

MARKET WATCH

Discretionary Decision Making By Primary Care Physicians And The Cost Of U.S. Health Care

Clinical judgment, not clinical guidelines, should be the focus of policy efforts to improve the quality of care and address disparities in spending.

by **Brenda Sirovich, Patricia M. Gallagher, David E. Wennberg, and Elliott S. Fisher**

ABSTRACT: Efforts to improve the quality and costs of U.S. health care have focused largely on fostering physician adherence to evidence-based guidelines, ignoring the role of clinical judgment in more discretionary settings. We surveyed primary care physicians to assess variability in discretionary decision making and evaluate its relationship to the cost of health care. Physicians in high-spending regions see patients back more frequently and are more likely to recommend screening tests of unproven benefit and discretionary interventions compared with physicians in low-spending regions; however, both appear equally likely to recommend guideline-supported interventions. Greater attention should be paid to the local factors that influence physicians' clinical judgment in discretionary settings. [*Health Affairs* 27, no. 3 (2008): 813–823; 10.1377/hlthaff.27.3.813]

HEALTH CARE SPENDING in the United States is the highest in the world and continues to grow at a rate of 7 percent per year.¹ Such liberal health care spending fails, however, to provide the country with the best health in the world.² Even within the United States—where per capita spending varies more than twofold between the lowest- and highest-spending regions—higher spending appears to result, if anything, in slightly lower quality and worse outcomes.³ U.S. regions with the highest spending levels do not achieve lower mortality, nor

do they show greater improvements in mortality over time.⁴ Higher spending is also not associated with better access to care, patient satisfaction, or physicians' ability to provide high-quality care.⁵ These findings underscore the serious problem of wasteful—and possibly harmful—overuse within the U.S. health care system.

Efforts to improve both the quality and the cost of U.S. health care have focused largely on fostering physician adherence to evidence-based clinical practice guidelines and reducing frank medical errors.⁶ These approaches are

.....
Brenda Sirovich (Brenda.Sirovich@dartmouth.edu) is a research associate in the Outcomes Group, Veterans Affairs Medical Center, in White River Junction, Vermont, and an assistant professor of medicine at Dartmouth Medical School, in Hanover, New Hampshire. Patricia Gallagher is a senior research fellow at the Center for Survey Research, University of Massachusetts Boston. David Wennberg is president and chief operating officer of Health Dialog Analytic Solutions, in Portland, Maine. Elliott Fisher is director of the Center for Health Policy Research, Dartmouth Institute for Health Policy and Clinical Practice, in Hanover, New Hampshire.

useful when it is possible to precisely define and reliably measure the correct action in a specific clinical situation. However, such explicit approaches are not applicable to the many discretionary decisions that physicians face. Of 2,500 treatments for a variety of conditions reviewed by *BMJ Clinical Evidence*, more than half fell into this gray zone.⁷ Current approaches also make it difficult to measure overuse of clinical services; very few of the explicit measures available today focus on overuse, including fewer than 10 percent of the 439 quality indicators in the RAND Quality Tools measurement set.⁸

To the extent that higher utilization rates depend on physicians' discretionary clinical decisions—for which the evidence does not point clearly to a right answer—current quality improvement and performance measurement initiatives are unlikely to address rising spending. To examine whether alternative approaches are therefore necessary, we investigated the relationship between health care use and primary care physicians' discretionary decision making about medical interventions, by comparing physicians across different areas of the country. Instead of comparing actual practice—which might unfairly contrast physicians caring for widely varying panels of patients—we studied discretionary decision making via physicians' responses to identical hypothetical patients presented as part of a national physician survey. Previous work suggesting a direct relationship between utilization (spending) and physicians' tendency to intervene was limited to six discrete questions representing simplified discretionary decisions.⁹ The current survey was developed specifically to reflect the complex array of decisions faced daily in primary care practice and to include decisions about both evidence-based and discretionary interventions.

Study Data And Methods

We conducted a mail survey of primary care physicians to examine the tendency of physicians practicing in regions with different levels of health care spending to intervene (to order tests, referrals, or treatment) in specific

clinical situations. This project was approved by the Institutional Review Boards at Dartmouth Medical School and the University of Massachusetts.

■ **Survey development.** To learn how physicians make decisions in the primary care setting and to pilot specific clinical vignettes, we conducted focus groups with primary care physicians in two cities. The focus groups and all subsequent survey development were done in collaboration with the Center for Survey Research, a professional survey research firm affiliated with the University of Massachusetts Boston. Focus groups concentrated on development and wording of realistic clinical vignettes that provided adequate detail and described a patient about whom physicians might disagree on management, and response categories that included an appropriate range of options, of which any number (including none) might be chosen at the time of the visit described.

The draft survey instrument was revised based on the results of cognitive interviews, conducted to ensure that the questions were well understood and the answers meaningful. The final survey consisted of questions about the physician (for example, board certification), his or her practice (for example, setting), and his or her clinical care (including routine follow-up, cancer screening, and evaluation and management decisions).

■ **Survey administration.** *Sample selection.* Using the Masterfiles of the American Medical Association and American Osteopathic Association, we obtained a random sample of primary care physicians (self-identified as family practice, general practice, or internal medicine) practicing at least twenty hours per week in the United States. Residents and retired physicians were ineligible.

Using computer-assisted telephone interviewing (CATI) software, trained professional phone interviewers checked to ensure that each sampled physician met eligibility requirements. A maximum of three call attempts were made during daytime and evening hours, weekdays and weekends, to try to speak to the sampled physician or an informant (such as a

receptionist). During the eight-week verification period, we identified 1,419 eligible physicians out of an original sample of 1,775, of whom 1,333 were randomly selected to receive the survey.

Mail protocols. Each potential participant was sent an initial questionnaire packet with a letter explaining the study, a \$20 cash incentive, a survey instrument, and a postage-paid return envelope. Two weeks after the initial mailing, all physicians who had not responded were sent another questionnaire packet, absent the cash incentive. Of 1,333 physicians who were mailed the initial survey, 58 were found ineligible; of the remaining 1,275, 801 responded (response rate 63 percent). Nonresponders did not differ from responders in terms of sex, primary specialty, practice type, or years in practice.

■ **Measures.** *Local spending.* For our measure of local health care spending, we used the End-of-Life Expenditure Index (EOL-EI), a measure based on Medicare expenditures in the last six months of life that we have described in detail elsewhere.¹⁰ The advantage of using this index to compare health care spending across different U.S. regions is that it is closely correlated with overall spending but unrelated to illness—because all patients included in calculating the measure have a life expectancy of six months.¹¹ The greater-than-twofold differences in EOL-EI across U.S. regions are not related to patients’ preferences or illness levels.¹² Mean EOL-EI was calculated for each of 306 U.S. Hospital Referral Regions (HRRs), which were then grouped into quintiles (of equivalent population size); mean

EOL-EI for each quintile ranged from \$11,347 to \$17,809 (see Exhibit 1 for selected HRRs within each quintile). Each physician was located within an HRR and a quintile based on his or her practice address. We display results using three categories for spending: low spending (the lowest quintile of EOL-EI), moderate spending (the middle three quintiles combined), and high spending (the highest quintile).

Physician practice intensity. Physician practice intensity was measured using physicians’ responses to three types of survey questions: questions about routine follow-up intervals; questions about whether or not the physician routinely recommends screening patients for each of three cancers (mammography, for which evidence-based recommendations support screening; prostate-specific antigen, or PSA, for which recommendations are equivocal; and spiral computed tomography, or CT, for which recommendations do not support routine screening); and clinical vignettes in which the physician was asked how often he or she would arrange for specific interventions (such as test, referral, or hospitalization) for patients with common clinical conditions. We analyzed dichotomized responses to all questions. For follow-up interval, we used a cut-off of three months. For clinical vignettes, we combined the responses “always or almost always” and “most of the time,” compared to the grouping of “some of the time,” “rarely,” and “never.” One vignette option was excluded because of ambiguous wording.

■ **Analysis.** We examined the relationship between local spending (EOL-EI) and physi-

**EXHIBIT 1
Selected Hospital Referral Regions (HRRs) In Each Quintile Of End-Of-Life Spending**

Quintile				
Lowest	2	3	4	Highest
Boise, ID	Amarillo, TX	Hartford, CT	Ann Arbor, MI	Las Vegas, NV
Sacramento, CA	Indianapolis, IN	Toledo, OH	Lexington, KY	Monroe, LA
Rochester, NY	Wichita, KS	Savannah, GA	Wilmington, DE	New York, NY
Asheville, NC	Cincinnati, OH	Sun City, AZ	Ventura, CA	Miami, FL

SOURCE: Authors’ calculations using Medicare claims data.

cian practice intensity using individual item responses as measures of practice intensity. Tests for trend were based on logistic regression in which the individual physician's (dichotomized) response was the dependent variable and the independent variable was spending in the physician's region, expressed as a continuous variable.

■ **Overall practice intensity.** We used factor analysis to summarize patterns of correlations among response variables and explore the possibility of collapsing a number of the observed variables into a factor or factors representing practice intensity.¹³ Factor analysis identified a single factor that appeared to measure physician practice intensity; using this factor (which weights the different component variables according to their apparent intensity), a summary intensity score was derived for each physician ($n = 593$) who responded to all questions included in the factor analysis. For analyses examining the relationship between local spending and summary intensity score, we used linear regression in which the physician's intensity score was the dependent variable, and the independent variable was spending in the physicians region (continuous).

We also conducted a multivariable analysis to control for other personal (age, sex, race), professional (specialty, U.S. medical graduate, board certification), and practice-level (set-

ting, number of managed care contracts, proportion of capitated patients) factors that could influence practice intensity.

An alternative approach to deriving summary intensity scores—using weights representing the intensity of each vignette option, derived by modified Delphi technique—yielded results that were nearly identical.¹⁴ All analyses were carried out in STATA 9.1.

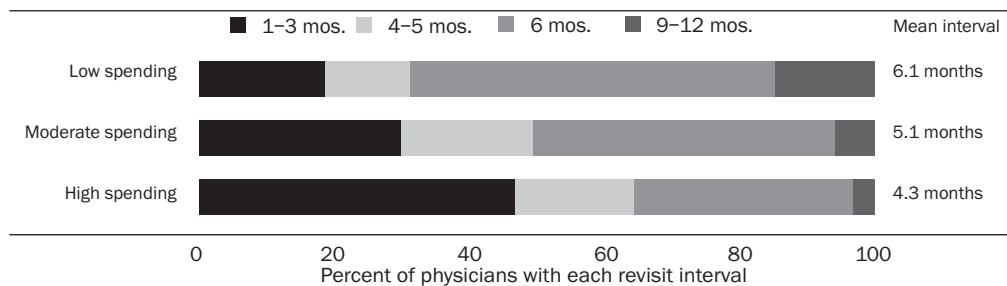
Study Results

■ **Respondents' characteristics.** Respondents were primarily male (75 percent) and white (74 percent) and had been in practice a median of twenty-one years. There were somewhat more internists (52 percent) than family practitioners (45 percent). Nearly all (86 percent) were board certified in their specialty; 23 percent had attended medical school outside the United States and Canada.

■ **Routine care.** Local spending level was a strong predictor of routine follow-up interval for patients with well-controlled hypertension. In high-spending regions, 47 percent of physicians schedule hypertensive patients every three months or more often, while only 19 percent of physicians in low-spending regions do so ($p_{\text{trend}} < 0.001$; Exhibit 2). Annual follow-up was almost nonexistent in high-spending regions.

■ **Routine screening for cancer.** Physicians were asked whether they routinely recommend three cancer screening tests to pa-

EXHIBIT 2
Routine Revisit Intervals For Patients With Well-Controlled Hypertension, Reported By Physicians In Regions With Low, Moderate, And High Spending

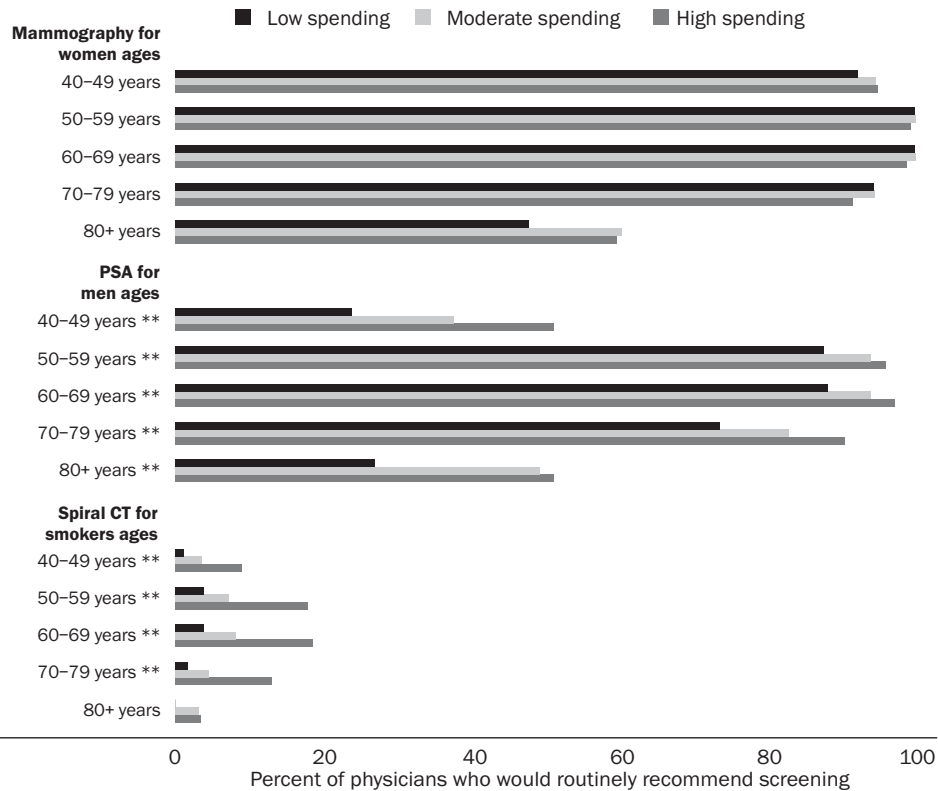


SOURCE: Authors' calculations using survey data and Medicare claims data.
NOTES: Low spending includes the quintile of lowest per capita Medicare spending (based on the 2003 End-of-Life Expenditure Index); moderate spending includes the three intermediate quartiles; and high spending includes the quintile of highest spending.

tients in different age groups (Exhibit 3). Physicians were equally likely to recommend mammographic screening regardless of where they practiced. PSA screening, on the other hand, was more likely to be recommended by physicians in high- compared with low-spending regions for all men age forty and older. The disparity was largest for men age eighty and older. Few physicians recommended routine spiral CT screening for lung cancer, but those in high-spending regions were more likely than others to recommend such screening (for smokers ages 40–79).

■ **Evaluation and management.** Typical GERD. When seeing a seventy-five-year-old woman with typical symptoms of gastroesophageal reflux disease (GERD) (Exhibit 4), similar proportions of physicians in high- and low-spending areas would order a number of interventions, from *Helicobacter pylori* testing to prescription of a proton pump inhibitor (PPI). However, physicians in high-spending areas were more likely than others to refer the patient directly for upper gastrointestinal (GI) endoscopy and much more likely to refer the patient to a gastroenterologist for further

EXHIBIT 3
Tendency To Screen For Various Cancers Among Physicians In Regions With Low, Moderate, And High Spending, By Patient Age



SOURCE: Authors' calculations using survey data and Medicare claims data.
NOTES: Physicians were asked about mammography for women at average risk of breast cancer, prostate-specific antigen (PSA) testing for men at average risk of prostate cancer, and spiral computed tomography (CT) screening for smokers, in each of the age categories shown. Low spending includes the quintile of lowest per capita Medicare spending (based on the 2003 End-of-Life Expenditure Index); moderate spending includes the three intermediate quartiles; and high spending includes the quintile of highest spending. Statistical significance denotes test for trend, based on logistic regression in which the physician's tendency to screen (yes/no) was the dependent variable and the independent variable was spending in the physician's region, expressed as a continuous variable.
 ** p < 0.05

EXHIBIT 4
Percentage Of Physicians Who Would Order Each Intervention “Always Or Almost Always” Or “Most Of The Time” For The Patient Specified In The Clinical Vignette, According To Spending Level In The Physician’s Region (HRR)

Spending in area	Interventions						
	More common in low-spending areas	Equally common across spending levels			More common in high-spending areas		
Vignette 1: 75-year-old woman presents with typical heartburn symptoms							
	H. pylori testing	Barium swallow	Behavior change	Pre-scribe PPI	Refer to GI for EGD only	Refer to GI for further management	Pre-scribe H2 blocker
Low	40%	13%	77%	89%	27%	7%	19%
Moderate	43	19	67	88	29	22	20
High	42	17	68	86	41	36	28
Vignette 2: 75-year-old man presents with chest pain on heavy exertion							
	Standard ETT	Stress imaging study	Curbside consult	Echo-cardiogram	Refer to cardi-ology	Admit to hospital	
Low	35	48	8	15	34	18	
Moderate	29	55	8	28	37	17	
High	24	59	10	40	45	26	
Vignette 3: 85-year-old man presents with an exacerbation of end-stage CHF							
	Discuss palliative care	Discharge patient home		Admit (non-ICU bed)	Admit (ICU bed)	PA cath-eter	Biven-tricular pacer
Low	65%	21		41	11	0	1
Moderate	48	22		48	22	3	2
High	46	15		57	35	6	3

SOURCE: Authors’ calculations using survey data and Medicare claims data.

NOTES: Low spending represents physicians practicing in Hospital Referral Regions (HRRs) in the lowest quintile of Medicare spending (measured by the End-of-Life Expenditure Index); moderate spending, in one of the three intermediate quintiles; and high spending, in the quintile of highest spending. Differences shown signify $p < 0.05$ for the logistic regression in which the physician’s response is the dependent variable and local spending the independent variable (continuous). In Vignette 2, standard exercise tolerance testing (ETT) reflects the proportion of physicians who would order a standard ETT but not a stress imaging study. PPI is proton pump inhibitor. GI is gastrointestinal. EGD is esophagogastroduodenoscopy. CHF is congestive heart failure. ICU is intensive care unit. PA is pulmonary artery. Full text of clinical vignettes is available in online Appendix B, <http://content.healthaffairs.org/cgi/content/full/27/3/813/DC1>.

management.

New-onset chest pain. For a seventy-five-year-old man with the new onset of chest pressure occurring upon heavy exertion (Exhibit 4), ordering patterns were similar across different spending levels for some interventions (stress testing and curbside consultation with a cardiologist). However, physicians in higher-spending areas were more likely than others to

order an echocardiogram, refer to a cardiologist, and admit the patient to the hospital.

End-stage congestive heart failure. For an eighty-five-year-old man with an exacerbation of end-stage (Class IV) congestive heart failure (CHF), local spending level was also predictive of the aggressiveness of a physician’s approach (Exhibit 4). Physicians in high-spending areas were more likely than others to admit

the patient to an acute medicine floor and much more likely to admit the patient to an intensive care unit. They were less likely to discuss palliative care with the patient.

■ **Overall practice intensity.** Summary practice intensity, derived from factor analysis that included all clinical decision-making variables, was significantly correlated with local health care spending at the individual physician level ($r = 0.22, p < 0.001$).¹⁵ When other factors at the physician and practice level that might be expected to influence practice intensity were controlled for, the correlation between local spending and physician intensity persisted ($r = 0.15, p = 0.001$).

When physicians were grouped into deciles of HRR-level spending (Exhibit 5), there was an extremely strong association between local spending and (factor-derived) summary intensity score ($r = 0.94, p < 0.001$).

Discussion

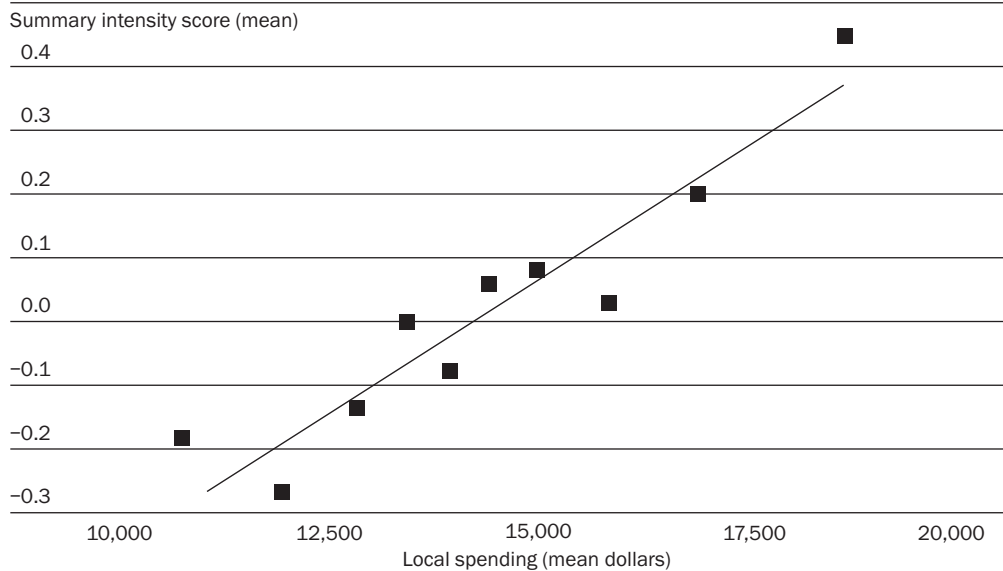
We found that widely varying levels of health care spending across the United States

are strongly correlated with the tendency of local physicians to recommend discretionary interventions. Primary care physicians in high-spending regions reported seeing patients back more frequently, recommending more screening tests of uncertain benefit, and opting for more-resource-intensive interventions than those practicing in low-spending regions. Compared with physicians practicing in the lowest quintile of spending, those in the highest quintile would recommend an additional eighty hypertension follow-up visits per year, fourteen spiral CT scans, twenty-five echocardiograms, twenty-four cardiac care unit admissions, and twenty-nine gastroenterology referrals (per 100 patients in each clinical category). In contrast, physicians in high- and low-spending regions are equally likely to recommend guideline-supported interventions.

■ **Study limitations.** Our study has several limitations. First, because we measured our exposure (local spending) using data on Medicare patients age sixty-five and older,

EXHIBIT 5

Association Between Physician Practice Intensity And Local Health Care Spending



SOURCE: Authors' calculations using survey data and Medicare claims data.

NOTES: Data are aggregated by decile of health care spending (End-of-Life Expenditure Index, or EOLEI). Physician practice intensity represents the mean factor-derived summary intensity score for physicians in Hospital Referral Regions (HRRs) within each decile of EOLEI. Data include only the 593 physicians who responded to all questions included in the factor analysis. Local health care spending represents mean local spending (EOLEI) for physicians in HRRs within each decile of EOLEI. $r = 0.94$.

some might be concerned that it does not represent overall U.S. health care spending (including on younger people). However, the correlation between practice patterns for U.S. populations under and over age sixty-five has been shown to be high.¹⁶ Similarly, state-level Medicare spending is closely correlated with overall per capita spending.¹⁷

Second, because we used clinical vignette responses to measure physicians' tendency to intervene, we cannot be certain that our outcome measure (practice intensity) accurately reflects physicians' ordering habits for actual patients. Comparing practice intensity based on physicians' actual practice, however, would introduce greater difficulties—because individual patients, and panels of patients, differ from provider to provider. Because clinical vignettes allow for each provider to manage the same patient, they provide a measure of practice intensity that is inherently case-mix-adjusted. For the present study, the survey methodology was revised considerably compared with previous work, to enrich clinical detail and provide a range of possible intervention options for each vignette that closely approximates the choices available in clinical practice.¹⁸ Hypothetical patient scenarios are used for other important provider comparisons (such as board-certifying examinations) and have been established as a good measure of providers' behavior with standardized patients.¹⁹

Finally, this paper does not address how individual physicians' attributes affect each of the outcome variables, or the effect of adjusting for physician-level variables on the relationship between spending and each individual measure of physician practice intensity. Addressing these questions is beyond the scope of the present work. Overall practice intensity, however, remained strongly correlated with health care spending even after all covariates were adjusted for.

“When physicians in different areas see the ‘same’ patients, they make different decisions—decisions that are strongly correlated with the local practice environment.”

■ **Geographic practice variation.** This study expands upon earlier work suggesting that physicians in different geographic regions practice differently. Physician decision making about breast cancer surgery; cardiac catheterization; and hospitalizing patients with asthma, diabetes, and CHF varies markedly across regions.²⁰ The tendency to intervene shows even more dramatic variation in international comparisons.²¹ Attempts to correlate local practice patterns with area-level variables such as physician supply, specialist supply, and hospitalization rates have produced mixed results.²² Recent work showed that physicians' tendency to intervene was strongly associated with health care spending in a physician's region but did not attempt to distinguish discretionary from guideline-directed interventions and was limited to six simplified scenarios.²³

■ **Relation of practice and local spending.** Our findings reveal that responses to a few clinical questions posed to primary care physicians are strongly correlated with the level of local spending in the Medicare population. On the face of it, this might not seem surprising. If utilization is higher in certain areas, many would assume that physician behavior is the cause. There are, however, other potential explanations. Higher levels of patient illness or demand could explain higher utilization. Or the fact that high-spending areas have many more physicians per capita—each of whom might act exactly the same regardless of where they practice—could explain spending differences. We have shown, however, that when physicians in different areas see the “same” patients, they make different decisions—decisions that are strongly correlated with the local practice environment (such as spending). Furthermore, physicians in regions of differing spending appear to differ only in their discretionary decision making. For the decisions we examined that are informed by evidence or

practice guidelines (such as screening mammography and standard exercise tolerance testing), physicians were equally likely to recommend interventions regardless of local spending levels.²⁴

Although we have shown that physicians who practice in areas of higher local health care spending are more prone to intervene in discretionary situations, we are unable to distinguish between a number of factors that could contribute to the development of these differing practice patterns. On the one hand, more-aggressive physicians might selectively choose to practice in certain areas. On the other, physicians might adapt to the practice style of the community in which they settle—a phenomenon that could have a number of explanations. First, the malpractice climate in high-spending (versus low-spending) regions might exert stronger effects on physicians' practice of defensive medicine, although other work suggests that this might be a small effect.²⁵ Second, differences in patients' expectations and demands across different regions might directly influence physicians' decision making (although in data not presented, there were no differences across regions in the proportion of physicians who cite patient pressures as an impetus to intervene). Third, it is possible, indeed likely, that physicians in the U.S. fee-for-service environment adapt their practice style to maintain their incomes in response to local market forces—including the local supply of services and physicians.²⁶ In higher-spending regions, characterized by a much greater supply of physicians per capita, physicians may (consciously or unconsciously) shorten their revisit intervals to keep their schedules full—and vice versa in lower-spending areas—and might modify their referral patterns based on the availability of specialists.²⁷ Hospitals and medical practices in high-spending regions might also feel more competitive stresses and be more likely to exert pressure on physicians to order profitable services (such as high-tech diagnostic imaging) for their patients; conversely, lower availability in low-spending regions could dissuade physicians from ordering such services. Finally, it is

very plausible that some or all of these factors could interact to create a culture of high-intensity specialty-oriented (or low-intensity primary care-oriented) practice, which then becomes the local “standard of care” and may thus be difficult for either patients or providers to resist.

■ Policy implications. We know that higher-spending regions do not achieve better health outcomes and that physicians practicing in these regions perceive greater difficulty providing high-quality care. Current policy efforts to improve the quality of care and address disparities in spending have focused largely on fostering adherence to clinical guidelines. This study suggests that greater attention to clinical judgment—and to the local factors that are likely to influence physician practice—will be required.

.....
An abstract based on this manuscript was presented at the 2005 annual meeting of the Society of General Internal Medicine in New Orleans, Louisiana, 11–14 May. Brenda Sirovich was supported by a Veterans Affairs Career Development Award in Health Services Research and Development at the time this work was initiated. This study was supported by the National Institute on Aging (Grant no. P01 AG19783) and the National Heart, Lung, and Blood Institute (Grant no. R01HL080437). Financial support was also provided by Research Enhancement Award no. 03-098 from the Department of Veterans Affairs and a grant from the Robert Wood Johnson Foundation. Sirovich had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The funders of this work had no role in the design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. None of the authors has a financial conflict of interest related to the paper. The views expressed herein do not necessarily represent the views of the Department of Veterans Affairs or the United States government.

NOTES

1. A. Catlin et al., "National Health Spending in 2006: A Year of Change for Prescription Drugs," *Health Affairs* 27, no. 1 (2008): 14–29.
2. J. Banks et al., "Disease and Disadvantage in the United States and in England," *Journal of the American Medical Association* 295, no. 17 (2006): 2037–2045; D.L. Hoyert et al., "Annual Summary of Vital Statistics: 2004," *Pediatrics* 117, no. 1 (2006): 168–183; and J.E. Weiss and M. Mushinski, "International Mortality Rates and Life Expectancy: Selected Countries," *Statistical Bulletin—Metropolitan Life Insurance Company* 80, no. 1 (1999): 13–21.
3. 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. The same is true when comparing leading academic medical centers. See J.E. Wennberg et al., "Use of Hospitals, Physician Visits, and Hospice Care during the Last Six Months of Life among Cohorts Loyal to Highly Respected Hospitals in the United States," *British Medical Journal* 328, no. 7440 (2004): 607; and E.S. Fisher et al., "Variations in the Longitudinal Efficiency of Academic Medical Centers," *Health Affairs* 23 (2004): var19–var32 (published online 7 October 2004; 10.1377/hlthaff.var.19).
4. Fisher et al., "The Implications of Regional Variations, Part 2"; and 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 (2006): w34–w47 (published online 7 February 2006; 10.1377/hlthaff.25.w34).
5. Fisher et al., "The Implications of Regional Variations, Part 1"; Fisher et al., "The Implications of Regional Variations, Part 2"; and B.E. Sirovich et al., "Regional Variations in Health Care Intensity and Physician Perceptions of Quality of Care," *Annals of Internal Medicine* 144, no. 9 (2006): 641–649.
6. Institute of Medicine, *Medicare's Quality Improvement Organization Program: Maximizing Potential* (Washington: National Academies Press, 2006).
7. "How Much Do We Know?" *BMJ Clinical Evidence*, 2006, <http://www.clinicalevidence.com/ceweb/about/knowledge.jsp> (accessed 29 January 2008).
8. E.A. McGlynn et al., "The Quality of Health Care Delivered to Adults in the United States," *New England Journal of Medicine* 248, no. 26 (2003): 2635–2645.
9. B.E. Sirovich et al., "Variation in the Tendency of Primary Care Physicians to Intervene," *Archives of Internal Medicine* 165, no. 19 (2005): 2252–2256.
10. *Ibid.*; and Sirovich et al., "Regional Variations."
11. Fisher et al., "The Implications of Regional Variations, Part 1"; and Fisher et al., "The Implications of Regional Variations, Part 2."
12. R.S. Pritchard et al., "Influence of Patient Preferences and Local Health System Characteristics on the Place of Death: Study to Understand Prognoses and Preferences for Risks and Outcomes of Treatment (SUPPORT)," *Journal of the American Geriatrics Society* 46, no. 10 (1998): 1242–1250; and J.E. Wennberg, E.S. Fisher, and J.S. Skinner, "Geography and the Debate over Medicare Reform," *Health Affairs* 21 (2002): w96–w114 (published online 13 February 2002; 10.1377/hlthaff.21.w96).
13. Results are available in online Appendix A at <http://content.healthaffairs.org/cgi/content/full/27/3/813/DC1>.
14. *Ibid.*
15. *Ibid.*
16. J.E. Wennberg and M.M. Cooper, eds., *The Dartmouth Atlas of Health Care in Michigan* (Detroit: Blue Cross/Blue Shield of Michigan, 2000); and J.E. Wennberg and M.M. Cooper, eds., *The Dartmouth Atlas of Health Care in Pennsylvania* (Chicago: American Hospital Publishing, 1998).
17. A.B. Martin, L.S. Whittle, and K.R. Levit, "Trends in State Health Care Expenditures and Funding: 1980–1998," *Health Care Financing Review* 22, no. 4 (2001): 111–140.
18. Sirovich et al., "Variation in the Tendency"; and F. Potter et al., "Community Tracking Study Physician Survey, Round 2," Survey Methodology Report (Washington: Center for Studying Health System Change, 2001).
19. Regarding board certification, see J.J. Norcini et al., "The Relationship between Features of Residency Training and ABIM Certifying Examination Performance," *Journal of General Internal Medicine* 2, no. 5 (1987): 330–336; and J.J. Norcini et al., "An Analysis of the Knowledge Base of Practicing Internists as Measured by the 1980 Recertification Examination," *Annals of Internal Medicine* 102, no. 3 (1985): 385–389. Regarding standardized patients, see J.W. Peabody et al., "Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of Three Methods for Measuring Quality," *Journal of the American Medical Association* 283, no. 13 (2000): 1715–1722; T.R. Dresselhaus et al., "An Evaluation of Vignettes for Predicting Variation in the Quality of Preventive Care," *Journal of General Internal Medicine* 19, no. 10 (2004): 1013–1018; and J.W. Peabody et al., "Measuring the Quality of Physician

- Practice by Using Clinical Vignettes: A Prospective Validation Study," *Annals of Internal Medicine* 141, no. 10 (2004): 771-780.
20. J.S. Mandelblatt et al., "Measuring and Predicting Surgeons' Practice Styles for Breast Cancer Treatment in Older Women," *Medical Care* 39, no. 3 (2001): 228-242; D.E. Wennberg et al., "Do Physicians Do What They Say? The Inclination to Test and Its Association with Coronary Angiography Rates," *Journal of General Internal Medicine* 12, no. 3 (1997): 172-176; and M. Komaromy et al., "Physician Practice Style and Rates of Hospitalization for Chronic Medical Conditions," *Medical Care* 34, no. 6 (1996): 594-609.
 21. C.B. Forrest et al., "Referral of Children to Specialists in the United States and the United Kingdom," *Archives of Pediatric and Adolescent Medicine* 157, no. 3 (2003): 279-285; and C.M. Koeck et al., "Using a Hypothetical Case to Measure Differences in Treatment Aggressiveness among Physicians in Canada, Germany, and the United States," *Wiener Klinische Wochenschrift* 110, no. 22 (1998): 783-788.
 22. D. Hemenway and D. Fallon, "Testing for Physician-Induced Demand with Hypothetical Cases," *Medical Care* 23, no. 4 (1985): 344-349; M. Roland and R. Morris, "Are Referrals by General Practitioners Influenced by the Availability of Consultants?" *British Medical Journal* 297, no. 6648 (1988): 599-600; and Komaromy et al., "Physician Practice."
 23. Sirovich et al., "Variation in the Tendency."
 24. Regarding screening mammography, see U.S. Preventive Services Task Force, *Screening for Breast Cancer: Recommendations and Rationale*, 3 September 2002, <http://www.ahrq.gov/clinic/3rduspstf/breastcancer/brcanrr.pdf> (accessed 29 January 2008); and R.A. Smith, V. Cokkinides, and H.J. Eyre, "American Cancer Society Guidelines for the Early Detection of Cancer, 2006," *CA—A Journal for Clinicians* 56, no. 1 (2006): 11-25. Regarding exercise tolerance testing, see V. Snow et al., "Evaluation of Primary Care Patients with Chronic Stable Angina: Guidelines from the American College of Physicians," *Annals of Internal Medicine* 141, no. 1 (2004) 157-164; and R.J. Gibbons et al., "ACC/AHA 2002 Guideline Update for the Management of Patients with Chronic Stable Angina: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1999 Guidelines for the Management of Patients with Chronic Stable Angina)" (Bethesda, Md., and Dallas: American College of Cardiology and American Heart Association, 2002).
 25. J.G. Elmore et al., "Does Litigation Influence Medical Practice? The Influence of Community Radiologists' Medical Malpractice Perceptions and Experience on Screening Mammography," *Radiology* 236, no. 1 (2005): 37-46; and L.M. Baldwin et al., "Defensive Medicine and Obstetrics," *Journal of the American Medical Association* 274, no. 20 (1995): 1606-1610.
 26. B.F. Kiker and M. Zeh, "Relative Income Expectations, Expected Malpractice Premium Costs, and Other Determinants of Physician Specialty Choice," *Journal of Health and Social Behavior* 39, no. 2 (1998): 152-167; and M.P. Rosenthal et al., "Influence of Income, Hours Worked, and Loan Repayment on Medical Students' Decision to Pursue a Primary Care Career," *Journal of the American Medical Association* 271, no. 12 (1994): 914-917.
 27. Sirovich et al., "Regional Variations."