Miami are treated more intensively than similar patients in Chicago, and there are diminishing marginal benefits to treatment, a reallocation of resources from Miami to Chicago would result in an improvement of health outcomes among Medicare beneficiaries overall.

To test efficiency empirically, they develop an econometric specification in which survival is a function of expenditures, which is a proxy for intensity. The problem with expenditures (or intensity) is that they are undoubtedly endogenous and influenced by unmeasured health characteristics. As instruments for intensity, they use a few measures, most prominently physician visits per decedent in the last six years of life within the unit area. A secondary set of instruments consists of expenditures for patients admitted for hip fracture.

The interpretation of the empirical results depends in large part upon one's view of the validity of the instruments. There are few examples of good instruments in studies of health care and health outcomes. One important exception is myocardial infarction, in which location of residence strongly influences the choice of treating hospital. Myocardial infarction, or heart attack, is considered a true medical emergency, so there is little opportunity for selection effects. Patients with symptoms of heart attacks are usually admitted to the nearest hospital, whether it is a small community hospital or a major referral center. Thus, proximity to an advanced hospital (capable of performing cardiac catheterization and heart surgery) can serve as a measure of intensity of care that is plausibly independent of patient characteristics (McClellan, McNeil, and Newhouse 1994).

One can argue that area-wide intensity measures, such as treatment within the last six months of life, are independent of individual patient characteristics and are therefore valid instruments, though it is also easy to posit circumstances in which such an instrument is not valid. The interpretation of the instrument depends upon the sources of variation in treatment intensity at the end of life and one's views about the difference between the intensity of end-of-life care and the intensity of care overall.

If the correlation between physician visits at the end of life and the general intensity of care is perfect, we need not worry about which measure is used. But if the two measures are not highly correlated, physicians in regions with high intensity at the end of life might be allocating (relative to other areas) more intensive treatment to people who won't benefit. In itself, this would be an interesting finding, but its interpretation is different from the interpretation that one would attach to a relationship between general intensity and outcomes. Intensive care in general might have a positive marginal effect on survival and other health outcomes, while intensive care (or physician visits) at the end of life might have no effect on outcomes. That is, there is no contradiction between the statement that Medicare beneficiaries who live in regions that treat dying patients intensively do not have better health outcomes, and the statement that regions that generally treat Medicare beneficiaries more intensively have better outcomes. Consequently, policies that reduced overall intensity might lead to worsened health outcomes, even if there was no relationship between end-of-life intensity and health.

The definition of efficiency is a crucial consideration for this work. By asking whether treatment intensity is equated at the margin across regions, Skinner and colleagues are implicitly addressing movement along a production possibility frontier. But they note that there is substantial variation in "process quality" of care, or in compliance with generally accepted guidelines for care. Although such variation may in part reflect intensity and represent different points on a production possibility frontier, variation in process quality can be a manifestation of x-inefficiency. That is, poor performance on process quality measures may mean that the care in some areas is interior to the production possibility frontier.

Movement along a production possibility frontier and elimination of x-inefficiency can both lead to better health outcomes for given expenditures. However, they are not likely to be equally acceptable, nor do they have the same economic implications. Elimination of x-inefficiency means that no Medicare beneficiary need be made worse off—that is, an actual Pareto improvement is possible. Movement along a production possibility frontier may satisfy the Kaldor-Hicks criterion, meaning that it will generate a *potential* Pareto improvement, but that does not mean that an actual Pareto improvement will occur. Residents of areas that use care more intensively will have to receive less care to equate treatment intensity at the margin across areas. If they do not receive compensation, they will be made worse off. Any policy that led to this shift of resources would meet with strong political resistance. Elimination of x-inefficiency would be an important contribution to health and health care, while movement along the production possibility frontier may be more feasible in the lecture hall than in the real world of medical practice.

This paper asks the right questions and analyzes a rich data set to provide some tentative yet provocative answers. It establishes the importance of learning more about the causes of variation in intensity and about the consequences of intensity variation for health. Until such information is forthcoming, though, the work by Skinner and colleagues casts doubt on the belief that more care inevitably leads to better health outcomes.

References

McClellan, M. B., B. J. McNeil, and J. P. Newhouse. 1994. Does more intensive treatment of acute myocardial infarction reduce mortality? *Journal of the American Medical Association* 272 (11): 859–66.

Phelps, C. E., and C. Mooney. 1993. Variations in medical practice use: Causes and consequences. In *Competitive approaches to health care reform,* ed. R. J. Arnould, R. F. Rich, and W. D. White, 139–78. Washington, DC: Urban Institute.

Wennberg, J. E., B. A. Barnes, and M. Zubkoff. 1982. Professional uncertainty and the problem of supplier-induced demand. *Social Sciences and Medicine* 16: 1811–24.

Wennberg, J. E., and A. Gittelsohn. 1973. Small area variations in health care delivery. *Science* 182:1102.