1 March, 2013

Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Spending

David Cutler
Harvard University
National Bureau of Economic Research
dcutler@fas.harvard.edu

Jonathan Skinner
Department of Economics, Dartmouth College
Dartmouth Institute for Health Policy & Clinical Practice
National Bureau of Economic Research
jon.skinner@dartmouth.edu

Ariel Dora Stern
Harvard University
arieldorastern@fas.harvard.edu

David Wennberg
Dartmouth Institute for Health Policy & Clinical Practice
david.e.wennberg@dartmouth.edu

Version 1.1: Please do not quote

Abstract

There is considerable controversy about the causes of risk-adjusted regional variations in healthcare expenditures. We use three surveys to test whether demand-side factors such as patient preferences, and supply-side factors such as financial incentives, organizational factors, and professional beliefs, can potentially explain observed regional variations in Medicare expenditures. The first survey asked elderly Medicare enrollees about preferences for end-of-life care. The second and third asked cardiologists and primary care physicians about organizational and financial pressures, and presented them with vignettes on how they would treat specific patients. The surveys were linked to Medicare end-of-life utilization data at the hospital-referral region (HRR) level. Our results show that physician beliefs about appropriate care have a large impact on end-of-life spending and matter much more than do patient preferences. While organizational factors influence physician beliefs, financial considerations appear not to. The best explanation for large observed variations in views of appropriate treatment is physician beliefs about their own productivity. Finally, many physician beliefs are more aggressive than warranted by professional guidelines, highlighting the likelihood for waste in U.S. health care.

Comments by Amitabh Chandra, Nancy Morden, Allison Rosen, Gregg Roth, Tisamarie Sherry, Sara Nadel, Pascal St.-Amour, and seminar participants at the NBER Summer Institute, Tinbergen Institute, Cornell University, Tilburg University, Erasmus University, CRISP Health Econometrics, University of Venice, Lausanne University, NBER's Health Care Program Meeting and Harvard University were exceedingly helpful. We are grateful to F. Jack Fowler and Patricia Gallagher of the University of Massachusetts Boston for developing the patient and physician questionnaires. Funding from the National Institute on Aging (T32-AG000186-23 to the National Bureau of Economic Research and P01-AG019783 to Dartmouth) and LEAP at Harvard University (to Skinner) is gratefully acknowledged. Survey data are available at www.intensity.dartmouth.edu.

Regional variations in rates of medical treatments are large in the United States and other countries. For example, in the U.S. Medicare population over age 65, rates of back surgery vary by a factor of more than four (Dartmouth Atlas, 2012), while in England rates of tonsillectomies exhibit seven-fold variation across regions (Suleman et al., 2010). Price-adjusted annual per-patient Medicare expenditures range from $6,424 in Salem, OR to $15,571 in Miami, FL, with most of the variation unexplained by regional differences in patient illness or poverty (Sutherland, et al., 2010; Zuckerman et al., 2010).

What drives such variation in treatment and spending? One possibility is patient demand. Many studies have been done in situations where all patients have a similar and fairly generous insurance policy,[1] so price and income differences are unlikely to be large. Still, heterogeneity in patient preferences for care may play a role. In the case of severe coronary disease, for example, some patients may prefer to try all possible measures, while others may prefer palliation and an out-of-hospital death. If patients with similar preferences group together geographically – for example, if people who value life-prolonging treatments live in areas with "world-class" interventional physicians – patient preference heterogeneity could lead to regional variation in equilibrium outcomes (Mandelblatt et al., 2012; Anthony et al., 2010).

Another source of variations arises from the supply side, with the most obvious explanation being differences across regions in profit margins for medical service providers. In a principle-agent framework (McGuire and Pauly, 1991), if prices are high enough (and income scarce), the patient's physician-agent recommends care beyond what is medically appropriate, leading to the canonical "supplier-induced demand." While physician utilization has been shown

[1] This is generally true in the U.S. Medicare program. The presence of supplemental insurance coverage differs across the country, but most studies do not find that these differences affect utilization by more than a small degree. Similarly, income elasticities of spending in insured populations are modest at best (Newhouse et al., 2003; McClellan and Skinner, 2006)

to be sensitive to prices (Jacobson et al., 2006, Clemens and Gottlieb, 2012), it would be difficult to explain observed Medicare variations using profit margins alone, since reimbursement rates are set administratively and do not vary greatly across areas.

At its broadest, “supplier-induced demand” describes a situation in which a health care provider shifts a patient’s demand curve. Variation in desired supply may also result from non-monetary incentives. Physicians may respond to organizational pressure or peer pressure to perform more procedures (de Jong, 2008), even if their current income is no higher as a consequence. For example, cardiologists may infer that inserting stents keeps referring primary care physicians happy and recommend more stents as a result. Independent of these factors, physicians may have different beliefs about appropriate treatments, particularly for conditions where there are few professional guidelines (Wennberg and Gittelsohn, 1975; Wennberg et al., 1982). These differences in physician beliefs may arise because of differences in where physicians were trained (Epstein and Nicholson, 2009) or their personal experience with different treatments (Levine-Taub et al., 2011). If this variation is correlated spatially – for example, if intensive physicians are more likely to hire other physicians with similar views – the resulting regional differences in perceptions about optimal treatment could explain regional variations in equilibrium spending outcomes.

It has proven difficult to separate out physician preferences, patient preferences, and other factors as they affect care decisions (Dranove and Wehner, 1994). To a great extent, this is because both patient and physician preferences are unknown and the two cannot be differentiated using just equilibrium outcomes. We address this problem using “strategic surveys” (as in Ameriks et al., 2011) that ask both providers and patients about motivations, clinical beliefs, and care preferences, and are in turn linked to average equilibrium measures of utilization at the regional level.

Patient preferences are measured by a survey of Medicare enrollees age 65 and older asking (N=1,413) about their preferences for a variety of aggressive care interventions.[2] We focus on the tradeoff between invasive procedures with potential longevity benefits versus palliative care and comfort at the end of life. Physician beliefs are captured by two surveys, one of cardiologists (N=516) and the second of primary care physicians (N=807). Both sets of physicians were presented with vignettes about four (largely overlapping) elderly individuals with chronic health conditions, and asked how they would manage and/or treat each one. These hypothetical patients ran the gamut from less severe stable angina to very severe late-stage congestive heart failure. Based on their responses, we characterize physicians along two dimensions: those who consistently and unambiguously recommended intensive care beyond that suggested by the medical literature and industry guidelines (or "cowboys"), and those who consistently recommended palliative care for the very severely ill ("comforters"). Physicians could be classified (and some were) along both margins.

We first use these surveys to examine whether patient or physician preferences are more important in explaining regional variations in care. Our results are very clear: the impact of physician preferences is significantly greater than the impact of patient preferences in explaining regional utilization patterns. In some models, we can explain over half of the variation in end-of-life spending across areas knowing only how a small sample of physicians in an area would treat hypothetical patients. In contrast, patient preferences explain little of the cross-area variation.

[2] The survey, funded by the National Institute on Aging, was conducted in 2005 by The Center for Survey Research (CSR) at the University of Massachusetts, Boston and Dartmouth (Geisel) Medical School under the direction of F. Jack Fowler and Patricia Gallagher.

We then try to understand why physicians have the treatment preferences they do, relating physicians' views about optimal treatment to questions about financial arrangements (fraction of Medicaid patients, fraction of capitated patients), organizational pressures (the frequency with which the cardiologist had recently accommodated colleague or referring physicians expectations, provided treatments for patients who expected but didn't need them), and malpractice concerns. We find that only a small fraction of physicians claim to have made recent intervention decisions as a result of financial considerations. On the other hand, we find that organizational factors reflected by "pressure to accommodate" either patients (by providing treatments that are not needed) or referring physicians (doing procedures to keep them happy and meet their expectations) were much more common among responding cardiologists, exerting a small but significant impact on physician beliefs about appropriate care.

Ultimately, the largest degree of regional variation in appears to be explained by exogenous 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, and these views are not highly correlated with demographics, background, or practice characteristics, and are often not consistent with then-current professional guidelines for appropriate care. These results are strongly suggestive of considerable waste in U.S. healthcare: high levels of expenditures for treatments with no evidence of medical value and no relationship to patient preferences.

I. A Model of Variation in Utilization

We develop a simple model of patient demand and physician supply and test the implications of the model using individual-level survey data from both patients and physicians, linked by geography and patterns of service utilization. The demand side of the model is a standard one; the indirect utility function is a function of out-of-pocket prices (p, which are generally low among Medicare beneficiaries), income (Y), and preferences for care (η); $V = V(p, Y, \eta)$. Solving this for optimal intensity of care, x, yields x^D. We assume that x^D is the patient's demand for the quantity of procedures when they are first stepping into the physician's office, prior to any demand "inducement." We approximate x^D empirically by the use of our patient survey with hypothetical questions about preferences for more aggressive care near the end of life.

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

Individual physicians are assumed to be price-takers (once their networks have negotiated prices with insurance companies), but face a wide range of reimbursement rates from private insurance, Medicare, and Medicaid. The model is therefore simpler than models in which hospital groups and physicians jointly determine quantity, quality, and price (Pauly, 1980) or where physicians exercise market power over patients to provide them with "too much" health care (McGuire, 2011). Following Chandra and Skinner (2012), we can write a physician's overall utility as:

(1) $$U = \Psi s(x) + \Omega(W + \pi x - R) - \phi(|x - x^D|) - \varphi(|x - x^O|)$$

where Ψ is the perceived social value of improving health, Ω is the utility function of the physician's own income, comprising her fixed payment W (in a salaried setting, for example) net of fixed costs R, and including the incremental "profits" from each additional test or procedure performed, π.[3] Note that π may be negative or positive depending on the type of procedure and who is paying for it.

The third term represents the loss in provider utility arising from the deviation between what the provider prescribes (x) and what the patient wants (x^D). This function could reflect classic supplier-induced demand – from the physician's point of view, x^D is too low relative to the physician's belief in optimal x – or it may reflects the extent to which physicians are acting as the agent of the (possibly misinformed) patient, for example when the patient wants a procedure that the physician does not feel is medically appropriate. The fourth term reflects a parallel influence on physician decision making from organizational factors that do not directly affect financial rewards, such as (physician) peer pressure or concerns about staying in referring physicians' good graces. For example, Moliter (2011) showed that when cardiologists move to a new region, their practice style adjusts partially to resemble that of their new peers.

The first-order condition for (1) is:

(2) $$\Psi s'(x) = -\Omega'\pi + \phi' + \varphi' \equiv \lambda$$

Physicians provide care up to the point where the choice of x reflects a balance between the perceived marginal value to health, $\Psi s'(x)$, and factors summarized by λ: (a) the incremental change in net income π, weighted by the importance of financial resources Ω′, (b) the incremental disutility from moving patient demand away from where it was originally, and (c) the incremental disutility from how much the physician's own choice of x deviates from her organization's perceived optimal level of intervention.

---

[3] We ignore here capacity constraints, such as the supply of hospital or ICU beds, or the availability of MRIs and catheterization laboratories; see for example Wennberg, et al. (2002) and Wennberg (2010). These are likely to be more important in non-U.S. health care systems with a greater degree of centralized planning for specialized facilities.

In this model, there are two ways to define conventional "supplier-induced demand." The broadest definition is simply the presence of any equilibrium quantity of care provided greater than *ex ante* patient preferences, i.e. $x > x^D$. This is still quite benign; after all, the patient seeks out the physician's opinion for a reason, so x^D may optimally change after the visit. More relevant is the sign of $s'(x^D)$; does the change in demand "induced" enhance or diminish health outcomes for the patient? Supplier-induced demand would then more narrowly be defined as $s'(z) \leq 0$ as z ranges from x^D to x; that patients gain no improvement in health outcomes (and may even experience a decline), presumably spending more through copays, deductibles, and time-costs because they followed their physician's advice (McGuire, 2011).

To develop an empirical model, we adopt a simple closed-form solution of the utility function for region i, which is written:[4]

(1′) $$U_i = \Psi s_i(x_i) + \omega[W_i + \pi_i x_i - R_i] - \frac{\phi}{2}(x_i - x_i{}^D)^2 - \frac{\varphi}{2}\left(x_i - x_i{}^O\right)^2$$

Note that ω/Ψ reflects the relative tradeoff between the physician's income and the value of improving patient lives, and thus might be viewed as a measure of "professionalism" in the sense of Campbell and colleagues (Campbell et al., 2007). The first-order condition is therefore:

(2′) $$\Psi s_i'(x_i) = \lambda \equiv -\omega\pi_i + \phi(x_i - x_i{}^D) + \varphi(x - x_i{}^O)$$

Figure 1 shows both $\Psi s'(x)$ and λ for this solution to the model. Note that λ is linear in x with an intercept equal to $-(\omega\pi_i + \phi x_i{}^D + \varphi x_i{}^O)$. Note also the key assumption that patients are sorted in order from most appropriate to least appropriate for treatment, thus describing a downward sloping $\Psi s'(x)$ curve. The equilibrium is where $\Psi s'(x) = \lambda$, at point A in Figure 1. Note that a shift in the intercept, which depends on reimbursement rates for procedures π, taste for income ω, regional demand x^D, and organizational or peer effects x^O, would yield a different λ*, and hence a

[4] We are grateful to Pascal St.-Amour for suggesting this approach.

different utilization rate. Note that all of these factors affect the intensity of treatments by a movement *along* the marginal benefit curve, $\Psi s'(x)$.

Alternatively, it may be that $s_i'(x)$ differs across physicians and regions – representing shifts in the productivity curve rather than movements along the curve. For example, if $s_i'(x) = \alpha_i\, s'(x)$, where $s'(x)$ is average physician productivity and α varies across regions, we can instead capture these differences by thinking of a shift in the marginal benefit curve. In Figure 1, point C corresponds to greater level of intensity than point A and arises naturally when the physician is and/or believes that she is more productive at any level of x.[5] 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). Because patients in regions with high intervention rates actually benefit differentially from these interventions, this scenario would not correspond to the narrow definition of "supplier-induced demand" because surgical patients are not being "over treated" or harmed.

The productivity shifter α in Figure 1 may also vary because of what Wennberg et. al. (1982) term "professional uncertainty," or differences in physician beliefs where – especially given how much professional disagreement exists among physicians regarding specific treatments – it's very likely that a given physician's perceived α differs from the true α. For example, physicians may also be overly optimistic with respect to their own ability to perform procedures (or to prescribe treatments), leading to expected or perceived benefits that exceed actual realized benefits.

Baumann et al. (1991) have documented the phenomenon of "macro uncertainty, micro certainty" in which physicians and nurses are very sure that their treatment benefited a specific

[5] This would also hold if either the patient or physician placed a greater relative weight on the social value of treatments, so that Ω differed across physicians or patients.

patient (micro certainty) even when there is no general consensus on which procedure is more clinically effective (macro uncertainty). Much of the evidence from psychology and the medical literature points to overconfidence in one's own ability, leading to a natural bias towards physicians doing more.[6] To see this in Figure 1, if the physician's perceived benefit is g′(x) but the actual benefit is s′(x) in Figure 1, then at point C, the incremental treatment harms the patient – even though the physician does not believe she is doing so. In equilibrium, this supplier behavior would appear consistent with the classic supplier-induced demand hypothesis – higher expenditures that are not associated with better outcomes – but the cause is quite different; in this case the physician believes she is maximizing utility for both herself and for her patient, and does not perceive any tradeoff.

Empirical Specification. To examine these theories empirically, we consider variation in practice at the regional level (for reasons explained below). Taking a first-order Taylor-series approximation of Equation (2′) for region i yields a linear equation that groups equilibrium outcomes into two basic components, demand factors Z^D and supply factors Z^S:

(4) $$x_i = \bar{x} + Z_i^D + Z_i^S + \varepsilon_i.$$

The demand-side component is:

(5) $$Z_i^D = \frac{\phi}{\mathrm{M}}(x_i^D - \bar{x}^D)$$

where $M = -\Psi s''(\bar{x}) + \phi + \varphi$. This first element of Equation (5) reflects the higher average demand for health care, multiplied by the extent to which physicians accommodate that demand, ϕ. The supply side component is written as:

(6) $$Z_i^S = \frac{1}{M}\{\omega\Delta\pi_i + \pi\Delta\omega_i + \phi({x_i}^O - \bar{x}^O) + \Psi s'(\bar{x})\Delta\alpha_i\}$$

[6] There are often clear psychological biases towards doing more; if the patient gets better anyway, the physician gets the credit, but if the patient gets worse, the physician has done everything she could to save the patient's life (e.g., Ransohoff et al., 2002).

The first term in brackets of Equation (6) reflects how differences in region i relative to the national average ($\Delta\pi$) might affect utilization; this simply reflects the upward sloping supply curve. We expect that these effects are not large given that Medicare prices are standardized, although relative prices between Medicare and non-Medicare patients may affect the quantity supplied (for example if the physician treats a large fraction of capitated or Medicaid patients). The second term reflects the extent to which physicians weight income more heavily – or their greater willingness to perform procedures even when expected medical benefits are modest. The third term captures organizational goals in region i relative to national averages $({x_i}^O - \bar{x}^O)$. The final term captures the impact of different physician beliefs about productivity of the treatment ($\Delta\alpha_i$); this term shifts the marginal productivity curve.

Equation (4) can be expanded to capture varying parameter values as well – for example, in some regions physicians may be more responsive to patient demand, so that ϕ_i is larger than average. These interactive effects therefore reflect both supply and demand – that is, when patients want a specific treatment and physicians respond to those preferences, there can be larger effects on measured utilization.

Finally, note that these effects are all scaled by 1/M, which in turn depends on –s″. That is, the magnitude of the effects will be larger the smaller is the degree of diminishing returns to treatment. Effectively, if there are many patients who might (barely) benefit from the treatment, strongly-held physician opinions can lead to highly variable treatment levels and intensities (Chandra and Skinner, 2012).

II. Data and Estimation Strategy

In general, it is very difficult to distinguish among these different explanations for treatment variation; even highly detailed clinical data make it difficult to second-guess the patient's condition as the treating physician generally knows far more about the patient than the researcher does, and patient preferences and physician beliefs are unknown when using *ex post* clinical data. In studying motives for household saving, Ameriks et al. (2011) implemented "strategic surveys" to identify their otherwise difficult-to-identify models.[7] We follow this approach here, using surveys asking potential patients about preferences for hypothetical end-of-life choices (that is, x^D before their interaction with the physician), and asking physicians about how they would treat a set of hypothetical patients with varying disease severity, as well as questions about their financial and organizational constraints.

In an ideal world, we would have patient surveys matched with surveys from their treating physicians. Thus, we would have all parts of the puzzle for a set of patients and physicians. Our surveys were not matched, however, and scattered across the country. As an alternative, we match supply and demand at the area level (HRR, or Hospital Referral Region[8]). In Equation (4), we therefore define x to be the regional average spending measure. While our primary measure is the natural logarithm of local risk-adjusted and price-adjusted Medicare expenditures in the last two years of life, we also consider several others such as one-year risk- and price-adjusted expenditures for Medicare enrollees following hip fracture, risk-adjusted expenditures in the last six months of life, and overall price-adjusted expenditures.

[7] As they noted, it was nearly impossible to distinguish among saving motives based solely on how much households saved. They used the survey questions to ask about how respondents valued outcomes that could be affected by savings, for example the disutility of having to qualify for Medicaid if they failed to save enough for retirement.

[8] These HRRs are defined in the *Dartmouth Atlas of Health Care*, which divides the United States into 306 HRRs, defined by an algorithm that takes into account commuting patterns and the location of major referral hospitals.

Thus our first estimation, based on Equation (4), asks whether area-level supply or demand factors can better explain actual regional expenditures. Our second set of estimates then seeks to understand why physicians hold the beliefs they do (Equation 6). For the latter, we exploit our survey data at the individual physician level. We relate physicican beliefs to financial and organizational factors. The residual, which we do not measure, is treated as a set of physician beliefs, or shifts in the perceived marginal treatment curve from $\Psi s'(x)$ to $\Psi g'(x)$.

We use three sources of survey data: from patients, cardiologists, and primary care physicians (PCPs). Each survey was conducted by the Center for Survey Research (CSR), University of Massachusetts Boston, with clinical vignettes presented in Table 2.

Patient Survey. The survey sampling frame was all Medicare beneficiaries in the 20% denominator file who were age 65 or older on July 1, 2003 (see Barnato et al., 2009, for details).[9] A simple random sample of 4,000 people was drawn, and beneficiaries were contacted both by phone and mail; the response rate was 65%, with the final sample used in the analysis (limited to respondents who provided all variables of interest) containing a total of 1,413 Medicare beneficiaries. The final sample of respondents reside in 64 HRRs, all of which have sufficient physician observations to be included in the empirical model.

We used responses to 6 survey questions, the means of which are shown in Table 1.[10] Two of the questions relate to unnecessary care, asking people if they would like a test or cardiac

[9] The survey was conducted in both English and Spanish, and was piloted with 15 seniors in intensive one-on-one interviews to test construct validity.

[10] Following Barnato et al (2009), answers other than "yes" or "no" (e.g., "not concerned" or "I don't know") are treated as missing data. Item non-response was less than 1% among eligible respondents for each outcome measure.

referral even if their primary care physician did not think they needed one.[11] Overall, 71 percent of patients wanted such a test and 54 percent wanted a cardiac referral. There is wide variation across regions in averages responses to these question. Figure 2 shows the distribution of the share of patients responding that they wanted an unnecessary specialist referral.[12]

Other questions concerned receiving too little or too much care in the last year of life, preference for life-prolonging drugs with side-effects, for palliative drugs for comfort with potential for life-shortening, and for mechanical ventilation with potential for severe discomfort. We group these variables into two binary indices, one reflecting patients' desire for aggressive care at the end of life (in which patients respond that they would want to be put on a respirator if it would extend their life for either a week, or a month), and a second reflecting desire for comfort at the end of life (in which patients responded that if they reached a point at which they were feeling bad all of the time, they would want drugs to make them feel better, even if those drugs might shorten their life). In each case, we observe significant variation across areas (Table 1).

Patients' preferences are generally positively correlated across related items. For example, the correlation coefficient between wanting an unneeded cardiac referral and wanting an unnecessary test is 0.43 (the highest absolute value of a correlation coefficent among patients' preferences), whereas the correlation coefficient between wanting palliative care and wanting to be on a respirator was -0.02.

Since the questions patients are asked are hypothetical and typically do not reflect experiences that have not yet happened, we think of them as x^D – preferences not affected by

[11] This question thereby captures pure patient demand independent of what the physician wants. However, patients could still answer no, that they would not seek an additional referral, if they were unwilling to disagree with their physician.

[12] This figure uses the sample of 64 HRRs discussed below.

physician advice. Naturally, they will reflect the economic characteristics of the individual, such as copayment, household income, and health status. Yet survey responses may vary systematically by demographic covariates such as race and ethnicity. To address this potential for omitted variable bias, we created selection-adjusted HRR-level measures of preferences by adjusting for all of the predictive information that can be estimated from observed patient demographics (race, age and gender). Thus, the resulting variation in patient preferences is only that component of the variation that is uncorrelated with known demographic differences in such preferences.

Physician Surveys. A total of 999 cardiologists were randomly selected to receive the survey. Of these, 614 cardiologists responded, for a response rate of 61%. Seventeen physicians did not self-identify as cardiologists, and 81 physicians were missing crucial information (such as practice type) or practiced in HRRs with too few local respondents to include in the analysis, leaving us a final sample of 516 cardiologists. These cardiologists practice in 64 HRRs, all of which have 3 or more cardiologists represented.

The primary care physician responses come from a parallel survey of PCPs (family practice (FP), internal medicine (IM), or internal medicine/family practice (IFP)). A total of 1,333 primary care physicians were randomly selected to receive the survey.[13] Of the 1,333 physicians contacted, 967 responded, for a response rate of 73%. A total of 807 PCPs had complete responses to the survey and practiced in HRRs with enough local respondents to include in the analysis; our final PCP sample includes these 807 physicians in the same 64 HRRs noted above.

---

[13] The original sampling strategy included oversamples of physicians in four targeted cities, two in areas of low intensity (Minneapolis, MN; Rochester, NY) and two in areas of high intensity (Manhattan, NY; and Miami, FL). Each of these HRRs is represented in our final sample.

The physician surveys asked about a number of clinical vignettes, which we discuss in the next section. In addition, it asked about a variety of practice characteristics, which we will use to test the theories above. Two measures of financial circumstances are reported: the share of the physician's patients that is capitated – that is, a fixed amount of money is received for a given patient, regardless of the quantity of services provided – and the share on Medicaid. Both of these factors are associated with lower marginal reimbursement. As Table 1 shows, these shares respectively average about 16 and 10 percent in our combined sample of physicians.

A second set of questions ask about the organization of the practice by ownership and number of physicians: solo/2-person practice; single or multi-specialty group practices; and group or staff model HMO or a hospital-based practice. Physicians are least likely to share in ownership rewards in larger organizations. 29 percent of physicians are in small practices, 60 percent are in single or multi-specialty group practices, and 11 percent are in HMOs or hospital-based practices.

Third, the survey asks about physician responsiveness to external incentives, including how frequently, if ever, in the past 12 months they have intervened for non-clinical reasons – e.g. due to malpractice concerns, because their colleagues would do so in the same situation, or to meet the expectations of the referring doctor. For example, 8% of all physicians in the sample reported that they would "frequently" recommend cardiac catheterization (cardiologists) or make a cardiologist referral (PCPs) because their "colleagues would do so in the same situation," while 41% of all physicians would make the same set of recommendations at least "sometimes" (i.e. "sometimes or "frequently") in response to colleague expectations.

Finally, we observe a number of characteristics about the physician, including age, gender, whether the physician is board certified, and the number of weekly patient days practiced.

Medicare Utilization Data. We merge the survey responses with expenditure data in the last two years of life among Medicare enrollees over age 65 with the following life-threatening illnesses: congestive heart failure (the focus of some vignettes below), 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. All HRR-level measures are adjusted for age, sex, race, differences in Medicare reimbursement rates, and the type of disease (including an indicator for multiple diseases). This measure implicitly adjusts 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 in West Virginia or Oregon.[14] These end-of-life measures are commonly used to instrument for health care intensity, (e.g., Fisher et al., 2003a,b; Doyle, 2011; Silber et al., 2010; Romley et al., 2011) and are also highly correlated with other medical expenditure measures such as one-year expenditures following a heart attack (Skinner et al., 2010). In sensitivity analysis, we also consider a data set of "forward-looking" measures for a cohort of price- and risk-adjusted hip-fracture patients, price-adjusted six-month end-of-life expenditures and overall price-adjusted expenditures.

We focus on the 64 HRRs in the combined sample with a minimum of 3 cardiologists (average =5.4) and 2 primary care physicians (average = 7.9) surveyed. Of the 2,515 responding Medicare enrollees in the patient survey, 1,413 enrollees live in one of those 64 HRRs, for an average of 22 respondents per HRR.

III. Clinical Vignettes from the Physician Surveys

[14] If more intensive spending saves lives, then in regions with more intensive spending, there would be fewer people who die, and hence the end-of-life measure could be biased (e.g., Bach et al., 2004). However, the bias would be small and in an uncertain direction – adding or subtracting just one or two percent of the denominator would have little impact on per-capita expenditures.

Since the clinical vignettes are crucial for our analysis, we describe them in some detail. We note first the obvious: the response to the vignette may not be what the physician would actually do in practice. Empirical evidence, however, strongly indicates that vignettes in clinical settings closely predict how physicians actually intervene (Peabody et al., 2004; Mandelblatt et al., 2012; Dresselhaus et al., 2004).

We assume that the physician's responses to the vignettes are "all in" measures (Z^S, as in Equation 6), reflecting physician beliefs as well as the variety of financial, organizational, and capacity-related constraints she faces. Alternatively, one could interpret the physician's responses to the vignettes as a pure reflection of beliefs (for example, how one might answer for qualifying boards), and not as representative of the day-to-day realities of their practice. We tested this alternative explanation by including the organizational and financial variables in our estimation equations in addition to the vignette estimates. This did not appreciably increase the explanatory power of these equations.

One might alternatively argue that physicians in regions with (say) very sick patients may "fill in" missing characteristics of the vignettes, for example that their representative patients are also ones with poor support systems at home. Biases from such "fill-ins" could go in either direction; procedures like bypass surgery require considerable rehabilitation which might not always be available for patients with problems of access to care.

The detailed clinical vignette questions as well as possible responses are presented in Table 2; [15] summary statistics are presented in Table 1.[16] We begin with the vignette for Patient A, which asks how frequently the physician would schedule routine follow-up visits for patients with

[15] An additional vignette was presented in the cardiologist survey and not in the PCP survey; a patient with a typical clinical presentation of new onset angina - pain or pressure in the chest arising with exertion.

[16] We order these questions differently from their chronological appearance in the survey.

stable angina whose symptoms and cardiac risk factors are now well controlled on current medical therapy (cardiologists) or patients with hypertension (primary care physicians). The response is unbounded, and expressed in months, which in practice ranged from 1 month to 24 months. Figure 3 presents a HRR-level histogram from the cardiology survey for all regions with at least 3 cardiologists. In the regression analysis, we adopt an inverse function of this follow-up measure -- the average number of prescribed follow-up visits per year -- as our first of three indices summarizing Z^S.

How do these responses correspond to the guidelines on managing chronic stable angina? While diagnosis and management of coronary artery disease (the cause of angina) is the most common clinical issue faced by cardiologists on a day-to-day basis, there are no hard data to support any recommendation. The 2005 American College of Cardiology/American Heart Association [ACC/AHA] guidelines (Hunt et al., 2005) – what most cardiologists would have considered the "Bible" in the field at the time the survey was fielded – were very imprecise: they recommended follow-up every 4-12 months. However, even with these broad recommendations, we find that nearly one fifth (19%) of cardiologists in the sample recommend follow-up visits more frequently than every 4 months. The equivalent follow-up measure for the primary care physicians is for a hypothetical patient with well-controlled hypertension and there was a similarly wide range of responses. We define an aggressive physician with regard to follow-up frequency as one who is at least one standard deviation above the mean, which encompasses 20 percent of the cardiologist sample and 31 percent of the PCP sample. Less aggressive follow-up physicians are those who are more than one standard deviation below the mean; this accounts for 18 percent of the cardiologist sample and 8 percent of the PCP sample.

Office visits are not a large component of physicians' income (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 more highly remunerated tests and interventions (such as echocardiography, stress imaging studies, and so forth) that further set in motion the "diagnostic-therapeutic cascade" resulting in subsequent diagnostic tests, treatments, and follow-up visits, albeit with uncertain patient benefit (Black and Welch, 1993; Wennberg, et al., 1996; Lucas, et al., 2008). The next two vignettes concern patients with heart failure (or congestive heart failure), a much more expensive setting.

There are several types of heart failure. Systolic failure arises when the heart muscle loses its ability to contract vigorously, which means that there is less oxygenated blood pushed into the aorta. Diastolic failure occurs when the left ventricle becomes thickened and stiff, and thus does not allow enough blood to fill during the "resting" stage. In both cases, it is a chronic disease that can only be managed; there is no quick fix. Furthermore, heart failure is common, has a poor prognosis and is expensive in the elderly.

In Table 2, vignettes are presented for Patients B and C who have Class IV heart failure, the most severe classification and one in which patients have symptoms at rest. It is important to note that in both scenarios the patient is on maximal (presumably optimal) medications, and neither are candidates for revascularization: Patient B already had a coronary stent placed without symptom change, and Patient C is noted to not be a candidate for this procedure. The key differences between the two scenarios are patients' ages (75 in the first, 85 in the second), the presence of asymptomatic non-sustained ventricular tachycardia (VT) in the younger man, and severe symptoms that resolve partially with increased oxygen in the older man. Regardless, prognosis is poor for both patients under almost any type of management. In a recent study using

Medicare data, one-year mortality for all patients with HF following hospitalization was 30% (Chen et al., 2011). In those with medication refractory (that is, continued symptoms despite treatment) Class IV heart failure (Patients B and C) the one-year mortality rate for class IV HF is nearly 50 percent (Horwich et al., 2004).

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], and pacemaker insertion), one involving medication (anti arrhythmic therapy), and palliative care. Patient C also has 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 diuresis) and palliative care. In each case, cardiologists ranked their likelihood of recommending each intervention on a range from "never" to "always / almost always." Physicians could indicate strong or weak support for more than one option, for example, for both palliative care and an intervention.

We start with the obvious: regardless of the religious, political or moral persuasion of the cardiologist, these two men deserve a frank conversation about their prognosis and an ascertainment of their preferences and values for end-of-life care. Studies have shown that patients, physicians and family members are often not on the "same page" when it comes to advanced directive planning (Connors, et al., 1995). A clinical strategy that does not explicitly incorporate patient values for end-of-life care can result in patients being subjected to unwanted intensive, invasive, and expensive care. If compliant with the guideline, therefore, every one of the cardiologists should have answered “always/almost always”, or at least “most of the time,” to initiating or continuing discussions about palliative care.[17]

For Patient B, only 30 percent of cardiologists responded that they would take this course of action “most of the time” or “always/almost always.” For Patient C, 43 percent of cardiologists and 50 percent of primary care physicians were likely to recommend this course of action most of the time or always/almost always. In both cases, physicians’ recommendations fall short of the guidelines. We define our second index of physicians to reflect this. We classify the doctor as a “comforter” if the physician would discuss palliative care with the patient “always / almost always” for both Patients B and C (cardiologists) or just for patient C (primary care physicians). A total of 28 percent of cardiologists and 50 percent of primary care physicians met the requirement for being a comforter.

We now turn to more controversial aspects of the patient management questions. The language in the vignettes was carefully constructed relative to the contemporaneous guidelines. Several key aspects of Patient B rule out both the ICD and pacemaker insertion[18] and indeed the ACC-AHA guidelines explicitly recommend against the use of an ICD for Class IV patients potentially near death.[19] On the other hand, both treatments are highly reimbursed (for the cardiologist) and expensive for Medicare.

[17] According to the AHA-ACC directives, “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 at the end of life.” (Hunt et al., 2005, p. e206) Note that the patient can always say “do everything you can to prolong my life” even if empirical evidence suggests most don’t ask for this.

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

[19] Hunt et al., 2005; p. e206. In part, ICDs are not recommended because of the risk that the procedure itself will be fatal. But even if the cardiologist were successful in implanting the ICD, other problems might arise: when the heart is failing near death but the ICD continues to deliver repeated

Similarly, since patient C is already on maximal medications and is not a candidate for revascularization, the management goal is to make him as comfortable as possible and discuss palliative end-of-life care (again). This goal should be accomplished in the least invasive manner possible, for example 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.[20] In fact, 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 CHF, and could even cause harm (ESCAPE, 2005; Battacharya et al., 2008).

Despite these guideline recommendations, physicians in our data show a surprising degree of enthusiasm for additional interventions for patients B and C (observed among cardiologists for both patients and among primary care physicians for patient C). For patient B, 6 percent of cardiologists recommends a repeat angiography most of the time, almost always, or always and an additional 24 recommended a repeat angiography "some of the time," meaning that nearly 1/3 of the cardiologists surveyed could recommend a procedure for which there is no clinical evidence.

Moreover, 65 percent of cardiologists recommend an ICD most of the time, always or almost always while 47 percent recommend a pacemaker with the same frequency. For patient C, 18 percent recommend an ICU/CCU admission, 2 percent recommend a pulmonary artery catheter and 15 percent recommend a pacemaker at least most of the time. Among primary care

---

shocks, or when the ICD prevents the common, and usually peaceful, way to die for patients with dementia or other severe illnesses, pain and suffering is likely.

[20] The clinical improvement with a simple intervention (increasing his oxygen) also argues against more intensive interventions.

physicians, the shares of intensive recommendations for patient C are 21 percent for ICU/CCU admission, 2 percent for a pulmonary artery catheter, and 2 percent for a pacemaker.

Our next measure of Z^S is based on a summary of these intensity recommendations. We start with the three most intensive interventions for both patients.[21] Cardiologists' responses on aggressiveness are correlated across these two patients. Of the 28% (N=143) of cardiologists in the sample who would "frequently" or "always/almost always" recommend at least one of the above-listed high-intensity procedures for patient C, 93% (N=133) would also frequently or always/almost always recommend at least one high-intensity intervention for patient B. We use this overlap (the highest in our data) to define a "cowboy" cardiologist – a cardiologist who recommends one of the three possible intensive treatments to both patients B and C most of the time or always/almost always. Because Vignette B was not presented to the primary care physicians, we use only their response to Vignette C to categorize them. In total, 26% of the cardiologists are classified as cowboys, as are 24% of the primary care physicians.

All told, we have multiple measures of Z^S: high or low frequency of follow-up visits, a dummy variable for being a cowboy, and a dummy variable for being a comforter. A first question is how these measures of treatment preferences are related. Table 3 shows the bivariate relationships between the three variables. Among both PCPs and Cardiologists in our data, chi-squared tests strongly reject the null that there is no association between follow-up frequencies recommended for vignette patients and status as a "cowboy" or "comforter." Although these sets of characteristics were defined based on non-overlapping survey questions, our physicians' responses to questions about treatment preferences appear to be related. Figure 4 provides

[21] Note that each of these interventions would score at least 8 points out of 10 on an intensity score developed by Lucas et al. (2008).

additional visual evidence of the relationship between cowboy status and recommended follow-up frequency for the HRRs with the greatest number of respondents.

There is a good deal of geographic variation in these physician belief variables, as Table 1 shows. Relative to the mean rate of high follow-up frequency across areas (28.5% for all physicians), the standard deviation is 19%. Similar variation is present for the cowboy and comforter ratios across HRRs. The p-values in the last column show that we can reject the null hypothesis of a random distribution of physicians across areas with high confidence. Generally, the areas in which physicians are more aggressive include HRRs such as Manhattan and East Long Island in New York, while the areas where physicians are less aggressive include HRRs such as Birmingham, Alabama (Figure 4).

IV. Model Estimates

We now proceed with our estimates of the models presented above. We first consider Equation (4), the relationship between area-level spending and local patient and physician preferences. We then turn to the Equation (6), modeling the factors leading physicians to be more and less aggressive.

Do Survey Responses Predict Regional Medicare Expenditures?

We start with the basic relationship between area spending, patient preferences and physician preferences for the 64 HRRs with at least 3 cardiologists and 2 primary care physician surveys. Figure 5 shows scatter plots between area-level end of life spending and our measures of supply and demand for care. The measures we include 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 than one standard deviation above the norm (panel c), and the share of patients who desire more aggressive care at the end of life (panel d). Each circle is an HRR, and the size of the circle is proportional related to the sample size in the HRR.

In the case of the three supply-side variables, the results are consistent with the theory: despite the small sample sizes of physicians per HRR, end of life spending is positively related to the cowboy ratio, negatively related to the comforter ratio, and positively related to the desired frequency of follow-up is high. The demand variable, in contrast, is not related to spending; the data points form a cloud more than a line.

Table 4 explores this result more formally and also presents magnitudes. The first two columns present area-level regressions relating the natural logarithm of 2-year end-of-life spending to physician preferences, pooling responses of all primary care physicians and cardiologists at the area level. The regressions use sampling weights that denote the inverse of the probability that the observation is included because of the sampling design in order to take into account differences in the number of observations per HRR.

As the first column shows, the local proportion of cowboys and comforters predicts local spending levels well, explaining 36 percent of the regional variations in risk-adjusted end-of-life spending. Further, the magnitudes are large. If an area went from no cowboys to all cowboys, end of life spending is predicted to increase by 75 percent. Making all doctors comforters, in contrast, would be predicted to reduce end of life spending by 42 percent. The relationship between spending and cowboys and comforters is true for both cardiologists and primary care physicians analyzed separately, as shown in Appendix Tables A and B. The coefficients are somewhat larger for primary care physicians than for cardiologists, but not statistically significantly so.[22]

[22] Note that this might be a true difference, or might reflect a greater number of primary care physicians in the survey, and hence greater precision in the cowboy and comforter ratios.

Column 2 of Table 5 shows that recommended follow-up frequency – in particular, our indicator for high frequency follow-up recommendations – is also a meaningful predictor of HRR-level spending. In areas where 100% of doctors prefer to see patients much more frequently than the norm, end of life spending would be expected to be 61 percent higher. There is no difference in spending associated with less frequent than average follow-up times. The combination of just these three statistically significant supply-side preferences explains over 59 percent of the observed end-of-life spending variation in our 64 HRRs.

The next three columns add measures of patient preferences to the regressions: the share of patients wishing to have unneeded tests, the share wanting to see an unneeded cardiologist, the share preferring aggressive end-of-life care, and the share preferring comfortable end-of-life care. In almost all cases, these coefficients are small and statistically insignificant. The one exception to this is the share of patients who prefer comfort care at the end of life, which is negatively associated with end-of-life spending. The magnitude of this coefficient is large (a 43 percent reduction in spending if all patients prefer comfort care), but the share of the variation explained by this variable is relatively small. Even excluding the physician preference variables entirely, as in column 7, the R^2 from the patient preference variables as a whole is only 7.5 percent, far below the R^2 from the physician preference variables.

Tables 4(b) repeats the analysis in Table 4(a), but uses total average per beneficiary expenditures, adjusted for prices, age, sex, and race/ethnicity, as the dependent variable, as this may be a more relevant measure of intensity among primary care physicians. In both the combined sample (shown) and for PCPs only (not shown), the fraction of cowboys in an HRR is a consistently strong predictor of spending across models. Moreover, although R^2 values are smaller in these models, supply-side factors explain more of the variation in spending than demand-side factors. We also note that among PCPs only, higher follow-up frequencies are associated with higher overall Medicare spending at the HRR level.

Finally, we consider risk-adjusted one-year expenditures for a "forward looking" cohort of hip fracture patients. These use price-adjusted expenditures as the dependent variable, but control at the individual level for hierarchical condition categories (HCCs) for the 6 months prior to admission, as well as comorbidities and type of hip fracture at the time of admission (Skinner et al., 2013). The estimated coefficients, reported in Appendix Table C, exhibit results very similar to those in Table 4(a).

It is possible that there is an interaction effect between patient and physician preferences. Such interactions might occur in multiple ways. One theory is that aggressive physicians will interact with aggressive patients to drive even more spending. A second theory is that comforter physicians will cater more to patients with comforter preferences. A third theory is that patients and physicians with different preferences will mute each other – aggressive physicians will drive more care only if patients are not in need of comfort, and vice versa.

We test these theories in Table 5. The first column is the same as column 6 in Table 4(a); the second column adds two interactions: one between the share of physicians who are cowboys and the share of patients who desire aggressive care, and the second between the share of physicians who are cowboys and the share of patients who desire comfort care. The third column repeats the interactions, using the share of physicians who are comforters instead. All of the interaction effects estimated are statistically insignificant, however.[23]

[23] One might also wonder whether the results are driven by broad differences in (e.g.) east-coast versus west-coast practice styles caused by (for example) assortative migration among physicians. However, we did not find any meaningful differences between physicians in the east and west of the United States and our estimated coefficients are similar to those in Table 4(a) when the east and west coasts of the US are modeled separately.

What factors predict physician responses to the vignettes?

To this point, we have shown that physician preferences matter enormously for spending, and that physician preferences vary across areas more than would be expected given random variation. The obvious question is: what explains this variation in physician preferences? In this section, we estimate the model in equation (2) to test our hypotheses as to the importance of a variety of financial and organizational factors in explaining physician recommendations. Specifically, we ask which physician-level factors predict being a high follow-up, cowboy or comforter physician. We present results for the pooled sample in Tables 6(a) – 6(c) using a logistic model with HRR level random effects. Results for cardiologists and primary care physicians are presented separately in Appendix D.[24]

As noted earlier, we divide our explanatory variables into four categories. Financial considerations include the share of patients that are capitated and the share on Medicaid. Each of these is associated with less generous reimbursement for the treating physician. Thus, if we believed that financial incentives drive treatment recommendations, we would expect each to be negatively related to the share of cowboy physicians and positively related to the share of comforters. Organizational form is measured by practice in larger groups – single/multi-specialty practices or HMOs/hospital-owned practice, relative to the baseline of being in a solo or two person practice. In larger groups, we expect less autonomy for physicians to select their own practice styles, and hence greater adherence to guidelines, which in this case means fewer cowboys but more comforters. The third class of variables refers to expectations responsiveness – the degree to which physicians report recommending intensive care for non-clinical reasons, including meeting patient expectations, following what colleagues would do, or meeting the

[24] In combined regressions, we control for the fraction of cardiologists in the sample by HRR to adjust for secular differences between primary care physicians and cardiologists in making intervention recommendations.

expectations of referring physicians (cardiologists only). We include malpractice concerns in this category – specifically, whether the physician reports having made cardiologist referrals (PCPs) or recommended PCI (cardiologists) in the past year to protect against a possible malpractice suit.

Tables 6(a) – 6(c) show the results of this analysis. Table 6(a) considers factors that predict which physicians are cowboys, Table 6(b) predicts which physicians are comforters, and Table 6(c) predicts high follow-up frequency recommendations. Within each panel, the first column includes financial considerations, the second column includes organizational factors, the third column includes expectations responsiveness, and the final column includes all factors together. In each column, we include basic physician demographics: age, gender, board certification status, and days per week spent seeing patients. We also control for cardiologists per 100,000 Medicare beneficiaries to account for the role of local competition.

The demographic factors included reveal that older physicians are more likely to recommend high rates of follow-up and are more likely to be cowboys, but age is not a significant predictor of comforter status. Women are more likely to be comforters, and there is some evidence that men are more likely to be cowboys and to recommend higher rates of follow-up[25]. Interestingly, factors typically associated with skill and quality of care are negatively associated with cowboy status and positively associated with being a comforter: both board certification and the number of days per week a physician spends seeing patients are negatively predictive of cowboy status and days per week spent seeing patients is positively associated with being a comforter. Board certification is also negatively associated with high follow-up recommendations. One interpretation of this is that uncertainty breeds aggressiveness – the more

[25] We note that over 80% of our sample is male (including over 90% of our cardiologists). Given the relatively small number of women per HRR, we are not comfortable making strong statements about the role of gender in predicting physician type.

a physician knows about the field or has experience with a particular type of patient, the less aggressive they are.

At the same time, a greater number of cardiologists per 100,000 Medicare beneficiaries is associated with a higher likelihood of a physician being a cowboy or high follow-up doctor and with a lower likelihood of the physician being a comforter. In a linear model, moving from the 25th to the 75th percentile of cardiologists per capita is equivalent to an 8.2% increase in 2-year EOL spending, while moving from the 10th to the 90th percentile is equivalent to a 13.2% increase in 2-year EOL spending.

The financial considerations with respect to the composition of the patient population do not conform to the theory. Physicians who see more Medicaid patients and more capitated patients are more likely to be cowboys, and neither greater capitated patient fractions nor greater Medicaid patient fractions are associated with being a comforter. As further evidence that financial considerations may not play as great a role as traditional supplier-induced demand models might suggest, we asked the simple question: do cowboys make more money? Using evidence on Cardiologist income from the 2012 *MedScape* Cardiologist Compensation Report[26], we matched our data to average cardiologist incomes in 8 U.S. regions. The correlation coefficient between average cardiologist salary and the fraction of cowboys among cardiologists in our sample was -0.08, suggesting that if anything, a smaller fraction of cowboys practice in regions with higher cardiologist salaries.

Organizational factors are associated with physician beliefs about appropriate practice. Physicians in solo or 2-person practices are far more likely to be aggressive than physicians in single or multi-specialty group practices or physicians who are part of an HMO or a hospital-based practice. Yet physicians in group or staff model HMOs or hospital-based practices are no

[26] http://www.medscape.com/features/slideshow/compensation/2012/cardiology

more likely to be comforters, and physicians in single or multi-group practices are somewhat less likely to be comforters.

Physician beliefs about the correct frequency and intensity of intervention are in some cases related to their self-assessed responses to expectations. Cardiologists who report being most responsive to referring doctor expectations are also most likely to recommend aggressive interventions, while physicians who report responding to patient expectations are more likely to be comforters. One way to reconcile these results is that physicians respond to referring physicians' expectations by doing more interventions, but when they respond to patient wishes, they are more likely to provide palliative care.

Forty-two percent of physicians respond that they had taken clinical action in response to malpractice concerns over the past year. Despite the high level of concern regarding malpractice issues, this variable plays no role in explaining intervention intensity or frequency in any of our models. Sins of omission (as opposed to sins of commission) are less likely to be a key consideration for the types of patients described in Vignettes B and C, particularly where guidelines exist.

Overall, our results show a number of factors that influence physician beliefs, the most important of which are organizational factors and the number of days practiced each week. Each of these significantly affects whether the physician is a cowboy, and to a somewhat lesser extent, whether they are a comforter. Still, the overall explanatory power of these variables is limited. The pseudo-R^2 (or actual R^2 in the linear model) from the combination of practice characteristics is typically 0.15 or less, meaning that there is a large component of physician beliefs unexplained by these combined factors. And while we interpret this residual variation as reflecting differences in beliefs about the efficacy of the treatment or the technical skill of the physician, we cannot decisively rule out other explanations.[27]

V. Conclusion and Implications

While there is a good deal of regional variation in medical spending and care utilization in the U.S. and elsewhere,[28] there is little agreement about the causes of such variations. Do they arise from variations in patient demand, from variations the preferences of physicians or both? In this paper, we find that patient demand, as proxied by patients' responses to a nationwide survey, has modest predictive association with regional Medicare end-of-life treatment patterns. Conversely, individual physician heterogeneity regarding treatment options can explain a substantial degree of regional variations in utilization among the U.S. Medicare population, a result that is consistent with Sirovich et al. (2008), Lucas et al. (2008), and Wennberg et al. (1997).

We then turn to the factors that lead physicians to have different preferences. We find that the traditional factors in supplier-induced demand models, such as the fraction of patients paid through capitation (or on Medicaid), or the responsiveness to financial factors, play a relatively small role in explaining equilibrium variations in utilization patterns. Organizational factors such as accommodating colleagues help to explain some but not all individual intervention decisions.

But the largest factor explaining these regional variations appears to be physician beliefs about the effectiveness of treatments that are orthogonal to perceived organizational and financial

[27] Peer effects could lead to doctors doing less in low-utilization regions, or lead doctors to do more in high-utilization regions, with a net impact of roughly zero. (That is, x^o could vary across regions.) We tested for these effects by identifying regions with above-average PCI rates, and introducing a dummy for being in a high PCI region with a dummy for responsiveness to colleague expectations (or responsiveness to all non-clinical factors). We did not find support for this hypothesis in the data.

[28] See for example http://www.wennbergcollaborative.org/Publications.htm.

factors. Our results are therefore consistent with Epstein and Nicholson (2009), who find large variations in Cesarean section surgical rates among obstetricians within the same practice, even after adjusting for where the physicians trained.

One concern about the interpretation of the vignette responses as "overuse" is that they may reflect the true productivity of physicians (in the sense of Chandra and Staiger, 2007). A doctor who is particularly skilled in intensive procedures may rationally be a "cowboy," since his outcomes would be better than the average outcome that serves as the basis for guidelines. While we cannot rule this out, we note that physicians with greater objective qualifications, such as board certification and the number of days per week spent practicing, are no more likely to be cowboys. Nor do the updated 2009 heart failure guidelines recommend more aggressive care, (Hunt et al., 2009).[29]

Another possibility is that while "cowboys" may over-treat patients along some dimensions, they may also avoid the underuse of effective care along other dimensions (e.g., Landrum et al., 2008). Our survey did not ask about whether the physician provides effective care or not. But other evidence suggests that overall end-of-life expenditures are at best poorly correlated with health outcomes (Silber et al., 2010; Fisher and Skinner, 2010), and at worst negatively associated with both outcomes and effective care (Baicker and Chandra, 2004; Fisher et al., 2003a,b).

The major unknown in our results is how physician beliefs arise. Simple heterogeneity in physician beliefs cannot explain regional variation in expenditures; the regional pattern of beliefs cannot be explained by simple random variation. Rather, spatial correlation in beliefs is required

[29] A more subtle point is that guidelines are based on randomized trials conducted in high-quality academic settings. Yet those same procedures conducted in community settings could yield worse outcomes (Wennberg, et al., 1998).

in order to explain the regional patterns we see. We do find that physicians' propensity to intervene for non-clinical reasons is related to the expectations of physicians with whom they regularly interact, a result consistent with Burke et al (2006), who show that network and peer effects models can naturally lead to such spatial correlations. But we are still left with a question as to why some regions become more aggressive than others, and why such characteristics could be related to factors such as social capital (Skinner and Staiger, 2007).

Our results do not mean to imply that economic incentives are unimportant. Clearly, changes in payment margins have a large impact on behavior, as in Clemens and Gottlieb (2012). But in a system where the effects of price margins and reimbursement rates are blunted, such as in many European countries, physician beliefs have a much larger scope to drive the dramatic degree of variation in healthcare systems, particularly for medical conditions where there is some uncertainty about the value of medical treatments (National Health Service, 2010; McPherson et al., 1981). Understanding how best to measure and, if necessary, shape these beliefs is a key challenge for future research.

References

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) (April).

Anthony, D. L., M. B. Herndon, P. M. Gallagher, A. E. Barnato, J. P. Bynum, et al., 2009. "How much do patients' preferences contribute to resource use?" *Health Affairs* 28, May-Jun, 864-73.

Bach, Peter B., Deborah Schrag, and Colin B. Begg, 2004. "Resurrecting Treatment Histories of Dead Patients: A Study Design That Should Be Laid to Rest," *JAMA* 292(22) (December 8): 2765-2770.

Baicker, K., and A. Chandra, 2004. "Medicare spending, the physician workforce, and beneficiaries' quality of care," *Health Affairs* (April 7).

Barnato, A.E, M.B. Herndon, D.L Anthony, P. Gallagher, JS Skinner, et al., 2007. "Are regional variations in end-of-life care intensity explained by patient preferences? A study of the US Medicare population," *Medical Care* 45, 386-393.

Battacharya, Jay, Azeem M. Shaikh, and Edward Vytlacil, 2008. "Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization," *American Economic Review: Papers & Proceedings* 98:2, 351–356.

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(1), 167-74.

Black, William C., and H.G. Welch, "Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy," *NEJM* 1993;328:1237-1243.

Burke, Mary, Gary Fournier, and Kislaya Prasad, "The Emergence of Local Norms in Networks," *Complexity*, 11(5) May/June 2006: 65-83.

Campbell, Eric G., Susan Regan, Russell L. Gruen, Timothy G. Ferris, Sowmya R. Rao, et al., 2007. "Professionalism in Medicine: Results of a National Survey of Physicians," *Annals of Internal Medicine* 147(11):795-802.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks," *J Political Economy* 115, 103-140.

Chandra, Amitabh, and Jonathan Skinner, 2012, "Productivity growth and expenditure growth in U.S. Health care," *Journal of Economic Literature* 50(3) (September): 645-80.

Chen, Jersey, Sharon-Lise T. Normand, Yun Wang, and Harlan M. Krumholz, 2011. "National and Regional Trends in Heart Failure Hospitalization and Mortality Rates for Medicare Beneficiaries, 1998-2008," *JAMA*, October 19, 306(15): 1669-78.

Clemens, Jeffrey, and Joshua Gottlieb, 2012. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" Stanford University (SIEPR), http://siepr.stanford.edu/publicationsprofile/2447.

Connors AF, Jr, Dawson NV, Desbiens NA, et al. 1995: "A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT)." *JAMA.* 274(20):1591-1598.

Dartmouth Atlas, 2012. www.dartmouthatlas.org .

de Jong, Judith D., 2008. *Explaining medical practice variation: Social organization and institutional mechanism* (Nivel, Utrecht, The Netherlands).

Doyle, Joseph, 2011. "Returns to local-area healthcare spending: Using 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) (March): 61–73.

Dresselhaus, Timothy R., John W. Peabody, Jeff Luck, Dan Bertenthal, 2004. "An Evaluation of Vignettes for Predicting Variation in the Quality of Preventive Care," *Journal General Internal Medicine*, October; 19(10): 1013–1018.

Epstein, Andrew J. & Nicholson, Sean, 2009. "The formation and evolution of physician treatment styles: An application to cesarean sections," *Journal of Health Economics,* 28(6): 1126-1140, December.

(ESCAPE) Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness: The ESCAPE Trial. *JAMA.* 2005; 294(13): 1625-1633

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, et al., 2003a, The implications of regional variations in medicare spending. Part 1: The content, quality, and accessibility of care, *Ann Intern Med* 138, Feb 18, 273-87.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, et al., 2003b, The implications of regional variations in medicare spending. Part 2: Health outcomes and satisfaction with care, *Ann Intern Med* 138, Feb 18, 288-98.

Fisher, Elliott, and Jonathan Skinner, 2010. "Comment on Silber et al.: Aggressive Treatment Styles and Surgical Outcomes, *Health Services Research* 45: 1908-11 (December).

Gruber, J., and M. Owings. 1996. Physician financial incentives and cesarean section delivery. *Rand Journal of Economics* 27 (1):99-123.

Horwich, Tamara B, W. Robb MacLellan, and Gregg C. Fonarow, 2004. "Statin Therapy Is Associated With Improved Survival in Ischemic and Non-Ischemic Heart Failure," *Journal of the American College of Cardiology* 43(4): 642-48.

Hunt, Sharon Ann, William T. Abraham, Marshall H. Chin, Arthur M. Feldman, Gary S. Francis, et al., 2005. "ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult," *Circulation*. 2005;112:e154-e235; (September 13).

Hunt, Sharon Ann, William T. Abraham, Marshall H. Chin, Arthur M. Feldman, Gary S. Francis, et al., 2009. "2009 Focused Update Incorporated Into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Chronic Heart Failure in the Adult," *Circulation* 119:e391-e479 (March 26).

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): 437-443.

Landrum, Mary Beth, Ellen R. Meara, Amitabh Chandra, Edward Guadagnoli and Nancy L. Keating, 2008. "Is Spending More Always Wasteful? The Appropriateness of Care and Outcomes among Colorectal Cancer Patients," *Health Affairs* 27(1):159-168 (January).

Levine-Taub, Anna A., Anton Kolotilin, Robert S. Gibbons, Ernst R. Berndt, "The Diversity of Concentrated Prescribing Behavior: An Application to Antipsychotics," NBER Working Paper No. 16823 (February 2011).

Lucas, FL, Siewers, AE, Malenka, DJ, Wennberg, DE. Diagnostic-therapeutic cascade revisited. Coronary angiography, coroabnry artery bypass surgery, and percutaneous coronary intervention in the modern era. *Circulation*. 2008;118:2797-2802.

McClellan, Mark, and Jonathan Skinner, 2006, "The Incidence of Medicare," *Journal of Public Economics* 90, 2006/1, 257-276.

McGuire, Thomas C., 2011, Physician agency and payment for primary medical care, in Sherry Glied, and Peter C. Smith, eds.: *The Oxford handbook of health economics* (Oxford University Press).

McGuire, T.G., Pauly, M., 1991. "Physician response to fee changes with multiple payers." *Journal of Health Economics*, 10 (4), 385-410.

McPherson, Kim, PM Strong, Arnold Epstein, and Lesley Jones, 1981, Regional Variations in the Use of Common Surgical Procedures: Within and Between England and Wales, Canada, and the United States," *Social Science & Medicine. Part A: Medical Sociology* 15 May, 273-88.

Mandelblatt JS, Faul LA, Luta G, Makgoeng SB, Isaacs C, 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* July 20;30(21): 2609-14.
Moliter, David, "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration," MIT, 2011. http://economics.mit.edu/files/7301.

National Health Service, 2010. *The NHS Atlas of variation in healthcare* (November). http://www.rightcare.nhs.uk/atlas/qipp_nhsAtlas-LOW_261110c.pdf.

Pauly, Mark, 1980. *Doctors and their workshops: Economic models of physician behavior* (University of Chicago Press, Chicago).

Peabody JW, Luck J, Glassman P, Jain S, Hansen J, Spell M, Lee M., 2004. "Measuring the quality of physician practice by using clinical vignettes: a prospective validation study," *Annals of Internal Medicine* Nov 16;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," *The American Journal of Medicine* 113(8): 663-67.

Romley, J. A., A. B. Jena, and D. P. Goldman, 2011, Hospital spending and inpatient mortality: Evidence from california: An observational study, Annals of Internal Medicine 154: 160-7 (February 1).

Silber, J. H., R. Kaestner, O. Even-Shoshan, Y. Wang, and L. J. Bressler, 2010, Aggressive treatment style and surgical outcomes, *Health Services Research* 45: 1872-92 (December).

Sirovich, B., P. M. Gallagher, D. E. Wennberg, and E. S. Fisher, 2008. "Discretionary Decision Making by Primary Care Physicians and the Cost of U.S. Health Care," *Health Affairs* 27, May-Jun, 813-23.

Skinner, Jonathan, 2012. "Causes and Consequences of Geographic Variation in Health Care" in T. McGuire, M. Pauly, and P. Pita Baros (eds.) Handbook of Health Economics Vol. 2, North Holland.

Skinner, Jonathan, and Douglas Staiger, 2009, Technology diffusion and productivity growth in health care, Working Paper Series (National Bureau of Economic Research, Cambridge MA).

Skinner, Jonathan, Weiping Zhou, Daniel Gottlieb, and Elliott Fisher, "Measuring Hospital-Level Healthcare Expenditures and Outcomes with Disease-Specific Cohorts" Report to the Institute of Medicine Committee on Geographic Variation in Health Care Spending, January 2013.

Suleman M, Clark MP, Goldacre M, Burton M., 2010. "Exploring the Variation in Paediatric Tonsillectomy Rates between English Regions: a 5-year NHS and Independent Sector Data Analysis. *Clin Otolaryngol*. 2010 Apr;35(2):111-7.

Sutherland, J. M., E. S. Fisher, and J. S. Skinner, 2009 “Getting past denial--the high cost of health care in the United States”, *New England Journal of Medicine* 361, Sep 24: 1227-30.

Wennberg, David E., et al. 1998. “Variations in Carotid Endarterectomy Mortality in the Medicare Population: Trial Hospitals, Volume, and Patient Characteristics.” *JAMA*, 279(16): 1278-81.

Wennberg, John E., 2010. *Tracking medicine: A researcher's quest to understanding health care* (Oxford University Press, New York).

Wennberg, John E., and Alan Gittelsohn, 1975. “Health Care Delivery in Maine I: Patterns of Use of Common Surgical Procedures,” The Journal of the Maine Medical Association 66(5): 123-130 & 149 (July).

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): 811–824.

Wennberg, David E; Kellet, MA; Dickens, Jr., JD; Malenka, DJ; Keilson, LM; Keller, RB: The Association Between Local Diagnostic Testing Intensity and Invasive Cardiac Procedures. *JAMA,* Vol. 275, No. 15; p1161-1164; April 17, 1996.

Wennberg, David E; Dickens, Jr., JD; Biener, L; Fowler, Jr., FJ; Soule, DN; Keller, RB., 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):172-176.

Wennberg, John E., Elliott S. Fisher, and Jonathan S. Skinner, 2002, “Geography and the debate over Medicare reform,” *Health Affairs* (www.healthaffairs.org), W96-W114.

Zuckerman, S., T. Waidmann, R. Berenson, and J. Hadley, 2010. “Clarifying Sources of Geographic Differences in Medicare Spending.” *New England Journal of Medicine* 363 (1):54-62.

Figure 1: Variations in Equilibrium: Differences in λ and Differences in Actual or Perceived Productivity

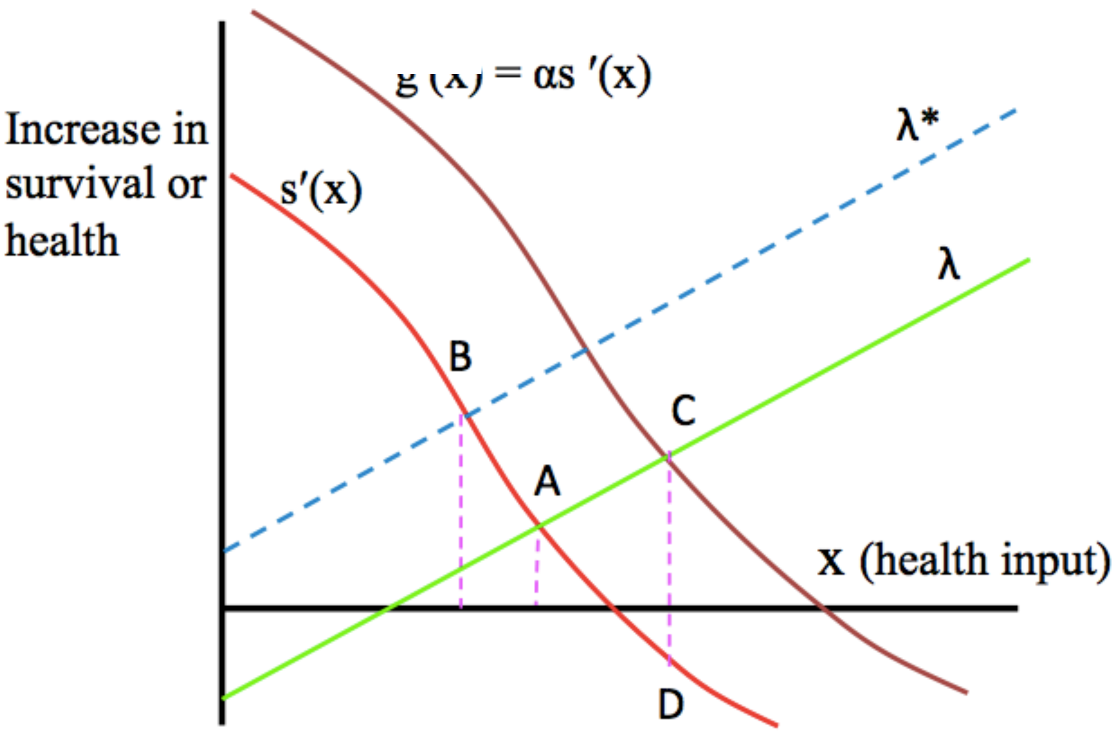


Figure 2: Fraction of Patients Who Would See Unneeded Cardiologist (HRR-Level Distribution)

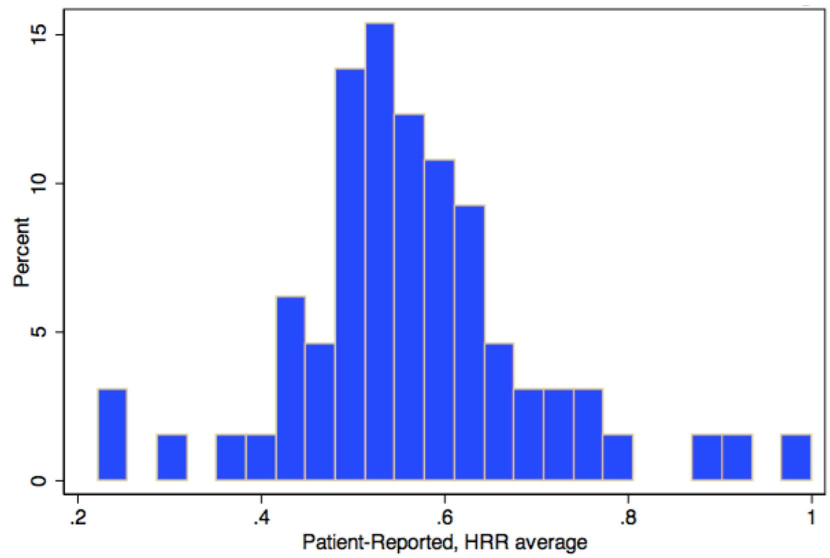


Figure 3: Distribution of Length of Time before Next Visit for Patient with Well-Controlled Angina (Cardiologist HRR-Level Distribution)

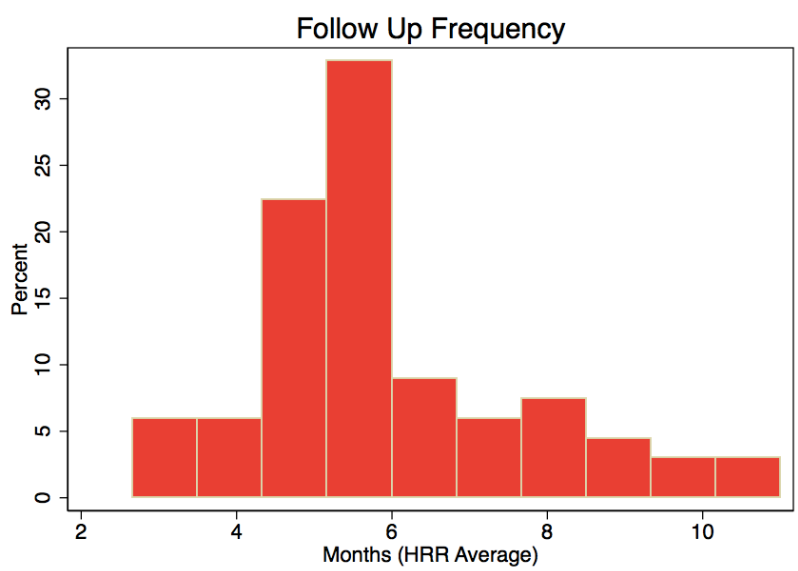
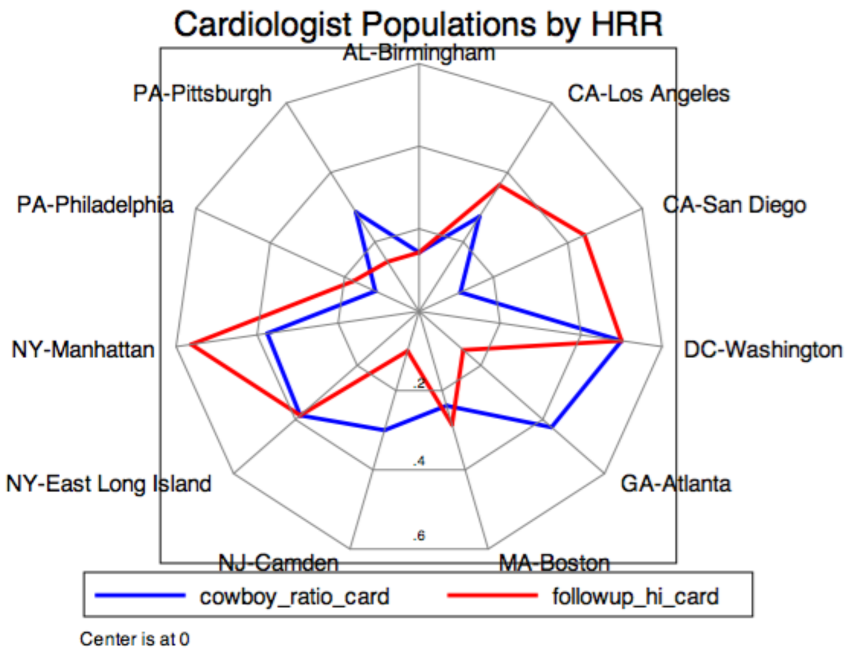


Figure 4: High Follow-up Frequency and Cowboy Prevalence by HRR:



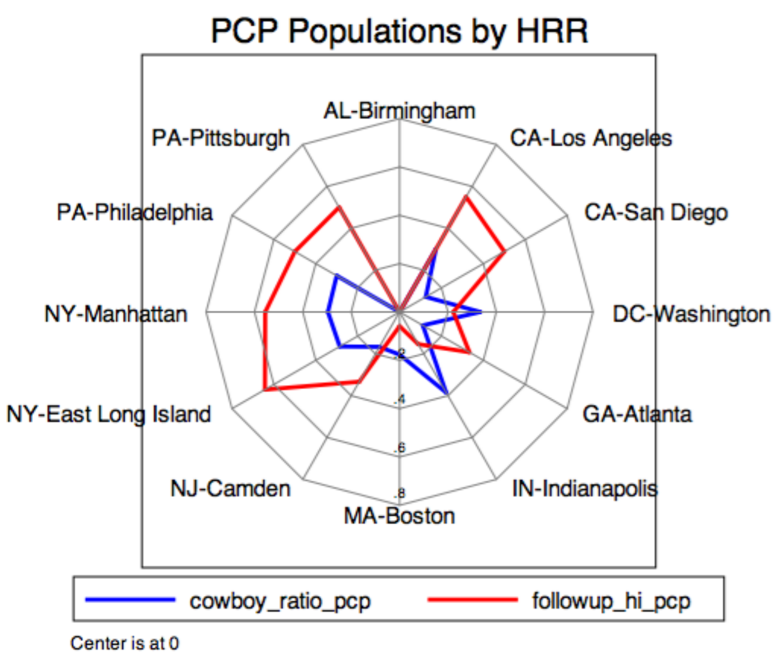


Figure 5: Log of Inpatient 2-year End-of-Life Regional Spending vs. Various Independent Variables

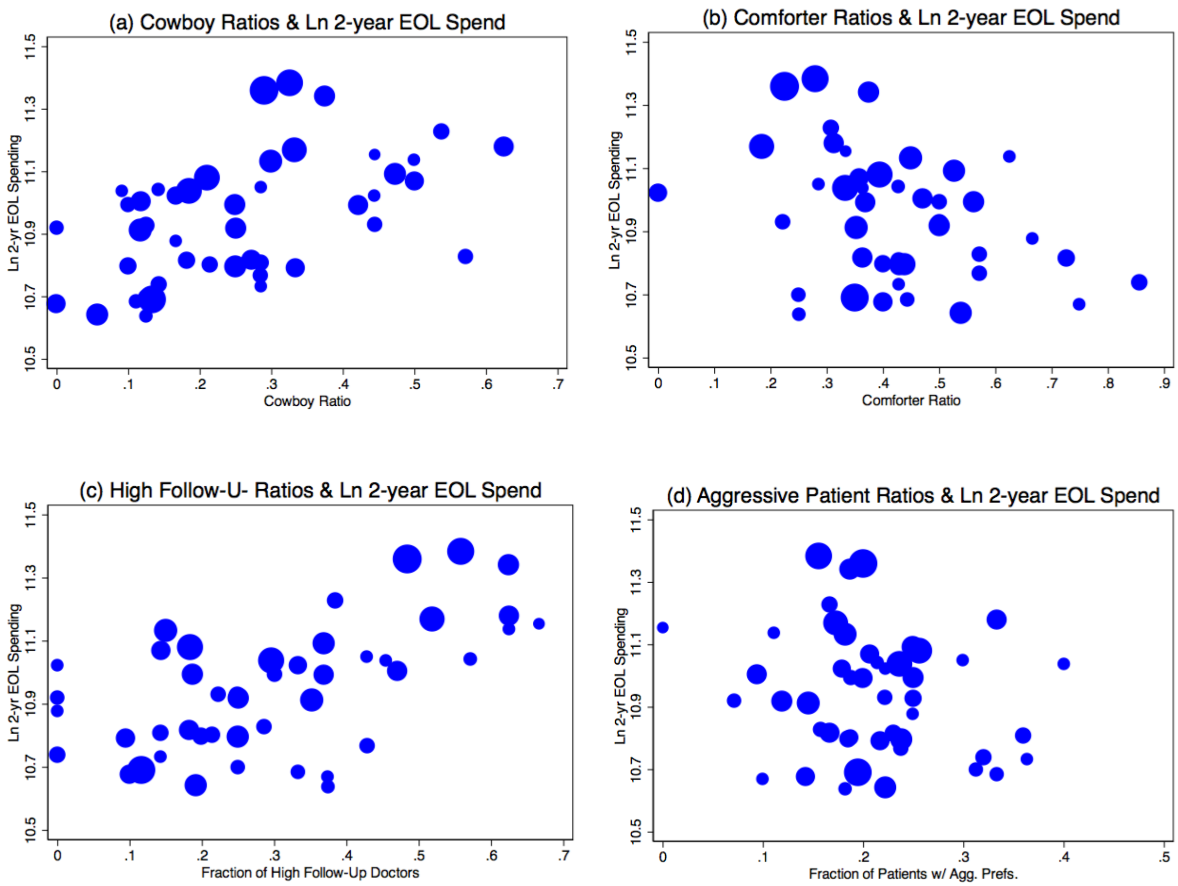


Table 1: Summary Statistics

Variable	Mean	Individual SD	Area Average SD
Patient Variables			
Have unneeded tests	73%	44%	10%
See unneeded cardiologist	56%	50%	13%
aggressive patient prefs. ratio	8%	27%	6%
comfort patient prefs. ratio	48%	50%	15%
Primary care physician variables			
cowboy ratio	22%	41%	26%
comforter ratio	50%	50%	29%
followup lo	8%	28%	14%
followup hi	31%	46%	27%
Cardiologist variables			
cowboy ratio	25%	44%	21%
comforter ratio	28%	44%	23%
followup lo	18%	38%	22%
followup hi	19%	40%	22%
Spending and utilization			
2-year EOL spending	$56,219	-	$10,715
6-month EOL spending	$14,272	-	$2,660
total per patient spending	$7,837	-	$1,032
hip fracture patient spending	$52,574	-	$4,996
Organizational			
fraction capitated patients	16%	25%	17%
fraction Medicaid patients	10%	13%	8%
weekly patient days	3.1	1.5	0.7
physician age	57.5	9.8	4.0
board certified	89%	31%	12%
Cardiologists per 100k	6.7	1.9	1.6
responds to patient expect.	60%	48%	-
responds to colleague expect.	40%	49%	-
responds to refrrer expect.	10%	30%	-
responds to malpract. concerns	44%	50%	-

Note: The table shows means for the sample living or practicing in one of the 64 HRRs with at least 3 cardiologists and 2 primary care physicians. The area average standard deviation is weighted by the number of observations in the HRR.

Table 2: Clinical Vignettes and Response Options for Cardiologists and Primary Care Physicians

In the next set of questions, you will be presented with brief clinical descriptions for three different patients. For each, you will be asked a series of questions regarding how you would be likely to treat that patient were he or she in your care.

PATIENT A - CARDIOLOGIST: *For this question, think about a patient with stable angina whose symptoms and cardiac risk factors are now well controlled on current medical therapy. In general, how frequently do you schedule routine follow-up visits for a patient like this?*

*Answer recorded in number of months

PATIENT A - PCPs: *In general, how frequently do you schedule routine follow-up visits for a patient with well-controlled hypertension?*

*Answer recorded in number of months

PATIENT B: *A 75 year old man with severe (Class IV) congestive heart failure from ischemic heart disease, is on maximal medications and has effective disease management counseling. His symptoms did not improve after recent angioplasty and stent placement and CABG is not an option. He is uncomfortable at rest. He is noted to have frequent, asymptomatic nonsustained VT on cardiac monitoring. He has adequate health insurance to cover tests and medications. At this point, for a patient presenting like this, how often would you arrange for each of the following?*

CARDIOLOGIST SURVEY
a - Repeat angiography
b - Initiate antiarryghmic therapy
c - Recommend an Implantable Cardiac Defibrilator (ICD)
d - Recommend biventricular pacemaker for cardiac resynchronization
e - Initiate or continue discussions about palliative care

POSSIBLE RESPONSES
1 Always/Almost always
2 Most of the time
3 Some of the time
4 Rarely
5 Never
9 NA

PATIENT C: *An 85 year old male patient has severe (Class IV) congestive heart failure from ischemic heart disease, is on maximal medications, and is not a candidate for coronary revascularization. He is on 2 liters per minute of supplemental oxygen at home. He presents to your office with worsening shortness of breath and difficulty sleeping due to orthopnea. Office chest xray confirms severe congestive heart failure. Oxygen saturation was 85% and increased to 94% on 4 liters and the patient is more comfortable. He has adequate health insurance to cover tests and medications. At this point, for a patient presenting like this, how often would you arrange for each of the following?*

PCP and CARDIOLOGIST SURVEY
a - Allow the patient to return home on increased oxygen and increased diuretics
b - Admit to the hospital for aggressive diuresis (not to the ICU/CCU)
c - Admit to the ICU/CCU for intensive therapy and monitoring
d - Place a pulmonary artery catheter for hemodynamic optimization
e - Recommend biventricular pacemaker for cardiac resynchronization
f - Initiate or continue discussions about palliative care

POSSIBLE RESPONSES (both surveys)
1 Always/Almost always
2 Most of the time
3 Some of the time
4 Rarely
5 Never
9 NA

Table 3(a): PCPs

	Cowboy		
Follow-up frequency	Yes	No	
Low	16	61	8.42%
Medium	98	452	60.18%
High	87	200	31.40%
	21.99%	78.01%	

p(χ2): **0.000**

Comforter	
Yes	No
39	38
300	250
115	172
49.67%	50.33%

p(χ2): **0.019**

	Comforter		
Cowboy	Yes	No	
Yes	87	114	21.99%
No	367	346	78.01%
	49.67%	50.33%	

p(χ2): **0.145**

Table 3(b): Cardiologists

	Cowboy		
Follow-up frequency	Yes	No	
Low	17	76	18.02%
Medium	85	238	62.60%
High	31	69	19.38%
	25.78%	74.22%	

p(χ2): 0.000

Comforter	
Yes	No
27	66
94	229
22	78
27.71%	72.29%

p(χ2): 0.000

	Comforter		
Cowboy	Yes	No	
Yes	39	94	25.78%
No	104	279	74.22%
	27.71%	72.29%	

p(χ2): 0.000

Table 4(a): Estimates for Explaining Ln 2-Year End-of-Life Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Combined Sample of PCPs and Cardiologists						
cowboy ratio, all docs	0.7565***	0.4940***	0.4926***	0.4704***	0.4840***	0.4401***	
	(0.1662)	(0.1276)	(0.1145)	(0.1229)	(0.1317)	(0.1153)	
comforter ratio, all docs	-0.4112*	-0.2670*	-0.2666*	-0.2250+	-0.2718*	-0.2109+	
	(0.1646)	(0.1257)	(0.1267)	(0.1161)	(0.1185)	(0.1128)	
followup lo, all docs		0.0000	-0.0002	0.0194	-0.0062	-0.0047	
		(0.2045)	(0.2068)	(0.1995)	(0.2157)	(0.2131)	
followup hi, all docs		0.6438***	0.6449***	0.6595***	0.6644***	0.6891***	
		(0.1471)	(0.1423)	(0.1395)	(0.1472)	(0.1279)	
have unneeded tests			-0.0066			-0.0471	-0.0543
			(0.2359)			(0.2527)	(0.3400)
see unneeded card.				0.3444**		0.3987*	0.5397+
				(0.1249)		(0.1576)	(0.2855)
aggressive prefs. patient ratio					0.0508	-0.0390	-0.5395
					(0.4014)	(0.3832)	(0.7526)
comfortable prefs. patient ratio					-0.2321	-0.2838	-0.1917
					(0.1678)	(0.1850)	(0.2499)
N	64	64	64	64	64	64	64
R^2	0.3613	0.5962	0.5962	0.6211	0.6100	0.6412	0.0750

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent patient race, age & heart disease history; sampling weights take into account differences in the number of observations per HRR

Table 4(b): Estimates for Explaining Ln Per Beneficiary Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Combined Sample of PCPs and Cardiologists						
cowboy ratio, all docs	0.4425**	0.3362**	0.3193**	0.3151**	0.3312**	0.2716*	
	(0.1623)	(0.1109)	(0.1044)	(0.1043)	(0.1163)	(0.1029)	
comforter ratio, all docs	-0.1456	-0.0875	-0.0822	-0.0501	-0.0967	-0.0334	
	(0.0949)	(0.1097)	(0.1106)	(0.1135)	(0.1083)	(0.1248)	
followup lo, all docs		0.0158	0.0136	0.0331	0.0212	0.0201	
		(0.1628)	(0.1601)	(0.1650)	(0.1567)	(0.1583)	
followup hi, all docs		0.2680	0.2806	0.2820+	0.2865	0.3223+	
		(0.1773)	(0.1771)	(0.1673)	(0.1745)	(0.1702)	
have unneeded tests			-0.0807			-0.1380	-0.1599
			(0.1789)			(0.1955)	(0.2042)
see unneeded card.				0.3067		0.3698	0.4322
				(0.2219)		(0.2419)	(0.2775)
aggressive prefs. patient ratio					0.0938	0.0038	-0.2220
					(0.3614)	(0.3687)	(0.4624)
comfortable prefs. patient ratio					-0.1738	-0.2024	-0.1593
					(0.1491)	(0.1606)	(0.1880)
N	64	64	64	64	64	64	64
R^2	0.2133	0.2945	0.2964	0.3354	0.3123	0.3659	0.0931

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent race,age & heart disease history; sampling weights take into account differences in the number of obs. per HRR

Table 5: Estimates for Ln 2-year End-of-Life Expenditures with Interaction Terms

	(1)	(2)	(3)
Combined Sample of PCPs and Cardiologists			
cowboy ratio, all docs	0.4401***	0.4363***	0.4344***
	(0.1153)	(0.1140)	(0.1209)
comforter ratio, all docs	-0.2109+	-0.2124+	-0.2385+
	(0.1128)	(0.1128)	(0.1260)
followup lo, all docs	-0.0047	-0.0136	-0.0354
	(0.2131)	(0.2160)	(0.2201)
followup hi, all docs	0.6891***	0.6750***	0.6684***
	(0.1279)	(0.1379)	(0.1296)
have unneeded tests	-0.0471	-0.0353	-0.0200
	(0.2527)	(0.2490)	(0.2531)
see unneeded card.	0.3987*	0.3591*	0.3766*
	(0.1576)	(0.1743)	(0.1514)
aggressive prefs. patient ratio	-0.0390	-0.0646	0.4746
	(0.3832)	(0.5076)	(0.9575)
comfortable prefs. patient ratio	-0.2838	-0.4462*	-0.0282
	(0.1850)	(0.2149)	(0.3470)
cowboy ratio * agg. prefs. patient ratio		0.1744	
		(2.0010)	
cowboy ratio * comf. prefs. patient ratio		0.6870	
		(0.9229)	
comforter ratio * agg. prefs. patient ratio			-1.2884
			(2.1294)
comforter ratio * comf. prefs. patient ratio			-0.6247
			(0.7451)
N	64	64	64
R^2	0.6412	0.6436	0.6443

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent patient race, age & heart disease history; sampling weights take into account differences in the number of observations per HRR

Table 6(a): Predictors of Cowboy Type

Predictors of being a Cowboy (all doctors)				
age	0.0313***	0.0253***	0.0283***	0.0279***
	(0.0074)	(0.0074)	(0.0074)	(0.0076)
male	0.3836+	0.3309	0.2507	0.3688+
	(0.2043)	(0.2025)	(0.2036)	(0.2098)
weekly patient days	-0.0776+	-0.0955*	-0.0795+	-0.0655
	(0.0440)	(0.0443)	(0.0448)	(0.0455)
board	-0.3939+	-0.4339*	-0.5383**	-0.3970+
	(0.2068)	(0.2075)	(0.2060)	(0.2119)
cardiologists per 100k	0.1172**	0.1139*	0.1090*	0.1282**
	(0.0452)	(0.0448)	(0.0455)	(0.0461)
fraction capitated patients	0.4663+			0.6123*
	(0.2710)			(0.2775)
fraction Medicaid patients	1.4366**			1.7123**
	(0.5233)			(0.5452)
single/multi speciality group		-0.2947+		-0.3611*
		(0.1520)		(0.1576)
group/staff HMO or Hospital		-0.7501**		-0.9934***
		(0.2775)		(0.2891)
responds to pat. expect.			-0.0518	-0.1294
			(0.1580)	(0.1612)
responds to colleague expect.			0.1357	0.0899
			(0.1520)	(0.1546)
responds to refrrer expect.			0.4896*	0.5685**
			(0.2123)	(0.2158)
responds to malpract. concerns			0.0010	-0.0247
			(0.1506)	(0.1518)
N	1349	1349	1349	1349
sigma u	0.5034	0.4895	0.5136	0.5051
rho	0.0715	0.0679	0.0742	0.0720

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

All logit regressions include a constant & HRR-level random effects

Table 6(b): Predictors of Comforter Type

Predictors of being a Comforter (all doctors)				
age	-0.0014	-0.0035	0.0012	0.0006
	(0.0062)	(0.0063)	(0.0063)	(0.0064)
male	-0.4903**	-0.4520**	-0.3576*	-0.3429*
	(0.1542)	(0.1538)	(0.1563)	(0.1584)
weekly patient days	0.1150**	0.1207**	0.0742b+	0.0845*
	(0.0377)	(0.0378)	(0.0383)	(0.0389)
board	-0.1240	-0.0693	-0.0576	-0.0126
	(0.1874)	(0.1888)	(0.1883)	(0.1919)
cardiologists per 100k	-0.1078**	-0.1187***	-0.0991**	-0.1051**
	(0.0334)	(0.0337)	(0.0340)	(0.0344)
fraction capitated patients	-0.0151			-0.1178
	(0.2310)			(0.2360)
fraction Medicaid patients	0.4327			0.2121
	(0.4604)			(0.4715)
single/multi speciality group		-0.2756*		-0.1548
		(0.1298)		(0.1334)
group/staff HMO or Hospital		0.0651		0.1403
		(0.2121)		(0.2186)
responds to pat. expect.			0.5068***	0.4929***
			(0.1334)	(0.1349)
responds to colleague expect.			0.0448	0.0363
			(0.1252)	(0.1259)
responds to refrrer expect.			-0.3058	-0.2785
			(0.2026)	(0.2037)
responds to malpract. concerns			0.1766	0.1764
			(0.1241)	(0.1243)
N	1349	1349	1349	1349
sigma u	0.2263	0.2237	0.2432	0.2394
rho	0.0153	0.0150	0.0177	0.0171

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

All logit regressions include a constant & HRR-level random effects

Table 6(c): Predictors of High Follow-Up Type

Predictors of being a High Follow-Up Physician (all doctors)				
age	0.0374***	0.0249***	0.0343***	0.0306***
	(0.0073)	(0.0073)	(0.0072)	(0.0075)
male	-0.3794*	-0.2968	-0.3619*	-0.1684
	(0.1817)	(0.1823)	(0.1805)	(0.1890)
weekly patient days	0.0202	0.0201	-0.0297	0.0176
	(0.0434)	(0.0436)	(0.0439)	(0.0451)
board	-0.9227***	-0.8206***	-0.9429***	-0.6664**
	(0.2001)	(0.2010)	(0.1995)	(0.2061)
cardiologists per 100k	0.2160***	0.1990***	0.2144***	0.2286***
	(0.0429)	(0.0417)	(0.0435)	(0.0429)
fraction capitated patients	0.4562b+			0.5544*
	(0.2666)			(0.2749)
fraction Medicaid patients	2.3044***			2.2064***
	(0.5249)			(0.5415)
single/multi speciality group		-1.1575***		-1.0872***
		(0.1468)		(0.1517)
group/staff HMO or Hospital		-0.8845***		-1.1129***
		(0.2428)		(0.2579)
responds to pat. expect.			0.2920b+	0.1084
			(0.1567)	(0.1619)
responds to colleague expect.			0.3384*	0.2770b+
			(0.1480)	(0.1526)
responds to refrrer expect.			-0.8101**	-0.6303*
			(0.2654)	(0.2704)
responds to malpract. concerns			-0.0227	-0.0382
			(0.1474)	(0.1509)
N	1349	1349	1349	1349
sigma u	0.4482	0.3854	0.4710	0.3938
rho	0.0576	0.0432	0.0632	0.0450

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

All logit regressions include a constant & HRR-level random effects

Appendix Table A: Estimates for Explaining Ln 2-year End-of-Life Expenditures (Cardiologists Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Cardiologists				
cowboy ratio, cards	0.1825+	0.1521+	0.1978*	0.1804*	0.1387+	0.2082*	
	(0.1027)	(0.0842)	(0.0832)	(0.0892)	(0.0825)	(0.0867)	
comforter ratio, cards	-0.1261	0.0002	0.0212	0.0189	-0.0008	0.0298	
	(0.1100)	(0.0867)	(0.0898)	(0.0905)	(0.0868)	(0.0879)	
followup lo, cards		-0.1863+	-0.1846+	-0.1747	-0.1992+	-0.1892+	
		(0.1046)	(0.1019)	(0.1079)	(0.1055)	(0.1052)	
followup hi, cards		0.4794***	0.4699***	0.4839***	0.4777***	0.4750***	
		(0.1157)	(0.1119)	(0.1136)	(0.1136)	(0.1094)	
have unneeded tests			0.3138+			0.2927	0.2343
			(0.1865)			(0.1932)	(0.2302)
see unneeded card.				0.2783		0.2681	0.2411
				(0.1899)		(0.1841)	(0.2083)
aggressive prefs. patient ratio					-0.2581	-0.2678	-0.2870
					(0.2843)	(0.2870)	(0.4397)
comfortable prefs. patient ratio					-0.0585	-0.1270	0.0120
					(0.1363)	(0.1353)	(0.1559)
N	64	64	64	64	64	64	64
R^2	0.0535	0.4272	0.4476	0.4481	0.4324	0.4728	0.0406

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent patient race, age & heart disease history; sampling weights take into account differences in the number of observations per HRR

Appendix Table B: Estimates for Explaining Ln 2-year End-of-Life Expenditures (PCPs Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	PCPs						
cowboy ratio, PCPs	0.6689***	0.4385***	0.4163***	0.3723**	0.4402***	0.3166**	
	(0.1687)	(0.1239)	(0.1143)	(0.1269)	(0.1193)	(0.1105)	
comforter ratio, PCPs	-0.2489+	-0.2633*	-0.2506*	-0.2436*	-0.2450*	-0.2096*	
	(0.1380)	(0.1081)	(0.1096)	(0.1005)	(0.1016)	(0.0979)	
followup lo, PCPs		-0.2202	-0.2211	-0.1716	-0.2793	-0.1812	
		(0.2575)	(0.2404)	(0.2601)	(0.2454)	(0.2239)	
followup hi, PCPs		0.4122*	0.4372**	0.4438**	0.3961**	0.4860***	
		(0.1571)	(0.1569)	(0.1529)	(0.1411)	(0.1338)	
have unneeded tests			-0.2236			-0.3111	-0.2371
			(0.3027)			(0.3034)	(0.3941)
see unneeded card.				0.3972*		0.5368**	0.7422*
				(0.1752)		(0.1934)	(0.3350)
aggressive prefs. patient ratio					-0.3592	-0.2056	-0.6638
					(0.6067)	(0.5193)	(0.9768)
comfortable prefs. patient ratio					-0.3021	-0.3067	-0.3864
					(0.2425)	(0.2433)	(0.3348)
N	64	64	64	64	64	64	64
R^2	0.3430	0.4932	0.4995	0.5209	0.5086	0.5545	0.1290

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent patient race, age & heart disease history; sampling weights take into account differences in the number of observations per HRR

Appendix Table C: Estimates for Regression Explaining Log Expenditures for Forward Looking Cohort of Hip Fracture Patients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Combined Sample of PCPs and Cardiologists						
cowboy ratio, all docs	0.4725**	0.3332**	0.2935**	0.3145*	0.3187**	0.2316*	
	(0.1429)	(0.1199)	(0.0929)	(0.1185)	(0.1121)	(0.0882)	
comforter ratio, all docs	-0.1427	-0.0726	-0.0577	-0.0383	-0.0631	0.0115	
	(0.1061)	(0.1086)	(0.0974)	(0.1093)	(0.0949)	(0.0932)	
followup lo, all docs		0.0313	0.0351	0.0324	0.0121	0.0148	
		(0.1609)	(0.1536)	(0.1632)	(0.1372)	(0.1460)	
followup hi, all docs		0.2827*	0.3116*	0.2831**	0.2834*	0.3258**	
		(0.1201)	(0.1210)	(0.1054)	(0.1181)	(0.0980)	
have unneeded tests			-0.1657			-0.2462	-0.2809
			(0.1942)			(0.2178)	(0.2765)
see unneeded card.				0.2072		0.3025*	0.4274+
				(0.1369)		(0.1437)	(0.2239)
aggressive prefs. patient ratio					-0.1061	-0.1277	-0.3071
					(0.3593)	(0.3243)	(0.5621)
comfortable prefs. patient ratio					-0.1346	-0.1378	-0.1867
					(0.1556)	(0.1715)	(0.2025)
N	64	64	64	64	64	64	64
R^2	0.3488	0.4645	0.4753	0.4908	0.4745	0.5290	0.1565

+p<0.10, * p<0.05, ** p<0.01, *** p<0.001

2-year EOL Spending is price, age, sex and race adjusted spending; results for 64 HRRs in which we have at least three patients and cardiologists surveyed; all regressions include a constant and controls for respondent patient race, age & heart disease history; sampling weights take into account differences in the number of observations per HRR