

## The Debate over Regional Variation in Health Care Spending

*The regional variations in health care spending that are documented by the Dartmouth Atlas of Health Care have been cited by many as a justification, and possible basis, for changes in provider payment rates. The articles below — and the responses that follow them — address various concerns about the Dartmouth data.*

### A Map to Bad Policy — Hospital Efficiency Measures in the Dartmouth Atlas

Peter B. Bach, M.D., M.A.P.P.

In showing that regional spending variations do not correlate with differences in disease burden or outcome, the Dartmouth Atlas of Health Care has felled the notion that higher health care spending necessarily leads to improved health outcomes. Policymakers have seized on two possible ways to wring savings out of this information. The health care reform bill passed by the House of Representatives contains provisions for identifying regions where Medicare spending appears disproportionately high and adjusting payment rates accordingly. And some policymakers, including President Barack Obama, have proposed that the features of high-performing, “efficient” health care systems should be identified and their lower-cost practices emulated.

Dartmouth Atlas researchers have also begun attempting to convert their observations into cost-saving policies. By analyzing Medicare claims for

### Looking Back, Moving Forward

Jonathan Skinner, Ph.D., Douglas Staiger, Ph.D., and Elliott S. Fisher, M.D., M.P.H.

The recent Senate election in Massachusetts may reshape or delay health care reform, but we still face the twin challenges of unsustainable cost increases and uneven quality that plague U.S. health care. Recent controversies have left many people confused about how we might wisely move forward.

One such controversy is the debate over the “value index,” a reimbursement approach that would adjust providers’ payments on the basis of regional performance on quality and cost measures. Legitimately concerned that careless implementation of a value index might hurt some preeminent teaching institutions, some leaders of academic medical centers have responded to this proposal by questioning the validity of existing measures of cost performance, many of which have been generated from Medicare data by our Dartmouth research group.

*Bach continued*

people who have been treated in hospitals and have died, they aim to identify and rank high- and low-efficiency hospitals.<sup>1</sup> Because the Atlas is so influential, their rankings could have broad effects on policy. They could affect hospitals' payments and prestige, and *Consumer Reports* is already publicizing them to consumers ([www.consumerhealthreports.org](http://www.consumerhealthreports.org)). Given their potentially far-reaching implications, it is concerning that the rankings are unsound, both conceptually and methodologically.

The conceptual problem lies in the fact that in Atlas analyses all health care costs that are incurred by patients over the 2 years before their death are attributed to the hospital where they were admitted most frequently during that period. This method assumes that the hospital controls all, or at least most, patient care, even if it occurs outside the hospital or in another hospital. It thus seems to presuppose a system in which hospitals are accountable for all care — perhaps a noble long-term objective, but not a current reality.

Hospitals can sometimes influence the immediate follow-up care of discharged patients, but much care is beyond their control. According to the Center for Studying Health System Change, less than 20% of U.S. physicians have a financial connection with any hospital. Home health agencies and inpatient facilities for subacute care — which tend to have high costs — are designed to serve multiple hospitals, not a single hospital that can exert control over them. Frequently, even the hospital to which a patient is assigned has very limited contact with that patient. In one Atlas analysis, one third of the patients who were included in an assessment of hospital efficiency had been admitted to the hospital in question only once.<sup>2</sup>

Atlas-based analyses are also hampered by methodologic problems, starting with their implicit definition of efficiency. A true analysis of efficiency would ask “whether healthcare resources are being used to get . . . improved health,”<sup>3</sup> weighing both resources consumed and outcomes. Yet Atlas efficiency rankings consider only costs (i.e., resources consumed).

Conceptually, this approach would be appropriate only if outcomes were the same in all hospitals, so that costs equaled efficiency. But since outcomes vary among hospitals and providers, both costs and outcomes must be assessed in evaluating efficiency. Atlas researchers might correctly argue that costs

*Skinner et al. continued*

Perhaps the loudest criticisms have focused on the Dartmouth end-of-life analyses, which measure spending and utilization in either the last 6 months or 2 years of life (often referred to as “look-back” measures). Criticisms have ranged from thoughtful concerns that even among people in their last months of life, health status could vary systematically among hospitals<sup>1</sup> to misinformed claims, reported in the *New York Times* and elsewhere, that of course spending more won't help the end-of-life patients — because they're all dead.

One group argues that the only valid approach is a “look-forward” method that begins with a well-defined, clinically homogeneous cohort (e.g., patients with acute myocardial infarction) and follows their costs for a fixed period.<sup>2</sup> We would counter, however, that the whole debate over how best to measure costs is a distraction, since look-forward and look-back measures yield very similar results, as a 2003 Dartmouth study showed.<sup>3</sup>

We have now updated this result at the hospital level, using two measures of costs. The first comes from a sample of all Medicare enrollees who were admitted to a hospital with a diagnosis of acute myocardial infarction. Starting from the day of admission, we measured Part A (inpatient) expenditures for 1 year forward, as well as 1-year mortality. These look-forward measures have been fully risk-adjusted at the individual level for the anatomical location of the myocardial infarction, income level in the patient's ZIP Code, and coexisting conditions. The look-back measure of spending from the 2008 Dartmouth Atlas of Health Care focused on the last 2 years of life for people with at least one serious chronic illness; this measure, too, has been risk-adjusted for the type of illness and the presence or absence of multiple diseases. All individual-level measures of spending have been adjusted for regional price differences.<sup>4</sup> These spending measures, as well as rates of death for patients with acute myocardial infarction, have been aggregated to the hospital level.

The graph shows the correlation between the look-forward and look-back measures for 144 of the largest U.S. hospitals. We would not expect a perfect correlation; end-of-life spending largely reflects the use of the hospital as a site of care and how much care is provided in the intensive care unit, whereas costs for acute myocardial infarction in part reflect the relative intensity of invasive proce-

*Bach continued*

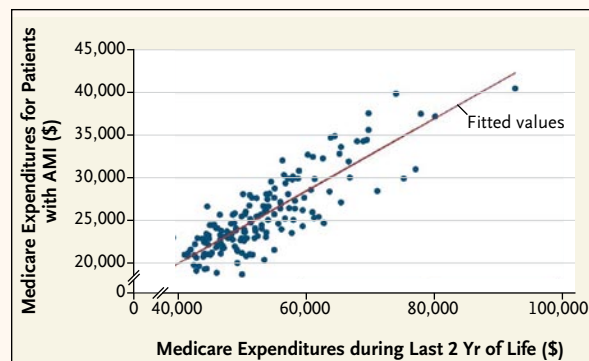
correlate poorly with outcomes. But poor correlation does not imply that outcomes are homogeneous, but rather that there are high-spending hospitals that use resources in a manner that improves outcomes and others that squander resources, failing to improve health. The same goes for low-spending hospitals. Figuring out which is which is the purpose of efficiency assessment, which therefore requires consideration of both costs and outcomes.

Say Hospital A and Hospital B each has a group of patients with a fatal disease. Hospital A gives each patient a \$1 pill and cures half of them; Hospital B provides no treatment. An Atlas analysis would conclude that Hospital B was more efficient, since it spent less per decedent. But all the patients die at Hospital B, whereas only half of the patients do at Hospital A, where the cost per life saved is a bargain at \$2. Although \$1 cures are rare, changing the price or efficacy of the pill does not alter the fundamental problem with examining costs alone when cost differences are sometimes associated with outcome differences.

Another methodologic problem is that Atlas analyses assess hospital efficiency overall on the basis of costs incurred for nonrepresentative patients — decedents who were enrolled in fee-for-service Medicare. This group varies among hospitals in terms of severity of illness and is not representative of a given hospital's overall spending pattern.

Regarding illness severity, Atlas researchers note on their Web site ([www.dartmouthatlas.org/faq/hospital.shtm](http://www.dartmouthatlas.org/faq/hospital.shtm)) that they focus on “patients who died so that [they can be sure] that patients were similarly ill across hospitals,” further explaining that “by definition, the prognosis of all patients [who died was] identical — all were dead . . . therefore, variations [in resource use] cannot be explained by differences in the severity of illness.” But since some hospitals take care of sicker patients than others, the average severity of illness of patients who die also varies among hospitals. This fact is being ignored when all spending differences are attributed to differences in efficiency.<sup>4</sup>

My analysis of data from the 2006 Nationwide Inpatient Sample on decedents 65 years of age or older illustrates how far off the Atlas assumption is (see graph). Using “All-Patient-Refined Diagnosis-Related Groups,” I found that among such dece-

*Skinner et al. continued*

#### Look-Forward and End-of-Life Measures of Health Care Expenditures at 144 Large U.S. Hospitals.

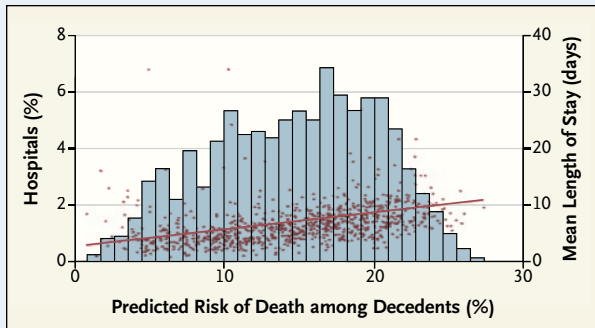
The x axis shows Medicare expenditures during the last 2 years of life for a sample of 4.7 million fee-for-service Medicare enrollees 65 years of age or older with at least one chronic disease. Hospital-level spending measures are created by aggregating individual spending measures, with adjustment for the type of chronic illness and the presence or absence of more than one chronic illness. The y axis shows Part A (inpatient) Medicare expenditures for patients with acute myocardial infarction (AMI) according to hospital during the period from 2000 through 2005. The sample comprised the 144 hospitals with an average of at least 200 annual admissions for AMI. Risk adjustment included the anatomical location of the infarction, coexisting conditions, and median income and poverty rate in the patient's ZIP Code of residence. All expenditures were adjusted for regional differences in average Medicare reimbursement rates, according to the methods in Gottlieb et al.<sup>4</sup> Data regarding Medicare expenditures are available at [www.dartmouth.edu/~jskinner/documents/SkinnerStaigerw14865.pdf](http://www.dartmouth.edu/~jskinner/documents/SkinnerStaigerw14865.pdf).

dures. Nevertheless, the two measures are very highly correlated (correlation coefficient, 0.85;  $P < 0.001$ ). Even among all 2360 hospitals in the sample with an annual average of at least 10 admissions for acute myocardial infarction, the correlation coefficient is 0.73 ( $P < 0.001$ ). That is, hospitals ranking high on one index rank high on the other, a correlation also found in a study of six hospitals in California.<sup>2</sup>

Furthermore, both spending measures are modestly but positively associated with 1-year mortality from acute myocardial infarction, meaning that, on average, higher-spending hospitals have worse outcomes among patients with acute myocardial infarction. This does not mean that all high-cost hospitals are low-quality; some of the higher spending may be devoted to beneficial treatments, but some may well be devoted to unnecessary or cost-inefficient care, such as discretionary use of the hospital as a site of care and more frequent use of physician visits, specialist referrals, diagnostic tests, and minor procedures.<sup>5</sup> And no amount of risk, price, or poverty ad-

*Bach continued*

dents, the predicted risk of death at the time of admission varied widely among hospitals. At the average hospital, the average risk was 15%. But the severity of illness was far lower in hospitals at the



**Distribution among Hospitals of the Average Predicted Risk of Death at Admission among Patients 65 Years of Age or Older Who Died.**

The histogram, with bars graphed on the left y axis, shows the distribution of the average predicted risk of death at admission in various hospitals among patients 65 years of age or older who died. The average length of stay for these patients in each of those hospitals (red dots) and the linear relation between the average predicted risk of death in the hospital and the average length of stay for those decedents (red line,  $P < 0.001$ ) are shown on the right y axis. Data are from the 2006 Nationwide Inpatient Sample.

10th percentile (6% risk of death) and far higher in those at the 90th percentile (22% risk of death). Differences in illness severity result in differences in resource consumption. For example, the average length of stay for this population was 4.4 days in hospitals at the 10th percentile but 9.3 days in those at the 90th percentile. In other words, the average decedent in a hospital with a low average severity of illness started out much less sick than the average decedent in a hospital with high average severity, and differences in resource consumption should be expected. The Atlas, however, assumes that all decedents in all hospitals were equally sick before death, an error that tends to make low-severity hospitals look more efficient than high-severity hospitals even if the hospitals are equally efficient.

Resources that are consumed in the care of fee-for-service Medicare patients who died are also a poor surrogate for costs of care for other patients at the same hospital. If resource consumption is measured by the number of hospital days, the correlations are only 0.40 with patients in Medicare Ad-

*Skinner et al. continued*

justment — or argument about whether to look forward or back — will correct this problem.

Health care reformers seek to help providers reduce costs and improve the quality of care. There are two necessary conditions to attaining these goals. The first is accurate measurement of costs and quality — otherwise, incentive systems that are designed to reward specific providers are worse than useless. The second is the right incentives to encourage low-cost, high-quality care.

With regard to measuring cost, we're doing better than many people think. The striking empirical finding is that a common cost factor seems to exist for hospitals and regions, no matter whether costs are measured in terms of expenditures for cohorts with acute myocardial infarction, those cared for at the end of life, or those with colon cancer or hip fracture.<sup>3</sup> Measures of end-of-life costs do better than any of the individual look-forward measures in predicting this common cost factor, in part because the sample sizes are so large.

Why are these cost measures so strongly correlated? As we have shown,<sup>2</sup> spending does not appear to be consistently related to patient outcomes or satisfaction levels. Instead, the common component seems to be something systemic about the hospital-physician network; the factor, whatever it is, drives up the numbers of hospital days and physician visits and the use of imaging and other services for all types of patients. (Interestingly, surgical rates do not follow this pattern; Miami, the country's most expensive region, has low rates of hip and knee replacements.) The implication of these results is that excessive health care costs arise at the level of the hospital-provider network. Thus, incentives that are designed to reduce costs should be targeted to specific networks, rather than regions or states.

Several approaches are being considered for addressing high costs with the right incentives. One involves "bundled payments," under which hospital-physician systems receive a single payment for all services provided during a given episode, such as congestive heart failure. Thus, physicians and hospitals that reduce the rates of complications and readmissions capture some of the savings.

Another approach is the accountable care organization (ACO) model, in which payers identify the primary care patients of a physician-hospital net-

*Bach continued*

vantage plans and 0.34 with patients covered by commercial insurance.<sup>5</sup> If most care and resources went to dying fee-for-service Medicare patients, then maybe these weak correlations wouldn't matter. But I estimate that these patients account for less than \$3 of every \$100 spent on health care.

Only scientifically valid measures of efficiency should guide policy. Several organizations are developing such measures — for example, Medicare's Physician Group Practice Demonstration recently released data incorporating measures of cost and quality with respect to the management of specific conditions; the Centers for Medicare and Medicaid Services is developing "value scores" for physicians on the basis of quality performance and resource use; the National Committee for Quality Assurance publishes health-plan-level "Relative Resource Utilization" measures for various conditions; and Prometheus Payment has developed approaches to prospective payment that reward efficiency by enabling providers to share in savings for avoiding complications.

Of course, more work is required. But there are dangers in using scientifically shaky measures of efficiency to identify high-performing providers. Patients may be misled in selecting providers. Providers may base delivery changes on the wrong models. And the whole system may waste precious time and resources pursuing a path that fails to lead toward quality improvement and cost containment.

Financial and other disclosures provided by the author are available with the full text of this article at NEJM.org.

From Memorial Sloan-Kettering Cancer Center, New York.

1. Wennberg JE, Fisher ES, Baker L, Sharp SM, Bronner KK. Evaluating the efficiency of California providers in caring for patients with chronic illnesses. *Health Aff (Millwood)* 2005;Suppl Web Exclusives: W5-526–W5-543.
2. Wennberg JE, Fisher ES, Stukel TA, Skinner JS, Sharp SM, Bronner KK. Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States. *BMJ* 2004;328:607.
3. Palmer S, Torgerson TJ. Economic notes: definitions of efficiency. *BMJ* 1999;318:1136.
4. Bach PB, Schrag D, Begg CB. Resurrecting treatment histories of dead patients: a study design that should be laid to rest. *JAMA* 2004; 292:2765-70.
5. Baker LC, Fisher ES, Wennberg JE. Variations in hospital resource use for Medicare and privately insured populations in California. *Health Aff (Millwood)* 2008.27(2):w123-w134.

*Skinner et al. continued*

work that is willing to take responsibility for the full continuum of care. A spending target is set for these patients, and if the ACO meets quality benchmarks and reduces per-beneficiary spending below the target, providers receive a share of the savings. Ideally, a virtuous cycle would emerge in which continued improvements in care would lead to reduced utilization and lower costs.

We are more concerned about other proposals that may not get the incentives right. For example, under the current conception of the "value index," rewards would be based on regional averages rather than the performance of the particular physician-hospital network; such a system would inappropriately punish good providers in low-performing regions and reward poor providers in high-performing ones. And it would drive providers whose fees were cut to seek other ways to increase their revenues, such as by increasing the frequency of visits or the volume of profitable services — trapping them on the treadmill of "hamster medicine."

So what's next? An "industry standard" of hospital-level health care cost measures for the Medicare population is within our grasp. This standard would adjust for many relevant factors — such as the average severity of illness in a system's population, its poverty level, and numbers of patients undergoing organ transplantation — and direct funding would be provided to compensate academic medical centers for mission-related activities. Such measures are a necessary first step to getting the incentives right, and hospitals and physician groups that are interested in preparing for such a shift could do worse than to re-examine how they treat patients near the end of life.

Financial and other disclosures provided by the authors are available with the full text of this article at NEJM.org.

From the Dartmouth Institute for Health Policy and Clinical Practice (J.S., D.S., E.S.F.) and the Department of Medicine, Dartmouth Medical School (E.S.F.) — both in Lebanon, NH; and the Department of Economics, Dartmouth College, Hanover, NH (J.S., D.S.).

1. Bach PB, Schrag D, Begg CB. Resurrecting treatment histories of dead patients: a study design that should be laid to rest. *JAMA* 2004;292: 2765-70.
2. Ong MK, Mangione CM, Romano PS, et al. Looking forward, looking back: assessing variations in hospital resource use and outcomes for elderly patients with heart failure. *Circ Cardiovasc Qual Outcomes* 2009; 2:548-57.
3. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 2: health outcomes and satisfaction with care. *Ann Intern Med* 2003;138:288-98.

*Bach continued***DRS. SKINNER, STAIGER, AND FISHER REPLY:**

Dr. Bach makes the important point that even people near death may vary with respect to underlying illness and medical expenses. But he is mistaken when he claims that the 2008 Dartmouth Atlas end-of-life measures do not adjust for such differences. The Atlas sample comprises Medicare enrollees with at least one life-threatening chronic disease in their last 2 years of life. It further adjusts for the type of chronic disease and the presence of multiple diseases. As we show, appropriately risk-adjusted “look-forward” and “look-back” measures are very highly correlated.

We agree with Dr. Bach that fragmentation of care — resulting in the admission of patients to multiple hospitals and nursing homes — can explain why some hospitals appear so expensive in the Dartmouth data. But patients need to know about such fragmentation. Do they really want to be cared for in a hospital–physician network where patients are bounced from one hospital to another? Furthermore, accountable care organizations are a promising approach to discouraging such poorly coordinated care.

Finally, we agree that measuring hospital quality accurately is both necessary and difficult. Unlike Medicare cost measures, quality measures for specific clinical conditions are often poorly correlated within a given hospital. For this reason, the National Quality Forum has recommended evaluating quality and efficiency specific to care for a given clinical condition. Indeed, we would be remiss if we did not develop reliable measures of costs and quality — including assessments of whether patients’ preferences are aligned with treatment choices — for the large population of patients who have life-threatening chronic illnesses and are near death.

Copyright © 2010 Massachusetts Medical Society.

*Skinner et al. continued*

4. Gottlieb DJ, Zhou W, Song Y, Andrews KG, Skinner JS, Sutherland JM. Prices don't drive regional Medicare spending variations. *Health Affairs*, January 28, 2010. (Accessed February 1, 2010, at <http://content.healthaffairs.org/cgi/content/abstract/hlthaff.2009.0609>.)
5. Sirovich B, Gallagher PM, Wennberg DE, Fisher ES. Discretionary decision making by primary care physicians and the cost of U.S. health care. *Health Aff (Millwood)* 2008;27:813-23.

**DR. BACH REPLIES:**

Until now, the way costs have been measured for research purposes has not mattered that much. Prospectively gathered costs for all patients and retrospectively gathered costs for decedents yield the same general conclusion: spending is poorly correlated with outcome.

Going forward, however, the method will matter. Assessing costs of care prospectively for all patients, not only for those who die, will provide hospital and practice CEOs with the right action item, not a perverse incentive. They will seek to streamline care for all patients rather than withhold care from the seriously ill.

“Bundling” gets this effort right. It provides prospective payment to an organization on the basis of the average costs of care for patients, not decedents.<sup>1</sup> Accountable care organizations do this, too — bonuses are for savings per patient, not per decedent.<sup>2</sup>

Regionally adjusting payment rates to counteract differences in regional spending is less appealing. Nowhere is care regionally organized. When regions have overall spending that is too high and, as a result, rates are cut, greater finger pointing, not greater care coordination, will follow.

In any case, spending differences among regions are a lot smaller than we originally thought. The spending ranks of regions change a lot from year to year, so regions are not simply “low-spending” or “high-spending.”<sup>3</sup> Indeed, McAllen, Texas, is not even the highest-spending region in Texas, according to the most recent analysis.<sup>4</sup> Moreover, differences in spending among regions, though still present, are not nearly as dramatic after adjustment for regional differences in severity of illness and input prices.<sup>4</sup>

1. Miller HD. From volume to value: better ways to pay for health care. *Health Aff (Millwood)* 2009;28:1418-28.
2. Fisher ES, McClellan MB, Bertko J, et al. Fostering accountable health care: moving forward in Medicare. *Health Aff (Millwood)* 2009;28(2):w219-w231.
3. Chernew ME, Sabik L, Chandra A, Newhouse JP. Ensuring the fiscal sustainability of health care reform. *N Engl J Med* 2010;362:1-3.
4. Medicare Payment Advisory Committee. Report to Congress: measuring regional variation in service use. Washington, DC: MEDPAC, 2009.

Copyright © 2010 Massachusetts Medical Society.