Hospital Productivity and the Misallocation of Healthcare Inputs

Amitabh Chandra
Kennedy School of Government, Harvard University,
Harvard Business School,
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

February 23, 2020

Very Preliminary – Please do not quote or distribute

Abstract

There is growing evidence for wide variation in productivity across hospitals, with large differences in risk-adjusted health outcomes as well as expenditures. In this paper, we consider the contribution of *misallocation* in input choices – the underuse of effective inputs and overuse of ineffective ones -- to explain why some hospitals get better outcomes at lower cost. We use a sample of 1.7 million patients in the Medicare fee-for-service population with acute myocardial infarction (AMI), or heart attacks, during 2007-17. The problem of confounding health factors is addressed in several ways, including the use of tourists, whose assignment to hospitals resembles random assignment (Doyle, 2011), and ZIP-code fixed effects. Briefly, we find that misallocation accounts for as much as 22 percent of risk-adjusted survival rates across hospitals. Greater use of highly effective inputs, such as beta blocker, statin, and ACE/ARB drug treatments, primary care support, and stenting are predictive of highly-productive hospitals, while an excess of unnecessary scans and potentially fraudulent home health care are generally predictive of low-productivity hospitals.

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, and seminar participants at the University of Chicago for very helpful suggestions. Weiping Zhou and Ben Usadi have provided extraordinary programming and data assistance.

I. Introduction

There is increasing evidence of widespread differences across regions and hospitals in both intensity of care as well as health care quality (e.g., Finkelstein et al., 2016, 2019; Fisher et al., 2003a,b, Wennberg, 2010; Baicker and Chandra, 2004a; Doyle et al., 2015, 2017; Yasaitis et al., 2014; Deryugina and Molitor, 2018; Romley et al., 2011). Much of the debate has surrounded whether there is a positive, negative, or zero association between overall spending and health outcomes, with the implicit interpretation that the coefficient is stimating the slope of the production function for health care, thus addressing the question of whether U.S. health care is allocatively efficient (Garber and Skinner, 2008). Two recent studies noted that these wide variations in spending and outcomes might be better understood as hospital-based differences in total factor productivity (TFP), finding evidence of strong differences in productivity across hospitals, as well as rising market share for the most productive hospitals (Chandra et al., 2016a,b).

In this paper, we adopt this productivity framework, and consider both allocative and productive efficiency in health care production by allowing for the *misallocation* of inputs (or treatments), whether through underuse of effective inputs or the overuse of ineffective inputs, or both (Restuccia and Rogerson, 2017; David et al., 2016). In a classic study, Hsieh and Klenow (2009) found that input misallocation (relative to a US benchmark) explained between 30-60% of TFP differences in Indian and Chinese manufacturing; Restuccia and Rogerson (2008) similarly found between 30-50% of productivity differences accounted for by (implicit) price distortions leading to suboptimal input choices. In health care, Skinner and Staiger (2015) showed that differences in input choices for highly effective beta blockers across hospitals could similarly explain long-term productivity gaps in outcomes, and the “Choosing Wisely” campaigns in health policy focus on identifying low-value health services (Colla et al., 2015). While the detection of misallocation has proven controversial in the productivity literature (e.g., Haltiwanger et al., 2018), our measurement of physical inputs and well-measured outcomes (e.g., survival) allows for a novel test of misallocation in the health care sector.

To formalize the idea of misallocation in health care, we draw on Chandra et al. (2016a) but expand the model to allow for misallocation because of informational asymmetries, physician beliefs, or other types of distortions regarding highly effective or ineffective care. We develop an estimation equation that implies that, under the null hypothesis of optimized inputs, constant Medicare prices, and a common production function, the choice of factor inputs are orthogonal to TFP *conditional on total expenditures*. Thus including specific input choices, as well as expenditures, on the right-hand side of the equation provides information on the degree of input misallocation across hospitals.[1] One clear implication of this model is that in the presence of misallocation, conventional regression approaches that seek to estimate "the" association between spending and outcomes (e.g., Doyle et al., 2015, 2017; Doyle, 2011; Wennberg et al., 2002; Fisher et al., 2003a,b) are misspecified, with the bias depending on the correlation across different categories of inputs.

The model is tested by considering the entire population of 1.7 million elderly (age 65+) fee-for-service Medicare enrollees with acute myocardial infarction (AMI), or heart attacks during 2007-2017, along with a subset of 123,984 "tourists," or people admitted to hospital far from home, as in Doyle (2011). We choose inputs based on clinical evidence regarding their effectiveness in improving health outcomes in a cost-effective way. As in Chandra and Skinner (2012) (see also Wennberg, 2010), we consider cost-effective "Category I" inputs such as statins, primary-care within 14 days of discharge, or same-day stents. "Category II" inputs are those with heterogeneous treatment effects – for example, the first physician visit or scan is probably more valuable than the 47th, while "Category III" treatments that

[1] Another potential explanation for a finding that input choice affects outcomes is that hospitals differ with regard to their production functions, as in Chandra and Staiger (2007); we consider this case in more detail below.

are both costly and unlikely, based on clinical evidence, to offer health benefits, such as excessive and potentially fraudulent post-acute or home health care (Doyle et al., 2017) and "Choosing Wisely" treatments deemed by professional physician groups to be wasteful (Colla, et al., 2015). We hypothesize that higher use of the often underused "Category I" treatments, and lesser use of the "Category III" treatments, will be associated with higher measured productivity, with ambiguous predictions for the "Category II" treatments.

We also address several econometric challenges related to unmeasured patient health and endogenous patient selection. First, there is the well-known problem of inadequate risk adjustment; hospitals may appear to be highly productive when in fact they are treating healthier patients. We include a wide set of risk-adjusters that include both individual measures such as the type and location of the AMI (e.g., a subendocardial MI), whether the patient is eligible for low-income drug subsidies and/or Medicaid (as in Lewis et al., 2019) , ZIP-level socioeconomic status, and the share of the hospital referral region enrolled in Medicare managed care.

But a more subtle bias may result; perhaps patients with unmeasured health status seek care at one hospital but not another. While AMI is a disease where every minute counts for treatment, leading to patients generally being taken to the nearest hospital (at least conditional on ambulance service; see Doyle et al., 2015, 2017), there may be biases with regard to geographic regions if (for example) people living in some regions are more likely to smoke than others. We address this potential problem in two ways; we first include ZIP-code fixed effects (e.g., Garthwaite et al., 2019) which remove any place-based unmeasured factors that may be expected to affect health outcomes; thus the comparison is for patients from the same ZIP code admitted to different hospitals (as might happen, for example, in the case of ambulance services as in Doyle et al., 2015; 2017). [2] Second, we consider the subset of patients

[2] Another approach is to consider movers, as in Song et al. (2010) and Finkelstein et al. (2016); while this sample is likely to be small, we plan ton consider the subset of movers in the future.

whose heart attack occurred far from home, and are therefore considered to be "tourists" as in Doyle (2011).

Briefly, we find, like previous studies, a wide range across hospitals in both one-year risk-adjusted survival rates (with a patient-weighted range from 0.68 in the 10th percentile of hospitals to 0.75 in the 90th percentile of hospitals), and in price-adjusted spending (between $42,133 in the 10th percentile to $52,676 in the 90th percentile). Similar results are found for Medicare "tourists" who were in different hospital referral regions (HRRs) when admitted for their AMI.

When we run conventional regressions as in Doyle et al. (2015, 2017) and Doyle (2011), our coefficients on spending are roughly consistent, albeit sensitive to regression specification. But a key assumption necessary to interpret these estimates as "the return to health care spending" or the slope of the production function, is that input choices are made optimally, an assumption that is strongly rejected by the data both in the full sample and for the tourists; hospital-specific inputs have a strong impact on productivity measures conditional on total expenditures. The signs of the coefficients are also consistent with the model; Category I or highly effective input rates are positively correlated with productivity, Category III are negatively correlated, while Category II treatments are in-between, with some categories positively associated (late stenting) and others negatively associated (e.g., being in the top quartile of hospitals with respect to MRI/CT scans, suggestive of fragmented care). For tourists, the effects are similar although the results are not always significant. A further implication of the model is that 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.[3] That Doyle et al. (2015a) should find positive effects of spending on outcomes for acute-care hospital treatments within the first 30 days,

[3] 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).

but in Doyle et al. (2017) find no impact of overall spending on outcomes for a one-year horizon, where post-acute care is a major component of subsequent spending, is consistent with these results.

Quantitatively, we find that hospital-level misallocation can explain as much as 22 percent of total variation across hospitals in risk-adjusted one-year survival rates. In a hypothetical counterfactual, moving all hospitals to at least the 90th percentile with regard to efficient use of inputs would lead to a reduction of 7.6 percent in mortality. Policies to reduce misallocation are also quite different from conventional approaches that involve trying to measure "value" or total expenditures, since such policies would lead to a greater emphasis on monitoring specific inputs, treatments, and other markers for productivity to improve health outcomes across hospitals.

II. The Model

Following Chandra et al. (2016), we define a hospital-level production function for patient i in hospital h as

$$\begin{aligned} &Y_{ih} = \tilde{Y}_{ih} e^{\varepsilon_i} \\ &where\ \tilde{Y}_{ih} = A_h \left[\prod_w Z_{iw}^{\omega_w} \right] \left[\prod_k X_{ik}^{\beta_{hk}} \right] \quad (1) \end{aligned}$$

and the health outcome Y_{ih} is defined (for technical reasons) as the exponent of survival for (e.g.) 30 days or one year. Survival in turn is a function of "produced" survival $\tilde{Y}_{ih}$, which depends on hospital-specific productivity A_h, health and socioeconomic factors Z_{iw}, w = 1,…,W for patient i, and healthcare inputs X_{ik}, k = 1,…,K; the actual outcome is equal to $\tilde{Y}_{ih}$ times the (exponential) error term ε_i.[4]

[4] At the individual level, the actual outcome may be binary, in which case the condition is that Y = 1 if $\tilde{Y} \geq 0$ and is zero otherwise. Generally, we consider outcomes at the hospital level in which case the outcome will be an average.

Because we are studying the Medicare program, we assume that prices are constant (except for cost-of-living differences and other vagaries of the Medicare program) across hospitals. For the moment, we also assume that the production parameters β_k are constant across hospitals.

In this model, the hospital system (including physicians, nurses, ancillary health employees, and administrators) is the relevant decision-making entity; thus productivity A_h will reflect this complex mix of physician skills, diffusion of best practices and guidelines, organizational structure and coordination, and other institution-specific factors. Hospitals will obviously differ with regard to such decision-making, so we allow below systematic errors in decision-making of the institution.

Assume that the relevant objective function is:

$$\Omega_h = \varphi Y_h - \sum_k \phi p_k X_{hk} \qquad (2)$$

where φ is the monetary value to society of raising the outcome variable by one unit (which we assume is constant across hospitals), Ω_h is the hospital-specific benefit for hospital h, and because we are focusing here on hospital-level averages, we drop the i subscripts.[5]

A more difficult question is to determine what is the cost function for hospitals? First assume that p_k is the social price of the kth input (assumed constant across hospitals), and that (for the moment) $\phi = 1$. Then the hospital is solving for the social optimum, leading to allocative efficiency in the sense that the point chosen on the production function exhibits a slope equal to the social value of health.

For example, Figure 1 shows a production function for health with an index of health inputs on the horizontal axis and survival (or quality-adjusted life years) on the vertical axis. The optimum

---

[5] Note that this measure of net social benefit, which captures both allocative and productive efficiency, is more general than the focus purely on productivity A_h, or output given a vector of inputs.

between A and A' is shown where the slope of the tangent line is equal to $1/\varphi$; other points are allocatively inefficient, representing either too little investment in health (to the left of A') or too much (e.g., H) or even where more spending is harmful to health (J).

In the Medicare data during this period, it is not an unreasonable assumption that Medicare prices are roughly equal to marginal cost (or average variable cost). But it may be a less tenable assumption that $\phi = 1$ given that Medicare spending by the government represents revenue for the hospital.[6] For example, when $\phi = 0$ the hospital ignores cost (or Medicare revenue – or equivalently, Medicare reimbursements are equal to marginal cost so hospitals are in financial equipoise) and simply maximizes health, leading to Point H in Figure 1; $\phi < 0$ corresponds to classic supplier-induced demand where hospitals spend more to the detrement of health, leading to Point J.[7] Intuitively, conventional estimates of the marginal returns to expenditures involves estimating a regression with (exogenous) measures of X on the right-side of the equation, and survival or other outcome on the left-side, leading to a slope like that shown in Figure 1; depending on the coefficient (and the corresponding cost-effectiveness), the results are viewed as shedding light as to whether the U.S. is spending too much or too little on health care. But in all cases, variation in ϕ moves outcomes along a common production possibility frontier, which corresponds to measuring allocative inefficiency.

We now allow either for differences across hospitals in TFP, or for non-optimizing behavior that could move hospitals off of the production function in Figure 1; either generalization will lead to invalid estimates of "the" slope of the production function. Under optimizing behavior, the first-order

[6] This approach follows Skinner and Staiger (2015).

[7] The Cobb-Douglas specification of the production function we use in (1) does not allow for spending more to lead to worse outcomes.

conditions equate the marginal incremental value of each input X_k, $\frac{\varphi}{\phi}\partial Y_h / \partial X_{hk}$, with its (adjusted) price. But suppose that there were regulations, or informational barriers, or internal resistance to using effective treatments with clear benefits; then:

$$\frac{\varphi}{\phi}\partial Y_h / \partial X_{hk} - p_k = \mu_{hk} > 0 \qquad (3)$$

In other words, these non-price barriers lead to "underuse" of otherwise effective treatments.[8] 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 include poor organizational or management structure (e.g., Bloom et al., 2014; McConnell et al., 2013) the lack of leaders championing their use (Bradley et al., 2005), or systematic differences in training environments (Chan, 2020). We cannot quantify the specific causes of underuse but can capture the implicit costs (defined broadly) that would have generated the behavior we observe (Westfall et al., 2007).

Alternatively, suppose that there were strong financial incentives to "overuse" specific treatments because the private gain to the hospital or the providers was so large (e.g., supplier-induced demand) or because of incorrect physician beliefs about the marginal productivity of the input (Cutler et al., 2019). In that case, $\mu_{hk} < 0$, and there is systematic overuse of the low-productivity treatment.[9] For analytical purposes, it is easier to define the distortion as proportional to the price, so that

$$p_k(1+\lambda_{hk}) \equiv p_k + \mu_{hk} \qquad (4)$$

[8] Also see Díaz-Hernández et al. (2008) for a similar "shadow price" approach.

[9] Note that this is input-specific, and does not represent general overuse, as would be the case if $\phi < 0$ for example.

Thus the proportional input distortions λ_{hk} are defined implicitly based on actions or decisions by the hospital, so they will differ across hospitals (and presumably over time, although in this model we assume temporal stability).

We first derive our estimating equation in the special (nested) case where $\lambda_{hk} = 0$, so there is no misallocation of factor inputs. Letting $p_1 = 1$ as the numeraire and referencing production parameters β_k we can write:

$$X_{kh} = X_{1h}\left[\frac{\beta_k}{p_k\beta_1}\right] \qquad (5)$$

which allows us to write total expenditures *M* solely as a function of X_1:

$$M_h = \sum_k p_k X_{hk} = X_{1h}(\sum_k \left[\frac{\beta_k}{\beta_1}\right]) \qquad (6)$$

Using the same first-order condition in (5), we can similarly express output solely as a function

$$\tilde{Y}_h = A_h\left[X_{h1}\prod_k\left[\frac{\beta_k}{p_k\beta_1}\right]\right] \qquad (7)$$

where $\tilde{Y}_h$ is "produced" health by the hospital, but is expressed for an average patient (whose risk-adjustment product $\left[\prod_w Z_{iw}^{\omega_w}\right]$ is normalized to 1).

This means that normalized output can be written:

$$\tilde{Y}_h = A_h\frac{\left[M_h\prod_k[\eta_k]\right]}{\sum_k[\eta_k]} \qquad (8)$$

where $\eta_k = \left[\frac{\beta_k}{p_k \beta_1} \right]$.

Finally, we take the log of output Y and represent logged values by lower-case letters:

$$y_h = \alpha_h + \beta' m_h + \left[\sum_k \ln \eta_k - \ln(\sum_k \eta_k) \right] + \varepsilon_h \qquad (9)$$

and $\beta' = \sum_k \beta_k$. We are unlikely to separately identify the terms in the brackets, but we note that all of these terms that depend solely on (fixed) Medicare prices and production parameters – are assumed for the moment to be constant across hospitals (and independent of TFP), and would therefore be absorbed in the constant term. Under the assumption that hospitals are choosing input prices optimally, m_h or logged total expenditures, along with the conventionally defined total factor productivity (TFP) parameter α_h summarizes the predictable (non-random) component of spending. Small fluctuations in inputs choices will, by the envelope theorem, affect both y and m equally, leading to an orthogonality condition that when inputs are chosen optimally, and the β_k coefficients are the same across hospitals, the specific choice of inputs should not predict outcomes. In the empirical section, we therefore consider an F-test for the joint hypothesis that all input variables are zero.

In Figure 1, we previously illustrated the inefficiency arising from allocative inefficiency; Figure 2 illustrates how either differences in TFP or productive inefficiency, can lead to arbitrary correlations between spending and outcomes. Assume two hospitals, A and B, but where we consider expenditures (on the horizontal axis) and survival and more generally quality of life (y) on the vertical axis. In the case we consider, where each hospital is at Point A and Point B, expenditures are identical but they differ with respect to survival; A gets better outcomes for the same spending level. There are two at least reasons why that might be so. The first is that TFP explains the entire gap between $S_A(X^*)$ and

$S_B(X^*)$, explained by more highly skilled physicians (conditional on inputs), nurses, and support staff, better administration, and cleaner facilities at Hospital A (e.g., de Vries et al., 2010; Bloom et al. (2014); these are shown by the hypothetical production functions that are drawn through A and B, but in the case of A, it is shifted up by the difference in TFP.

The second explanation is independent of TFP; assume for the moment that the two hospitals are identical in that respect. For simplicity, assume there are two distinct technologies for treatment; the Category I "green" input is more cost-effective (the green line 0C in Figure 2; at point C the effectiveness of the treatment ceases), and the alternative "red" Category III input with little or no net health benefit. While Hospital A uses the first (green) technology up to its maximum potential value, at Point C, and then spends additionally on the second technology (the red horizontal line), Hospital B does not use the first technology to its fullest extent (only to Point D) before spending more on the less effective technology (DB), leading to the same spending level, but with worse outcomes. In this case, misallocation would explain the *entire* difference in outcomes between Hospitals A and B.

Figure 2 also illustrates why regressions that attempt to regress outcomes on expenditures may not be estimating the slope of the production function (Diaz-Hernandez et al., 2008). For example, suppose that one's empirical sample comprised of 4 hospitals corresponding to the points D, C, B, and D* in Figure 2 – e.g., different hospitals with respect to their adoption of inputs 1 and 2. The different points could be explained either by TFP differences, by misallocation, or by both. In either case, a conventional regression falsely suggests higher spending "causes" patients to die (e.g., Point J in Figure 1), even though no patient is being harmed by spending more. Conversely, a sample comprising points C, D, A, and D* would yield a positive regression line, suggesting that any efforts to scale back spending would be deleterious to health; again the regression coefficient says nothing about the shape of the production function, only the correlation between Category I and Category III treatments at the hospital level. This is consistent with the findings in Colla et al. (2015) that regional use of low-value (Category III) care is positively correlated with regional spending overall. To recover valid estimates, one must either allow for hospital-specific productivity measures (as in Skinner and Staiger, 2015, or Chandra et al., 2016a) or allow for the presence of misallocation, or both.

We return to Equation 9 to rewrite in order to allow for misallocation as well as differences in TFP across hospitals.

$$y_h = \alpha_h + \beta' m_h + \left[\sum_k \ln \eta^*_{hk} - \ln(\sum_k \eta^*_{hk}) \right] + \varepsilon_h \qquad (10)$$

where $\eta^*_{hk} = \left[\frac{(1+\lambda_{hk})\beta_k}{(1+\lambda_{h1}) p_k \beta_1} \right]$.

To develop the intuition, consider Hospital B in Figure 2 under the assumption that its lower output is the consequence of the underuse of Input 1 (the green technology) and overuse of Input 2 (the red technology). The implicit shadow price for Input 2 is negative, so that the ratio ($(1+\lambda_{h2})/(1+\lambda_{h1})$) is less than one. It is straightforward to show in the two-input case that when there is a preexisting distortion, an increase in the relative price distortion between the two inputs will reduce output conditional on expenditures. Given that higher shadow prices are implied by lower utilization (and conversely), we use factor inputs (relative to other hospitals) as measures of misallocation or appropriate allocation. We acknowledge that these estimated coefficients may also capture some component of TFP, in the sense that more skilled physicians may also make better input choices. Thus comparing coefficients in our regressions with the corresponding estimates from clinical literature provides some bounds on what the coefficients are capturing.

Another potential concern is that hospitals may differ with regard to the productivity of their inputs; for example as in Chandra and Staiger (2007) where hospital cardiologists skilled in the use of percutaneous coronary interventions (PCI) exhibited (optimally) higher rates of use. In the context of our production function, this would correspond to a larger β for PCI. In this case, we could observe a positive association between high PCI use and outcomes conditional on spending, but it's not because other hospitals with low PCI use are lagging behind or behaving sub-optimally.

In the context of the Cobb-Douglas model, we can consider what might the impact be of varying values of β across inputs. Holding the output elasticity or the sum $\beta*$ (that is, the overall log return to doubling inputs) constant, it turns out that varying the individual β s has little impact on outcomes, even when the quantities of inputs are allowed to vary optimally.[10] One can provide a rough test this hypothesis, however, by considering whether hospitals with higher measured TFP gain a higher marginal return (with regard to survival) for the three input categories (or to $\beta*$ more generally); we consider this below.

III. Data

Medicare claims. We created a cohort of patients hospitalized with acute myocardial infarction (AMI) in the fee-for-service Medicare population during 2007-June 2017, with follow up data through December 31, 2017. An AMI is based on the first diagnosis code, which is 410.x0 or 410.x1, not including 410.x2, in ICD9 coding (prior to October 2015) and I21.x in subsequent ICD10 coding beginning October 1, 2015. We have considered issues regarding the transition elsewhere (Mainor et al., 2019) and were not able to detect coding-induced jumps around October 1, 2015. In the hospital-level analysis for the entire sample, we limit to hospitals with at least 50 admissions for AMI during the combined years 2007-17.

---

[10] For example, in a simple case where $Y = X_1^{.1} X_2^{.1}$, total expenditures are $100, and prices of each input are equal to $1, output will be 2.187. Shifting the production parameters to 0.0125 and 0.0075 (so the sum is still .2) leads to a change of 33 percent in optimal input use (to 66.7 and 33.3, respectively) but only a 1 percent increase in outcomes, to 2.080.

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 quintiles based on the American Community Survey (2010-2014 five-year estimates) , and age-sex 5-year cells (e.g., women aged 70-74), and race (African-American, Hispanic, Asian, Native American). 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 admission, and weights them for severity. However, we note that the use of HCC measures leads to biases in conventional regressions of spending on outcomes because physicians who see patients more often and look harder for diseases will tend to code more diseases, thus making them appear sicker (Song et al., 2010; Finkelstein et al., 2016), meaning that when they do survive, the hospital will get credit as a highly productive institution. We also adjust for the fraction of the hospital referral region enrolled in Medicare Advantage to capture the idea that the fee-for-service population could exhibit greater unmeasured health deficits if healthier enrollees select into managed care.

Tourists. We follow Doyle (2011) by considering tourists under the reasonable assumption that few tourists consider whether their vacation hotel is near a high- or low-intensity hospital. We define tourists in the following way – that they received their treatment in an HRR that is different from their HRR of residence, and that they received less than 20% of their healthcare in the HRR in which the AMI was treated.

ZIP Codes. Another approach is to sweep out all neighborhood variation by including ZIP[11] code fixed effects; this will of course absorb common health behaviors, average socioeconomic status, environmental health effects, and other neighborhood factors. One limitation of a ZIP code is that there

[11] ZIP stands for zone improvement plan.

can still remain considerable variation within a ZIP code, but there is clear evidence of considerable variation in health behaviors and socioeconomic status across ZIP codes.

Clinically relevant inputs: Category I, II, and III. There are a wide array of different treatments for AMI patients, both in the acute setting, and subsequently post-discharge. We consider a range of such treatments or procedures where our initial hypothesis of effectiveness is based on existing clinical evidence. Because these measures are derived from the claims data, there is a potential bias in measuring such rates across hospitals when mortality rates differ. Suppose that Hospital B has a higher mortality rate than Hospital A; then any measures of inputs could be systematically biased; either spending (if people who die cost more – although empirically, people who survive post-AMI account for more spending, not less), or (e.g.) PCI rates or primary care follow-ups, which do not (generally) occur when people have already died. For this reason, when creating input measures, we consider both spending measures based on everyone in the sample (including those who die), as well as people who survived for the relevant length of time, whether 30 days (for the corresponding 30-day spending measure), 6 months, or 1 year; these are then applied at the hospital level to the entire dataset.[12]

To help organize the data analysis, we follow Wennberg et al. (2002), Wennberg (2010), 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" or Category I inputs which are distinguished by their high cost-effectiveness and limited scope for expensive overuse. Examples are beta blocker, statin, and ACE/ARB 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, but there are a significant minority of people who do not tolerate

[12] When we use public measures from (e.g.) Hospital Compare, these may include people who died.

such treatments well, so the optimal compliance is not 100%.[13] Similarly, the integration between the hospital and physicians in the community is potentially important for post-acute survival (Sharma et al., 2010; Hernandez et al., 2010); thus we use the fraction of patients discharged from the hospital for a medical condition that is seen by any physician within 14 days.[14] Finally, we include 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). In the 1990s, there was evidence that this procedure exhibited diminishing returns as physicians reached into less appropriate patients, but current evidence for same-day PCI suggests less heterogeneity in treatment effects.

The “Category II” treatments are hypothesized to exhibit a greater degree of heterogeneity in incremental benefits across different types of patients. While same-day PCI has well-established benefits, subsequent PCI is often viewed as potentially less beneficial, and in the post-acute setting may exhibit diminishing returns working further into the distribution of patients (Chandra and Staiger, 2019). 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 (Baicker and Chandra, 2004b.) A final example is the number of

[13] Teaching hospitals may exhibit higher levels of TFP, as in (Burke et al., 2018), but we do not consider teaching hospitals as an input.

[14] We use this broader measure instead of post-acute physician visits for AMI patients because the sample sizes are so much larger; however, the two series are highly correlated, with a rho greater than .7 when weighted by sample size.

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). For this reason, we hypothesize an inverse-U-shape influence of the number of different physicians, or MRI and CT scans, on survival rates; we address this by including dummy variables for hospitals if they are in the (weighted) top or bottom quartiles of these inputs.

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 fit this description.[15] 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 AMI-specific home health care spending among those who receive home health care on a post-acute basis ($10,687 in 2017 dollars). Under the null of no hospital variation, all hospitals would report a fraction of .10. This measure is not necessarily higher in a sicker population, since it captures only these "outlier" utilization measures, rather than the number of home health patients *per se*.[16]

We also consider services that are indirectly related to AMI patients but reflect the practice and management styles of physicians at the hospital. We use three of these "Choosing Wisely" measures involving the use of "double CT" scans of the chest and abdomen, 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).

---

[15] 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/

[16] 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."

To describe the basic patterns of the data, we begin with two linear regressions: One that estimates risk-adjusted one-year survival, and the other estimating risk-adjusted one-year expenditures. (Estimating two separate regressions avoids the biases noted above that arise from putting spending on the right-hand side of an outcomes regression.) The regressions use individual-level risk adjusters and ZIP-level measures of socioeconomic status with hospital-level random effects. These random-effects models "shrink" hospital-specific effects towards the predicted values when hospitals are small, and thus avoid over-fitting as in fixed-effects models. And while the random effects assume independence of the error term with the X and Z variables, in practice random effects estimates for the larger hospitals are nearly identical to fixed effects models (the correlation coefficient is 0.99 for hospitals with N > 500 patients). The regression coefficients and risk-adjusters are included in Appendix Table A.1. All estimates are clustered by hospital.

We then turn to estimating the productivity model using a similar regression model as the survival model described above including patient and ZIP-code levels, but also including hospital-level measures of log average price-adjusted Medicare expenditures, and hospital-specific Category I through Category III inputs. Rather than just tossing in all the separate components, we use three distinct principal components models, one for each of the categories, to create a common component for ease of interpretation and greater statistical prediction.

V. Results

In Table 1, summary statistics are presented for the entire sample of 1.7 million individuals with AMI, and for the 123,984 out-of-HRR "tourists." Tourists were slightly younger, were less likely to be Black, and were somewhat healthier based on comorbidities at admission.

In Figure 3A, we use the regression models from Table A.1 of the predicted hospital-specific random-effects for expenditures (the horizontal axis) and one-year survival (the vertical axis) for the full sample evaluated at mean values of the independent variables. While the previous literature has focused on whether the slope of the regression line is positive or negative, the more interesting finding of these tables are the large differences in productivity (and net benefit); hospitals in the Northwest corner of the graph are achieving above-average outcomes at low costs, while those in the Southeast corner are suggestive of low-productivity hospitals. And while other industries also exhibit wide variation in productivity (Chandra et al., 2016b), the difference in survival rates range over 10 percentage points.

Figure 3B presents a similar graph for the tourist sample (given the smaller overall sample, we include all hospitals with at least 50 AMIs during the period of analysis). Like Figure 3A, there is wide variability in both spending and outcomes, with at best only a weak correlation between the two.[17]

Our first set of regressions are presented in Table 2, where we report regession estimates that follow standard conventions of placing log expenditures on the right-hand side of the regression equation; all regressions include a full set of risk adjusters (not reported) and in some cases ZIP code fixed effects. Model 1 reports coefficients of 30-day mortality on the log of 30-day expenditures; with a highly significant coefficient of 0.0368, it implies a positive association between spending and survival, in magnitude roughly between the Doyle (2011) and the Doyle et al. (2015) estimates. Model 2 differs from Model 1 in limiting the cohort for estimating hospital-level spending to only those who survived 30 days; this reduces the coefficient to 0.00977 (but still significant). Model 3 moves to a one-year horizon, with a coefficient of 0.0541 when spending is not limited to survivors, and 0.00674 (and marginally significant) for survivors only in Model 4; however, including ZIP code fixed effects increases the estimated effect to 0.0157 in Model 5. We next move to tourists in Models 6 through 8, for a variety of specifications, but where the estimates are not significantly different from zero. In sum, the estimates are quite sensitive to the specification of the regression, with the magnitude of the

[17] For the record, the bivariate regression coefficient is -9.015e-08 (s.e. = 1.526e-07) for the full sample and -5.487e-07 (s.e. = 3.011e-07) for the tourist sample in the sample of hospitals reported in Figures 3A and 3B.

association varying based on whether spending truncated by death is included, and with ZIP code fixed effects strengthening the association.

In Table 3, we account for the allocation of inputs within the hospital, including individual independent input variables in the equation in Model (6), alongside the risk adjustment tools described above. Each of the first three models include only the first principal components of each of the categories, as well as the log of total price-adjusted expenditures; all regressions reject the null hypothesis of orthogonality with regard to inputs.

Proxying for the use of efficient, Category I treatments with the first principal component score, the principal component score on Category I services is significant and positive across specifications, as hypothesized. Each of the individual inputs is also positive or non-significant in the deconstructed regressions. The coefficients on Category II, or treatments with heterogeneous effects depending on patient or provider, are mixed. The coefficient on the principal component for Category II is positive or zero. Late PCI (after the admission date) is consistently protective of survival, as are the hospital having fewer different physicians treat each patient, and not being in the top quartile for use of MRI and CT, indicating diminishing returns to multiple physicians and scans. The coefficient on the Category III principal component score is consistently negative and significant. However, the deconstructed measures are non-significant or close to zero when provider-level random effects are included.

Generally across specifications, the results are consistent with our model, in the sense that the effective Category I treatments are highly and positively associated with productivity; for example in Model 1, a one-standard deviation rise in the principal component Category I treatment is predicted to increase survival by 1.5 percentage points, or roughly half of the interquartile range across hospitals. The average Category II component has little impact, while Category III use is negatively associated with productivity (a one-standard deviation in this measure is predicted to reduce survival by 0.3

percentage points). These results are insensitive to the exact specification of the model, as shown in the first three columns (Model 1 through 3).

Finally, we compare a random effects model with and without the detailed inputs in Models 4 and 5. Here we have broken out high-quartile and low-quartile utilization of Category II treatments to test the hypothesis that both high-use and low-use of multiple physicians could lead to worse outcomes, with a similar pattern for scans. We find some evidence consistent with this hypothesis for scans, but in general a greater degree of physician specialization did not exhibit higher returns relative to the lowest quartile (Model 5). And while the specific inputs are often highly significant, the coefficients are sometimes counterintuitive, as in the case of home health care outlier behavior (showing a positive coefficient), a result at odds with Doyle et al. (2017) who find consistently negative effects of home health care on health outcomes (also see McKnight, 2006). This may reflect the interaction of our different measures of Category III inputs; the bivariate correlation between home health care and risk adjusted survival is strongly negative; the corresponding correlation for spending is positive. Finally, the difference in the estimated standard deviation of the estimated random effects are only slightly lower with inputs included.[18] Results are similar with included ZIP code fixed effects.

We repeat the analyses on our sample of tourists (Table 4). While the precision of the estimates is diminished, the results are broadly similar. The Category I principal component score is consistently, significantly positively related to survival, while individual treatment indicators at the hospital level are either positive and significant or zero. Late PCI remains positive and significant in the tourist sample, while the principal component score for Category II is non-significant. Similar to the analyses on the full sample, the coefficient on the Category III principal component score is consistently negative and

[18] This may also be the consequence of the smaller hospital sample size in Model 5 owing to the lack of reliable data on inputs for all hospitals.

significant.[19] In sum, the coefficients on the three principal components are remarkably stable across different specifications, and for tourists and the entire sample.

We return to our concern above that the reason why hospitals may choose different input combinations is because they have different production functions; some are better at PCI, while others are not as Chandra and Staiger (2007) found for hospitals in the mid-1990s when such procedures were being newly developed. However, we do not find substantive differences in coefficients on Category I, II, or III when the sample is split into terciles based on their estimated TFP productivity.

How important is misallocation? While we recognize that the coefficients may overstate the true effects of the treatments, we create a predicted risk-adjusted measure of survival based solely on values of inputs and the estimated coefficients for the entire sample; call this $\hat{S}_h$; we normalize the mean to the overall survival mean. Values of $\hat{S}_h$ are shown in Figure 4, along with risk-adjusted survival rates from Figure 3A, again for hospitals with at least 400 AMIs during the period of analysis. Predicted mortality based solely on input choices ranges by 10 percentage points across hospitals, and is positively associated with risk-adjusted mortality; a bivariate regression explains 22 percent of the variance.

Are there characteristics of hospitals that are closely associated with low levels of input misallocation? One obvious feature would be hospital volume; the larger the hospital, the greater is the incentive to invest in providing high-quality inputs to improve health outcomes (Skinner and Staiger, 2015). This hypothesis appears to be consistent with the data; we illustrate predicted survival ($\hat{S}_h$) by ventiles of hospitals as ranked from lowest to highest volume hospitals in Figure 5; there is a strong association between high-volume hospitals and adoption of efficient input practices.

[19] The estimated standard error of the random effects are *higher* for Model 5.

VII. Conclusions

In this paper, we have revisited the literature on health care productivity, introducing the idea that systematic misallocation in inputs can contribute to substantial differences across hospitals in survival rates. We estimate our model of productivity for the specific case of AMI, and find that misallocation can explain considerable differences across hospitals in productivity, although a very important role for TFP in explaining hospital-level differences in survival remains.

While there is a general recognition that there is wide variation in TFP across a variety of industries (e.g., Syverson, 2011), it has been more difficult to identify exactly what causes such differences. In the case of misallocation, however, it is more clear; measurable inputs that in theory could be acted upon, either by the hospital seeking to improve outcomes, or by patients who might more quickly find a more productive hospital. In this paper, we include several of these measures, but also others that have not previously been estimated, and we find that some – but not all – of these measures are predictive of differences across hospitals in productivity.

We acknowledge that the parameters used in the model may not capture causal factors, so that simply lowering (e.g.) double CT scans may not have a direct impact on hospital productivity if such scans are a symptom of poor organizational structure rather than a causal measure *per se*. However, many if not most of our measures do have a causal interpretation; our estimated effects of same-day PCI for example is just a bit larger than estimates from randomized trials. Indeed, one could simply include estimates of survival effects based on randomized trial evidence rather than estimating them, as we do here, to infer the potential gains from reducing misallocation.

What are the policy implications? Current alternative payment models seek to reduce some of the financial incentives simply to do more, but the fundamental question is whether we might expect an improvement – in the sense of a reduction in the magnitude of λ over time. 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. (2019).

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 misallocation, there is a real potential for improving productivity across hospitals that might yield real benefit to patients.

REFERENCES

Baicker, Katherine, and Amitabh Chandra, 2004a. Medicare Spending, the Physician Workforce, and Beneficiaries' Quality of Care." *Health Affairs* Apr 7.

Baicker, Katherine, and Amitabh Chandra, 2004b. "The Productivity of Physician Specialization: Evidence from the Medicare Program," *American Economic Review* 94(2):357-61.

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

Bekelis Kimon. Elliott Fisher, N. Labropoulos, Weiping Zhou, and Jonathan Skinner, 2014. "Variations in the Intensive use of Head Computed Tomography for Elderly Hemorrhagic Stroke Patients in Leading U.S. hospitals," *Radiology* 275(1): 188-195.

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., 2001. "A Qualitative Study of Increasing β-Blocker Use After Myocardial Infarction: Why Do Some Hospitals Succeed?," *JAMA* 285(20): 2604-2611.

Burke, Laura, Dhruv Khullar, E. John Orav, Jie Zheng, Austin Frakt, and Ashish K. Jha, 2018. "Do Academic Medical Centers Disproportionately Benefit the Sickest Patients?" *Health Affairs* 37(6): 864-872.

Chan, David, 2020. "Influence and Information in Team Decisions: Evidence from Medical Residency," *American Economic Journal: Economic Policy* (forthcoming).

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, 2016a. "Health Care Exceptionalism? Performance and Allocation in the US health care sector." *American Economic Review* 106(8): 2110-44.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, 2016b. "Productivity Dispersion in Medicine and Manufacturing." *American Economic Review* 106(5): 99-103.

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

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

Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB, 2015. Choosing wisely: prevalence and correlates of low-value health care services in the United States. *J General Internal Med.* 2015 Feb;30(2):221-8.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg, 2019. "Physician Beliefs and Patient Preferences: a New Look at Regional Variation in Health Care Spending." *American Economic Journal: Economic Policy* 11(1): 192-221.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran, 2016. "Information, Misallocation, and Aggregate Productivity." *Quarterly Journal of Economics* 131(2): 943-1005.

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. *New England J Medicine* 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. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy*, 123(1): 170–214.

Doyle Jr, Joseph J., John A. Graves, and Jonathan Gruber, 2017. "Uncovering Waste in US healthcare: Evidence from Ambulance Referral Patterns." *Journal of Health Economics* 54: 25-39.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The Quarterly Journal of Economics* 131(4): 1681-1726.

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. *Annals of Internal 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. *Annals of Internal 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.

Garber, Alan M., and Jonathan Skinner, 2008. "Is American Health Care Uniquely Inefficient?." *Journal of Economic Perspectives* 22(4): 27-50.

Garthwaite, Craig, John A. Graves, Tal Gross, Zeynal Karaca, Victoria R. Marone, and Matthew J. Notowidigdo, 2019. "All Medicaid Expansions Are Not Created Equal: The Geography and Targeting of the Affordable Care Act," National Bureau of Economic Research Working Paper No. 26289.

Gottlieb, Daniel. J., Weiping Zhou, Yunjie Song, K. G. Andrews, Jonathan Skinner, 2010. "Prices Don't Drive Regional Medicare Spending Variations." *Health Affairs* (Millwood) 29, 3, Mar-Apr, 537-43.

Haltiwanger, John, Robert Kulick, and Chad Syverson, 2018. "Misallocation Measures: The Distortion that Ate the Residual." National Bureau of Economic Research Working Paper No. 24199.

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.

Hsieh, Chang-Tai, and Peter J. Klenow. "Misallocation and manufacturing TFP in China and India, 2009." The *Quarterly Journal of Economics* 124(4): 1403-1448.

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* 27, 1, Jan-Feb, 159-68.

Lewis, Valerie A., Karen Joynt Maddox, Andrea M. Austin, Daniel J. Gottlieb, and Julie PW Bynum, 2019. "Developing and Validating a Measure to Estimate Poverty in Medicare Administrative Data." *Medical Care* 57(8): 601-607.

Mainor, Alexander J., Nancy E. Morden, Jeremy Smith, Stephanie Tomlin, and Jonathan Skinner, 2019. "ICD-10 Coding Will Challenge Researchers." *Medical Care* 57(7): e42-e46.

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

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

Deryugina, Tatyana, and David Molitor, 2019. "Does When you Die Depend on Where You Live? Evidence from Hurricane Katrina" No. w24822. National Bureau of Economic Research, 2018.

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.

____ and ____, 2017. "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, 31(3):151–174.

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

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, et al., 2010. "Regional Variations in Diagnostic Practices." *New England J Medicine* 363(1):45-53.

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

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

Westfall, John M., James Mold, and Lyle Fagnan, 2007. "Practice-based Research—'Blue Highways' on the NIH Roadmap." JAMA 297(4): 403-406.

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): w566-72.

Figure 1: Allocative Efficiency and Inefficiency in a Health Care Production Function

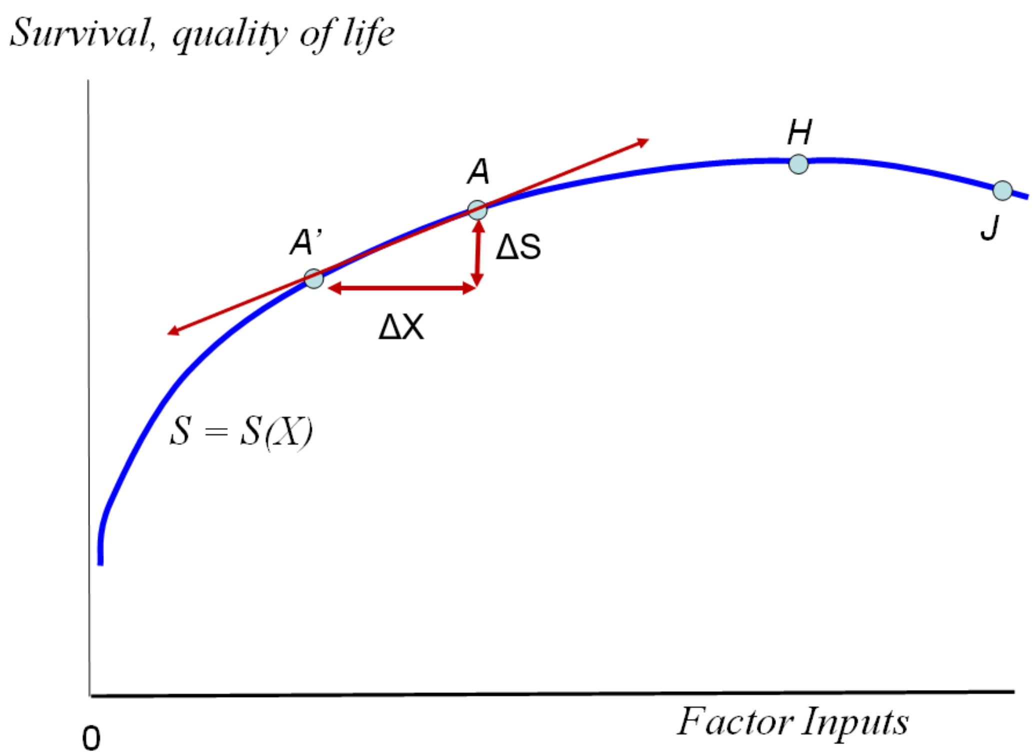

Figure 2: Sources of Differences in Outcomes: Total Factor Productivity and Input Misallocation

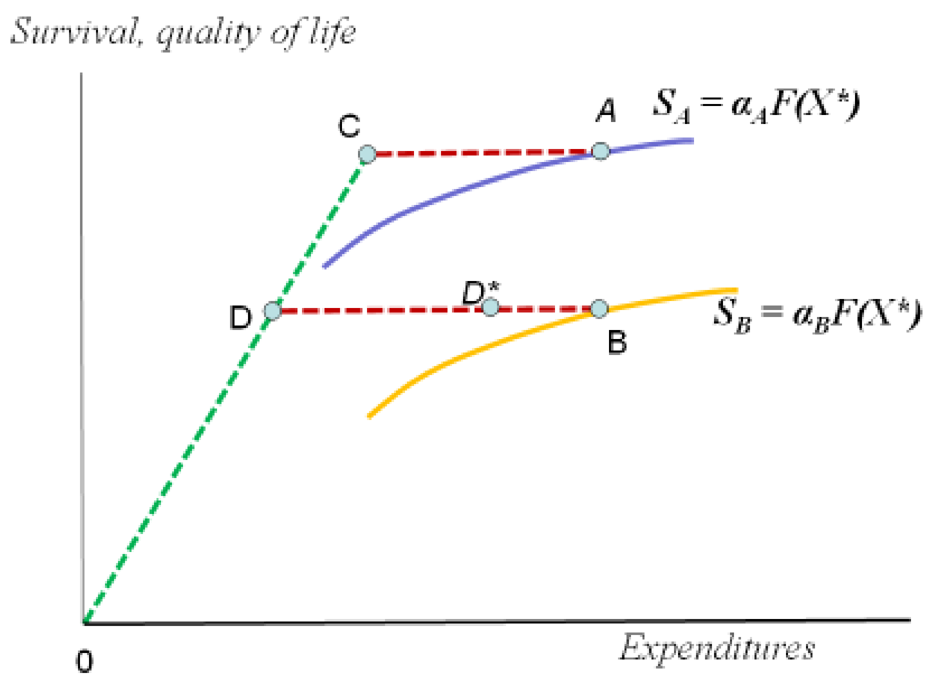

Figure 3a: Association between Risk-Adjusted Spending and Survival: 2007-2017

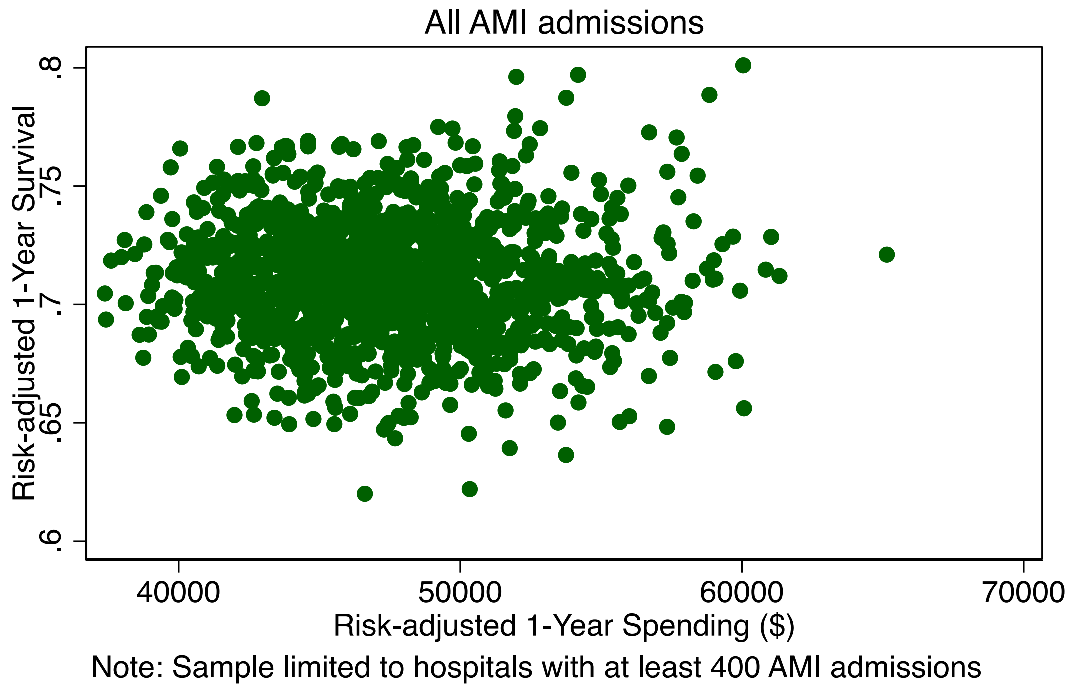

Figure 3b: Association between Risk-Adjusted Spending and Mortality for Tourists: 2007-2017

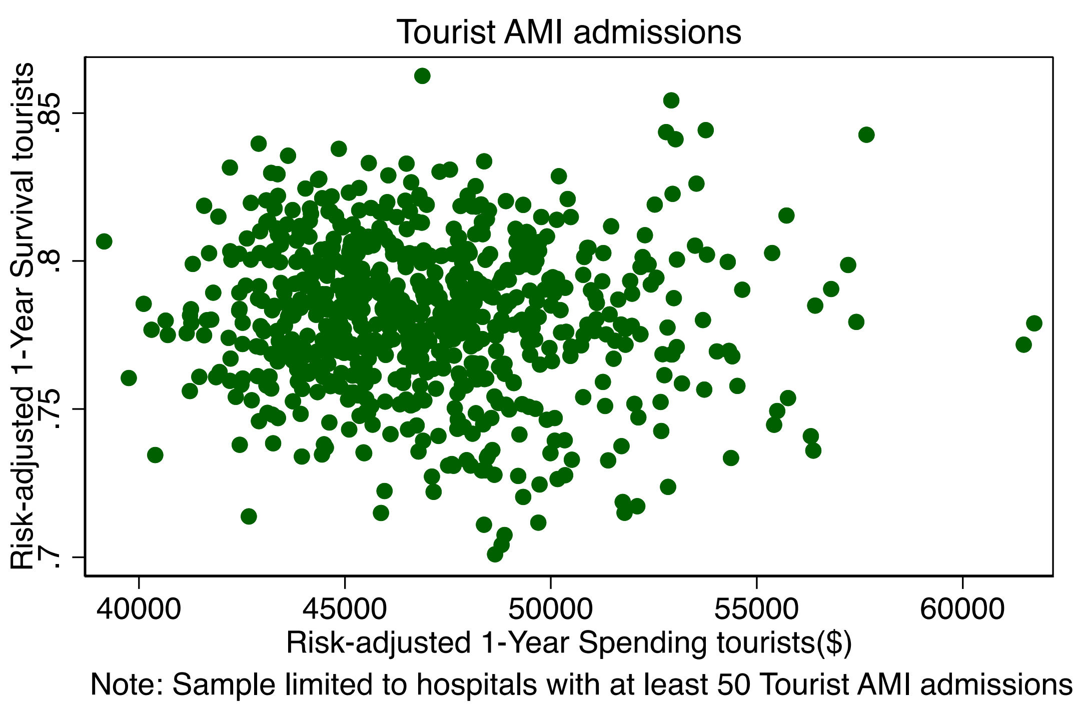

Figure 4: Association between Predicted Survival Based on Misallocated Inputs and Risk-Adjusted Survival

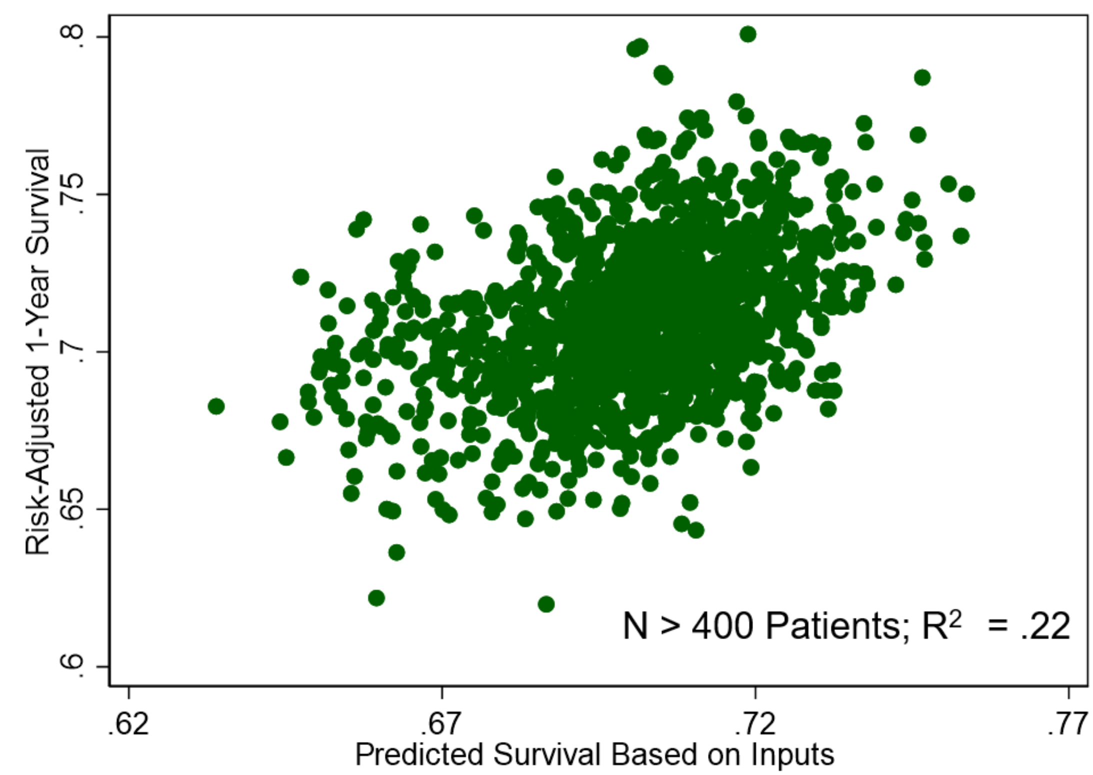

Figure 5: Association between Predicted Survival Based on Misallocated Inputs and Hospital Volume (by Ventile of Volume)

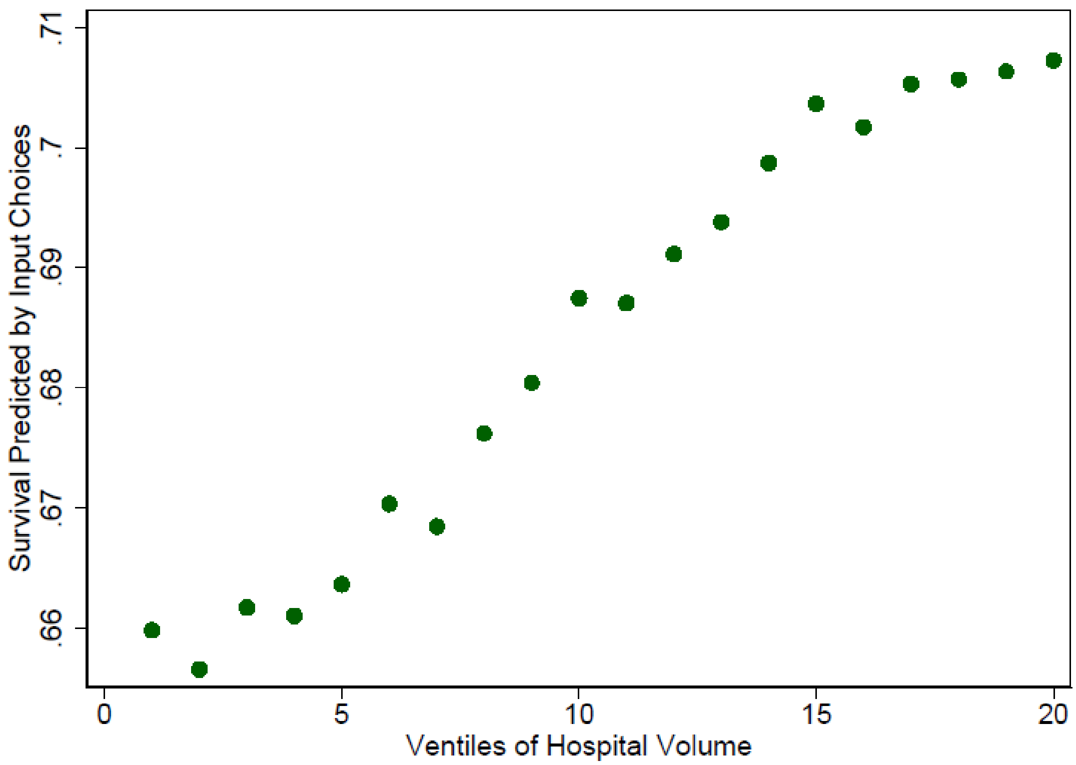

Table 1: Descriptive Characteristics of Medicare Beneficiaries with AMI: 2007-15

	All AMI Admissions Mean (SD)	Tourist AMI Admissions Mean (SD)
Survival		
1 year	0.703 (0.457)	0.779 (0.415)
6 months	0.775 (0.418)	0.833 (0.373)
30 days	0.862 (0.345)	0.890 (0.313)
7 days	0.922 (0.269)	0.935 (0.247)
Price Adjusted Spending		
1 year	46702 (43524)	46704 (44196)
6 months	37765 (34958)	38888 (37102)
30 days	24307 (19856)	26696 (22207)
age at index admitted date	78.219 (8.356)	76.142 (7.902)
female	0.483 (0.500)	0.416 (0.493)
Comorbidities		
Peripheral Vascular Disease	0.081 (0.273)	0.073 (0.261)
Chronic Non-Asthmatic Lung Disease	0.178 (0.382)	0.151 (0.358)
Dementia	0.038 (0.192)	0.027 (0.161)
Chronic Renal Failure	0.189 (0.392)	0.142 (0.349)
Cancers (various)	0.050 (0.218)	0.040 (0.197)
Metastatic Solid Tumor	0.013 (0.111)	0.008 (0.088)
Congestive Heart Failur	0.397 (0.489)	0.343 (0.475)
HIV/AIDS	0.001 (0.023)	0.000 (0.020)
Hemi/Paraplegia	0.003 (0.052)	0.003 (0.051)
Liver Disease	0.005 (0.070)	0.004 (0.064)
Diabetes	0.274 (0.446)	0.258 (0.437)
Peptic Ulcer Disease	0.008 (0.090)	0.008 (0.089)
Rheumatologic Disease	0.017 (0.130)	0.014 (0.117)
HCC score 6m before index admission	1.343 (1.091)	1.086 (0.931)
Race	0.000 (0.000)	0.000 (0.000)
Native	0.005 (0.070)	0.008 (0.090)
Hispanic	0.015 (0.121)	0.012 (0.108)
Other	0.011 (0.103)	0.011 (0.105)
Asian	0.012 (0.111)	0.011 (0.103)
Black	0.074 (0.261)	0.056 (0.231)
White	0.884 (0.321)	0.902 (0.298)
Medicare Advantage Proportion in residence HRR	0.261 (0.127)	0.261 (0.131)
Dual Eligible for Medare and Medicaid	0.165 (0.371)	0.130 (0.336)
Median Household Income of zipcode	55500 (25402)	54933 (26133)
AMI subtype		
Anterior	0.086 (0.280)	0.106 (0.308)
Inferior	0.104 (0.305)	0.133 (0.339)
Right	0.012 (0.109)	0.016 (0.124)

Subend	0.731 (0.443)	0.684 (0.465)
Other sites	0.019 (0.135)	0.020 (0.139)
Unspecified	0.049 (0.216)	0.042 (0.200)
Number of observations*	1,713,345	123,984

*Sample size for 1 year follow-up variables: (number of observations for 1 year follow up: 1,637,468 and 117,880, respectively.

Table 2: Association between Spending and Survival: Conventional Regression Approaches

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6 (Tourists)	Model 7 (Tourists)	Model 8 (Tourists)
Log 1-year Spending			0.0541***	0.00674*	0.0157***			-0.0124
			(0.0031)	(0.0033)	(0.0027)			(0.0079)
			17.35	2.044	5.785			-1.567
Log 30-Day Spending	0.0368***	0.00977***				0.0124	0.00766	
	(0.0027)	(0.0026)				(0.0070)	(0.0071)	
	13.42	3.746				1.756	1.086	
Outcome	30-day survival	1-year survival	1-year survival	1-year survival	1-year survival	30-day survival	30-day survival	1-year survival
Spending limited to survivors		x		x	x		x	x
Zip code fixed effects					x			
Observations	1,801,733	1,801,540	1,721,178	1,722,000	1,721,178	128,565	128,561	122,186
R-squared	0.085	0.085	0.182	0.201	0.201	0.073	0.073	0.154

Robust standard errors in parentheses, t-statistics below standard errors.

*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 3: Regressions Estimates for 1-Year Survival Allowing for Misallocation: Entire Sample

	Model 1	Model 2	Model 3	Model 4	Model 5
Log 1-Year Hospital Spending		0.0310***	0.0106**	0.0214***	0.0262***
		(0.0044)	(0.0038)	(0.0024)	(0.0030)
		6.987	2.783	8.901	8.634
Log 30-Day Hospital Spending	-0.00929*				
	(0.0043)				
	-2.155				
Category 1 Principal Component	0.0131***	0.0129***	0.0131***		
	(0.0006)	(0.0006)	(0.0006)		
	22.14	21.86	21.97		
Category 2 Principal Component	0.00265**	-0.00189	0.000549		
	(0.0010)	(0.0010)	(0.0010)		
	2.674	-1.827	0.54		
Category 3 Principal Component	-0.00308***	-0.00390***	-0.00342***		
	(0.0006)	(0.0006)	(0.0006)		
	-5.04	-6.394	-5.619		
Category I Inputs:					
Beta Blockers Post-Discharge					0.00747*
					(0.0034)
					2.23
Statins Post-Discharge					0.00555
					(0.0030)
					1.832
ACE/ARB Post-Discharge					-0.00259
					(0.0028)
					-0.932
Same-Day (Early) PCI					0.176***
					(0.0045)
					38.83
MD Visits within 14 Days Discharge					0.0188**
					(0.0069)
					2.708
Category II Inputs:					
Late (> 1 Day) PCI					0.233***
					(0.0095)
					24.56
Lowest Quartile Different MDs					0.00240*
					(0.0012)
					1.991
Highest Quartile Different MDs					-0.00161
					(0.0012)
					-1.303
Lowest Quartile MRI/CT Scans					-0.0014
					(0.0010)
					-1.432
Highest Quartile MRI/CT Scans					-0.00395***
					(0.0011)
					-3.616
Category III Inputs:					
Cardiac Imaging for Low-Risk Preop.					0.00014
					(0.0003)
					0.522
Fraction in Top Decile Home Health					0.0188***
					(0.0055)
					3.42
Double CT Scan (Abdomen)					-3.50E-05
					(0.0000)
					-0.836
Double CT Scan (Chest)					-0.000247***
					(0.0001)
					-3.614
Spending limited to survivors	x		x	x	x
Provider random effects				x	x
Observations	1,638,272	1,638,272	1,638,168	1,721,178	1,472,697
R-squared	0.182	0.182	0.182		
Number of providers				3,062	2,507
sigma_u				0.0264	0.0258

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table 4: Regressions Estimates for 1-Year Survival Allowing for Misallocation: Tourists

	Model 1	Model 2	Model 3	Model 4	Model 5
Log 1-Year Hospital Spending		-0.0014	-0.011	-0.0022	0.0149
		(0.0111)	(0.0103)	(0.0074)	(0.0103)
		-0.126	-1.072	-0.296	1.44
Log 30-Day Hospital Spending	-0.000645				
	(0.0106)				
	-0.061				
Category 1 Principal Component	0.0138***	0.0137***	0.0136***		
	(0.0017)	(0.0017)	(0.0017)		
	8.206	8.23	8.138		
Category 2 Principal Component	0.00172	0.00184	0.003		
	(0.0020)	(0.0023)	(0.0023)		
	0.847	0.799	1.323		
Category 3 Principal Component	-0.00364**	-0.00361**	-0.00339*		
	(0.0014)	(0.0014)	(0.0014)		
	-2.642	-2.596	-2.449		
Category I Inputs:					
Beta Blockers Post-Discharge					-0.00832
					(0.0124)
					-0.673
Statins Post-Discharge					0.0250*
					(0.0116)
					2.148
ACE/ARB Post-Discharge					-0.00725
					(0.0097)
					-0.744
Same-Day (Early) PCI					0.186***
					(0.0140)
					13.31
MD Visits within 14 Days Discharge					-0.00532
					(0.0188)
					-0.284
Category II Inputs:					
Late (> 1 Day) PCI					0.221***
					(0.0329)
					6.729
Lowest Quartile Different MDs					-0.00417
					(0.0036)
					-1.168
Highest Quartile Different MDs					1.40E-05
					(0.0039)
					0.00363
Lowest Quartile MRI/CT Scans					-0.00358
					(0.0033)
					-1.099
Highest Quartile MRI/CT Scans					-0.000774
					(0.0037)
					-0.212
Category III Inputs:					
Cardiac Imaging for Low-Risk Preop.					9.14E-05
					(0.0008)
					0.115
Fraction in Top Decile Home Health					-0.0365*
					(0.0183)
					-1.99
Double CT Scan (Abdomen)					-5.17E-05
					(0.0001)
					-0.441
Double CT Scan (Chest)					-0.000231
					(0.0002)
					-1.115
Spending limited to survivors	x		x	x	x
Provider random effects				x	x
Observations	117,905	117,905	117,903	122,186	107,115
R-squared	0.154	0.154	0.154		
Number of provider				2,751	2,237
sigma_u				0.039	0.0478

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Appendix Table A.1: Risk Adjustment Models for 1-Year Spending and Outcomes

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
	All Patients	All Patients	Tourists	Tourists
	Price-Adjusted Expenditures	One-Year Survival	Price-Adjusted Expenditures	One-Year Survival
Peripheral Vascular Disease	1,807***	0.0103***	2,447***	0.000323
	(131.9)	(0.00126)	(525.0)	(0.00444)
	13.70	8.133	4.661	0.0727
Chronic Non-Asthmatic Lung Disease	285.9**	-0.0300***	-100.6	-0.0315***
	(100.0)	(0.000971)	(365.8)	(0.00356)
	2.859	-30.92	-0.275	-8.838
Chronic Renal Failure	8,081***	-0.0585***	7,367***	-0.0531***
	(130.2)	(0.000995)	(414.1)	(0.00352)
	62.08	-58.77	17.79	-15.09
Cancers	-1,562***	-0.0699***	-958.4	-0.0506***
	(157.1)	(0.00172)	(602.0)	(0.00627)
	-9.944	-40.62	-1.592	-8.073
Cancer: Metastatic Solid Tumor	-9,700***	-0.346***	-8,266***	-0.360***
	(287.7)	(0.00326)	(1,397)	(0.0163)
	-33.71	-106.0	-5.917	-22.12
Congestive Heart Failure	9,869***	-0.120***	13,612***	-0.104***
	(105.9)	(0.000873)	(312.7)	(0.00276)
	93.17	-137.3	43.53	-37.69
HIV	2,994	0.00423	-2,194	0.0148
	(1,684)	(0.0136)	(4,963)	(0.0566)
	1.778	0.310	-0.442	0.262
Stroke	13,746***	-0.169***	20,020***	-0.199***
	(908.6)	(0.00701)	(3,774)	(0.0274)
	15.13	-24.05	5.305	-7.287
Liver disease	-2,365***	-0.0865***	-4,720**	-0.0744***
	(542.2)	(0.00506)	(1,614)	(0.0190)
	-4.363	-17.08	-2.925	-3.910
Diabetes	-62.07	0.0439***	-1,298***	0.0341***
	(104.7)	(0.000820)	(307.4)	(0.00271)
	-0.593	53.52	-4.222	12.58
Peptic Ulcer Disease	6,556***	0.000717	9,574***	-0.00717
	(398.6)	(0.00367)	(1,692)	(0.0130)
	16.45	0.195	5.658	-0.552
Rheumatologic Disease	-2,010***	0.0261***	-2,768**	0.0217*
	(227.1)	(0.00244)	(954.3)	(0.00904)
	-8.851	10.68	-2.900	2.403
Native-American	3,436***	0.000803	3,326*	-0.00410
	(497.7)	(0.00470)	(1,489)	(0.0142)
	6.903	0.171	2.233	-0.289
Hispanic	1,453***	0.0287***	1,002	0.0145
	(336.6)	(0.00282)	(1,220)	(0.0108)

	4.316	10.19	0.821	1.340
Other Race/Ethnicity	1,490***	-3.72e-06	2,373	-0.00888
	(385.8)	(0.00323)	(1,429)	(0.00967)
	3.861	-0.00115	1.661	-0.919
Asian	1,198**	0.0173***	2,143	-0.00949
	(406.4)	(0.00319)	(1,470)	(0.0123)
	2.947	5.437	1.458	-0.772
Black	3,706***	-0.00144	4,091***	-0.0150**
	(190.2)	(0.00151)	(712.7)	(0.00548)
	19.48	-0.956	5.740	-2.736
Location of MI: Anterior	6,674***	0.0931***	4,539***	0.0925***
	(183.2)	(0.00213)	(750.1)	(0.00735)
	36.43	43.75	6.051	12.58
Location of MI: Inferior	4,665***	0.133***	2,345***	0.126***
	(176.0)	(0.00204)	(707.7)	(0.00715)
	26.50	65.31	3.314	17.63
Location of MI: Right	5,770***	0.112***	3,600**	0.0983***
	(308.9)	(0.00340)	(1,130)	(0.0115)
	18.68	32.94	3.186	8.541
Location of MI: Subendocardial	6,108***	0.163***	2,450***	0.162***
	(160.2)	(0.00188)	(671.0)	(0.00665)
	38.12	87.10	3.651	24.28
Location of MI: Other	4,632***	0.0900***	2,966**	0.0829***
	(281.4)	(0.00295)	(1,075)	(0.00994)
	16.46	30.47	2.760	8.333
2nd Quintile ZIP Income	-274.3*	0.00286**	-728.4	0.00613
	(109.8)	(0.00106)	(384.5)	(0.00343)
	-2.498	2.702	-1.894	1.783
3rd Quintile ZIP Income	-595.2***	0.00507***	-1,051**	0.00869*
	(115.4)	(0.00110)	(389.1)	(0.00369)
	-5.160	4.602	-2.701	2.353
4th Quintile ZIP Income	-467.9***	0.00612***	-353.3	0.0150***
	(118.9)	(0.00118)	(430.8)	(0.00372)
	-3.935	5.185	-0.820	4.014
5th Quintile ZIP Income	-547.2***	0.0137***	-158.6	0.0299***
	(127.1)	(0.00125)	(429.0)	(0.00378)
	-4.307	10.92	-0.370	7.910
year2007	-26,189***	-0.0336***	-26,518***	-0.0260**
	(302.8)	(0.00292)	(1,029)	(0.00971)
	-86.50	-11.51	-25.78	-2.672
year2008	-23,583***	-0.0354***	-24,758***	-0.0276**
	(308.4)	(0.00293)	(1,025)	(0.00954)
	-76.48	-12.12	-24.16	-2.897
year2009	-23,073***	-0.0318***	-24,413***	-0.0252**
	(307.2)	(0.00295)	(1,036)	(0.00957)
	-75.10	-10.77	-23.56	-2.631
year2010	-23,369***	-0.0278***	-24,585***	-0.0217*
	(310.2)	(0.00289)	(1,041)	(0.00963)
	-75.34	-9.612	-23.62	-2.255
year2011	-21,513***	-0.0248***	-22,645***	-0.0212*
	(308.7)	(0.00293)	(1,035)	(0.00969)
	-69.70	-8.456	-21.88	-2.183

year2012	-21,271***	-0.0258***	-22,327***	-0.0215*
	(307.6)	(0.00291)	(1,026)	(0.00975)
	-69.15	-8.850	-21.76	-2.208
year2013	-20,732***	-0.0205***	-21,781***	-0.0190
	(304.2)	(0.00292)	(1,058)	(0.00996)
	-68.16	-7.033	-20.59	-1.905
year2014	-18,922***	-0.0259***	-20,078***	-0.0183
	(308.1)	(0.00290)	(1,031)	(0.00962)
	-61.41	-8.908	-19.47	-1.901
year2015	-15,180***	-0.0173***	-15,556***	-0.0164*
	(242.7)	(0.00237)	(887.4)	(0.00773)
	-62.54	-7.309	-17.53	-2.120
icd10 indicator (adjust for coding shift)	-18,021***	-0.00516*	-19,555***	-0.00488
	(263.5)	(0.00251)	(898.6)	(0.00833)
	-68.40	-2.057	-21.76	-0.586
Dementia (ICD10)	-5,828***	-0.145***	-5,062***	-0.129***
	(162.8)	(0.00239)	(795.3)	(0.0105)
	-35.79	-60.67	-6.365	-12.29
Dementia (ICD10 interacted)	1,676***	0.00604	921.6	-0.0231
	(337.4)	(0.00436)	(1,426)	(0.0190)
	4.968	1.385	0.646	-1.213
Medicare Advtg proportion (0 to 1)	3,086***	-0.00222	764.8	0.0213*
	(607.2)	(0.00448)	(1,169)	(0.00970)
	5.083	-0.495	0.654	2.194
Any month enrolled in Medicaid prior 6 months	209.3	-0.0306***	-126.9	-0.0276***
	(121.2)	(0.00119)	(450.5)	(0.00410)
	1.726	-25.65	-0.282	-6.746
Second Quintile HCC	7,643***	-0.0290***	6,371***	-0.0229***
	(112.8)	(0.000944)	(363.7)	(0.00287)
	67.73	-30.73	17.52	-7.974
Third Quintile HCC	11,416***	-0.0617***	10,499***	-0.0502***
	(134.0)	(0.00110)	(470.3)	(0.00392)
	85.18	-55.95	22.32	-12.80
Fourth Quintile HCC	15,307***	-0.117***	14,890***	-0.0966***
	(150.1)	(0.00118)	(526.4)	(0.00427)
	102.0	-99.05	28.29	-22.64
Fifth Quintile HCC	21,099***	-0.259***	22,197***	-0.233***
	(205.4)	(0.00135)	(677.3)	(0.00523)
	102.7	-192.4	32.78	-44.64
Constant	50,934***	0.813***	55,168***	0.810***
	(401.4)	(0.00364)	(1,274)	(0.0120)
	126.9	223.7	43.31	67.39
Observations	1,722,000	1,722,000	122,207	122,207
Number of provider	3,062	3,062	2,754	2,754
sigma_u	3597	0.0264	4683	0.0391

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05