Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending[†]

By DAVID CUTLER, JONATHAN S. SKINNER, ARIEL DORA STERN,
AND DAVID WENNBERG*

There is considerable controversy about the causes of regional variations in health care expenditures. Using vignettes from patient and physician surveys linked to fee-for-service Medicare expenditures, this study asks whether patient demand-side factors or physician supply-side factors explain these variations. The results indicate that patient demand is relatively unimportant in explaining variations. Physician organizational factors matter, but the most important factor is physician beliefs about treatment. In Medicare, we estimate that 35 percent of spending for end-of-life care and 12 percent of spending for heart attack patients (and for all enrollees) is associated with physician beliefs unsupported by clinical evidence. (*JEL* D83, H75, I11, I18)

Regional variations in rates of medical treatments are large in the United States and other countries (Skinner 2011). In the US fee-for-service Medicare population, price-adjusted, per enrollee 2014 spending varies nearly two-fold across areas. Similar price-adjusted variations are found in the under-65 population (Cooper et al. 2015), and do not appear to be explained by patient illness or poverty (Newhouse et al. 2013).

What drives such variation in treatment and spending? One possibility is patient demand. Most studies of variations have been conducted in settings where all patients have similar and generous insurance policies, so price differences and income effects are likely to be small. Still, heterogeneity in patient preferences for care may play a role. In very acute situations, some patients may prefer to try all

possible measures to (potentially) prolong life, while others may prefer palliation and an out-of-hospital death. If people who value and demand life-prolonging treatments tend to live in the same areas, patient preference heterogeneity could lead to regional variation in equilibrium outcomes (Anthony et al. 2009; Baker, Bundorf, and Kessler 2014; Mandelblatt et al. 2012).

Another possible source of variation arises from the supply side. In "supplier-induced demand," a health care provider shifts a patient's demand curve beyond the level of care that the fully informed patient would otherwise want. This would be true in a principal-agent framework (McGuire and Pauly 1991) if prices are high enough or income is scarce. While physician utilization has been shown to be sensitive to prices (Jacobson et al. 2006; Clemens and Gottlieb 2014), it would be difficult to explain the magnitude of US Medicare spending variations using prices alone, since federal reimbursement rates do not vary greatly across areas.

Variations in supply could also occur if physicians respond to organizational pressure or peer pressure to treat patients more intensively, or maintain differing beliefs about appropriate treatments—particularly for conditions where there are few professional guidelines (Wennberg, Barnes, and Zubkoff 1982). If this variation is spatially correlated—for example, if physicians with more intensive treatment preferences hire other physicians with similar practice styles—the resulting regional differences in beliefs could explain regional variations in equilibrium spending.

It has proven difficult to separately estimate the impact of physician beliefs, patient preferences, and other factors as they affect equilibrium healthcare outcomes, largely because of challenges in identifying factors that affect only supply or demand (Dranove and Wehner 1994). Finkelstein, Gentzkow, and Williams (2016) address these challenges by following Medicare enrollees who moved to either higher or lower spending regions, finding that roughly half of the difference in spending across areas can be attributed to supply, and approximately half to demand. Their study, however, cannot identify the specific components of demand-side factors (e.g., preferences) nor supply-side factors (income and substitution effects, or physician beliefs).

In this study, we use "strategic" survey questions of physicians and patients (in the sense of Ameriks et al. 2011) to avoid several concerns related to risk-adjustment and endogeneity in estimating a unified model of regional Medicare expenditures. Strategic surveys have been used to estimate underlying preference parameters such as risk aversion (Barsky et al. 1997), the demand for precautionary saving (Ameriks et al. 2011), and long-term care insurance (Ameriks et al. 2016). Vignettes have been shown to predict actual physician behavior in clinical settings (e.g., Peabody et al. 2000, 2004; Dresselhaus et al. 2004; Mandelblatt et al. 2012; Evans et al. 2015). In prior work, our physician surveys have been used by Sirovich et al. (2008) and Lucas et al. (2010) to find that Medicare spending predicted physician treatment intensity, but these studies did not adjust for patient demand, nor could they answer the question of whether (and which) supply-side factors explain observed variation in Medicare expenditures.[1]

---

[1] In these studies, because spending was on the right-hand side of the equation and physician survey responses on the left, they could not estimate whether survey responses explain regional variations in spending, nor could the

In our study, physician incentives and beliefs are captured using detailed, scenario-based surveys that present physician-respondents with questions about their financial and practice organization, and with vignette-based questions about how the physician would manage elderly individuals with specific chronic health conditions and a given medical and treatment history.[2] Patient preferences are measured by a different survey of Medicare enrollees age 65 and older asking about preferences among a variety of aggressive and/or palliative care interventions. Previous studies using this survey (Anthony et al. 2009; Barnato et al. 2007; Baker, Bundorf, and Kessler 2014) found a small but discernible influence of patient preferences on regional variations in health care spending.[3]

We develop an equilibrium model that specifies health care intensity (as measured by expenditures) as a function of a variety of factors specific to health care providers' and patients' preferences. To estimate this requires a multivariate model in which supply and demand factors are on the right-hand side of the equation, and overall intensity (or expenditures) are on the left-hand side. We use Hospital Referral Region[4] (HRR)-level measures that account for both physician and patient responses within each HRR, and estimate models at the HRR level for 74 of the largest areas (covering roughly half of the Medicare population). We also test our model at the individual level for Medicare fee-for-service patients with acute myocardial infarction (AMI) admitted to the hospital during 2007–2010. This latter sample allows us to control more accurately for individual health status and poverty, and to examine the association between physician characteristics and quality of care.

In our empirical exercises, we characterize physicians along two nonexclusive dimensions: those who consistently and unambiguously recommended intensive care beyond what would be indicated by current clinical guidelines (whom we refer to as "cowboys"), and those who consistently recommended palliative care for the very severely ill (whom we refer to as "comforters"). At the HRR level, we explain roughly 60 percent of the variation in end-of-life spending by knowing only these two composite measures of physician treatment intensity combined with the frequency with which physicians recommend that their patients return for routine office visits. Our results are consistent with those in Finkelstein, Gentzkow, and Williams (2016) whose estimated supply-side variations are similar to (or a bit larger than) those reported below. Conditional on the supply side, we find that in this setting, demand-side Medicare patient preferences explain much less of total spending, consistent with other previous studies (Anthony et al. 2009; Barnato et al. 2007; Baker, Bundorf, and Kessler 2014; Bederman et al. 2011; and Gogineni et al. 2015).

---

different factors (e.g., financial incentives, vignette responses) be considered in a multivariate framework.

[2] However, our surveys do not capture "deep" preference parameters as in (e.g., Barsky et al. 1997).

[3] The Baker, Bundorf, and Kessler (2014) study used the same patient-level survey data to create regional measures of patient preferences, and attempted to explain differences in spending using these measures. They included potentially endogenous hospital bed and physician supply to adjust for the supply-side, rather than the physician survey data we use.

[4] The Dartmouth Atlas of Health Care has defined 306 hospital referral regions based on patient migration data from the Medicare claims data in 1992–1993. Each region required a tertiary hospital providing cardiovascular surgical procedures and for neurosurgery (Dartmouth Atlas of Health Care 2018).

Prior studies have not been able to explain why some physicians practice very aggressively (cowboys) while others are more likely to treat end-of-life patients with palliative care (comforters). We use the full sample of surveyed physicians to examine factors associated with the likelihood of being a cowboy or a comforter (few physicians were both). We find that only a small fraction of physicians report making recent decisions purely due to financial considerations. We also find that "pressure to accommodate" either patients (by providing treatments that are not needed) or referring physicians' expectations (doing procedures to keep them happy) have a modest but significant relationship with physician beliefs about appropriate care. While many physicians report making clinical interventions as a result of malpractice concerns, these responses do not explain variation in treatment recommendations.

Ultimately, the largest degree of residual variation appears to be explained by differences in physician beliefs about the efficacy of particular therapies. Physicians in our data have starkly different views about how to treat the same patients. These views are not strongly correlated with demographics, financial incentives, background, or practice characteristics, and are often inconsistent with evidence-based professional guidelines for appropriate care. Extrapolating from the sample of HRRs, we calculate that as much as 35 percent of end-of-life fee-for-service Medicare expenditures, 12 percent of Medicare expenditures for heart attack patients, and 12 percent of overall Medicare expenditures are explained by supply-side (physician) beliefs that cannot be justified either by patient preferences or by evidence of clinical effectiveness.

This paper proceeds as follows. Section I provides a graphical interpretation of the theory (with derivations in online Appendix A); Section II provides data and estimation strategies, while Section III presents the results, with particular attention paid to methodological challenges to the vignette approach raised by Sheiner (2014). Section IV concludes.

I. A Model of Variation in Medical Intensity

We develop a simple model of patient demand and physician supply, presented in detail in online Appendix A.[5] On the demand side, the patient's indirect utility function is a function of out-of-pocket prices, income, health, and preferences for care. Solving this for optimal intensity of care yields patient demand $\mathbf{x}^D$. As in McGuire (2011), we assume that $\mathbf{x}^D$ is the fully informed patient's demand for the quantity of care or procedures prior to any demand "inducement."

On the supply side, we assume that physicians seek to maximize the perceived health of their ith patient, $s(\mathbf{x}_i)$, by appropriate choice of inputs $\mathbf{x}$, subject to patient demand $\left(\mathbf{x}_i^D\right)$, financial considerations, and organizational factors. Note that the function $s(\mathbf{x})$ captures both patient survival and patient quality of life, for example, as measured by quality-adjusted life years (QALYs).

---

[5] This model draws on Chandra and Skinner (2012); also see Finkelstein, Gentzkow, and Williams (2016).
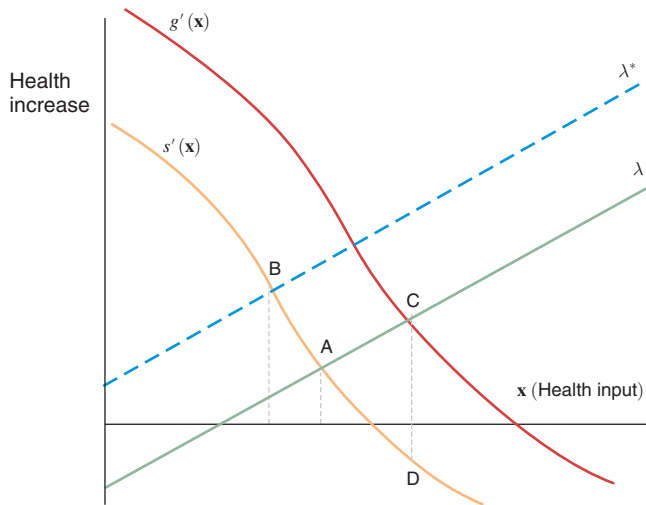
FIGURE 1. THEORETICAL MODEL OF EQUILIBRIUM HEALTH CARE INTENSITY

Notes: The model that corresponds to this figure is presented in detail in online Appendix A. The curves $s'(\mathbf{x})$ and $g'(\mathbf{x})$ represent two distinct marginal benefit curves—either real or perceived. The lines λ^* and λ represent two different sets of constraints. In the model, physicians assess the expected marginal benefit of treatment and the constraints that they face and deliver care up to the point where the two curves intersect.

Individual physicians are assumed to be price-takers, but face a wide range of reimbursement rates from private insurance providers, Medicare, and Medicaid.[6] We summarize the model graphically in Figure 1, with further details in online Appendix A. The downward sloping curve, $s'(\mathbf{x})$, represents the true value of the treatment $\mathbf{x}$ where patients are lined up in decreasing order of appropriateness for treatment. This is not a demand curve per se, but instead reflects diminishing returns to treatment by moving further into the population of patients less likely to benefit. The "supply" curve, λ, reflects rising costs to physicians, converting dollars to QALYs by taking the inverse of the social cost-effectiveness ratio (e.g., \$100,000 per life-year). The optimum is therefore the intersection of the supply and demand curves, at point A.

Why might the equilibrium quantity of care provided differ from A? In traditional supplier-induced demand models, reimbursement rates for procedures, taste for income, and organizational or peer effects would cause the supply curve to shift to (say) λ^*, leading to (in this case) higher utilization. Traditional supplier-induced demand effects occur via movement *along* the marginal benefit curve, $s'(\mathbf{x})$. Alternatively, patient demand could affect the quantity consumed. But so long as patients perceive their true medical benefit curve, $s'(\mathbf{x})$, demand factors such

---

[6] The model is therefore simpler than models in which hospital groups and physicians jointly determine quantity, quality, and prices, (Pauly 1980) or in which physicians exercise market power over patients to provide them with "too much" health care (McGuire 2011).

as preference, income, and prices will shift the λ curve, so that λ^*, for example, corresponds to a patient with a lower income or a higher copay.

Less conventionally, the marginal value curve may differ across physicians and their patients, for example, as shown in Figure 1 by $g'(\mathbf{x})$, which could represent a more skilled physician. Thus, point C in Figure 1 corresponds to greater intensity of care than point A and arises naturally when the physician is more productive. For example, heart attack patients experience better outcomes from cardiac interventions in regions with higher rates of revascularization, consistent with a Roy model of occupational sorting (Chandra and Staiger 2007; 2017). Conditional on characteristics, patients treated in regions where the true productivity curve is $g'(\mathbf{x})$ will experience better health outcomes.

Alternatively, physicians may be overly optimistic with respect to their ability to perform procedures, leading to expected benefits that exceed realized benefits (Comin, Skinner, and Staiger 2018). Baumann, Deber, and Thompson (1991) have documented the phenomenon of "macro uncertainty, micro certainty" in which physicians and nurses exhibit overconfidence in the value of their treatment for a specific patient (micro certainty) even in the absence of a general consensus as to which procedure is more clinically effective (macro uncertainty). Ransohoff, McNaughton Collins, and Fowler (2002) has noted a further psychological bias toward more aggressive treatment decisions. If the patient gets better, the physician gets the credit, but if the patient gets worse, the physician is able to say that she did everything possible.

To see this in Figure 1, suppose the actual benefit is realized along the curve $s'(\mathbf{x})$ but the physician's *perceived benefit* is along the curve $g'(\mathbf{x})$. The equilibrium achieved will then be at point D: in this scenario, the marginal treatment harms the patient, even though the physician expects the opposite, incorrectly believing they are targeting point C. In equilibrium, this type of supplier behavior would appear consistent with classic supplier-induced demand, but the cause is quite different. Most importantly, here the physician would not need to consciously "over-treat" (and thus harm) the patient, as in standard supplier-induced demand models.

How can we distinguish among these scenarios? Using the estimation equation derived in online Appendix A, we use information about patient income and preferences for treatment to capture demand effects. For supply effects, we document traditional supplier-induced demand measures through detailed questions to individual physicians regarding financial pressure, peer expectations, and willingness to cater to patient demand. To measure beliefs, we use vignettes describing well-defined clinical cases. We can further distinguish between a Roy model of treatment (as in Chandra and Staiger 2007) and one of overconfidence by testing whether patients experience higher quality care in regions with a greater share of cowboys and/or more frequent follow-up visits.

II. Data and Estimation Strategy

In this section, we discuss the practicalities of estimating supply and demand using detailed survey data. Even detailed clinical data reveal only a subset of what a physician knows about her patients' health and reveal virtually nothing

about non-clinical drivers of patients' demand for health care services. We use the results of a vignette-based physician survey to proxy for physician beliefs about appropriate care (supply), and a complementary vignette-based patient survey to capture patient's preferences for care (demand). Ameriks et al. (2011) and Ameriks et al. (2016) use "strategic surveys" to address similar challenges in identifying supply and demand factors, although the approach is relatively novel in health care economics.

We assume that the physician's responses to the vignettes are "all in" measures, reflecting physician beliefs as well as the variety of financial, organizational, and capacity-related constraints physicians face. For example, if a physician responds to a vignette by proposing a specific treatment that in turn generates revenue for her practice, we assume that any potential financial benefit is already reflected in her response. An alternative assumption is that the physician's response purely reflects clinical beliefs (for example, how one might answer for qualifying boards), and is not representative of the day-to-day realities of their practice. While we cannot fully address this limitation, we consider this alternative assumption about behavior by including organizational and financial variables in the estimation equations in addition to the vignette estimates.[7] That the organizational and financial variables add so little explanatory power to the regression is consistent with our "all-in" assumption.

In an ideal world, patient survey responses would be matched with those of their physicians. Because our data do not directly link physician and patients responses (nor are we permitted to match physician responses to their claims data), we instead match supply and demand at the area level using HRRs.[8]

A. *Physician Surveys*

Two subsamples were drawn from the American Medical Association Database: cardiologists and primary care physicians (PCPs).[9,10] A total of 999 cardiologists were randomly selected to receive the survey. Of these, 614 cardiologists responded, for a response rate of 61.5 percent. Some physicians surveyed did not self-identify as cardiologists, and others were missing crucial information such as practice type, leaving us a final sample of 598 cardiologists with complete survey responses.

---

[7] See Table 8 of online Appendix C.

[8] Spending measures are based on area of patient residence, not where treatment is actually received, to avoid biases arising from large "destination" medical centers in regions like Boston, MA or Rochester, MN. For the analysis of urgent AMI, patients are unlikely to seek care outside of the HRR since every minute counts for survival. For these we link physician and patient responses to the HRRs in which the *hospitals* are located.

[9] The document outlining "Survey Methods and Findings" (from The Center for Survey Research, University of Massachusetts Boston [CSR/UMass] and Dartmouth Medical School) is available upon request; our discussion follows Sirovich et al. (2008). A mail survey of PCPs and cardiologists was conducted by CSR/UMass. To construct the survey, focus groups concentrated on the development and wording of realistic clinical vignettes to ensure that the questions were well understood and the answers meaningful. Next, Masterfiles of the American Medical Association and American Osteopathic Association were used to obtain a random sample of domestic PCPs (self-identified as family practice, general practice, or internal medicine) or cardiologists practicing at least twenty hours per week. Residents and retired physicians were ineligible. After telephone verification of the physician's offices, each physician was mailed a survey along with $20, with a telephone follow-up.

[10] The PCP survey oversampled in four targeted cities: two areas of low intensity (Minneapolis, MN; Rochester, NY) and two areas of high intensity (Manhattan, NY; Miami, FL).

The PCP responses come from a parallel survey of self-identified PCPs (family practice, internal medicine, or internal medicine/family practice). A total of 1,333 PCPs were randomly selected to receive the survey and 967 individuals had complete responses to the survey for a response rate of 72.5 percent.[11]

B. *Patient Survey*

The patient survey sampling frame included all Medicare beneficiaries in the 20 percent denominator file who were age 65 or older on July 1, 2003 (Barnato et al. 2008). A random sample of 4,000 individuals was drawn; the response rate was 72.1 percent, leading to 2,882 total survey responses.

C. *Utilization and Quality Data*

We match the survey responses to Medicare expenditure data from the Dartmouth Atlas of Health Care (2018). The Atlas includes detailed data on how medical resources are distributed and used in the United States, using Medicare data to provide information and analysis about national, regional, and local markets, as well as hospitals and their affiliated physicians.

Our primary dependent variable is the natural logarithm of price, age, sex, and risk-adjusted HRR-level, per patient Medicare expenditures in the last two years of life for enrollees over age 65 with a number of fatal illnesses.[12] These measures implicitly adjust for differences across regions in health status: an individual with renal failure who subsequently dies is likely to be in similar (poor) health regardless of whether she lives. End-of-life measures are commonly used to instrument for health care intensity (e.g., Fisher et al. 2003a, b; Doyle 2011; Romley, Jena, and Goldman 2011). (In addition, we consider overall HRR-level Medicare spending, adjusted only for price, age, sex, and race differences.) Initially, we limit our attention to the 74 HRRs with a minimum of 3 patients, 3 cardiologists (average $=$ 5.6), and at least 1 PCP (average $=$ 7.7) surveyed, leading to an attenuated sample of physicians ($N =$ 985 in the restricted sample) and patients ($N =$ 1,516 in the restricted sample with an average of 20.5 patient respondents per HRR).[13]

We also consider a third measure of utilization: one-year Medicare expenditures after admission for AMI during 2007–2010 in the fee-for-service population. The advantage of this patient-level data is that we are better able to address concerns

[11] Physician and patient survey responses could vary systematically by demographic covariates, such as age, sex, and race or ethnicity, leading to greater sampling variability if we happen to sample a few physicians, for example, in one group. However, our main regression results (Table 3) were very similar with such adjustment (presented in online Appendix Tables C3 and C4), so we did not pursue this approach.

[12] These include congestive heart failure, cancer/leukemia, chronic pulmonary disease, coronary artery disease, peripheral vascular disease, severe chronic liver disease, diabetes with end organ damage, chronic renal failure, and dementia, which account for 90 percent of end-of-life spending.

[13] To address potential measurement error of the constructed variables on the right-hand side of this regression specification, we bootstrap the entire estimation process 1,000 times to calculate all standard errors in HRR-level regressions, beginning with the construction of the HRR-level physician and patient samples, and proceeding through the HRR-level regression analysis.

that physician practice patterns are correlated with unobservable health status of patients (Sheiner 2014). By limiting our attention to AMI patients—all of whom are seriously ill—and by using extensive risk-adjustment based on comorbidities during admission, quintiles of ZIP code income, and an HCC (hierarchical condition categories) score based on previous outpatient and inpatient diagnoses, we can better adjust for differences across regions in underlying health.

We also consider quality measures. In the aggregated analysis, we use multiple Medicare measures from the "Hospital Compare" database aggregated to the HRR level. In the AMI data, we use two quality measures: risk-adjusted, one-year mortality and risk-adjusted percutaneous coronary interventions (PCI), or stents, on the day of admission. Clinical trials have shown that same-day PCI yield the greatest benefit for patients (Keeley, Boura, and Grines 2003; Boden et al. 2007).

D. *Patient Survey Vignette Detail*

We use responses to five survey questions that ask patients about their likelihood of wanting unnecessary tests or cardiologist referrals in the case of new chest pain as well as preferences for comfort versus intensive life-prolonging interventions in an end-of-life situation. The exact language used in these vignettes is reproduced in online Appendix B. Since the questions patients respond to are hypothetical and typically describe scenarios that have not yet happened, we think of them as preferences not affected by physician advice ($\mathbf{x}^D$).

Two of the questions on the patient survey relate to unnecessary care, asking respondents if they would like a test or cardiac referral even if their PCP did not think they needed one (Table 1).[14] Overall, 73 percent of patients wanted such an unneeded test and 56 percent wanted an unneeded cardiac referral. However, there is wide variation across regions in average responses to these questions. Figure 2 shows density plots of patient preferences for the main questions in the patient survey for the 74 HRRs considered (weighted by the number of patients per HRR). Superimposed on the figure is the simulated distribution based on 1,000 bootstrap samples, with replacement, under the null hypothesis that individuals were randomly assigned to areas. The p-values for over-dispersion are reported in the last column of Table 1; these indicate that the observed variation is significantly greater than can be explained by random variation.

Three other patient questions, grouped into two binary indicators each, measure preferences for end-of-life care. One reflects patients' desire for aggressive care at the end of life: whether they would want to be put on a respirator if it would extend their life for either a week (one question) or a month (another question). The next question asks if, in situations in which the patient reached a point at which they were feeling bad all of the time, the patient would want drugs to feel better, even if those drugs might shorten their life. In each case, there is statistically significant variation across HRRs (Table 1).

---

[14] This question captures pure patient demand independent of what the physician wants. Note, however, that patients could still answer they would not seek an additional referral if they were unwilling to disagree with their physician.

TABLE 1—PRIMARY VARIABLES AND SAMPLE DISTRIBUTION

Variable	Mean	Individual standard deviation	Area average standard deviation	*p*-value
Panel A. Spending and utilization				
Two-year end-of-life spending	$55,194	—	$10,634	—
Average EOL spending	$14,409	—	$3,983	—
Hip fracture patient spending	$52,282	—	$4,882	—
Panel B. Patient variables (N = 1,516)				
Have unneeded tests	73%	45%	11%	<0.01
See unneeded cardiologist	56%	50%	14%	<0.01
Aggressive patient preferences ratio	8.0%	27%	7.0%	<0.01
Comfort patient preferences ratio	48%	50%	13%	<0.01
Panel C. Primary care physician variables (N = 570 in restricted sample)				
Cowboy	23%	42%	27%	<0.01
Comforter	44%	50%	28%	<0.01
Follow-up low	8.8%	28%	19%	<0.01
Follow-up high	4.4%	20%	6.4%	<0.01
Panel D. Cardiologist variables (N = 415 in restricted sample)				
Cowboy	26%	44%	20%	<0.01
Comforter	27%	45%	23%	<0.01
Follow-up low	1.4%	12%	6.3%	—
Follow-up high	21%	41%	22%	<0.01
Panel E. All physicians (N = 985 in restricted sample)				
Cowboy	24%	43%	15%	<0.01
Comforter	37%	48%	18%	<0.01
Follow-up low	5.7%	23%	8.6%	<0.01
Follow-up high	12%	32%	12%	<0.01
Panel F. Organizational and financial variables				
Reimbursement				
Fraction Medicare patients	42%	22%	—	—
Fraction Medicaid patients	10%	13%	—	—
Fraction capitated patients	15%	25%	—	—
Organizational factors				
Solo or two-person practice	29%	46%	—	—
Single/multi speciality group practice	57%	50%	—	—
Group/staff HMO or hospital-based practice	13%	33%	—	—
General controls				
Physician age	53	10	—	—
Male	81%	39%	—	—
Board certified	88%	33%	—	—
Weekly patient days	2.9	1.5	—	—
Cardiologists per 100k (N = 74, HRR-level)	6.8	1.9	—	—
Responsiveness factors				
Responds to referrer expectations (cardiologists only)	31%	46%	—	—
Responds to colleague expectations	38%	49%	—	—
Responds to patient expectations	54%	50%	—	—
Responds to malpractice concerns	39%	49%	—	—
Responds to practice financial incentives	9%	29%	—	—

Notes: The table shows means for the (restricted) primary analysis sample. The *p*-value in the last column is for the null hypothesis of no excess variance across areas and is taken from a bootstrap of patient or physician responses across areas. For each of 1,000 simulations, we draw patients or providers randomly (with replacement) and calculate the simulated area average and the standard deviation of that area average. The empirical distribution of the standard deviation of the area average is used to form the *p*-value for the actual area average.
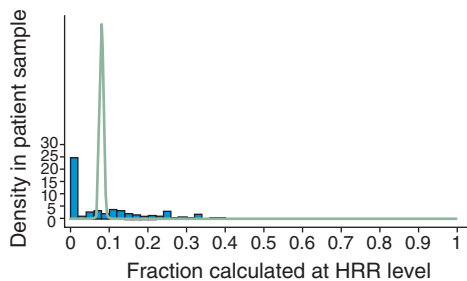
Panel A. Fraction who would have unneeded tests (simulated versus observed density; $p < 0.01$)
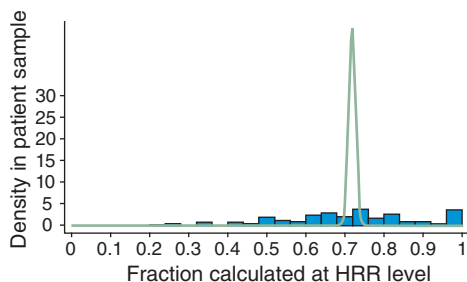
Panel B. Fraction who would see unneeded cardiologist (simulated versus observed density; $p < 0.01$)
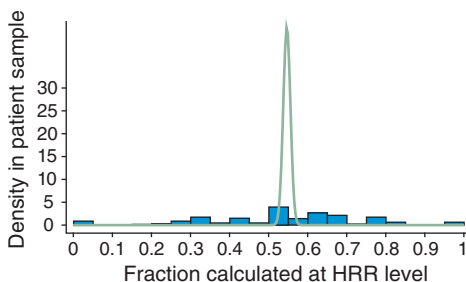
Panel C. Fraction of patients with aggressive EOL preferences (simulated versus observed density; $p < 0.01$)

Panel D. Fraction of patients with comfort EOL preferences (simulated versus observed density; $p < 0.01$)
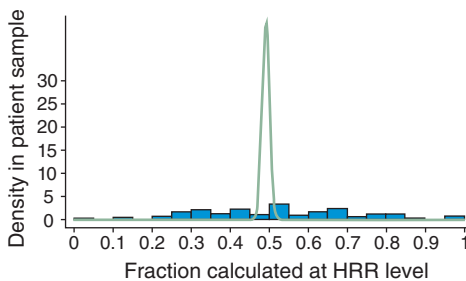

Figure 2. Distributions of Observed Patient Preferences versus Simulated Distributions (*based on 1,000 bootstrap samples with replacement*)

Notes: This figure shows density plots of patient preferences for the main questions in the patient survey for the 74 HRRs considered (weighted by observations per HRR). A simulated distribution based on 1,000 bootstrap samples under the null hypothesis that individuals were randomly assigned to areas is superimposed on the observed distributions. *p*-values for over-dispersion are reported in Table 1 and indicate that the observed variation is significantly greater than can be explained by random variation.

Patients' preferences are sometimes correlated across questions. For example, the correlation coefficient between wanting an unneeded cardiac referral and wanting an unnecessary test is 0.43 ($p < 0.01$). But other comparisons point to modest, negative associations, for example, a -0.28 correlation coefficient between wanting palliative care and wanting to be on a respirator at the end of life.

E. *Potential Biases in the Patient Survey*

One concern with these questions is that they are hypothetical and might not reflect true patient demand. However, Anthony et al. (2009) have linked survey respondents to the individuals' Medicare claims and found a strong link between preferences and care-seeking behavior (e.g., physician visits) at the individual level (also see Mandelblatt et al. 2012). Questions about end-of-life care may appear abstract until the patient actually faces decisions related to end-of-life care, but our

results are similar even with a subset of respondents age 80+ where end-of-life scenarios and treatment decisions are likely to be more salient. Moreover, Medicare enrollees are increasingly choosing to create advanced directives that specify their end-of-life choices well before they face those decisions.[15]

Another limitation of the study is the possibility that patient survey responses are measured with more error than physician responses, which may lead us to underestimate the role of patient preferences in determining utilization. Patient-level sample sizes are larger than physician sample sizes, but physicians treat multiple patients (and even those small samples are a larger fraction of the physician population), and so may better capture underlying physician beliefs. While we cannot rule out this hypothesis, we do not find a strong correlation between utilization and patient preferences even in the subset of HRRs with at least 20 patient surveys.[16]

Finally, because higher proportions of patients shared aggressive preferences relative to the proportion of "cowboy" physicians, the patient survey data may constitute a less sensitive test for aggressive tastes. As a result of all of the above factors, patient preferences may be measured with more error than physician beliefs in these surveys.

F. *Physician Survey Vignette Detail*

Prior research has established that in practice physicians "do what they say they'd do" in surveys. For example, physicians' self-reported testing intensity has been shown to be predictive of population-based rates of coronary angiography (Wennberg et al. 1997). The detailed clinical vignette questions used in the physician surveys are shown in online Appendix B and summary statistics are presented in Table 1.

We first consider the vignette for Patient A, which asks a physician how frequently they would schedule routine follow-up visits for a patient with stable angina whose symptoms and cardiac risk factors are well controlled on current medical therapy (for cardiologists) or a patient with hypertension (for PCPs). The response is unbounded, and expressed in months, but in practice individual physician responses ranged from 1 to 24 months. Figure 3, panel A presents a HRR-level histogram of averages from the cardiology survey for 74 HRRs.

How do these responses correspond to guidelines for managing such patients? The 2005 American College of Cardiology/American Heart Association [ACC/AHA] guidelines (Hunt et al. 2005)—what most cardiologists would have considered "the cardiology Bible" at the time the survey was fielded—recommended follow-up every 4–12 months. However, even with these broad recommendations, we find that 21 percent of cardiologists are "high follow-up" in the sense of recommending visits more frequently than every 4 months, and 1.4 percent are "low follow-up" and recommend visits less frequently than every 12 months.

---

[15] According to one study, more than half of elderly Americans have advanced care directives. The primary reason for not having one was being unaware of their existence (Rao et al. 2014).

[16] In addition, Newhouse et al. (2013) and Newhouse and Garber (2013) have noted that even within HRRs, there is considerable variation in utilization at the much smaller hospital service area (HSA) level. Assuming that our sample of physicians and patients are drawn randomly within HRRs from different HSAs, this would imply that the average HRR-level samples of (e.g.) physician beliefs would be measured with greater error. This makes it even more surprising that responses from just a handful of physicians in an HRR could provide such a strong prediction for that region's end-of-life spending.

Panel A. Distribution of length of time before next visit for patient with well-controlled angina (regional averages for cardiologists)

Recommended frequency of follow-up visits (months)
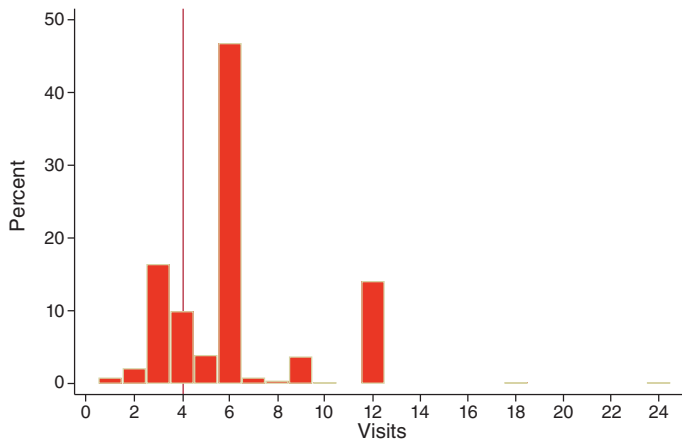
Panel B. Distribution of high follow-up physicians and geographic correlation (regional averages for cardiologists)

Fraction high follow-up cardiologists (simulated versus observed density; $p < 0.01$)
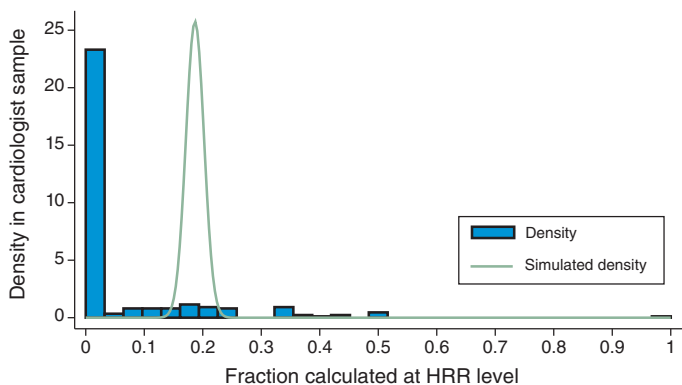
FIGURE 3

Notes: Panel A presents a HRR-level histogram of averages from the cardiology survey for the 74 HRRs in the main analysis sample. Panel B shows the distribution of high follow-up cardiologists in this sample. A simulated distribution based on 1,000 bootstrap samples under the null hypothesis that individuals were randomly assigned to areas is superimposed on the observed distributions. *p*-values for over-dispersion are reported in Table 1 and indicate that the observed variation is significantly greater than can be explained by random variation.

These physicians were geographically clustered in a subset of HRRs ($p < 0.01$ in a test of the null of no correlation) and the distribution of high follow-up cardiologists across HRRs is shown in Figure 3, panel B.

The equivalent measure for PCPs is recommended follow-up frequency for a patient with well-controlled hypertension. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (US Department of Health and Human Services 2004), which represented guideline-based recommendations at the time, suggests follow-up for this type

of patient every three to six months. Based on this range, 4.4 percent of PCPs in our sample are "high follow-up" physicians and 8.8 percent are "low follow-up" physicians.[17]

The next two vignettes focus on patients with heart failure, a much costlier setting. Heart failure is also a natural scenario to consider because it is common, the disease is chronic, its prognosis is poor, and treatment is expensive. Vignettes for Patients B and C ask questions about the treatment of Class IV heart failure, the most severe classification and one in which patients experience symptoms at rest. In both patient scenarios, the vignette patient is on maximal (presumably optimal) medications, and neither patient described is a candidate for revascularization: Patient B has already had a coronary stent placed without symptom change, and Patient C is explicitly noted to not be a candidate for this procedure. The key differences between the two scenarios are the respective patients' ages (75 for Patient B, 85 for Patient C), the presence of asymptomatic non-sustained ventricular tachycardia in Patient B, and severe symptoms that resolve partially with increased oxygen in Patient C.

Cardiologists in the survey were asked about various interventions as well as palliative care for each of these patients. For Patient B, they were given five choices: three intensive treatments (repeat angiography; implantable cardiac defibrillator [ICD] placement, and pacemaker insertion), one involving medication (antiarrhythmic therapy), and palliative care. For Patient C, each physician is also presented with three intensive options (admit to the ICU/CCU, placement of a coronary artery catheter, and pacemaker insertion), two less aggressive options (admit to the hospital [but not the ICU/CCU] for diuresis, and send home on increased oxygen and diuretics), and palliative care. In each scenario, cardiologists separately reported their likelihood of recommending each potential intervention on a five-interval range from "always/almost always" to "never." Each response is recorded independently of other responses. For example, physicians could "frequently" recommend both palliative care and one (or more) additional intervention(s).

Beginning with the obvious, both patients described in the survey deserve a frank conversation about their prognosis and an ascertainment of their preferences for end-of-life care. The one-year mortality rate for each patient is expected to be as high as 50 percent (Ahmed, Aronow, and Fleg 2006). If compliant with guidelines, therefore, all physicians should have answered "always/almost always," or at least "most of the time," when reporting their likelihood of initiating or continuing discussions about palliative care.[18]

For Patient B, only 29 percent of cardiologists responded that they would initiate or continue discussions about palliative care "most of the time" or "always/almost always." For Patient C, 44 percent of cardiologists and 47 percent of PCPs were likely to recommend this course of action "most of the time" or "always/almost

---

[17] Office visits are not a large component of physicians' incomes (or overall Medicare expenditures). Thus, any correlation between the frequency of follow-up visits and overall expenditures would most likely be because frequent office visits are also associated with additional highly remunerated tests and interventions (such as echocardiography, stress imaging studies, and so forth) that further set in motion the "diagnostic-therapeutic cascade" (Lucas et al. 2008).

[18] The AHA-ACC guidelines say that "patient and family education about options for formulating and implementing advance directives and the role of palliative and hospice care services with reevaluation for changing clinical status is recommended for patients with HF [heart failure] at the end of life." (Hunt et al. 2005, e206).

always." In both cases, physicians' recommendations fall far short of guidelines. We classify a doctor as a "comforter" if they would discuss palliative care with the patient "always/almost always" or "most of the time" for both Patients B and C (among cardiologists) or for Patient C only (among PCPs, who did not have Patient B's vignette in their survey). In our restricted analysis sample, 27 percent of cardiologists and 44 percent of PCPs met this definition of a comforter.

We now turn to more medically intensive aspects of patient management, where the language in the vignettes was carefully constructed to relate to clinical guidelines. Several key aspects of Patient B rule out both the ICD and pacemaker insertion[19] and indeed the ACC-AHA guidelines explicitly recommend against the use of an ICD for Class IV patients (Hunt et al. 2005). Since Patient C is already on maximal medications and is not a candidate for revascularization, the physician's goal should be to keep him as comfortable as possible. This should be accomplished in the least invasive manner possible (e.g., at home), and if that is not possible, in an uncomplicated setting, for example, during admission to the hospital for simple diuresis. According to the ACC/AHA guidelines, no additional interventions are appropriate for this patient.[20]

Despite these guideline recommendations, physicians in our data show a great deal of enthusiasm for additional interventions. For Patient B, over one quarter of the cardiologists surveyed (28 percent) would recommend a repeat angiography at least "some of the time." Similarly, 62 percent of cardiologists would recommend an ICD "most of the time" or "always/almost always," while 45 percent would recommend a pacemaker at least "most of the time." For Patient C, 18 percent of cardiologists would recommend an ICU/CCU admission, 2 percent would recommend a pulmonary artery catheter, and 14 percent would recommend a pacemaker at least "most of the time."

Cardiologists' responses on aggressive interventions are highly correlated across Patients B and C. Of the 28 percent ($N = 170$) of cardiologists in the full physician sample who would "frequently" or "always/almost always" recommend at least one of the high-intensity procedures for Patient C, 90 percent ($N = 153$) would also frequently or always/almost always recommend at least one high-intensity intervention for Patient B. We use this overlap to define a "cowboy" cardiologist as one who would recommend at least one of the three possible intensive treatments for both Patients B and C at least "most of the time." Because Vignette Patient B was not presented to PCPs, we use their responses to questions about Vignette Patient C to categorize them using the same criteria. In total, 26 percent of the cardiologists and 23 percent of PCPs in our sample are classified as cowboys.

All told, we consider four main measures of physician behavior (supply): high or low frequency of follow-up visits, a dummy variable for being a cowboy, and a dummy variable for being a comforter. How are these measures related to

---

[19] This includes his advanced stage, his severe (Class IV) medication refractory heart failure, and the asymptomatic non-sustained nature of the ventricular tachycardia.

[20] This patient's clinical improvement with increased oxygen (see vignette in online Appendix B) also argues against more intensive interventions. Moreover, even a simple but invasive test, the pulmonary artery catheter, has been found to be of no marginal value over good clinical decision making in managing patients with congestive heart failure (CHF), and could even cause harm (ESCAPE 2005).

TABLE 2—DISTRIBUTION OF PHYSICIANS BY VIGNETTE RESPONSES

Panel A. PCPs (N = 570 in restricted sample)

	Cowboy		Comforter		
Follow-up frequency	No	Yes	No	Yes	
Low	41	9	34	16	8.8%
Medium	387	108	270	225	88%
High	12	13	15	10	4.4%
	77%	23%	56%	44%	
$p(\chi^2)$:	**<0.001**		$p(\chi^2)$:	0.055	

	Comforter		
Cowboy	No	Yes	
No	236	204	77%
Yes	83	47	23%
	56%	44%	
$p(\chi^2)$:	**0.013**		

Panel B. Cardiologists (N = 415 in restricted sample)

	Cowboy		Comforter		
Follow-up frequency	No	Yes	No	Yes	
Low	6	0	5	1	1.4%
Medium	240	80	228	92	77%
High	61	28	69	20	21%
	74%	26%	73%	27%	
$p(\chi^2)$:	**<0.001**		$p(\chi^2)$:	**<0.001**	

	Comforter		
Cowboy	No	Yes	
No	227	80	74%
Yes	75	33	26%
	73%	27%	
$p(\chi^2)$:	**<0.001**		

Notes: This table shows the bivariate relationships between the guideline-defined indicators for recommended follow-up frequency, as well as "Cowboy" and "Comforter" status among both PCPs and cardiologists in the (restricted) analysis sample. χ^2-tests evaluate the null hypothesis that there is no association between pairs of indicators in the table.

one another? Table 2 shows that physicians with a low follow-up frequency are more likely to be comforters and less likely to be cowboys than physicians with a high follow-up frequency. Similarly, cowboy physicians are unlikely to be comforter physicians (although a small minority were classified as both).

III. Estimation Results

Our first set of empirical results runs a "horse race" between area-level supply and demand factors to assess which set of factors can better explain actual HRR-level expenditures measures. In our second set of estimates, we return to the full sample of physicians to understand why they hold the beliefs they do by relating individual physician vignette responses to physicians' financial and organizational incentives.
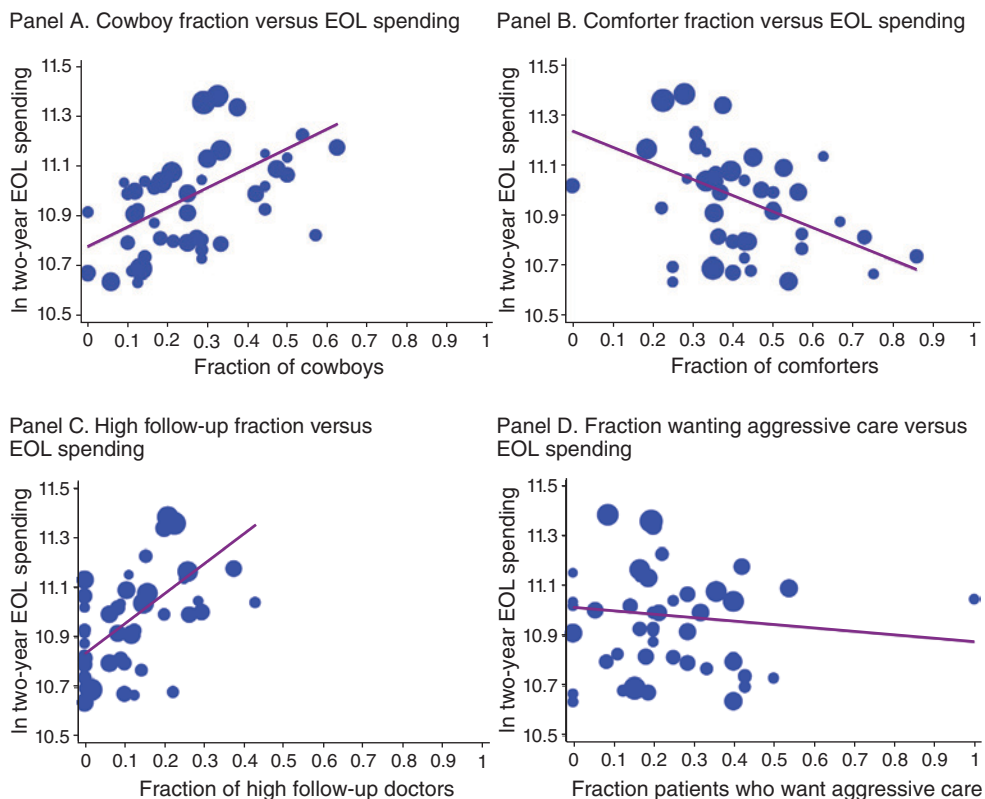
Panel A. Cowboy fraction versus EOL spending

Panel B. Comforter fraction versus EOL spending

Panel C. High follow-up fraction versus EOL spending

Panel D. Fraction wanting aggressive care versus EOL spending

FIGURE 4. NATURAL LOG OF TWO-YEAR END-OF-LIFE (EOL) SPENDING VERSUS SELECTED INDEPENDENT VARIABLES

Notes: This figure presents scatter plots of area-level end of life spending versus measures of supply and demand for care. Measures included are: A) fraction of all physicians who are cowboys; B) fraction of physicians who are comforters; C) fraction of high follow-up physicians; and D) the share of patients who desire aggressive care at the end of life. Each circle represents one HRR, with size proportional to the size of the survey sample represented.

A. *Do Survey Responses Predict Regional Medicare Expenditures?*

We begin with end-of-life spending, since it is a measure of utilization that is most likely to be sensitive to differences across regions for patients who, like those in our vignettes, are at a high risk of death.[21] Figure 4 presents scatter plots of area-level end-of-life spending versus measures of supply and demand for care. The measures included here are the fraction of all physicians in the area who are cowboys (panel A), the fraction of physicians who are comforters (panel B), the fraction of physicians who recommend follow-up more frequently than recommended guidelines (panel C), and the share of patients who desire more aggressive care at the end of life (panel D). Each circle represents one HRR, with its size proportional to the size of the survey sample represented.

---

[21] End-of-life spending is also highly correlated with prospective measures such as spending for advanced lung-cancer (Keating et al. 2016) and overall spending among those at high risk of death (Kelley et al. 2018).

TABLE 3—REGRESSION ESTIMATES OF MEDICARE EXPENDITURES

	Combined sample: PCPs and cardiologists						
	Medicare spending in last two years of life (columns 1–6)						Overall
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cowboy physician share	0.822	0.676	0.675	0.684	0.685		0.904
	(0.196)	(0.143)	(0.131)	(0.132)	(0.123)		(0.182)
Comforter physician share	−0.315	−0.200	−0.175	−0.161	−0.151		−0.274
	(0.144)	(0.098)	(0.098)	(0.099)	(0.096)		(0.136)
High follow-up physician share		1.060	0.964	1.033	0.917		1.066
		(0.176)	(0.172)	(0.165)	(0.166)		(0.215)
Low follow-up physician share		−0.307	−0.289	−0.402	−0.345		−0.673
		(0.220)	(0.217)	(0.263)	(0.239)		(0.289)
Have unneeded tests patient share			0.297		0.351	0.500	0.421
			(0.176)		(0.193)	(0.255)	(0.239)
See unneeded cardiologist patient share			0.150		0.132	0.543	−0.079
			(0.129)		(0.130)	(0.260)	(0.151)
Aggressive patient preferences share				−0.427	−0.294	−0.021	−0.695
				(0.408)	(0.379)	(0.518)	(0.404)
Comfort patient preferences share				−0.237	−0.282	−0.426	−0.639
				(0.137)	(0.144)	(0.217)	(0.173)
Observations	74	74	74	74	74	74	74
R^2	0.356	0.619	0.644	0.634	0.662	0.203	0.623

Notes: Two-year end-of-life spending (outcome in columns 1–6) and overall spending (outcome in column 7) are in natural log form and are price, age, sex, and race adjusted. Results shown are for the 74 Hospital Referral Regions (HRRs) with survey responses for at least three patients, one primary care physician, and three cardiologists. All regressions include a constant, control for the fraction of primary care physicians in the sample, and are robust to controlling for percent black and percent poverty at the HRR level. Survey sampling weights take into account differences in the number of physician observations per HRR.

In the case of the three supply-side variables in Figure 4, the results are consistent with a meaningful role in determining spending. Despite the relatively small sample sizes of physicians in each HRR, end-of-life spending is positively related to the cowboy ratio, negatively related to the comforter ratio, and positively related to the fraction of doctors recommending follow-up visits more frequently than clinical guidelines indicate. The demand variable, in contrast, is not strongly related to spending; the data points do not reveal a clear relationship.

Table 3 explores these results more formally with regression estimates of the natural log of end-of-life expenditures in the last two years of life (columns 1–6) and overall Medicare expenditures (column 7). The statistical significance of all results is robust to adjusting standard errors using 1,000 bootstrap samples of patient and physician surveys. As the first column shows, the local proportion of cowboys and comforters explains 36 percent of the observed regional variation in end-of-life spending. Further, the estimated magnitudes are large: increasing the percentage of cowboys by 1 standard deviation (15 percentage points) is associated with a 12 percent increase in 2-year end-of-life expenditures, while increasing the fraction of comforters by 1 standard deviation (18 percentage points) implies a 5.7 percent decrease in those expenditures. These relationships between spending and the local fractions of cowboys and comforters are robust

to a number of alternative specifications.[22] Results also hold when each physician sample is analyzed separately (online Appendix C, Tables 1 and 2) and when survey responses are age, sex, and race/ethnicity-adjusted (online Appendix C, Tables 3 and 4).

Column 2 of Table 3 shows that the indicator for the share of physicians recommending high frequency follow-up is also a predictor of HRR-level end-of-life spending; conditional on the fraction of cowboys and comforters, an increase of 1 standard deviation (12 percentage points) of high follow-up physicians is associated with a 13 percent increase in 2-year end-of-life spending. While the low frequency follow-up coefficient is meaningful in magnitude—implying roughly a 3 percent reduction in spending for a one standard deviation increase in low follow-up physicians—it is not statistically significant. Notably, the combination of just these 4 variables about supplier beliefs alone can explain over 60 percent of the observed end-of-life spending variation in the 74 HRRs in this sample.[23]

The next two columns add various HRR-level measures of patient preferences to the regressions: the share of patients wishing to have unneeded tests in a cardiac setting, the share of patients wanting to see an unneeded cardiologist, the share of patients preferring aggressive end-of-life care, and the share preferring comfortable end-of-life care. None of these variables are statistically significant at the 5 percent level. Excluding the physician belief variables entirely, as in column 6, the R^2 value from the regression incorporating patient preference variables alone is 0.20 (slightly larger than in Baker, Bundorf, and Kessler 2014). In further sensitivity analyses, we also considered using alternate and/or additional measures of patient demand recovered from the patient survey, or financial variables from the physician surveys. When predicting Medicare spending in the last two years of life, these alternate measures do not increase the explanatory power of our regression models, nor are the coefficients related to patient demand factors statistically significant at conventional levels.[24]

Regressions on the natural log of total per patient Medicare expenditures are presented in the final column of Table 3; coefficients are similar, but in this specification the role of patient preferences for comfort-oriented care emerges as statistically significant, consistent with Finkelstein, Gentzkow, and Williams (2016) who find that patient demand explains a significant share of Medicare spending in a non-end-of-life context.

We further consider the possibility that physicians' aggressiveness in end-of-life care may be related to overall quality. We analyze the range of quality measures from

---

[22] In robustness tests not presented, we further find that the observed patterns are robust to the exclusion of "small cells" (HRRs with fewer than eight physician surveys), to controlling for percent poverty and percent black at the HRR level, and to re-weighting HRRs by their patient population.

[23] As Black, Berger, and Scott (2000) note, the OLS estimate is a lower bound and under weak assumptions, the expected value of the OLS parameter estimate is of smaller magnitude than the true parameter. (The R^2 is also a lower bound owing to measurement error.)

[24] Nor were our results driven by differences between east coast versus west coast medical practices; similar results were obtained when east and west coasts were estimated separately. It is also possible that there is an interaction effect between patient preferences and physician beliefs—for example if aggressive physicians interact with patients with preferences for aggressive care—to generate even more utilization (or conversely for comforter physicians and patients who demand palliative care to interact and reduce care provided in a mutually reinforcing way). We test this in online Appendix C, Table C6, but do not find evidence for this hypothesis.

the CMS Hospital Compare data (December 2005 data release,[25] averaged at the HRR level), but find few statistically significant correlations between the fraction of aggressive physicians (cowboys) and quality measures. Of the 17 quality measures considered, only 2[26] have a statistically significant correlation with the fraction of cowboys in a region (and those were negative correlations); the remainder showed no significant correlation with the cowboy fraction.

B. *Expenditures and Outcomes in Acute Myocardial Infarction*

We next consider expenditures and outcomes in a sample of AMI patients. By focusing on physician vignette measures, which we think are of primary importance in acute and post-acute treatment decisions, we are able to expand the sample size to 137 HRRs accounting for 71 percent of the entire AMI sample or a total of 560,392 patients.[27] Spending measures are adjusted for differences across regions in prices using the Gottlieb et al. (2010) method,[28] and all regressions are clustered at the HRR level. Summary statistics are presented in Table 7 of online Appendix C.

Regression estimates are presented in Table 4, with expenditures in levels in column 1, and in natural logs in column 2. As expected, the risk-adjustment HCCs are large in magnitude and highly significant, but conditional on health status, ZIP income quintile coefficients are all generally small in magnitude. The coefficients for cowboy ($7,644) and high revisit rates ($16,560) are large relative to the mean value ($45,827) and highly statistically significant. There is a strong negative coefficient for comforters (−$2,911) and, while the coefficient for low revisit rates (−$3,269) is in the expected direction, the estimate is not statistically significant at conventional levels, as was the case in Table 3. Similar results are shown in the natural log model, albeit with coefficients that imply slightly smaller proportional effects. Regression estimates weighted by the number of physicians in the survey (columns 3 and 4) exhibit coefficients with larger magnitudes than the unweighted regressions; the coefficients imply that high-follow-up physicians are associated with roughly 36 percent greater one-year expenditures for AMI patients.

Finally, Table 4 provides evidence on the hypothesis that regions with a greater share of cowboy or high follow-up physicians experience higher quality care for AMI patients. Column 5 provides estimated marginal effects from probit models of one-year survival, with a mean of 68 percent. None of the physician characteristics

---

[25] Data available at: https://data.medicare.gov/data/archives/hospital-compare.

[26] The two quality measures were: (a) AMI patients with left ventricular systolic dysfunction (and without contraindications) who were prescribed an angiotensin-converting enzyme (ACE) inhibitor or an angiotensin II receptor blocker (ARB) at hospital discharge, and (b) Pneumonia patients who receive their first dose of antibiotic within four hours of arrival at the hospital.

[27] In sensitivity analysis, not reported, we also included patient demand variables, but none were significant in any of the regression analyses, except for one: The patient aggressiveness measure was associated with a *lower* probability of early PCI (stenting), the reverse of what one might expect.

[28] Gottlieb et al. (2010) use a price-adjustment method that sets a common national price for each individual Medicare service, and thus removes price differences across regions owing to cost-of-living adjustments, payments provided for graduate medical programs, and disproportionate-share programs, and other items.

TABLE 4—REGRESSION RESULTS FOR ONE-YEAR MEDICARE SPENDING AND QUALITY: ACUTE MYOCARDIAL INFARCTION PATIENTS, 2007–2010

	1-year medical expenditure (1)	ln 1-year medical expenditure (2)	1-year medical expenditure (3)	ln 1-year medical expenditure (4)	1-year survival (5)	Fraction with same-day PCI (6)
Cowboy physician share	7,644 (1,720)	0.142 (0.034)	10,870 (2,419)	0.196 (0.044)	0.006 (0.008)	0.016 (0.019)
Comforter physician share	−2,911 (1,137)	−0.050 (0.021)	−3,019 (1,350)	−0.046 (0.024)	0.009 (0.007)	−0.008 (0.016)
High follow-up physician share	16,560 (2,530)	0.270 (0.046)	22,310 (3,215)	0.360 (0.051)	-0.009 (0.013)	-0.063 (0.027)
Low follow-up physician share	−3,269 (2,855)	−0.084 (0.055)	−4,148 (3,463)	−0.115 (0.058)	0.015 (0.015)	0.089 (0.032)
HCC (second quintile)	5,297 (195.4)	0.106 (0.003)	5,423 (251.6)	0.108 (0.004)	−0.044 (0.002)	−0.027 (0.002)
HCC (third quintile)	8,535 (227.5)	0.176 (0.004)	9,098 (353.4)	0.187 (0.007)	−0.092 (0.002)	−0.053 (0.002)
HCC (fourth quintile)	11,660 (285.7)	0.242 (0.005)	12,180 (539.9)	0.252 (0.010)	−0.149 (0.003)	−0.084 (0.002)
HCC (highest quintile)	19,280 (506.7)	0.347 (0.008)	20,460 (934.3)	0.365 (0.015)	−0.258 (0.003)	−0.123 (0.003)
Black	4,823 (531.3)	0.052 (0.009)	4,543 (731.8)	0.046 (0.010)	−0.020 (0.003)	−0.039 (0.004)
Hispanic	4,921 (1,175)	0.079 (0.024)	4,770 (1,264)	0.080 (0.028)	0.016 (0.005)	−0.005 (0.005)
Asian	1,385 (1,141)	−0.010 (0.023)	858.5 (1,399.3)	−0.017 (0.026)	−0.005 (0.007)	−0.023 (0.011)
Income (second quintile)	−233.9 (342.2)	−0.007 (0.006)	−405.1 (591.6)	−0.007 (0.009)	0.010 (0.002)	0.007 (0.004)
Income (third quintile)	−286.0 (326.2)	−0.009 (0.006)	−455.5 (552.3)	−0.012 (0.008)	0.012 (0.002)	0.009 (0.004)
Income (fourth quintile)	160.0 (399.2)	−0.002 (0.008)	−363.8 (605.3)	−0.007 (0.011)	0.016 (0.002)	0.016 (0.005)
Income (highest quintile)	320.1 (500.8)	0.007 (0.009)	−678.7 (754.1)	−0.007 (0.013)	0.030 (0.003)	0.020 (0.006)
Observations	560,392	560,392	560,392	560,392	560,392	560,392
R^2	0.070	0.075	0.075	0.082	—	—

Notes: These results are based on a 100 percent sample of fee-for-service Medicare enrollees, 2007–2010 with mortality follow-up through 2011. Columns 1 and 2 are unweighted and columns 3 and 4 are weighted by the number of physician respondents included in the sample. All results are for the 137 HRRs sample, with risk adjustment measures included but not reported for the following: age (65–69, 70–74, 75–79, 80–84, 85+), fully interacted with sex, vascular disease, pulmonary disease, dementia, diabetes, liver failure, renal failure, cancer, plegia (stroke), rheumatologic disease, HIV, and location of the AMI: anterolateral, inferolateral, inferoposterior, all other inferior, true posterior walls, or subendocardial, other site, or not otherwise specified. All models include year fixed effects. Reported coefficients in columns 5 and 6 are marginal effects from probit regressions.

are individually significant predictors of survival, and the *p*-value for the *F*-test of all variables combined is 0.43.

Consider next the estimates for early PCI (mean, 26 percent) in column 6 of Table 4. Neither the cowboy nor comforter variables are associated with early PCI. Follow-up rates, by contrast, are associated *negatively* with early PCI;

the coefficient for the fraction of high follow-up rate physicians implies a 6.3 percent lower early PCI rate, with the converse true for low follow-up physicians (the *F*-test of all four variables is jointly significant at the 0.01 level). Along with results reported above, these estimates for early PCI provide no evidence for the hypothesis that "cowboy" physicians, as defined in this study, are more aggressive because of greater skill (e.g., Chandra and Staiger 2007). Instead, these results are consistent with our model in which physician beliefs that are unsupported by clinical evidence can increase utilization without benefiting patients.

C. *What Factors Predict Physician Responses to the Vignettes?*

To this point, we have shown that aggregated physician vignette responses explain regional patterns in spending, and that these responses vary across areas far more than would be expected to occur at random. But this finding raises the question: what explains observed variation in physician treatment beliefs? In this section, we use the entire sample of physicians to test for the relative importance of financial and organizational factors in explaining physician recommendations. We interpret the residual, conditional on financial and organizational factors, as "beliefs."

Various measures relating to reimbursement are reported in Table 1 for all physicians. We focus on three of these in particular: the share of patients for whom a physician is reimbursed on a capitated basis (on average, 15 percent), the share of a physician's patients whose primary health insurance is Medicare (42 percent), and the share of a physician's patients on Medicaid (10 percent), with capitated payment and Medicaid generally associated with lower marginal reimbursement. A second set of questions asks about characteristics of the physician and her practice: 29 percent work in small practices (solo or 2-person), 57 percent work in single or multi-specialty group practices, and 13 percent work in HMOs or hospital-based practices. We also observe a number of characteristics about the physician, including age, gender, whether she is board certified, and the number of days per week she spends seeing patients.

Third, the survey asks about a physician's self-assessed responsiveness to external incentives, including how frequently, if ever, she had intervened on a patient for nonclinical reasons in the past 12 months. We create a set of binary variables that indicates whether a physician responded to each set of incentives at least "sometimes" (i.e., "sometimes" or "frequently") over the past year. For example, among cardiologists, 31 percent reported that they had sometimes or frequently performed a cardiac catheterization because of the expectations of the referring physician and 38 percent of all physicians reported having intervened medically because of a colleague's expectations (Table 1).

Table 5 presents coefficients from a linear probability model with HRR-level random effects for three regressions at the physician level (results are similar with clustering at the HRR level). Our dependent variables are binary indicators for whether the physician is a cowboy (column 1), a comforter (column 2), or recommends high frequency follow-up (column 3), the three categories with predictive power in the main specifications (Table 3). In each model, we include

TABLE 5—PREDICTORS OF COWBOY, COMFORTER, AND HIGH FOLLOW-UP PHYSICIAN TYPES

	Cowboy physician (1)	Comforter physician (2)	High follow-up physician (3)
Panel A. General controls			
Cardiologist dummy	0.005	−0.092	0.222
	(0.042)	(0.048)	(0.029)
Age	0.003	0.000	0.004
	(0.001)	(0.002)	(0.001)
Male	0.027	−0.072	−0.078
	(0.037)	(0.042)	(0.025)
Weekly patient days	−0.011	0.024	0.008
	(0.009)	(0.011)	(0.006)
Board certified	−0.033	−0.035	0.010
	(0.043)	(0.049)	(0.030)
Panel B. Reimbursement			
Fraction capitated patients	0.001	-0.000	0.000
	(0.001)	(0.001)	(0.000)
Fraction Medicaid patients	0.004	0.001	0.001
	(0.001)	(0.001)	(0.001)
Fraction Medicare patients	0.001	0.000	0.000
	(0.001)	(0.001)	(0.000)
Panel C. Organizational factors			
(Baseline = Solo or two-person practice)	—	—	—
Single/multi speciality group practice	−0.101	−0.008	−0.164
	(0.032)	(0.037)	(0.022)
Group/staff HMO or hospital-based practice	−0.171	0.038	−0.153
	(0.047)	(0.053)	(0.032)
Panel D. Responsiveness factors			
Responds to patient expectations	−0.036	0.047	−0.007
	(0.036)	(0.041)	(0.025)
Responds to colleague expectations	−0.015	−0.005	0.026
	(0.030)	(0.035)	(0.021)
Responds to referrer expectations	0.061	−0.022	−0.033
	(0.046)	(0.053)	(0.032)
Responds to malpractice concerns	0.041	0.045	0.006
	(0.030)	(0.034)	(0.021)
Observations	978	978	978
R^2 (within)	0.048	0.046	0.150
R^2 (between)	0.075	0.009	0.117
R^2 (overall)	0.059	0.040	0.157

Notes: All logit regressions include a constant, HRR-level random effects, and general physician-level controls. Additional explanatory variables include financial, organizational and responsiveness factors. The question on responding to referring doctor expectations appeared in the cardiologist survey only, and therefore represents the preferences of cardiologists only; the cardiology dummy variable therefore reflects both the pure effect of being a practicing cardiologist, and a secondary adjustment arising from the referral question being set to zero for all primary care physicians. All results are robust to including a measure of capacity (cardiologists per 100k Medicare beneficiaries) among the explanatory variables.

basic physician demographics: age, gender, board certification status, whether the physician is a cardiologist, days per week spent seeing patients.

Notably, some physician characteristics are predictive of physician types; cardiologists are significantly more likely (by over 20 percentage points) to be high follow-up doctors, and older physicians in the sample are somewhat more likely to be cowboys and high follow-up doctors. A 20-year increase in age is associated with a 6 percentage point increase in the probability of being a cowboy and a 8 percentage point increase in the probability of being a high follow-up doctor. Physicians seeing patients on more days per week are more likely to be comforters.

A substitution effect would imply that lower incremental reimbursements associated with treating a larger share of Medicaid and capitated patients would lead to fewer interventions and more palliative care, while an income effect would suggest the opposite. Table 5 shows that physicians with a larger fraction of Medicaid (but not capitated patients) are more likely to be cowboys, supporting a stronger income effect in our sample.

Some organizational factors are clearly associated with physician beliefs about appropriate practice. Physicians in solo or two-person practices are far more likely to be both cowboys and high follow-up doctors than physicians in single or multi-specialty group practices or physicians who are part of an HMO or a hospital-based practice. Whether cardiologists accommodate referring physicians—a financial factor (since cardiologists benefit financially from future referrals) as well as an organizational one—is a positive predictor of cowboy status among cardiologists.[29] Self-reported responsiveness to malpractice concerns are high (nearly 39 percent of physicians report intervening for this reason), but are not statistically significant predictors of physician type.[30] While individual-level regressions with binary dependent variables often exhibit modest R^2 (the largest is just 0.16), the limited explanatory power of these regressions suggests a key role for physician beliefs regarding the productivity of treatments, rather than behaviors systematically related to financial, organizational, or other factors.

D. *Challenges to Identification*

One concern with the use of vignettes is the possibility that physicians who practice in areas with high rates of prevailing illness may recommend higher rates of follow-up care simply because their patients are sicker, on average. This would then lead to a spurious correlation between the fraction of high follow-up (or cowboy) physicians and spending (Sheiner 2014). Empirically, however, physicians in the quartile of HRRs with the greatest level of poverty are *less* likely to recommend a follow-up visit within three months than the average physician, although the results

---

[29] Note that the question on responding to referring doctor expectations appeared in the cardiologist survey only, and so reflects the preferences of cardiologists only; the cardiology dummy variable therefore reflects both the pure effect of being a practicing cardiologist, and a secondary adjustment arising from the referral question being set to zero for all PCPs.

[30] Perhaps because procedures performed on high-risk patients (such as Patients B and C) can increase the risk of a malpractice suit (Currie and MacLeod 2008).

are not significant. Similarly, there was a negative but statistically insignificant association between the fraction of a physician's Medicaid patients and the likelihood of a quick follow-up visit.[31] Finally, in separate, unreported regressions, we confirm that results in Table 3 are robust to controlling for percent poverty and the percent of black patients at the HRR level.

Nor is it the case that more intensive treatment is medically appropriate in the setting of congestive heart failure patients, whether for low-income or high-income patients. For example, one of the possible "cowboy" interventions for Patient B, a 75-year-old with Class IV heart failure, is an implantable defibrillator. Professional societies uniformly recommend against its use for Class IV CHF patients—and indeed, Medicare refuses to pay for it—because the risks (faulty leads, infections, painful death) outweigh the benefits for the average patient. It would be even less appropriate for patients with poor access to follow-up care and higher mortality risk, as one might find in low-income states.

Finally, Sheiner (2014) notes that better practice styles might be associated with higher social capital (e.g., Skinner and Staiger 2007), and thus proxy for unobservable patient health conditional on conventional risk adjustment.[32] If so, moving from a state with better practice styles to worse practice styles should have no impact on spending in the short-term. Yet Finkelstein, Gentzkow, and Williams (2016) found that our measure of "cowboy cardiologists" explains significant differences in Medicare expenditures for elderly enrollees who move. These findings are consistent with the view that physician beliefs exert a large impact on health care utilization.

IV. Discussion and Conclusion

While there is a good deal of regional variation in medical spending and care utilization in the United States and elsewhere, there is little agreement about the causes of such variations. Do differences arise from variation in patient demand, variation in physician behavior, or both? In this paper, we find that regional measures of patient demand, as characterized by responses to a nationwide survey, have modest predictive associations with regional end-of-life expenditures (Anthony et al. 2009; Baker, Bundorf, and Kessler 2014). We find a much stronger role for local physician beliefs in explaining Medicare expenditures, even after adjusting for conventional demographic, organizational, and financial incentives facing physicians.

Although practice type is associated with treatment recommendations, physician beliefs are not otherwise strongly correlated with conventional organizational and financial incentives for physicians that typically play a central role in most economic models of physician behavior. Traditional factors in supplier-induced

---

[31] Nor is there evidence that physicians respond differently in prescribing follow-up visits based on patient race or income, although there were significant differences by sex (Arber et al. 2006).

[32] Recall that the AMI regressions adjust for the HCC score, so any additional biases related to social capital would arise from unobservable health factors uncorrelated with the HCC score. Even the HCC "risk-adjustment" score is likely to capture physician intensity, as patients treated more intensively tend to be diagnosed with more diseases (Song et al. 2010; Finkelstein, Gentzkow, and Williams 2016).

demand models, such as the fraction of patients paid through capitation (or on Medicaid), or physicians' responsiveness to financial factors, play a relatively small role in explaining equilibrium variations in utilization patterns in our context. Organizational factors, such as accommodating colleagues, explain somewhat more of the observed variation in individual intervention decisions, leaving a considerable degree of unexplained differences in physician beliefs about the effectiveness of specific procedures and treatments.[33]

Unfortunately, we cannot shed light on how these differences in physician beliefs arise. Previous hypotheses have included variation in medical training (Epstein and Nicholson 2009) or personal experiences with different interventions (Berndt et al. 2015). Yet simple heterogeneity in physician beliefs alone cannot explain regional variation in expenditures, since the observed regional patterns in physician beliefs exhibit far greater inter-region variation than would be expected due to chance: spatial correlation in physician beliefs is clearly implied by the data. We do find that physicians' propensity to intervene for nonclinical reasons is related to the expectations of physicians with whom they regularly interact, a result consistent with network models of diffusion (Moen et al. 2016). Similarly, Molitor (2018) finds that cardiologists who move to more or less aggressive regions change their practice style to better conform to local norms.

How much would Medicare expenditures change in a counterfactual setting in which there were no cowboys, all physicians were comforters, and all physicians met guidelines for follow-up care? While we recognize the conjectural nature of this calculation and the fact that our estimates are limited to the Medicare fee-for-service population in the HRRs where we have survey data, our results imply that in total, these regions would see a decline of 35 percent in end-of-life Medicare expenditures, 12 percent for spending following AMI, and up to 12 percent in total Medicare expenditures.[34]

There are several limitations to our study. One is that our speculative inferences above regarding the importance of physician beliefs are based solely on Medicare fee-for-service data, and may not extend to other populations. Indeed, Cooper et al. (2015) found a very small regional correlation between Medicare expenditures and commercially-insured under-65 individuals living in the same area. This finding is explained in part by the negative correlation between commercial prices and Medicare spending (Romley et al. 2015), but there is also evidence on the correlation between utilization rates (that is, quantities) between commercial and Medicare enrollees of roughly 0.6 or less (Newhouse et al. 2013; Chernew et al. 2010). Most relevant for our analysis, Baker, Fisher, and Wennberg (2008) used California data to compare end-of-life spending between Medicare and commercially insured fee-for-service and preferred provider enrollees, and found a correlation of 0.64 in hospital days during the last 2 years of life, suggesting a stronger correlation across age groups in similar types of treatment. Nonetheless, these studies suggest caution in making

---

[33] This result is consistent with Epstein and Nicholson (2009), who find large variations in Cesarean section surgical rates among obstetricians within the same practice, and with Chassin's (1993) "Enthusiasm Hypothesis"— that regional differences in the use of health care services are caused by differences in the prevalence of physicians who are enthusiasts for those services.
[34] The AMI estimates use column 3 from Table 4.

inferences about *overall* regional intensity from the beliefs of physicians regarding the treatment of cardiovascular disease, as the intensity of beliefs could be quite different for (e.g.) obstetricians in the same region treating low-birthweight babies.

Our results do not imply that economic incentives are unimportant. Clearly, changes in payment margins have a large impact on behavior, as has been shown in a variety of settings (Clemens and Gottlieb 2014; Jacobson et al. 2006). But the prevalence of geographic variations in European countries, where economic incentives are often blunted (Skinner 2011), is consistent with the view that economic incentives cannot entirely explain observed geographic variations in health utilization, and that physician beliefs could plausibly play a large role in explaining such variations. A better understanding of both how physician beliefs form, and how they can be changed, is therefore a key challenge for future policy and research.

REFERENCES

Ahmed, Ali, Wilbert S. Aronow, and Jerome L. Fleg. 2006. "Higher New York Heart Association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function." *American Heart Journal* 151 (2): 444–50.

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti. 2016. "The Long-Term-Care Insurance Puzzle: Modeling and Measurement." National Bureau of Economic Research (NBER) Working Paper 22726.

Ameriks, John, Andrew Caplin, Steven Laufer, and Stijn van Nieuwerburgh. 2011. "The Joy of Giving or Assisted Living? Using Strategic Surveys to Separate Public Care Aversion from Bequest Motives." *Journal of Finance* 66 (2): 519–61.

Anthony, Denise L., M. Brooke Herndon, Patricia M. Gallagher, Amber E. Barnato, Julie P. W. Bynum, Daniel J. Gottlieb, Elliott S. Fisher, et al. 2009. "How Much Do Patients' Preferences Contribute to Resource Use?" *Health Affairs* 28 (3): 864–73.

Arber, Sara, John McKinlay, Ann Adams, Lisa Marceau, Carol Link, and Amy O'Donnell. 2006. "Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: A video-simulation experiment." *Social Science and Medicine* 62 (1): 103–15.

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler. 2014. "Patients' Preferences Explain a Small but Significant Share of Regional Variation in Medicare Spending." *Health Affairs* 33 (6): 957–63.

Baker, Laurence C., Elliott S. Fisher, and John E. Wennberg. 2008. "Variations in Hospital Resource Use for Medicare and Privately Insured Populations in California." *Health Affairs* 27 (2): w123–34.

Barnato, Amber E., M. Brooke Herndon, Denise L. Anthony, Patricia M. Gallagher, Jonathan S. Skinner, Julie P. W. Bynum, and Elliott S. Fisher. 2007. "Are Regional Variations in End-Of-Life Care Intensity Explained by Patient Preferences?: A Study of the US Medicare Population." *Medical Care* 45 (5): 386–93.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *Quarterly Journal of Economics* 112 (2): 537–79.

Baumann, Andrea O., Raisa B. Deber, Gail G. Thompson. 1991. "Overconfidence among physicians and nurses: The 'micro-certainty, macro-uncertainty' phenomenon." *Social Science and Medicine* 32 (2): 167–74.

Bederman, S. Samuel, Peter C. Coyte, Hans J. Kreder, Nizar N. Mahomed, Warren J. McIsaac, and James G. Wright. 2011. "Who's in the Driver's Seat? The Influence of Patient and Physician Enthusiasm on Regional Variation in Degenerative Lumbar Spinal Surgery: A Population-Based Study." *Spine* 36 (6): 481–89.

Berndt, Ernst R., Robert S. Gibbons, Anton Kolotilin, and Anna Levine Taub. 2015. "The heterogeneity of concentrated prescribing behavior: Theory and evidence from antipsychotics." *Journal of Health Economics* 40: 26–39.

Black, Dan A., Mark C. Berger, and Frank A. Scott. 2000. "Bounding Parameter Estimates with Nonclassical Measurement Error." *Journal of the American Statistical Association* 95 (451): 739–48.

Boden, William E., Robert A. O'Rourke, Koon K. Teo, Pamela M. Hartigan, David J. Maron, William J. Kostuk, Merril Knudtson, et al. 2007. "Optimal Medical Therapy with or without PCI for Stable Coronary Disease" *New England Journal of Medicine* 356 (15): 1503–16.

Chandra, Amitabh, and Jonathan Skinner. 2012. "Technology Growth and Expenditure Growth in Health Care." *Journal of Economic Literature* 50 (3): 645–80.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy* 115 (1): 103–40.

Chandra, Amitabh, and Douglas O. Staiger. 2017. "Identifying Sources of Inefficiency in Health Care." National Bureau of Economic Research (NBER) Working Paper 24035.

Chassin, Mark R. 1993. "Explaining Geographic Variations: The Enthusiasm Hypothesis." *Medical Care* 31 (Supplement 5): YS37–44.

Chernew, Michael E., Lindsay M. Sabik, Amitabh Chandra, Teresa B. Gibson, and Joseph P. Newhouse. 2010. "Geographic Correlation between Large-Firm Commercial Spending and Medicare Spending." *American Journal of Managed Care* 16 (2): 131–38.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review* 104 (4): 1320–49.

Comin, Diego, Jonathan Skinner, and Douglas Staiger. 2018. "Overconfidence and the Diffusion of Medical Technology." https://www.frbatlanta.org/~/media/Documents/research/seminars/2018/skinner-032118.pdf.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. Forthcoming. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *Quarterly Journal of Economics*.

Currie, Janet, and W. Bentley MacLeod. 2008. "First Do No Harm? Tort Reform and Birth Outcomes." *Quarterly Journal of Economics* 123 (2): 795–830.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending: Dataset." *American Economic Journal: Economic Policy*. https//doi.org/10.1257/pol.20150421.

Dartmouth Atlas of Health Care. 2018. Dartmouth Institute. http://www.dartmouthatlas.org/tools/downloads.aspx (accessed November 26, 2012).

Doyle, Joseph J., Jr. 2011. "Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far from Home." *American Economic Journal: Applied Economics* 3 (3): 221–43.

Dranove, David, and Paul Wehner. 1994. "Physician-induced demand for childbirths*." Journal of Health Economics* 13 (1): 61–73.

Dresselhaus, Timothy R., John W. Peabody, Jeff Luck, and Dan Bertenthal. 2004. "An evaluation of vignettes for predicting variation in the quality of preventive care." *Journal of General Internal Medicine* 19 (10): 1013–18.

Epstein, Andrew J., and Sean Nicholson. 2009. "The formation and evolution of physician treatment styles: An application to cesarean sections." *Journal of Health Economics* 28 (6): 1126–40.

ESCAPE Investigators and ESCAPE Study Coordinators (**ESCAPE**)**.** 2005. "Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness: The ESCAPE Trial." *JAMA* 294 (13): 1625–33.

Evans, Spencer C., Michael C. Roberts, Jared W. Keeley, Jennifer B. Blossom, Christina M. Amaro, Andrea M. Garcia, Cathleen Odar Stough, et al. 2015. "Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies." *International Journal of Clinical and Health Psychology* 15 (2): 160–70.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–1726.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Étoile 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): 273–87.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Étoile 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): 288–98.

Gogineni, Keerthi, Katherine L. Shuman, Derek Chinn, Nicole B. Gabler, and Ezekiel J. Emanuel. 2015. "Patient Demands and Requests for Cancer Tests and Treatments." *JAMA Oncology* 1 (1): 33–39.

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn Gilman Andrews, Jonathan S. Skinner, and Jason M. Sutherland. 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs* 29 (3): 537–43.

Hunt, Sharon Ann, William T. Abraham, Marshall H. Chin, Arthur M. Feldman, Gary S. Francis, Theodore G. Ganiats, Mariell Jessup, et al. 2005. "ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult." *Circulation* 112 (12): e154–235.

Jacobson, Mireille, A. James O'Malley, Craig C. Earle, Juliana Pakes, Peter Gaccione, and Joseph P. Newhouse. 2006. "Does Reimbursement Influence Chemotherapy Treatment for Cancer Patients?" *Health Affairs* 25 (2): 437–43.

Keating, Nancy L., Mary Beth Landrum, Haiden A. Huskamp, Elena M. Kouri, Holly G. Prigerson, Deborah Schrag, Paul K. Maciejewski, et al. 2016. "Dartmouth Atlas Area-Level Estimates of End-of-Life Expenditures: How Well Do They Reflect Expenditures for Prospectively Identified Advanced Lung Cancer Patients?" *Health Services Research* 51 (4): 1584–94.

Keeley, Ellen C., Judith A. Boura, and Cindy L. Grines. 2003. "Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: A quantitative review of 23 randomised trials." *Lancet* 361 (9351): 13–20.

Kelley, Amy S., Evan Bollens-Lund, Kenneth E. Covinsky, Jonathan S. Skinner, and R. Sean Morrison. 2018. "Prospective Identification of Patients at Risk for Unwarranted Variation in Treatment." *Journal of Palliative Medicine* 21 (1): 44–54.

Lucas, F. L., A. E. Siewers, D. J. Malenka, and D. E. Wennberg. 2008. "Diagnostic-Therapeutic Cascade Revisited: Coronary Angiography, Coronary Artery Bypass Graft Surgery, and Percutaneous Coronary Intervention in the Modern Era." *Circulation* 118 (25): 2797–2802.

Lucas, Frances Lee, Brenda E. Sirovich, Patricia M. Gallagher, Andrea E. Siewers, and David E. Wennberg. 2010. "Variation in Cardiologists' Propensity to Test and Treat: Is It Associated with Regional Variation in Utilization?" *Circulation: Cardiovascular Quality and Outcomes* 3 (3): 253–60.

Mandelblatt, Jeanne S., Leigh Anne Faul, George Luta, Solomon B. Makgoeng, Claudine Isaacs, Kathryn Taylor, Vanessa B. Sheppard, et al. 2012. "Patient and Physician Decision Styles and Breast Cancer Chemotherapy Use in Older Women: Cancer and Leukemia Group B Protocol 369901" *Journal of Clinical Oncology* 30 (21): 2609–14.

McGuire, Thomas G. 2011. "Physician Agency and Payment for Primary Medical Care." In *The Oxford Handbook of Health Economics*, edited by Sherry Glied and Peter C. Smith, 602–23. Oxford: Oxford University Press.

McGuire, Thomas G., and Mark V. Pauly. 1991. "Physician response to fee changes with multiple payers." *Journal of Health Economics* 10 (4): 385–410.

Moen, Erika L., Andrea M. Austin, Julie P. Bynum, Jonathan S. Skinner, and A. James O'Malley. 2016. "An analysis of patient-sharing physician networks and implantable cardioverter defibrillator therapy." *Health Services and Outcomes Research Methodology* 16 (3): 132–53.

Moliter, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy* 10 (1): 326–56.

Newhouse, Joseph P., and Alan M. Garber. 2013. "Geographic Variation in Medicare Services." *New England Journal of Medicine* 368 (16): 1465–68.

Newhouse, Joseph P., Alan M. Garber, Robin P. Graham, Margaret A. McCoy, Michelle Mancher, and Ashna Kibria, eds. 2013. *Variation in Health Care Spending: Target Decision Making, Not Geography*. Washington, DC: National Academies Press.

Pauly, Mark V. 1980. *Doctors and Their Workshops: Economic Models of Physician Behavior.* Chicago: University of Chicago Press.

Peabody, John W., Jeff Luck, Peter Glassman, Timothy R. Dresselhaus, and Martin Lee. 2000. "Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of 3 Methods for Measuring Quality." *JAMA* 283 (13): 1715–22.

Peabody John W., Jeff Luck, Peter Glassman, Sharad Jain, Joyce Hansen, Maureen Spell, and Martin Lee. 2004. "Measuring the Quality of Physician Practice by Using Clinical Vignettes: A Prospective Validation Study." *Annals of Internal Medicine* 141 (10): 771–80.

Ransohoff, David F., Mary McNaughton Collins, and Floyd J. Fowler, Jr. 2002. "Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback." *American Journal of Medicine* 113 (8): 663–67.

Rao, Jaya K., Lynda A. Anderson, Feng-Chang Lin, and Jeffrey P. Laux. 2014. "Completion of Advance Directives Among U.S. Consumers." *American Journal of Preventive Medicine* 46 (1): 65–70.

Romley, John A., Sarah Axeen, Darius N. Lakdawalla, Michael E. Chernew, Jay Bhattacharya, and Dana P. Goldman. 2015. "The Relationship between Commercial Health Care Prices and Medicare Spending and Utilization." *Health Services Research* 50 (3): 883–96.

Romley, John A., Anupam B. Jena, and Dana P. Goldman. 2011. "Hospital Spending and Inpatient Mortality: Evidence from California: An Observational Study." *Annals of Internal Medicine* 154 (3): 160–67.

Sheiner, Louise. 2014. "Why the Geographic Variation in Health Care Spending Cannot Tell Us Much about the Efficiency or Quality of our Health Care System." *Brookings Papers on Economic Activity* 44 (2): 1–52.

Sirovich, Brenda, Patricia M. Gallagher, David E. Wennberg, and Elliott S. Fisher. 2008. "Discretionary Decision Making by Primary Care Physicians and the Cost of U.S. Health Care." *Health Affairs* 27 (3): 813–23.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*, Vol. 2, edited by Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, 45–93. Amsterdam: North-Holland.

Skinner, Jonathan, and Douglas Staiger. 2007. "Technological Adoption from Hybrid Corn to Beta-Blockers." In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, edited by Ernst R. Berndt and Charles R. Hulten, 545–70. Chicago: University of Chicago Press.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E. Wennberg, and Elliott S. Fisher. 2010. "Regional Variations in Diagnostic Practices." *New England Journal of Medicine* 363 (1): 45–53.

U.S. Department of Health and Human Services. 2004. *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*. National Institute of Health. Washington, DC, August.

Wennberg, John E., Benjamin A. Barnes, and Michael Zubkoff. 1982. "Professional uncertainty and the problem of supplier-induced demand." *Social Science and Medicine* 16 (7): 811–24.

Wennberg, David E., John D. Dickens, Jr., Lois Biener, Floyd J. Fowler, Jr., David N. Soule, and Robert B. Keller. 1997. "Do Physicians Do What They Say? The Inclination to Test and Its Association with Coronary Angiography Rates." *Journal of General Internal Medicine* 12 (3): 172–76.