The Health Benefits of Medicare Expenditures: Evidence from the Healthcare Cost Slowdown

Amitabh Chandra
Kennedy School of Government, Harvard University,
and NBER

Carrie Colla
The Dartmouth Institute of Health Policy and Clinical Practice,
Geisel School of Medicine

Jonathan Skinner
Dartmouth College, Department of Economics,
The Dartmouth Institute of Health Policy and Clinical Practice,
Geisel School of Medicine, and NBER

December 2016
Preliminary – Please do not quote or distribute

Abstract

There is little consensus on the question of whether the incremental Medicare dollar spent yields health benefits to elderly Americans. In this paper, we develop a model of productive and allocative efficiency, and demonstrate that most empirical estimates are biased because they fail to account for differences in supply-side hospital productivity. We then test the model using a dataset of 890,080 Medicare enrollees, including 128,337 tourists, admitted to hospital with a heart attack, or acute myocardial infarction, during 2007-12. During this period, some hospitals experienced a slowdown in Medicare spending, and we further test whether reductions in spending had an adverse impact on health. We develop hospital-year measures of highly effective inputs (e.g., beta blocker and statin use, primary stenting) and ineffective inputs (e.g., home health care predictive of fraud). We explain a large fraction of hospital-level variability in both survival, and Medicare spending, solely by differences in hospital input choices. Hospitals that cut spending during 2007-12 experienced no significant difference in survival trends from those that increased spending, because the cutbacks occurred for inputs with little or no incremental health benefit. That is, it's less important how much money is spent; far more important is *how* it is spent.

We are grateful to the National Institute on Aging (PO1 AG19783) and the National Institutes of Health (U01 AG046830) for financial support, and John E. Wennberg, Douglas Staiger, Daniel Gottlieb, Weiping Zhou, and Greg Roth for very helpful comments and assistance. Daniel Gottlieb and Weiping Zhou provided extraordinary programming and data assistance.

I. Introduction

Even with the slowdown in health care spending, the federal government is projected to spend more than 7 trillion dollars in the Medicare program over the next 9 years (Medicare Trustees, 2014). From a public finance viewpoint, we know remarkably little about what U.S. taxpayers are getting for all that money. In cost-benefit analysis, the primary policy tool for judging this question is the marginal value in health benefits of the last Medicare dollar spent. In a world where physicians, patients, and third parties allocate healthcare resources efficiently, a summary statistic of overall expenditures would translate directly into a specific level of health outcomes, with the marginal medical value of healthcare expenditures always positive.

However, following on earlier research by John E. Wennberg (Wennberg, 2010), several studies in the mid-2000s suggested zero or negative average returns to wide variations in healthcare expenditures (Fisher et al., 2003a,b; Baicker and Chandra, 2004; Skinner, et al. 2005; Yasaitis et al., 2009). Most recently, a consensus report from the Institute of Medicine concluded that there was very little association between spending and health outcomes (IOM, 2013).

Other studies have found a positive associations between expenditures and health outcomes (Hadley, et al., 2011; Romley, et al., 2011; Silber, et al., 2010). Doyle (2011) used tourists far from home, and Doyle et al. (2015a) the loyalty of ambulances to hospitals, to create plausible natural randomizations that sidestepped potential biases arising from unmeasured health characteristics of patients.

In this paper, we develop a simple model of hospital productivity that draws on Wennberg et al. (2002), Chandra and Skinner (2012), and Díaz-Hernández et al. (2008). We demonstrate that in the real world of pervasive productive inefficiency, conventional estimates of the association between spending and outcomes cannot be interpreted in a meaningful way

(Garber and Skinner, 2008). When input choices are not optimized, the total level of spending is no longer a summary statistic for the intensity of care; instead what matters is how the money is spent. Our focus on choice of treatment, rather than how much money is spent, is motivated by the simple observation that the incremental dollar spent on beta blockers and statins after a heart attack has a highly cost effective impact on health, while the incremental "outlier" dollar spent for a given home health care patients is unlikely to improve health (as in McKnight, 2006), and to be a signal for fraud (OIG, 2012). Without knowledge of how the incremental dollar is being spent, the association between spending and outcomes could be positive, negative or zero, even when every treatment provides positive incremental value.[1] That Doyle et al. (2015a) should find positive effects of spending on outcomes for acute-care hospital treatments within the first 30 days, but that Doyle et al. (2015b) find no impact of spending on outcomes for a one-year horizon, where post-acute care is a major component of spending, is consistent with our model.

We test this model by considering the entire population of 890,000 elderly Medicare enrollees with acute myocardial infarction (AMI), or heart attacks during 2007-2011, with data on outcomes through December 2012. During this period, average real per-enrollee Medicare spending for heart attack patients was essentially flat, with many hospitals experiencing real declines in spending. The period is therefore well-suited to testing the hypothesis that more Medicare spending yields better health outcomes, and the converse that a decline in spending should adversely affect the health of the elderly.[2] In our analysis, we use both the full Medicare

[1] This result has a parallel in education, where there is a general lack of association between school spending and schooling outcomes, despite evidence that some specific interventions are highly effective (Hanushek, 2006).

[2] If every incremental Medicare dollar really does deliver positive benefits, then the recent Congressional Budget Office estimates that scale back Medicare spending projections by $715 billion from 2011 to 2020 (http://www.cbo.gov/publication/45581) could have consequences for the health of elderly people.

dataset, and a smaller sample of 128,000 "tourists," or AMI patients treated in hospitals outside of their Hospital Referral Region (HRR), following on Doyle (2011)'s study of Florida tourists.

We use the model to derive a two-equation estimation system in which both survival and expenditures depend on specific input choices based on Chandra and Skinner (2012), including "Category I" or highly effective treatments such as beta blockers and statins; "Category II" treatments with heterogeneous benefits such as percutaneous coronary interventions (PCI, commonly stents) or multiple physician visits, and "Category III" treatments with low or unknown benefit such as "double-CT" scans for the same part of the body that "needlessly exposes the patient to radiation" (Bogdanich and McGinty, 2011), outlier billings for home health care predictive of fraud, and "Choosing Wisely" treatments deemed by professional physician groups to be wasteful (Colla, et al., 2015).

We found that the estimated marginal value of some medical inputs are large. For example, the effective use of beta blockers within 6 months of discharge predicts a reduction in mortality of about 2.0 percentage points, while "early" stenting in the first day after the heart attack yielded large survival benefits. (Results for tourists were generally similar although sometimes more imprecisely estimated.) Yet other treatments were not estimated to be as valuable; an increased number of different physicians did not exhibit any benefit beyond the bottom quartile of use. Nor was there generally any salutary effect of our third category of treatments on health outcomes, although they were strongly predictive of high Medicare bills.

The model is used to develop measures of hospital-level productivity that are not based on actual risk-adjusted spending and survival, but instead are based solely on what hospitals (and physicians associated with those hospitals) do – that is, on treatment choices and factor inputs. Assigning $100,000 per quality-adjusted life-year to improved AMI survival, we find that the net social productivity of hospitals, relative to the national average, range from the Mayo Clinic ($33,114) in Rochester, Minnesota, and Intermountain ($41,710) in Salt Lake City, to Hialeah Hospital (-$44,897 per admission) in Hialeah, Florida.

Using the model estimated with hospital fixed-effects, we can further address the question of whether health care spending slowdowns has troublesome implications for health. We find that among hospitals with at least a 5 percent real reduction in health care costs during 2007-2011, roughly half of the declines were associated with cutbacks in potentially fraudulent home health care, multiple physician visits, and other services unlikely to adversely affect health; in contrast, spending for effective Category I/II services continued to grow. In hospitals with declining rates of spending, survival improved by 1.5 percentage points over this period; for those with more rapid increases in spending, survival improved by a nearly identical (and statistically indistinguishable) 1.7 percentage points. That is, most of the reductions in Medicare spending appear to be associated with Category II/III treatments with little or no impact on health outcomes.

II. The Model

There have been a variety of studies seeking to determine whether "more is better" with regard to health care expenditures. As noted above, some found a mixture of mostly zero or negative associations between spending and outcomes (Fisher et al. 2003a,b, Baicker and Chandra 2004, Skinner et al. 2005, Yasaitis, et al. 2009, Glance et al., 2008, Rothberg, et al. 2010; see Hussey et al., 2013 and IOM, 2013). These studies differ along a number of dimensions: the level of analysis (area-level, provider-level), the measure of quality, the measure of cost, and statistical approaches to risk adjustment. There are also a variety of studies that have found a positive association between spending and health outcomes (e.g., Doyle et al., 2015a, Hadley, et al., 2011; Romley, et al., 2011; Silber, et al., 2010). Barnato et al. (2010) is close in spirit to this paper, because they used specific inputs -- ICU use, mechanical ventilation, hemodialysis, tracheostomy and feeding tubes – to estimate the return to treatments for patients with a high predicted probability of death at hospital admission. On average, they found survival benefits of more intensive care: a roughly $14,000 increase in per capita expenditures for these treatments translated into only a 1.5 percent improvement in the chance of surviving an extra six months (with no clear evidence beyond 6 months).

Doyle (2011) focused on tourists to abstract from unmeasured health status leading to both higher costs and greater “risk adjusted” mortality. His intuition was that tourists with emergency illness (such as heart attacks) are allocated randomly to both high-cost and low-cost hospitals, and thus serves as a natural randomization of patients to hospitals. We follow on this insight to consider tourists separately in our study, defined as people who receive treatments in hospital referral regions (HRRs) outside their resident HRR. While we refer to these patients as “tourists” we recognize that there may be some cross-HRR-border travel even for acute care.[3]

Nearly all studies, regardless of regression results, find wide variation in both spending and health outcomes across institutions (Chandra, 2012), a finding that motivates our focus on productivity (also see Chandra, Syverson, Finkelstein and Scarny, 2012). We define a hospital-level “production” function $S_j = \alpha_j S(X, Z)$ for a patient with characteristics and health status Z, and a matrix of inputs $X = \{X_k\}$, $k = 1,\ldots,K$ that reflects the K different potential inputs into the production function. If we assume that the mean of $\alpha_j = 1$ across hospitals (without loss of

[3] We also tried imposing a 200 mile distance limitation between the residence and the hospital, but this resulted in very small sample sizes of just 33,000 (despite 5 years of 100% samples) and very imprecise estimates.

generality), then proportional variations in α correspond to differences in total factor productivity – that is, the proportional difference in outcomes given the same set of inputs. These differences are likely to reflect physician skills, organizational structure, and other institution-specific factors. Physician skills, of course, are also likely to affect input choices, so we cannot assume that α is uncorrelated with inputs.

We assume that physicians and other health professionals working at hospital j (or who see follow-up patients from hospital j) seek to maximize the social value of health $\Psi\alpha_j S(X,Z)$ where Ψ is the implicit dollar value of the quality-adjusted improvement in health outcomes. If we further assume that the (Medicare) prices P_k paid for each procedure reflects the true cost of the resource, and subscript the individual inputs and health status measures by individual i and hospital j, then the social welfare function for each patient can be written

$$\Omega_{ij} = \Psi\alpha_j\, S(X_{ij}, Z_{ij}) - PX_{ij} \qquad (1)$$

Note that the first-order condition for each input k is:

$$\Psi\alpha_j \partial S/\partial X_{ijk} - P_k = 0 \quad k = 1,\ldots,K \qquad (2)$$

When productive and allocative efficiency for each of the k first-order conditions in (2) holds, then by the smoothness of the production function, and the envelope theorem, each of the k different inputs yields equal incremental health per dollar spent.

This can be shown in a traditional production setting, as in Figure 1. Hospital A is more productive than hospital B, in the sense that they can "produce" the same level of survival S* at lower cost (hence $\alpha_A > \alpha_B$). The distance between Point D and Point C (measured in terms of input 1, $E_D - E_C$), is the difference in costs between the two hospitals of producing the same output. (If the lower quality hospital is unable to produce the quality provided by the high-quality hospital, at any cost, then the comparison is for an attainable S*.)

The second source of inefficiency arises from inefficient use of factor inputs. This is illustrated in Figure 1 by Point E, which is off of the tangency reflecting optimal input use. Thus the inappropriate use of inputs X (e.g., technical inefficiency) adds to overall inefficiency. Some macroeconomists have suggested that in non-health industries, this additional source can explain substantial fractions in cross-country productivity (Restuccia and Rogerson, 2008). We capture the idea of inefficient input choice through the use of a "shadow" price reflecting either barriers to use (a positive shadow price) or financial incentives to overuse (a negative shadow price).

Following Díaz-Hernández et al. (2008) and Skinner and Staiger (2015), we include an additional shadow constraint $\Gamma_j = \{\lambda_{jk}\}$, k = 1,…,K, to capture these additional factors. The Lagrangian is therefore written:

$$\mathcal{L} = \Psi\alpha_j S(X_{ij}, Z_i) - P_k - \Gamma_j \qquad (1')$$

The first-order condition for input k is now:

$$\Psi\alpha_j \partial S/\partial X_{ijk} - \mathrm{P}_k - \lambda_{jk} = 0 \qquad (2')$$

With just 2 inputs, X_1 and X_2, the social optimum occurs when $\lambda_1 = \lambda_2 = 0$. Figure 2 shows this point as "A" along the (Hospital j) production possibility frontier, with aggregate inputs $X = P_1X_1 + P_2X_2$ shown on the horizontal axis.

Allocative inefficiency occurs when hospitals remain on the production possibility frontier, but provide care beyond the marginal benefit, for example where the marginal cost of each factor does not justify the marginal benefit (given Ψ). In this case, $\lambda_1 < 0$ and $\lambda_2 < 0$, but $\lambda_1/P_1 = \lambda_2/P_2$, thus ensuring that the marginal rate of transformation between X_1 and X_2 is still equal to the ratio of the socially optimal prices, as shown in Point A* in Figure 2.

The shadow constraint λ captures the clinical ideas of underuse ($\lambda > 0$) and overuse ($\lambda < 0$). For example, suppose that X_1 represents beta blockers or aspirin on admission for a heart attack, with very high benefit and negligible financial cost. This is a discrete treatment – one either gives beta blockers, or one doesn't – so one can also think of it as a continuous measure at the hospital level of the fraction of patient likely to benefit from the treatment who actually get it. In this case, underuse ($\lambda_1 > 0$) leads to worse health outcomes, but with little or no impact on health care costs; this is shown in Figure 2 by Point B.

Why might this underuse occur? Information about the value of beta blockers may have been scarce in earlier years because of high search costs (e.g., Skinner and Staiger, 2015) or incorrect physician beliefs about their effectiveness. Other factors includes poor organizational or management structure (e.g., Bloom et al., 2014) and the lack of leaders championing their use (Bradley et al., 2005). We cannot quantify these reasons, but can capture the implicit costs (defined broadly) that would have generated the behavior we observe.

When X_2 is a treatment that has little or no value for patients relative to its price, but is overused, then $\lambda_2 < 0$ and Point C in Figure 1 occurs. Overuse would happen if providers are maximizing something other than patient health—for example, if they are trying to maximize profits. Either underuse or overuse may even happen if providers receive imperfect or incomplete information on patient outcomes (for example, physicians may observe survival over a week and try to maximize that, but use treatments with poor long-term effects on survival). Point D, which both costs more than the hospital at Point A, and yields worse results, can occur for two reasons.

Under one scenario, Hospital D in Figure 2 is worse than Hospital B when it's on the same production function $\alpha_h S(X)$, but where it exhibits both the underuse of highly effective beta blockers ($\lambda_{1D} > \lambda_{1B}$) and the overuse of low-value treatments ($\lambda_2 < 0$). Under a different scenario, Hospital D is making optimal input choices, but it is simply on a less productive production function, where $\alpha_l < \alpha_h$ as in Figure 2. One can also think of α_l in this case as capturing managerial or physician expertise independent of input choice (Bloom et al., 2010, 2014 and McConnell et al., 2013). We can distinguish between the two explanations for why Hospital D lags behind with the use of hospital fixed effects that sweep out any permanent differences in α, but do not adjust for systematic overuse or underuse of measured inputs.

Figure 2 illustrates why conventional estimation approaches are doomed, even when researchers have adjusted perfectly for underlying health of the patients. The researcher would like to estimate the slope of the production possibility frontier αS(X). As noted by Diaz-Hernandez et al. (2008), productively inefficient systems creates real problems for researchers seeking to estimate this slope. If we observe Hospitals A, A*, B, C, and D, with different degrees of productive inefficiency (either because they are on a different production function, or exhibit overuse or underuse), the regression line could be positive, negative, or zero regardless of the "true" slope of the production function. When there is a positive correlation among λ_k, then the conventional regression will tend to overstate the slope of the production function, and conversely for a negative correlation. A negative association between spending and outcomes (as in, for example, Fisher et al., 2003a,b) could therefore occur not because "more is worse," but because the correlation between Category I inputs (those that save lives) and Category III inputs (those that cost money) could be negative.

To characterize steady-state differences across hospitals in both survival and spending, we take the derivative of the first-order condition (2) with respect to changes in λ and changes in X_1, X_2, and the unobserved α and rearranging yields a first-order Taylor-Series of spending and outcomes for hospital j:

$$S_j = \bar{S} + \bar{S}\,\alpha_j + \sum_{k=1,\ldots,K} S'_k\,(X_{jk} - \bar{X}_k) \qquad \text{(3a)}$$

$$M_j = \bar{M} + \sum_{k=1,\dots,K} P_k \ (X_{jk} - \bar{X}_k) \quad \text{(3b)}$$

where outcomes, spending, and inputs are expressed for hospitals by aggregating over individual utilization: $X_{jk} = \sum_{i=1}^{N} X_{ijk}$, $S_j = \sum_{i=1}^{N} S_{ij}$, $M_j = \sum_{i=1}^{N} M_{ij}$, and a bar denoting the population average.[4]

Consider next a measure of hospital-specific net social benefit (or productivity) that depends both on hospital health outcomes, and expenditures. Taking a linear approximation of the first-order condition (2) yields an expression for the deviation of X_k from the average: $X_{jk} = \bar{X}_k + (\lambda_k - \bar{\lambda}_k)/S_k''$, where $S'' < 0$ is the second derivative of the survival production function (as in Chandra and Skinner, 2012), and again we assume that all hospitals face similar prices. The net social benefit is determined by combining 3(a), weighted by the social value of survival, and 3(b):

$$\Omega = (\Psi S_j - M_j) - \overline{[\Psi S} - \bar{M}] = \alpha_j \bar{S} + \sum_{k=1,2,u} (\Psi S_k' - P_k)(\lambda_k - \bar{\lambda}_k)/S_k'' \qquad (4)$$

The left-hand side of Equation (4) reflects the social value of Medicare spending for Hospital j relative to the average hospital value (denoted with a bar). It is written as the sum of the incremental values for each of the treatments. We assume input 1 exhibits high incremental value relative to its cost, that is, the value of the incremental life saved $\Psi S_k'$ minus the price of the service P_k. When there is relative underuse of input 1, so that $(\lambda_1 - \bar{\lambda}_1 > 0)$, this together with S" < 0 will imply that that hospital will lag behind the average hospital, primarily because of lower-than-average life-expectancy of the hospital's patients. Similarly, the overuse of input 2, where the incremental value of the service is negative, and $(\lambda_2 - \bar{\lambda}_2 < 0)$ leads to higher than average costs with minimal health benefits.

[4] Note that in this version of the model, the marginal value of the observable inputs are assumed constant across hospitals. Chandra and Staiger (2007) generalizes the model to one where differences in physician expertise lead to variations in input use for surgical treatment.

Note that we focus on the difference across hospitals in the *difference* between outcomes and costs, rather than the cost-effectiveness ratio $(M_j - \bar{M})/\Psi(S_j - \bar{S})$. If the ratio is positive, then we would expect to see that a decline in hospital-level Medicare spending would translate into a loss of lives. We test this hypothesis in the data section below during the slowdown in Medicare expenditure growth.

III. Estimation Strategy

We next turn to the empirical specification that captures potential health differences across patients. Writing γ_k for S_k' yields a simple estimation equation

$$S_{ij} = Z_{ij}\beta + \alpha_j\bar{S} + \sum_{k=1}^{K} X_{jk}\gamma_k + u_{ij} \qquad (6a)$$

Similarly,

$$M_{ij} = Z_{ij}\varphi + \sum_{k=1}^{K} X_{jk}\phi_k + \varepsilon_{ij} \qquad (6b)$$

We adopt an estimation equation, rather than simply counting up the price P_k of each input, because we think that treatment patterns will affect downstream costs as well. For example, a stent in the hospital could well reduce the likelihood of readmission, but it could just as well increase the likelihood of post-acute care.

The problem of estimating production functions in the presence of inefficiency is not new. There are two general approaches to the estimation of production frontiers, the data envelopment analysis and the stochastic frontier analysis (Fried, Lovell, and Schmidt, 2008). Data envelopment analysis picks off the dots along the Northwest of the productivity curve (for example, hospital A and A* in Figure 2), and estimates of inefficiency for all of the other hospitals depend on the distance of hospital j from this estimated production function. The problem with this approach is that it ignores the possibility of stochastic error terms. By contrast, stochastic frontier analysis allows for a random error term.[5] Since inefficiency can only be a negative outcome, stochastic frontier models distinguish inefficiency from random noise by assuming that the random noise is symmetric, so that any remaining skewness is the consequence of inefficiency (Skinner, 1994).

With cross-section time-series analysis, one can estimate institutional fixed-effects (which in theory identifies α_j), as in Kumbhakar, Lien, and Hardaker, (2014). In practice, we will not attempt to identify the fixed hospital effects α_j as productivity differences, since they are also likely to reflect permanent differences across hospitals in the unmeasured characteristics of patients. A more conservative measure of productive inefficiency is also estimated by ignoring the hospital fixed effects, but focusing on the differential use of factor inputs, as noted above.

Data: We created a cohort of patients hospitalized with acute myocardial infarction (AMI) in the fee-for-service Medicare population during 2007-2011, with follow up data through December 31, 2012. An AMI is based on the first diagnosis code (410.x1 or 410.x2), and not on the existence or type of actual diagnostic related group, which can often vary depending on how the patient is subsequently treated.

***Risk adjustment**:* The risk adjustment approach we use includes admission-level comorbidities such as cancer, diabetes, liver disease, peripheral vascular disease, congestive heart failure, the clinical location of the AMI (e.g., inferior, anterior, subendocardial), as well as zip-code-level income quartiles based on the American Community Survey (2010) , and age-sex 5-year cells (e.g., women aged 70-74), and race (African-American, Hispanic, Asian, Native American). In most cases, we also use Hierarchical Condition Categories (HCC), which counts the number of different diagnoses that patients have received in the 6 months prior to the index

[5] See Kuosmanen and Kortelainen (2012); they have also derived a model that combines both effects.

admission, and weights them for severity. However, we note that the use of HCC measures can lead to biases in conventional regressions of spending on outcomes. If a physician sees patients more often, and looks harder for diseases, her patients will more likely be coded as that much sicker (Song, et al., 2010), meaning that when they do survive, the hospital will get credit as a highly productive institution.

Despite these risk adjustments, the standard concern with observational studies is that the error terms are correlated with the treatments provided for that individual. One approach is to focus on a group of patients with similar risk characteristics who are distributed randomly across hospitals. Doyle (2011) considered tourists in Florida, under the reasonable assumption that few tourists consider whether their vacation hotel is near a high- or low-intensity hospital. Thus we also consider the subset of AMI patients in our national 100% database admitted to hospitals that were in a different hospital referral region (HRR); 128,237 were successfully matched. Summary statistics are presented in Table 1.

In the hospital-level analysis for the entire sample, we limit hospitals with at least 50 admissions for AMI during the combined years 2007-11 and at least 10 admissions in each year. For the tourist sample, we limit hospitals with at least 5 admissions per year.

Another concern is that people who live longer either could require less acute care, or (as is more often the case) will receive more post-acute care. (Even in high-fraud areas, few dead people are billed for Medicare services.) Aggregating individual treatments up to the hospital (or hospital-year) level, to focus on the hospital-based system of care also removes most of the bias. Even at the hospital level, if average survival rates rise, it could cause spending to change as well; we address this problem by measuring “exposure” or spending on specific inputs only among those who survive an entire year (operationally, this adjustment has little impact on results).

The "treatment effect" is therefore being admitted to a hospital that cares for its other patients in a specific way; this may include not only other AMI patients, but also patients in with other diseases (as for example feeding tubes for advanced dementia patients). This approach removes patient level correlation between patient inputs and illness, but cannot eliminate it at the hospital level, since some hospitals may have higher risk patients than others. As noted above, hospital fixed-effects models removes that bias when there are permanent differences across hospitals in their unobserved patient case-mix.

Our fourth approach is to focus on treatment measures that are justified (or unjustified) no matter what the health of the patient. For example, nearly all AMI patients should be taking beta blockers and statins after discharge, regardless of health status, while few if any patients should be receiving feeding tubes, double-CT scans, or extraordinarily high levels of home health care that the Office of the Inspector General has identified as a marker for fraud (OIG, 2012). This approach sidesteps the need for risk adjustment to judge the magnitude of productive inefficiency.

Clinically relevant inputs: Category I, II, and III. There are wide array of different treatments, with differing rates of effectiveness, for AMI patients. We consider a range of such treatments or procedures where our initial hypothesis of effectiveness is based on existing clinical evidence. To help organize the data analysis, we follow Wennberg et al. (2002) and Chandra and Skinner (2012) by appealing to clinical evidence to collapse this broad array of treatment effectiveness into three broad groups.

The first is "effective care" (Wennberg, 2010). These Category I technologies are distinguished by their high cost-effectiveness and limited scope for expensive overuse Examples are beta blocker and statin prescription fills for AMI patients during the 6 months after discharge from the hospital for AMI (Munson et al., 2013). Nearly everyone should get such treatments, regardless of health status. Another example is one that does not hold directly for AMI patients (because of lack of sample size), but that does reflects the degree of integration between the hospital and the physicians in the community: the fraction of patients discharged from the hospital for a medical condition that is seen by any physician within 14 days. This type of visit has been shown to be effective in reducing readmission rates and is often used as a quality measure (Sharma et al., 2010; Hernandez et al., 2010). We also hypothesize that teaching hospitals yield better outcomes at similar costs.[6]

The second category is somewhat less cost-effective, often because of considerable heterogeneity in benefits across different types of patients. Medical services may be placed in this group for a number of reasons: patient characteristics (patient age, other comorbidities, genomic makeup), patient preferences, or the timing of the medical service. One example is a coronary percutaneous intervention (PCI), in which a collapsed balloon is led by catheter into the blocked artery (or arteries) of the heart muscle, where it is inflated (and then withdrawn) to improve blood flow, typically in conjunction with a stent, a wire cylindrical mesh that helps to keep the artery open. It is highly effective in saving lives if administered for appropriate patients within 12 or 24 hours of a heart attack (Hartwell, et al., 2005). But the benefit is far more modest for patients with stable angina (Weintraub, et al., 2008), and for people who receive them

[6] In practice, teaching hospitals tend to charge much more per procedure, but since we measure only input use holding prices constant across hospitals, we test whether patients at such hospitals yield better outcomes for the same set of inputs, and whether teaching hospitals are associated with the greater use of factor inputs.

well beyond the initial heart attack. In many cases, these treatments yield sufficiently modest benefits, and entail sufficient risks, that the decision of whether to undergo the procedure should be determined by patient preferences; hence Wennberg et al.'s (2002) designation of such treatments as "preference-sensitive."

Another example of potential Category II treatment occurs when a larger number of different physicians treats the same patient. As demonstrated by Becker and Murphy (1992), more specialization can improve productivity, but at some point, there are diminishing returns to additional physicians, owing to rapidly rising costs of coordinating care. Thus we might expect a U-shape influence of the number of different physicians on mortality rates. A final example is the number of MRI and CT scans. Clearly the first few scans can save lives, but other studies have suggested that incremental CT scans for stroke patients are not associated with better outcomes, and carry significant radiation risk (Bekelis, et al., 2014). To capture the idea of heterogeneous benefits arising from these treatments, we created quartiles of treatment use and estimated each separately in both the survival and spending regressions.

Category III (low-value or potentially harmful) treatments are those for which marginal benefit is either small or unknown, but that have a large effect on spending. Services labelled as those "physicians and patients should question" by the Choosing Wisely program also fit this description.[7] As noted above, we use one measure from OIG (2012), the fraction of home health care patients with "outlier" payments that put the individual in the top 10 percent of all home

[7] These are a list of procedures created by national specialty groups where there is little or no evidence of benefit and often involve potential harm to patients. See http://www.choosingwisely.org/

health care spending, $11,219 in 2011 dollars. (The null would therefore be that all hospitals report 10 percent.) This measure is not necessarily higher in a sicker population.[8]

We also consider services that are not directly applied to heart attack patients, but reflect the practice and management styles of physicians at the hospital. For example, CMS has published rates of use for "double CT" scans of the chest, one with iodine contrast and the other without. This may be ordered by physicians in the mistaken belief that "more information is better," but it provides no additional clinical information, and is recognized as a marker of poor quality (Bogdanich and McGinty, 2011). Again, we do not believe that this measure per se will lead to higher costs for AMI patients, but instead that it acts as a proxy for a variety of other expensive Category III treatments.

Similarly, the use of feeding tubes for advanced dementia patients is recognized as a marker for poor quality, owing to its lack of effectiveness in improving health status and the associated rates of "agitation, increased use of physical and chemical restraints, and worsening pressure ulcers."[9] Based on Medicare claims data, we estimated hospital-year rates of feeding tube as the ratio of feeding tube use in a patient population of those (a) admitted to the hospital from a nursing home with a diagnosis of dementia, and (b) considering whether the patient subsequently received a feeding tube.

Finally, we note that healthcare is the function of observed inputs like stents and beta-blockers but also unobserved inputs like physician diligence, safety cultures, and nursing care,

---

[8] That is, a region may have many sick AMI patients requiring home health care, but that does not necessarily imply that among those who are receiving home health care, a higher fraction would be "outliers."

[9] This according to a May 2013 position paper by the American Geriatrics Association; see http://www.americangeriatrics.org/files/documents/feeding.tubes.advanced.dementia.pdf

reflected in our unobservable X_u. These factors will be picked up in fixed hospital effects, along with unobservable measures of health status.

V. Results

In Table 1, summary statistics are presented for the basic risk-adjustment model, for the entire sample of 897,008 individuals, and for the 129,289 out-of-HRR “tourists.” Tourists were slightly younger, were less likely to be Black, and were somewhat healthier based on comorbidities at admission.

Table 2 displays year-specific measures of outcomes and inputs categorized by year, for both the entire sample and for tourists. One-year survival grew, from 67.7 percent in 2007 to 70.5 percent in 2011, with a smaller increase for 30-day survival. Yet total inflation-adjusted spending during the same period grew by only 0.25% annually, from $46,234 to $47,088. A similar improvement in mortality was observed for tourists, while their expenditures were virtually flat for 2007-11. Rates at which AMI patients filled beta blocker prescriptions rose slightly from 76 percent in 2007 to 77 percent in 2011, and there was a similarly slow upward trend in the fraction of patients experiencing a follow-up visit within 2 weeks, from 62 to 64 percent.

The use of same-day PCI experienced a substantial increase, from 22 percent in 2007 to 27 percent in 2011, while later PCI was relatively steady at just over 8 percent. The average number of unique doctors grew during the period, from 11.8 to 12.3, while the average number of MRIs and CT scans actually fell. Finally, Category III measures exhibited only slow growth during this period, with modest growth in per-enrollee home health care spending and no change in the use of feeding tubes for dementia patients.

These aggregate spending measures mask substantial variation in average use across hospitals; rates of tube feeding among severely ill dementia patients were equal to 18 percent in one hospital, roughly 6 times the average of about 3 percent, while many hospital rates were zero. These aggregate statistics also mask differences across hospitals in growth rates, a topic to which we return below.

Conventional Regressions: Table 3 presents results from conventional regressions in which the left-hand side reflects the one-year or 30-day survival rate, and the right-hand side of the regression captures both a set of risk adjusters, and the log of price-adjusted average hospital-level spending. All standard errors are clustered by hospital. We present 12 distinct regressions reflecting different risk-adjustment strategies, hospital fixed effects, different survival periods (30-day versus 1 year), and with the regression results for tourists only. The results are all over the map. The thirty-day regression analysis exhibit a positive and significant association between spending and outcomes, whether HCC risk adjustment methods are used or not. The implied cost-effectiveness ratio is about $250,000 or more.[10] Surprisingly, the out-of-HRR estimates show, if anything, less pronounced benefits from spending more at 30-days (and at 1-year).

As in Doyle et al. (2015b) and (Silber et al., 2011), the longer-term one-year survival estimates are generally more modest than the 30-day estimates. Figure 3a shows risk-adjusted spending and risk-adjusted survival for hospitals with at least 400 AMI admissions (so we avoid most variation arising from statistical noise), where once again HCC adjustments are used. As can be seen, while there is a significant positive correlation coefficient equal to 0.09, the more interesting characteristic is the degree of variation for both survival and expenditures. This

[10] This assumes that the average elderly heart attack patient survives 5.25 years after the MI, with no discounting.

means that there are “high-productivity” hospitals with high survival rates and low costs (those in the Northwest corner of Figure 3a), and conversely the presence of “low-productivity” hospitals, with the opposite characteristics, in the Southeast corner of Figure 3a Similar results are shown for tourists, in Figure 3b, for hospitals with at least 100 admissions. In the regression analysis below, in Tables 4 and 5, we seek to explain high and low productivity hospitals based not on residuals or unexplained characteristics, but instead as the consequence of physician and hospital *choices* or the overuse of Category III treatments, the underuse of Category I treatments and the appropriate use of Category II treatments.

Table 4 presents six regression specifications for either one-year or 30-day survival. All regressions use the HCC risk adjustment approaches. The first three regressions do not include hospital fixed effects, but consider one-year survival in a least-squares model, marginal probabilities from a probit model, and for tourists. The results are all broadly consistent; the Category I treatments are estimated to confer positive benefits for survival, although the estimates for tourists are not always significant. Teaching hospitals exhibit slightly better outcomes (a 0.8 percentage point survival gain).

The subsequent three regressions with hospital fixed effects exhibit similar patterns, albeit with smaller effects of statins and no impact of the 14-day physician follow-up. The fifth column limits the sample to only larger hospitals, with an average of 100 AMI patients per year, to ensure that the constructed dependent variables are measured with less noise. These results taken together are suggestive of the unobservable managerial effects that may be reflected in (e.g.) 14-day follow-up visits, which apply to all patients and not solely to AMI patients, and thus may not explain year-to-year variation in AMI hospital outcomes.

The use of stents (PCI) in the first day are strongly predictive of better survival, with similar results for both the cross-sectional analysis and the hospital fixed-effects models. (Results in Table 4 are presented only for the fourth quartile, but the other quartile coefficients, not reported, increase monotonically.) The top-quartile hospitals with respect to early or primary stenting experience a roughly 6 percentage point improvement over hospitals in the bottom quartile of PCI use. Hospitals in the top quartile for late PCI also experience better survival, but these estimated effects are roughly half of the top-quartile early PCI coefficients.[11]

In contrast to our hypothesis, there was no apparent survival benefit associated with the number of unique doctors, and any incremental impact of MRIs and CT scans were small or sensitive to the composition of the sample. Nor did the category III treatments exhibit a consistent association with health outcomes; the fraction home health outlier was not associated with outcomes in the cross-section, and oddly associated positively with survival in the model with hospital fixed effects (Column 4) although this result doesn't hold for any of the other specifications. The use of feeding tubes was also associated in the cross-section analysis with worse outcomes, although again these results did not hold in the fixed-effect model, suggesting that feeding tube use may reflect management style that changes only gradually over time. Double CT scans were associated with adverse outcomes in the cross-section analysis, although the estimate was not significant for tourists.

Table 5 presents the corresponding regression results to Table 4, but with Medicare expenditures as the dependent variable. Category I treatments are either not associated significantly with expenditures, or if anything negatively associated. That the coefficients on the

[11] Because the percentage of early or primary PCI is so much higher, however, the linear estimates on PCI use are similar between early and late PCI.

variable measuring 14 day follow-up visits are so different between the cross-section and time-series (hospital fixed effect) model is again suggestive that this measure reflects a well-integrated health care system, rather than the physician visit *per se* reducing costs.

The top quartiles for early PCI and late PCI exhibit higher costs, and indeed the cost per PCI for the "late PCI" quartile is considerably larger per patient, as would be expected if the late PCI required readmission. Yet the largest influence on Medicare expenditures arises from the number of different physicians and the number of MRI scans; being in the top quartile of use for both of these measures leads to more than 20 percent more spending, holding health characteristics of the patient (as measured by the HCC and other comorbidities) constant. Similarly, the association between home health care outliers and expenditures, and feeding tubes, are remarkably large in the cross section. (Note that the large association between feeding tube use and spending in the cross section does not carry over to the fixed-effect model.) The use of double CT-scans is also associated with higher spending in the cross-section.

We can use these estimates to consider the following question: What fraction of variation shown in Figure 3a is associated with factor input choices made by physicians and hospitals? We can predict survival and spending based on the use of Category I, II, and III treatments, but where to be conservative, we impose that no treatment causes actual harm. In Figure 4, we demonstrate the extent of variations in both survival and expenditures that can be predicted using our cross-sectional regression model coefficients; the corresponding pattern when we use coefficients from the fixed-effects framework are more compressed with regard to both spending and survival. The pattern is once again a scatter-shot, with a faint positive correlation between predicted spending and predicted outcomes (the correlation coefficient is 0.07, $p < .01$).

The Spending Slowdown: 2007-11: As noted earlier, there was considerable variation in growth rates of Medicare expenditures across hospitals. For example, there were 380 hospitals with at least 200 AMI patients that experienced more than a 5 percentage point decline in average Medicare risk-adjusted per-enrollee expenditures; on average expenditures in these hospitals fell by 11 percent.[12] By contrast, among the 535 hospitals with at least 200 AMI patients where Medicare spending rose by 5 percent, average spending rose by 12 percent. In hospitals with a reduction in spending, survival improved by 1.5 percentage points; in hospitals where expenditures continued to grow rapidly, risk-adjusted survival rose by 1.7 percentage points – an insignificant difference. Using the fixed-effect equation coefficients from Column (4) in Table 5 indicated that roughly half of the $5200 per patient decline in spending for the low-growth hospitals could be explained by corresponding reductions in scanning rates, the number of different physicians, and outlier home health care payments. By contrast, spending associated with effective Category I/II treatments rose slightly more rapidly among the hospitals that reduced their overall expenditures.

VI. Measuring Productive and Allocative Efficiency at the Hospital Level

In this section, we calculate directly the degree of productive and allocative inefficiency Ω, as in Equation (4), for each hospital. We assume that the value for Ψ of $100,000 per quality-adjusted life year, and further assuming that the incremental AMI patient who survives a year after admission will live an additional 5.25 years, We normalize the degree of inefficiency to the mean value for the United States, so a positive number of (say) $10,000 implies that for each

[12] To reduce statistical noise, the change was specified as the log difference in average expenditures in 2007/08 compared to 2010/11.

AMI patient, the net social value (either lower costs or better health outcomes) is that amount about the national average.

Table 6 presents selected measures of the cumulative efficiency gain Ω, along with specific measures of inputs, for 5 hospitals, with a ranking from most efficient (Intermountain and the Mayo Clinic) to the least (Hialeah Hospital in Hialeah, Florida).[13] There is a nearly monotonic increase in Category III treatments, and a decrease in Category I/II treatments, as one moves from most to least efficient hospitals. As well, the predicted survival and expenditures – based only on treatment choices – do a reasonable job of matching the actual risk-adjusted survival and expenditures. (The correlation coefficients are 0.55 and 0.73, respectively, and are highly significant.) When the rates differ, for example when the Stanford medical center's risk-adjusted one-year of 74.3 percent is well above the predicted survival rate of 72.9 percent, there is likely the combination of unmeasured health status at the hospital level, and hospital-specific ability or expertise.

Using the estimated coefficients from the fixed-effects model suggests less dramatic variation across hospitals, but the correlation coefficient between these predicted measures of survival and spending, and actual survival and spending, is still quite high; 0.44 and 0.61, respectively. Another potential concern is that the quality measures could track the affluence of the community and the ability of patients to adhere to (e.g.) their beta blocker use. The Mayo Clinic is in rural Minnesota and Intermountain in Salt Lake City, while Hialeah Hospital is in a largely Spanish-speaking suburb of Miami. Yet our estimates adjust for both ZIP code income and Hispanic identification in the regression analysis. We note as well that the measure of Ω for

[13] As of February 15, 2015, the three leading Google reviews of the Hialeah Hospital are "Find another hospital if you can," "Worse hospital ever," and "Nurses don't inform family patients of status for loved one."

the University of Chicago Medical Center, \$26,963, is comparable to the Ω for Stanford Medical Center, \$16,192, and exhibits better predicted survival (and beta blocker use) to Stanford's despite its location in the South Side of Chicago. Still, we acknowledge the necessity of developing measures that are as unassociated with patient characteristics as possible.

VII. Conclusions

In this paper, we have attempted to measure empirically the value of Medicare expenditures during a period in which spending rates slowed substantially. First, we have shown that differences across hospitals with regard to the underuse of highly effective treatments, and the overuse of ineffective treatments, can explain a substantial fraction of the otherwise puzzling lack of correlation in spending and outcomes at the hospital level. Most of the explainable survival differences across hospitals are the consequence of differences in Category I and some Category II treatments across hospitals (especially primary PCI), while a large fraction of the spending differences are the consequence of treatment choices in the use of Category III and some Category II treatments across hospitals that yield little or no health benefits.

Second, despite the flattening of Medicare expenditures, one-year survival continued to rise from 2007 through 2011. Hospitals with a decline in expenditures during this period tended to cut back on the least valuable Category II and III treatments (and not the valuable Category I and II treatments), so it is less surprising that these hospitals experienced the same growth in one-year survival (1.7 percentage points) as did hospitals with an increase in expenditures. The results are therefore consistent with the idea that hospitals can scale back spending under shared savings programs such as Accountable Care Organizations (ACOs) without necessarily leading to worse outcomes for their patients.

We acknowledge that the parameters used in the model may not entirely capture causal factors. Restrictions on the use of (e.g.) double CT scans are unlikely to have an impact on AMI patient costs, unless there are changes in organizational structure that lead to both improvements in the quality of CT scanning, and better patient quality care. We plan to investigate further the use of the coefficients from the fixed-effects regression model, which take a more conservative stance on the causal influence of specific factors on outcomes. Still more conservative would be to use only coefficient estimates from randomized trials, which would therefore avoid potential biases in the estimate of marginal benefits.

What are the policy implications? Current payment reforms, such as accountable care organizations, encourage providers to shift their attention to improving use of effective care (Category I), discouraging use of low-value care (Category III) and making good choices with respect to heterogeneous care (Category II) though financial incentives based on cost and quality performance. Providers under these incentive schemes may choose to implement tools, such as shared decision making or appropriateness criteria prompts in an electronic medical record, in the clinic to support decision making in each of these circumstances. Yet the unanswered question is whether the shadow constraints leading to overuse or underuse can be changed in a fundamental way by the introduction of a new financial model. Certainly early results (e.g., Colla et al., 2012) are suggestive that basic Category I measures can be improved relatively easily, but that many institutions face challenges with the overuse of expensive treatments. It may also be more difficult to fundamentally change the way that health care is delivered if physicians hold strong beliefs about the use of specific treatments, even when there is little proven effectiveness of their value, as in Cutler et al. (2013).

Still, one may be more sanguine about the future of Medicare if it is possible to identify and measure systematically the degree of inefficiency in health care systems. As electronic health records become more sophisticated, allowing more accurate assessments of underuse and overuse, there is a real potential for stepped up productivity in the Medicare program that would ensure that the marginal dollar does in fact yield real benefit to patients.

Figure 1: Sources of Inefficiency

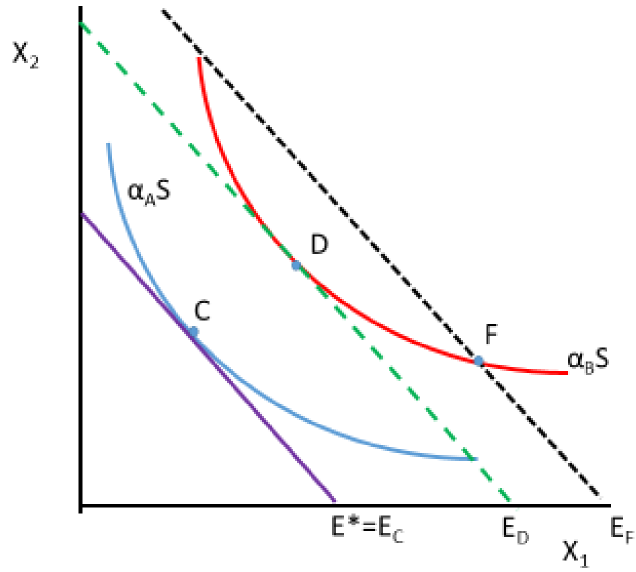

Figure 2: Hypothetical Hospital-Specific Production Functions

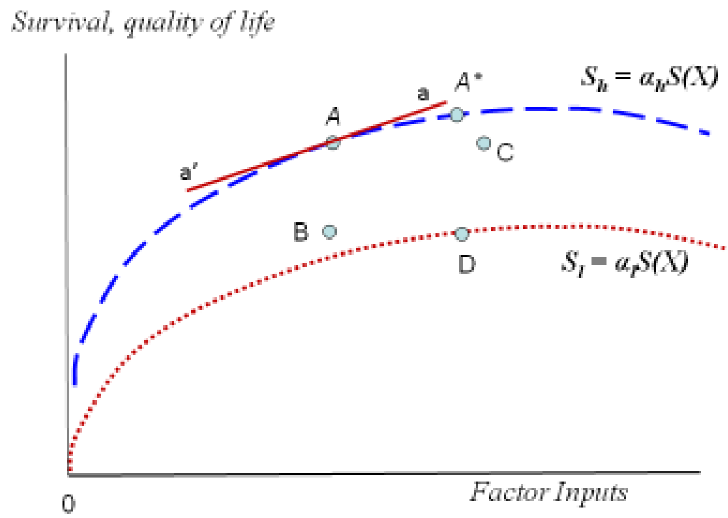

Figure 3a: Association between (HCC) Risk-Adjusted Spending and Mortality: Hospitals with at Least 400 AMI Patients, 2007-11

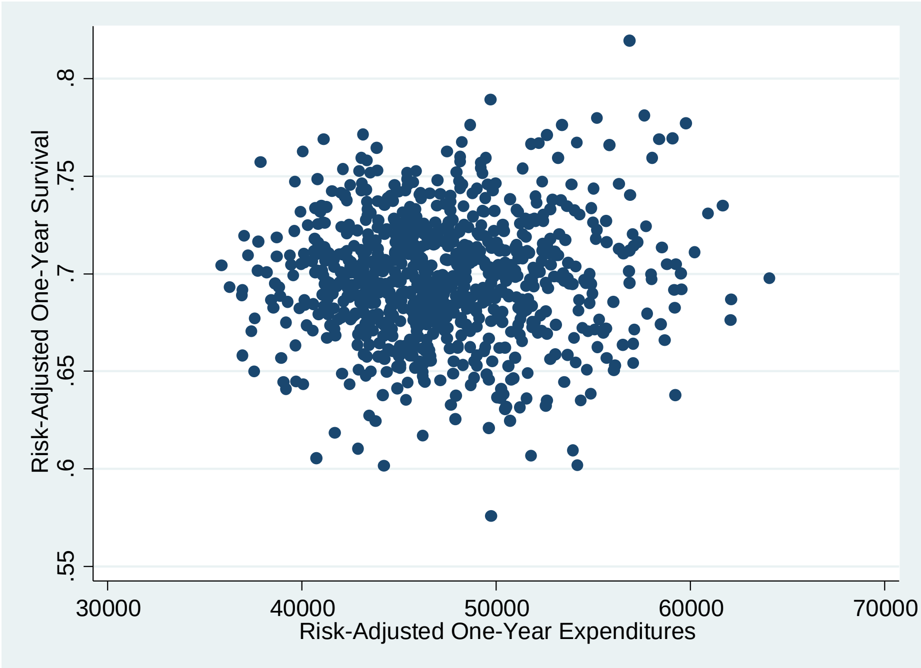

Figure 3b: Association between Risk-Adjusted Spending and Mortality for Tourists: Hospitals with at least 100 AMI (Tourist) Patients, 2007-11

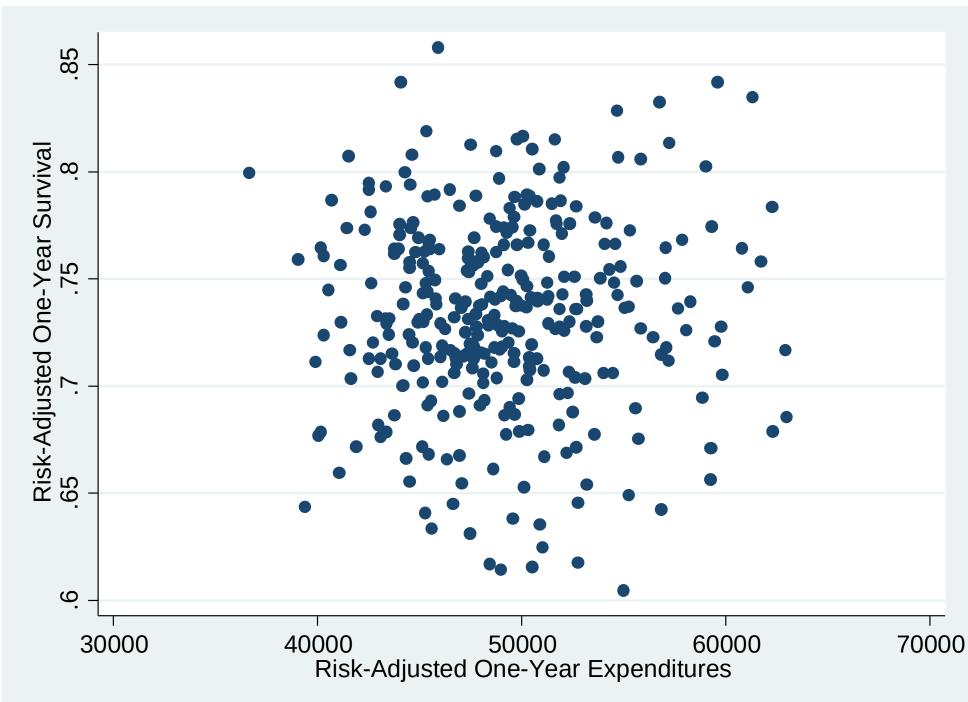

Figure 4: One-year Survival and One-year Expenditures for AMI Patients of Hospitals, Based Solely on Treatment Choices (N > 400 AMI patient per hospital)

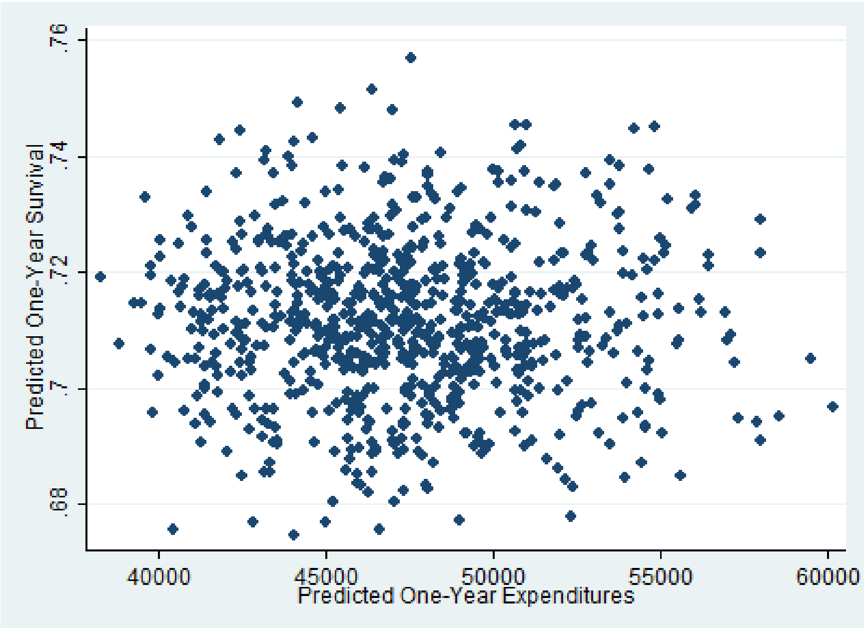

Table 1: Descriptive Characteristics of Medicare Beneficiaries with AMI 2007-2011

	All AMI	Tourists
N Beneficiaries	897,088	129,289
N Hospitals	2,257	2,257
Demographics		
Age	78.5	77.5
Female	.492	.461
Black	.073	.063
Native American	.004	.007
Hispanic	.016	.012
Asian	.012	.010
Other Race/Ethnicity	.009	.009
Income (Bene ZIP code)	$54,735	
Clinical characteristics		
Vascular	.084	.082
Pulmonary	.175	.162
Dementia	.031	.030
Diabetes	.265	.260
Liver Failure	.004	.003
Renal Failure	.181	.164
Cancer	.052	.050
Metastatic Cancer	.012	.011
AMI Location		
Anterolateral Wall	.020	.022
All Other Anterior Wall	.075	.080
Inferolateral Wall	.017	.018
Inferoposterior Wall	.011	.012
All Other Inferior Wall	.093	.102
True Posterior Wall	.003	.003
Subendocardial	.707	.697
Other Site	.009	.008
Not Otherwise Specified	.055	.044

Table 2: Outcome and Input Measures by Year (All)

Outcome Measures	2007	2008	2009	2010	2011
Survival within 1 year	.677	.686	.693	.700	.705
Survival within 30 Days	.848	.851	.856	.860	.861
Total Adjusted Spending 1 year	$46,234	$47,027	$48,296	$47,959	$47,088
Category I Measures (Hospital-Level)					
Received Beta Blockers within 6 Months	.763	.768	.767	.767	.770
Received Statins within 6 Months	.641	.659	.675	.688	.662
Follow-up within 14 days of discharge	.620	.621	.626	.630	.636
Category II Measures (Hospital-Level)					
PCI within 24 hours	.224	.231	.247	.258	.266
PCI post 24 hours	.088	.087	.086	.085	.084
Mean Number of Unique Doctors	11.8	11.8	12.1	12.1	12.3
Mean Number of CT + MRI Scans	1.8	1.8	1.8	1.6	1.3
Category III Measures (Hospital-Level)					
Home Health Care Spending (Avg.)	$1,694	$1,867	$1,922	$1,889	$1,730
Feeding Tubes Inserted - Dementia Patients	0.030	0.029	0.028	0.028	0.029

Tourist AMI Cohort

Outcome Measures	2007	2008	2009	2010	2011
Survival within 1 year	.717	.724	.727	.737	.741
Survival within 30 Days	.870	.873	.872	.878	.878
Total Adjusted Spending 1 Year	$47,905	$48,584	$49,423	$49,554	$48,188

Table 3: Regression Coefficient of the Association Between Risk-Adjusted Survival and the Log of Price-Adjusted Expenditures: Medicare AMI Cohort, 2007-11

	Full Sample (N =890,080)	t-statistic*	R^2	Tourists (N =128,237)	t-statistic*	R^2
1. 30-Day survival: Baseline Risk Adjustment	0.025	6.04	0.073	0.010	1.02	0.068
2. 30-Day survival: Baseline Risk Adjustment + HCCs	0.032	7.67	0.083	0.015	1.62	0.078
3. 30-Day survival: Baseline Risk Adjustment + HCCs + hospital fixed-effects	0.029	5.84	0.087	0.018	1.40	0.085
4. One-Year survival: Baseline Risk Adjustment	-0.021	4.11	0.145	-0.040	3.48	0.141
5. One-year survival: Baseline Risk Adjustment + HCCs	0.018	3.53	0.179	-.0004	0.04	0.175
6. One-year survival: Baseline Risk Adjustment + HCCs + hospital fixed-effects	0.028	5.56	0.184	0.017	1.30	0.183

Risk adjustment measures included but not reported: Age/sex cells, Black, Native American, Hispanic, Asian, other Race/Ethnicity, ZIP code income, vascular disease, pulmonary disease, dementia, diabetes, liver failure, renal failure, cancer, anterolateral, inferolateral, inferoposterior, all other inferior, true posterior walls, or subendocardial, other site, or not otherwise specified.

Table 4: Risk Adjusted Association of Category I, II, III Hospital Treatments on AMI Survival

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Survival 1 Year	Survival 1 Year	Survival 1 Year	Survival 1 Year	Survival 1 Year	Survival 30 Days
Cohort	All	All	Tourists	All	All	All
Hospital Fixed Effects	No	No	No	Yes	Yes	Yes
Hosps ≥500 AMIs	No	No	No	No	Yes	No
Model	OLS	Probit	OLS	OLS	OLS	OLS
	Coeff. (t-stat)	Marg Eff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)	Coeff. (t-stat)
Category I Treatments						
Beta Blockers Within 6M	0.015	0.014	0.008	0.014	0.018	0.002
	(3.46)	(3.46)	(0.75)	(3.24)	(2.29)	(0.65)
Statins Within 6M	0.020	0.018	0.014	0.006	0.008	-0.010
	(5.09)	(4.95)	(1.43)	(1.57)	(1.19)	(2.96)
Follow-up within 14 days of discharge	0.065	0.067	0.055	-0.011	0.033	0.012
	(5.39)	(5.70)	(2.55)	(0.41)	(0.87)	(0.51)
Teaching Hospital	0.011	0.010	0.009			
	(4.85)	(4.46)	(2.09)			
Category II Treatments						
Early PCI – Quartile 4	0.055	0.054	0.063	0.049	0.035	0.031
	(25.63)	(25.96)	(13.74)	(16.42)	(7.84)	(12.98)
Late PCI – Quartile 4	0.029	0.028	0.030	0.026	0.022	0.012
	(16.24)	(15.87)	(7.44)	(15.21)	(8.92)	(8.26)
# Unique Doctors – Quart. 4	0.001	0.001	-0.000	-0.011	-0.006	0.004
	(0.63)	(0.43)	(0.06)	(3.50)	(1.36)	(1.36)
# MRI/CT Scans – Quart. 4	0.008	0.007	0.009	0.008	0.004	0.001
	(3.67)	(3.19)	(1.89)	(3.59)	(1.35)	(0.47)
Category III Treatments						
Fraction Home Health Outlier	0.003	0.001	-0.013	0.063	0.024	0.027
	(0.39)	(0.11)	(0.77)	(6.75)	(1.44)	(3.49)
Feeding Tubes Inserted - Dementia Patients	-0.110	-0.117	-0.218	-0.013	0.040	0.001
	(2.40)	(2.58)	(2.99)	(0.29)	(0.58)	(0.02)
Double CT	-0.032	-0.032	-0.020			
	(3.62)	(3.70)	(0.97)			
N	847,609	847,609	116,099	884,472	503,524	884,472

All models include risk adjustment variables (see Table 3).

Table 5: Risk Adjusted Association of Category I, II, III Hospital Treatments and Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	Expenditures	Ln Expenditures	Expenditures	Expenditures	Expenditures	Expenditures
	1 Year	1 Year	1 Year	1 Year	1 Year	30 Days
Cohort	All	All	Tourists	All	All	All
Hospital Fixed Effects	No	No	No	Yes	Yes	Yes
Hosps ≥500 AMIs	No	No	No	No	Yes	No
	Coeff. (T-stat)	Coeff. (T-stat)	Coeff. (T-stat)	Coeff. (T-stat)	Coeff. (T-stat)	Coeff. (T-stat)
Category I Treatments						
Beta Blockers Within 6M	-2076	-0.029	-2570	-386	-328	-124
	(4.55)	(3.39)	(2.22)	(0.99)	(0.39)	(0.74)
Statins Within 6M	-1677	-0.031	-1061	-984	-1053	-275
	(4.00)	(3.86)	(1.08)	(2.74)	(1.36)	(1.79)
Follow-up within 14 days of discharge	-5201	-0.065	-7422	-408	1136	-146
	(3.80)	(2.34)	(3.45)	(0.14)	(0.26)	(0.12)
Teaching Hospital	-655	-0.014	-408			
	(2.28)	(2.71)	(0.99)			
Category II Treatments						
Early PCI – Q4	1816	0.058	2611	365	-155	-122
	(6.84)	(11.15)	(5.46)	(1.18)	(0.31)	(0.92)
Late PCI – Q4	1526	0.046	1654	671	160	125
	(7.54)	(12.00)	(4.30)	(3.98)	(0.62)	(1.73)
# Unique Doctors - Q4	6015	0.115	5681	7415	5569	4944
	(20.05)	(20.83)	(11.20)	(23.17)	(11.58)	(34.10)
# MRI/CT Scans – Q4	5866	0.107	5311	4557	3973	452
	(24.41)	(23.41)	(10.98)	(22.62)	(12.25)	(5.28)
Category III Treatments						
Home Health Outliers	13090	0.231	10135	8222	9287	-220
	(13.21)	(13.18)	(5.57)	(9.08)	(4.99)	(0.60)
Feeding Tubes Inserted - Dementia Patients	49443	0.733	34700	-1174	-3429	-23
	(11.14)	(9.34)	(4.40)	(0.27)	(0.47)	(0.01)
Double CT	3806	0.064	4455			
	(3.33)	(2.78)	(2.23)			

Table 6: Spending and Outcome Measures for Five Selected Hospitals

	Mayo Clinic	Inter-mountain	Stanford Med. Ctr.	U. of Chicago	University of Miami	Hialeah Hospital
Productivity Measure (Ω)	$33,114	$41,710	$19,967	$26,963	-$13,911	-$44,897
Predicted Survival	.757	.771	.729	.753	.696	.672
Actual (Risk-Adjusted) Survival	.745	.724	.743	.683	.689	.630
Predicted Spending	$48,157	$47,191	$45,713	$52,513	$63,060	$63,220
Actual (Risk-Adjusted) Spending	$45,225	$42,980	$44,278	$48,561	$73,044	$59,243
Number of Admissions (total)	1335	664	375	275	323	337
Early Stents (within 1 day)	.454	.553	.357	.400	.251	.015
Physician Visit within 14 Days of Discharge	.708	.660	.678	.615	.537	.516
Beta Blocker within 6 months	.872	.803	.752	.903	.776	.706
Double CT scans of the chest	.000	.009	.002	.000	.059	.104
Fraction Home Health Care Paid as Outlier	.031	.123	.082	.173	.413	.438
Sources: Medicare claims data (A&B), Medicare Part D data, Hospital Compare						

REFERENCES

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

Barnato, A. E., C. C. Chang, M. H. Farrell, J. R. Lave, M. S. Roberts, et al., 2010. Is survival better at hospitals with higher "end-of-life" treatment intensity? *Medical Care* 48, 2, Feb, 125-32.

Becker, Gary, and Kevin Murphy, 1992. The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics* 107, 4, 1137-60.

Bekelis K., Fisher E.S., Labropoulos N., Zhou W., Skinner J.S. “Variations in the Intensive use of Head Computed Tomography for Elderly Hemorrhagic Stroke Patients in Leading U.S. hospitals,” *Radiology*, 2014.

Bloom, Nicholas, Carol Propper, Stephan Seiler and John Van Reenen. 2010. “The Impact of Competition on Management Quality: Evidence from Public Hospitals.” National Bureau of Economic Research Working Paper No. 16032.

Bloom, Nicholas, Raffaella Sadun, Renata Lemos, Daniela Scur and John Van Reenen. ,2014. *The new empirical economics of management.* CEP Occasional Papers, CEPOP41. The London School of Economics and Political Science, Center of Economic Performance, London, UK.

Bogdanich, Walt, and Jo Craven McGinty, 2011, Medicare claims show overuse for CT scanning, *The New York Times*, June 17.

Bradley, Elizabeth H., Eric S. Holmboe, Jennifer A. Mattera. et al., “A Qualitative Study of Increasing β-Blocker Use After Myocardial Infarction: Why Do Some Hospitals Succeed?,” *JAMA* 285(20), 2001: 2604-2611.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny and Chad Syverson. 2013. “Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector.” National Bureau of Economic Research Working Paper No. 19200.

Chandra, Amitabh, and Douglas Staiger. 2007. “Productivity Spillovers in Healthcare: Evidence From the Treatment of Heart Attacks.” *Journal of Political Economy* 115(1):103-140.

Chandra, A., and J. Skinner, “Technology Growth and Expenditure Growth in Health Care,” *Journal of Economic Literature* 50(3) (September 2012): 645-80.

Clemens, J.P. and J.D. Gottlieb. “Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?” *American Economic Review*, 2014.

Colla C.H., Wennberg D.E., Meara E., Skinner J.S., Gottlieb D., Lewis V.A., Snyder C.M., Fisher E.S. “Spending Differences Associated With the Medicare Physician Group Practice Demonstration.” *JAMA*, 2012;308(10):1015-1023.

Colla CH, Sequist TD, Rosenthal MB, Schpero WL, Gottlieb DJ, Morden NE. Use of Non-indicated cardiac testing in low-risk patients: Choosing Wisely. *BMJ Quality & Safety* 2015;24:149-153.

Cutler, David, Jonathan Skinner, Areil Dora Stern and David Wennberg. 2013. “Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending.” National Bureau of Economic Research Working Paper No. 19320.

Díaz-Hernández, Juan José, Eduardo Martinez-Budría, and Serio Jara-Díaz, 2008. “The Effects of Ignoring Inefficiency in the Analysis of Production: The Case of Cargo Handling in Spanish Ports,” *Transportation Research* Part A, 42: 321-29.

de Vries, E. N., H. A. Prins, R. M. Crolla, A. J. den Outer, G. van Andel, et al., 2010. Effect of a comprehensive surgical safety system on patient outcomes. *N England J Med* 363, 20, Nov 11, 1928-37.

Doyle, Joseph, Steven Ewer and Todd Wagner. 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics* 29(6):866-882.

Doyle, Joseph. 2011. "Returns to Local-Area Healthcare Spending: Using Health Shocks to Patients Far From Home." *American Economic Journal: Applied Economics* 3(3):221-243.

Doyle, Joseph, John Graves, Jonathan Gruber and Samuel Kleiner. 2015a. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy*, 123(1): 170–214.

Doyle, Joseph, John Graves, and Jonathan Gruber, 2015b. "Uncovering Waste in U.S. Healthcare," NBER Working Paper No. 21050.

Finkelstein, Amy, Matthew Gentzkow and Heidi Williams. 2014. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." National Bureau of Economic Research Working Paper No. 20789.

Finucane, T. E., C. Christmas, and K. Travis, 1999. Tube feeding in patients with advanced dementia: A review of the evidence. *JAMA* 282, 14, Oct 13, 1365-70.

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, 4, 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, 4, Feb 18, 288-98.

Fried, Harold O., C. A. Knox Lovell and Shelton S. Schmidt, 2008 "Efficiency and Productivity," in H. O Fried, C.A. Knox Lovell, and S. C Schmidt (eds.) *The Measurement of Productive Efficiency and Productivity Growth*, Oxford University Press.

Gottlieb, D. J., W. Zhou, Y. Song, K. G. Andrews, J. S. Skinner, et al., 2010. Prices don't drive regional medicare spending variations. *Health Affairs* (Millwood) 29, 3, Mar-Apr, 537-43.

Hernandez AF, Greiner MA, Fonarow GC, Hammill BG, Heidenreich PA, Yancy CW, Peterson ED, Curtis LH. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA* 2010;303(17):1716-22.

IOM (Institute of Medicine). 2013. *Variation in health care spending: target decision making, not geography*. Washington, DC: The National Academies Press.

Jackson, Kirabo, Jonah Rockoff and Douglas Staiger. 2014. "Teacher Effects and Teacher-Related Policies," *Annual Review of Economics* Vol. 6: 801-825.

Kumbhakar, Subal C., Gudbrand Lien, and J. Brian Hardaker, 2014. "Technical efficiency in competing panel data models: A study of Norwegian grain farming," *Journal of Productivity Analysis* 41(2): 321-337

Kuosmanen, Timo, and Mika Kortelainen, 2012. "Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints," *Journal of Productivity Analysis* 38:11–28.

Landrum, M. B., E. R. Meara, A. Chandra, E. Guadagnoli, and N. L. Keating, 2008. Is spending more always wasteful? The appropriateness of care and outcomes among colorectal cancer patients. *Health Affairs* (Millwood) 27, 1, Jan-Feb, 159-68.

McConnell, K. John, Richard C. Lindrooth, Douglas R. Wholey, Thomas M. Maddox, and Nick Bloom, "Management Practices and the Quality of Care in Cardiac Units," *JAMA Internal Medicine*, 173(8), April 23, 2013: 684-692

McKnight, Robin, 2006. Home health care reimbursement, long-term care utilization, and health outcomes. *Journal of Public Economics* 90, 1-2, January, 293-323.

Munson J.C., Morden N.E., Goodman, D.C., Valle, L.F., Wennberg, J.E. T*he Dartmouth Atlas of Medicare Prescription Drug Use*. October 15, 2013. http://www.dartmouthatlas.org/downloads/reports/Prescription_Drug_Atlas_101513.pdf.

Office of Inspector General (OIG), 2012. Inappropriate and Questionable Billing by Medicare Home Health Agencies, OEI-04-11-00240. Washington DC: Department of Health and Human Services.

Restuccia, Diego, and Richard Rogerson, 2008. "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics*," 11(4) (October): 707-720.

Romley, J. A., A. B. Jena, and D. P. Goldman, 2011. Hospital spending and inpatient mortality: Evidence from California: An observational study. *Ann Intern Med* 154, 3, Feb 1, 160-7.

Rothberg, M. B., J. Cohen, P. Lindenauer, J. Maselli, and A. Auerbach, 2010a. Little evidence of correlation between growth in health care spending and reduced mortality. *Health Affairs* (Millwood) 29, 8, Aug, 1523-31.

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, 6 Pt 2, Dec, 1872-92.

Sharma G, Kuo YF, Freeman JL, Zhang DD, Goodwin JS. Outpatient follow-up visit and 30-day emergency department visit and readmission in patients hospitalized for chronic obstructive pulmonary disease. *Arch Intern Med* 2010;170(18):1664-70.

Skinner, Jonathan, 2012. Causes and consequences of geographic variations in health care," in in T. McGuire, M. Pauly, and P. Pita Baros (eds.) *Handbook of Health Economics* Vol. 2, North Holland; 1st edition.

Skinner, Jonathan, 1994. "What do Stochastic Frontier Cost Functions Tell Us About Inefficiency?" *Journal of Health Economics* 13(3): 323-328.

Skinner, J, ES Fisher, and JE Wennberg, 2005, The efficiency of medicare, in DA Wise, ed.: *Analyses in the economics of aging* (University of Chicago Press, Chicago).

Skinner, J, and D Staiger, 2015. Technology diffusion and productivity growth in health care, *Review of Economics and Statistics*.

Skinner, J. S., D. O. Staiger, and E. S. Fisher, 2006. Is technological change in medicine always worth it? The case of acute myocardial infarction. *Health Affairs* (Millwood) 25, 2, Mar-Apr, w34-47.

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, et al., 2010. Regional variations in diagnostic practices. *N Engl J Med* 363, 1, Jul 1, 45-53.

Syverson, Chad, 2011. What determines productivity? *Journal of Economic Literature* 44, 2, June, 326-65.

Teno, Joan M., Susan L. Mitchell, Pedro L. Gozalo, David Dosa, Amy Hsu, et al., 2010. Hospital characteristics associated with feeding tube placement in nursing home residents with advanced cognitive impairment. *JAMA* 303, 6, February 10, 2010, 544-550.

Weintraub, W. S., J. A. Spertus, P. Kolm, D. J. Maron, Z. Zhang, et al., 2008. Effect of PCI on quality of life in patients with stable coronary disease. *N Engl J Med* 359, 7, Aug 14, 677-87.

Wennberg, JE, ES Fisher, and JS Skinner, 2002. Geography and the debate over medicare reform. *Health Affairs* Web (www.healthaffairs.org), Feb 13, W96-W114.

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

Yasaitis, L., E. S. Fisher, J. S. Skinner, and A. Chandra, 2009. Hospital quality and intensity of spending: Is there an association? *Health Affairs* 28, 4, Jul-Aug, w566-72.