

P E R S P E C T I V E

The Elusive Connection Between Health Care Spending And Quality

Cooper may believe that he has challenged the basic scientific findings of the Dartmouth group. But he has not.

by Jonathan Skinner, Amitabh Chandra, David Goodman, and Elliott S. Fisher

ABSTRACT: Richard Cooper has shown a positive association between health care quality and “total spending” at the state level, but he does not appear to understand the limitations of this total spending measure; simply adjusting for median age causes the significant positive correlation to disappear. Cooper also finds that some third factor—we think that it is “social capital”—is the key to explaining health care quality. Cooper may believe that this result challenges three decades of research by the Dartmouth group. Instead, it supports the group’s view that improved efficiency—and not more doctors and hospital beds—is central to improving quality. [*Health Affairs* 28, no. 1 (2009): w119–w123 (published online 4 December 2008; 10.1377/hlthaff.28.1.w119)]

RICHARD COOPER has attempted to challenge Katherine Baicker and Amitabh Chandra’s finding of a negative correlation between average Medicare spending and quality measures.¹ He first replicates the Baicker and Chandra result using other quality measures from the Commonwealth Fund, and he finds a similar negative correlation. He then introduces a new and largely unproven alternative measure of per capita health care spending, and he demonstrates that for this measure, there is a positive correlation between spending and quality.

In this response we make two simple points. The first is that the author’s newly embraced measure of per capita health care spending is a poor measure of health care utilization, given its dependence on the age structure of the population and on indirect measures such as hospital business revenue from patient care and income-support payments to the disabled. Cooper appears to be unaware of these important limitations. Medicare data, which are used by the Dartmouth group, measure very accurately what doctors and hospitals do for their patients, allowing researchers to measure the association between the quan-

.....
Jon Skinner (jon.skinner@dartmouth.edu) is the John Sloan Dickey Third Century Professor in Economics in the Department of Economics, Dartmouth College, in Hanover, New Hampshire. Amitabh Chandra is a professor in the Kennedy School of Government, Harvard University, in Cambridge, Massachusetts. David Goodman is a professor of pediatrics and of community and family medicine at Dartmouth Medical School and associate director of the Center for Health Policy Research at Dartmouth. Elliott Fisher is director of the Center for Health Policy Research and a professor of medicine and of community and family medicine, Dartmouth Institute for Health Policy and Clinical Practice.

tity of health care and health outcomes.

Second, we demonstrate that even if one accepts these new data, the empirical analysis documented in Cooper's paper supports a conclusion entirely consistent with our own published research: What matters for health care quality is not traditional health care such as more hospital beds, intensive care unit (ICU) days, or specialists. Instead, quality appears to be associated with "social capital," or social networks resulting in well-functioning public and private organizations.² The policy implications are straightforward: the solution to poor quality is not to increase the supply of physicians or specialists or hospital beds, but instead to improve health care systems and incentives to ensure that existing physicians and hospitals provide the best possible quality at the lowest cost.

Measuring Per Capita Health Care Spending

Cooper relies on measures of per capita total health care spending developed by resourceful and imaginative researchers at the Centers for Medicare and Medicaid Services (CMS), doing the best they can with limited data sources.³ But it is important to understand exactly how limited these measures are. The Medicare data include detailed information about specific procedures, follow-up visits occurred, number of doctors who saw the patient, and whether they were specialists or primary care physicians. By contrast, the CMS research group is limited to one number in each year from every hospital and physician organization: business income based on self-reported survey data from the U.S. census, with supplemental information from the Internal Revenue Service. They further adjust as best they can for cross-border admissions and treatments (for example, when patients from Kentucky are admitted to a hospital in Cincinnati, Ohio).

There are two ways in which these state-level expenditures differ from the Medicare claims data. First, they include spending for the under-sixty-five population (and non-Medicare spending for the over-sixty-five);

second, they create estimates based on institution-level measures of business income, rather than patients' utilization.

We do not expect that the first explanation—including non-Medicare spending in the overall spending measure—should matter much for the ranking of state-level health care use and spending. Consider, for example, three reasons why Medicare spending and non-Medicare/under-sixty-five spending should be similar.

■ **Differences in cost of living.** Doctors and hospitals charge more in New York or Connecticut than in Oklahoma. Thus, non-Medicare spending in New York should be correlated with Medicare spending in New York simply because of differences in prices charged by providers.

■ **Illness levels.** On average, people in Louisiana experience worse health than people in Colorado.⁴ This is true for the under-sixty-five and over-sixty-five population. Simply on that basis, we would expect to find a correlation between state-level spending for the younger and older populations.

■ **Practice patterns.** Other studies have shown a correlation between the intensity of care for the over-sixty-five and the under-sixty-five populations within the same hospital, or between fee-for-service Medicare patients and nonelderly managed care patients.⁵ This makes sense—physicians in high-Medicare-spending regions recommend more aggressive treatments for identical (hypothetical) patients, and it seems likely that these practice styles would be reflected in the care for both a fifty-five-year-old and a seventy-year-old with congestive heart failure.⁶

Surprisingly, despite these three factors arguing for a strong positive correlation between CMS state-level non-Medicare spending estimates and actual Medicare spending data, the actual correlation is modest or nonexistent. Thus, one can conclude either that (1) the CMS data source is not a reliable indicator of age-sex-race-adjusted health care utilization, or that (2) state-level spending measures something entirely different from the health care utilization that has been the focus of

Dartmouth research—rates of surgery, physician visits, hospital days, screening and imaging rates, and so forth.

It is premature to conclude that item 1 is true; indeed, one might observe substitution between poor Medicaid reimbursements and a greater use of Medicare-funded services in response. But it is worth noting that the “total spending” measure reflects quite different factors from our measures of health care utilization in the Medicare claims data. First, it is virtually impossible to adjust for differences in age across states, given that one is working only with aggregated data from the hospital or physician practice. Older people account for far more health care than younger people, a point acknowledged in an earlier study by Anne Martin and colleagues.⁷ Thus, states with older populations should spend more; indeed, the correlation coefficient between

the median age of the state population and this new measure of total health care spending is positive: 0.45 ($p < 0.01$). Including median age in a regression of quality (from the Commonwealth Fund data) on total per capita spending (from Martin and colleagues) leads to a negation of Cooper’s basic result: there is no longer a significant correlation between quality and total spending ($p = 0.11$) after “adjusting” for age effects.⁸

Another problem in interpreting these estimated spending measures is that they include a wide variety of support programs, such as support for community centers and respite care for low-income elderly or disabled people, that we would not normally consider traditional health care. (This is one important reason why Maine ranks second among the fifty states in per capita total health care spending.) Also, these spending measures reflect higher reimbursement rates for the same procedure—or a “price” effect rather than a “quantity” effect. Unlike Medicare data, there is no way for Cooper to tease out price and quantity effects

in comparing cross-state data.

But let’s not reject the new data so quickly, and instead explore hypothesis 2 above—that the “total” per capita health care spending adopted by Cooper does in fact tell us something about the state, even if it has little to do with the use of health care per se.

What Can Be Learned From Cooper’s State-Level Correlations?

Recall that total health care spending is equal to price times quantity—or, in other words, how much is paid per procedure (a factor that varies with cost of living, profit margins, and local wages) times the number of procedures. Biomedical theories are based entirely on the medical procedure—does the stent improve functioning or survival—and not on how much the insurance company pays the hospital and cardiologist for placing the stent.

“States with high health care quality and high ‘total spending’ possess some third factor that is causal for both quality and generous reimbursements.”

Why, then, should the price of health care be associated with quality?

It seems improbable that spending more per stent, or providing more generous disability payments, directly “causes” an improvement in hospital quality measures. More likely, patterns of reimbursements, utilization, and the provision of quality are themselves generated by a third set of factors. Thus, states with high health care quality and high “total spending” (and modest Medicare utilization) possess some third factor that is causal for both quality and generous reimbursements. But what is this third factor?

In previous research, the Dartmouth group has suggested that “social capital” could play a role: the extent to which residents of a state participate in civic activities, are well-educated, trust others, or engage in philanthropic activities.⁹ These characteristics are also likely associated with what David Blumenthal and colleagues have referred to as “professionalism,” which would be associated with physicians’ providing high-quality care even in the

absence of explicit payments for such services.¹⁰

Evidence for this hypothesis can be seen from the very strong state-level correlation (0.68, $p < 0.01$) between a key measure of clinical quality—the use of beta-blockers at discharge for acute myocardial infarction (AMI) patients in 2000/01—and an index of social capital based on research by Robert Putnam.¹¹ States with high levels of social capital rank well along a number of dimensions, including life expectancy and the ability to adopt new innovations in fields outside of health care, such as tractors in the 1920s and hybrid corn in the 1940s.¹²

This “new view” provides a logical and cohesive way of interpreting Cooper’s ad hoc discussion of the state-level data. States with high levels of social capital tend to legislate more-comprehensive support programs for their residents and pay health care providers more, while the physicians who live in these high-social-capital states are more likely to adopt new and effective innovations rather than simply performing more tests and procedures with questionable medical efficacy. A correlation could well result between the provision of human services (including medical care) and medical quality, but this correlation tells us little about causality between total spending and better health quality, and nothing about the effects of more intensive health care on health outcomes.

Conclusion

Beginning with John Wennberg’s research in the 1970s, the Dartmouth research group has asked a basic and fundamental question: does more intensive health care spending yield better health outcomes? Most recently, a wide range of studies using detailed individual data on elderly patients with AMI, hip fractures, and colon cancer and in the general population have demonstrated that higher intensity of care does not improve health outcomes (and in fact might attenuate survival slightly); nor does it improve patients’ satisfaction or access to care.¹³

By ignoring three decades of research on the

vitality important issue of how to measure causal effects of health care utilization on health outcomes, Cooper may believe that he has challenged the basic scientific findings of the Dartmouth group—that greater health care intensity is not associated with better outcomes. But he has not. Instead, he has circuitously ended up providing another piece of evidence supporting the Dartmouth view by rediscovering the fundamental insight that what often leads to better health quality is superior “staffing decisions, the availability of information technology, preventive services, and other investments that contribute to quality and system performance.”¹⁴ Thus, improving health care quality is not a matter of simply adding more specialists or hospital beds; it is far more challenging and requires grappling with barriers to improving productivity in health care services—better quality at lower cost.

.....
This study was supported by Grant no. P01 AG19783 from the National Institute on Aging. The authors are grateful for invaluable assistance from Laura Yasaitis and Kathy Stroffolino.

NOTES

1. R.A. Cooper, “States with More Physicians Have Better-Quality Health Care,” *Health Affairs* 28, no. 1 (2009): w91–w102 (published online 4 December 2008; 10.1377/hlthaff.28.1.w91); R.A. Cooper, “States with More Health Care Spending Have Better-Quality Health Care: Lessons about Medicare,” *Health Affairs* 28, no. 1 (2009): w103–w115 (published online 4 December 2008; 10.1377/hlthaff.28.1.w103); and K. Baicker and A. Chandra, “Medicare Spending, the Physician Workforce, and Beneficiaries’ Quality of Care,” *Health Affairs* 23 (2004): w184–w197 (published online 7 April 2004; 10.1377/hlthaff.w4.184).
2. J. Skinner and D. Staiger, “Technological Diffusion from Hybrid Corn to Beta Blockers” in *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, ed. E. Berndt and C.M. Hulten (Chicago and Cambridge, Mass.: University of Chicago Press and National Bureau of Economic Research, 2007).
3. A.B. Martin et al., “Health Spending by State of Residence, 1991–2004,” *Health Affairs* 26, no. 6 (2007): w651–w663 (published online 18 September 2007; 10.1377/hlthaff.26.6.w651).

4. Commonwealth Fund, Table 6.1 (Healthy Living: Dimension Ranking), in "State Scorecard Data Tables, 2007," Supplement to *Aiming Higher: Results from a State Scorecard on Health System Performance* (New York: Commonwealth Fund, 2007).
5. M. Steifel, P. Feigenbaum, and E.S. Fisher, "The Dartmouth Atlas Applied to Kaiser Permanente: Analysis of Variation in Care at the End of Life," *Permanente Journal* 12, no. 1 (2008): 4–9; and L.C. Baker, E.S. Fisher, and J.E. Wennberg, "Variations in Hospital Resource Use for Medicare and Privately Insured Populations in California," *Health Affairs* 27, no. 2 (2008): w123–w134 (published online 12 February 2008; 10.1377/hlthaff.27.2.w123).
6. B. Sirovich et al., "Discretionary Decision Making by Primary Care Physicians and the Cost of U.S. Health Care," *Health Affairs* 27, no. 3 (2008): 813–823.
7. See the discussion on page 123 in A. Martin et al., "Health Care Spending during 1991–1998: A Fifty-State Review," *Health Affairs* 21, no. 4 (2002): 112–126.
8. We don't take this negation very seriously, however, as one should only trust individual-level adjustments for individual-level confounders. To do otherwise runs the risk of reporting spurious correlations. Dartmouth studies, by contrast, adjust for individual-level covariates such as age and race at the individual level, and socioeconomic status measures at either the individual or ZIP code level. Other variables, however, such as system-level effects (for example, a region may have a surplus of specialists or ICU beds) should be specified at the aggregated level of a region or hospital. Some of the confusion in Cooper's analysis arises from a misunderstanding of this distinction.
9. Skinner and Staiger, "Technological Diffusion." For the classic discussion of social capital, see R. Putnam, *Bowling Alone: America's Declining Social Capital* (New York: Simon and Schuster, 2000).
10. E.G. Campbell et al., "Professionalism in Medicine: Results of a National Survey of Physicians," *Annals of Internal Medicine* 147, no. 11 (2007): 795–802.
11. S.F. Jencks, E.D. Huff, and T. Cuerdon, "Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998–1999 to 2000–2001," *Journal of the American Medical Association* 289, no. 3 (2003): 305–312; and Skinner and Staiger, "Technological Diffusion." Because of a lack of data, Alaska, Hawaii, and the District of Columbia were excluded. There are newer data series on social capital; see <http://www.bowlingalone.com/data.htm> (accessed 27 October 2008).
12. Skinner and Staiger, "Technological Diffusion."
13. These studies have relied on hierarchical modeling methods to adjust for comorbidities at the individual level as well as "system" effects at the hospital or Hospital Referral Region (HRR) level. See E.S. Fisher et al., "The Implications of Regional Variations in Medicare Spending, Part 1: The Content, Quality, and Accessibility of Care," *Annals of Internal Medicine* 138, no. 4 (2003): 273–287; and E.S. Fisher et al., "The Implications of Regional Variations in Medicare Spending, Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine* 138, no. 4 (2003): 288–298. Additionally, semi-parametric instrumental variables techniques, analyses focusing on changes over time in region-specific spending and survival rates for AMI patients, and studies of low-birthweight babies yield similar conclusions. See also J. Skinner, E. Fisher, and J. Wennberg, "The Efficiency of Medicare," in *Analyses in the Economics of Aging*, ed. D. Wise (Chicago and Cambridge, Mass.: University of Chicago Press and NBER, 2005), 129–157; J.S. Skinner, D.O. Staiger, and E.S. Fisher, "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction," *Health Affairs* 25, no. 2 (2006): w34–w47 (published online 7 February 2006; 10.1377/hlthaff.25.w34); and D. Goodman et al., "The Relation between the Availability of Neonatal Intensive Care and Neonatal Mortality," *New England Journal of Medicine* 346, no. 20 (2002): 1538–1544.
14. Cooper, "States with More Health Care Spending," III.