Hospital Productivity and Input Misallocation

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 22, 2022

Preliminary – Please do not quote or distribute

Abstract

There is wide variation in productivity across hospitals, reflecting large differences in risk-adjusted health outcomes as well as expenditures, but less is known about the mechanisms by which some hospitals are more productive than others. We consider the contribution of *misallocation* in input choices – the underuse of effective inputs and overuse of ineffective ones – to explain the wide variation in productivity. Using 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, we find that misallocation accounts for as much as 18 percent of risk-adjusted survival rates across hospitals. Greater use of highly effective inputs, such as beta blockers, statins, and same-day stenting are predictive of highly-productive hospitals, while an excess of unnecessary scans and potentially fraudulent home health care are predictive of low-productivity hospitals.

We are grateful to the National Institute on Aging (PO1 AG19783) the National Institutes of Health (U01 AG046830), and the Stanford Institute for Economic Policy Research (SIEPR) for financial support, and John E. Wennberg, Douglas Staiger, Daniel Gottlieb, Peter Hull, Taylor Watson, Chad Syverson, Peter Klenow, and seminar participants at the Universities of Chicago and Wisconsin, UCLA, and Stanford for very helpful suggestions. Weiping Zhou and Ben Usadi provided extraordinary programming and data assistance.

I. Introduction

There are widespread differences across regions and hospitals in both intensity of care and health care outcomes.[1] These variations in spending and outcomes can be interpreted as hospital-based differences in total factor productivity (TFP) (Chandra et al., 2016a,b; Chandra and Staiger, 2020; Hull, 2020), and researchers have sought to explain these productivity differences by variation in information technology (Lee et al., 2013), management quality (Bloom et al., 2020, Tsai et al., 2015), physician skill (Abaluck et al., 2019; Chan, 2021; Chan et al., 2022; Mullainathan and Obermeyer, 2022; Burke et al., 2018) or process quality (Bradley et al., 2001). But one outstanding question remains: What are the mechanisms by which these inefficiencies affect productivity? Motivated by several classic studies from macroeconomics (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008) we suggest a new mechanism for productivity differences—the *misallocation* of inputs (treatments) across hospitals, whether through underuse of effective inputs, the overuse of ineffective inputs, or both. In their 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; with similar magnitudes found by Restuccia and Rogerson (2008). More recently Baqaee and Farhi (2020) estimated that eliminating misallocation would increase total U.S. factor productivity by 15 percent.[2]

In this paper, we measure the extent of misallocation in the health care sector using individual patient level data with accurately measured imputs and especially outputs: 30-day and 1-year survival following a heart attack (acute myocardial infarction, or AMI). We acknowledge that the assumption of

[1] For example, see Chandra and Staiger, 2020; Finkelstein et al., 2016, 2021; Fisher et al., 2003a,b, Wennberg, 2010; Baicker and Chandra, 2004a; Doyle et al., 2015, 2017; Deryugina and Molitor, 2018; Cutler et al., 2019; Romley et al., 2011, Rothberg et al., 2010, and many others.

[2] Bau (2019) estimated a high degree of misallocation (or mismatch) in Pakistan education and Bolhuis et al. (2021) in Indian agriculture, while Gollin and Udry (2021) found less evidence of misallocation given the presence of measurement error (a concern also raised in a different context by Haltiwanger et al. 2018), and heterogeneity across farmers. See Restuccia and Rogerson (2017) for an excellent survey.

profit-maximization may not always be satisfied in an industry so heavily dominated by not-for-profit firms, but the goal of minimizing production inefficiency – that is, saving the most lives for a given level of costs – is a relevant objective regardless of ownership status. Thus rather than testing whether marginal returns (or markups) are equalized across firms, as in much of the previous literature, we instead develop a new test for misallocation: Whether the incremental returns to inputs across firms or hospitals are different from zero, *conditional on total operating costs.*

The idea of misallocation has long been implicit in the health care literature, for example in the "Choosing Wisely" campaigns in health policy low-value health services (Colla et al., 2015) or the delineation of types of care into more or less effective (Wennberg et al., 2002; Chandra and Skinner, 2012). Our approach is to first draw on the model in Chandra et al. (2016a) for hospital measures of productivity, which are agnostic to the source of the productivity differences. We then expand the underlying model to allow input misallocation to affect productivity because of informational asymmetries, physician beliefs, or other types of distortions regarding highly effective or ineffective care. Next, under the null hypothesis of optimized inputs, fixed Medicare prices, and a common production function, we show that the choice of factor inputs should be orthogonal to the outcome, survival conditional on total expenditures. That is, including specific input choices on the right-hand side of a equation for survival yields information on the degree of input misallocation across hospitals. Under the null these inputs should have zero coefficients because in a well-optimized delivery system, the size of expenditures should be a sufficient statistic for all inputs.[3] One clear implication of the model is that in the presence of misallocation, conventional regression approaches that seek to estimate "the" association between expenditures and outcomes are highly problematic to interpret as the 'return

[3] 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 (2020); we consider this case in more detail below.

to more spending or intensity' since one must know whether the incremental expenditures were spent on effective or ineffective inputs.

We test this model using 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 115,282 "tourists," or people admitted to hospital far from home, as in Doyle (2011). In principal, there is a large set of inputs that we could include for our test, but to reduce the dimensionality of our enterprise we focus on three types of input choices that have been highlighted in the literature as as affecting productivity, and validated based on clinical evidence regarding their effectiveness in improving health outcomes. As in Chandra and Skinner (2012), we consider cost-effective "Category I" inputs such as statins and other effective drug treatments, 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 are both costly and unlikely, based on clinical evidence, to offer health benefits, such as excessive and potentially fraudulent post-acute care (Doyle et al., 2017; McKnight, 2006) and treatments deemed by professional physician groups to be wasteful (Colla, et al., 2015).[4] 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 heterogenous treatment effects hypothesized for "Category II" treatments. The key assumption to the use of these three types of inputs, instead of a vector of potentially hundreds of inputs, is that each principal component summarizes the types of inputs that get systematically misallocated. They are therefore not causal, but surrogates for a variety of smaller misallocations.

[4] These are closely related to the "effective," "preference-sensitive," and "supply-sensitive" categories of inputs as in Wennberg et al. (2002) and Wennberg (2010).

There are of course, several empirical challenges related to unmeasured patient health and endogenous patient selection that might look like productivity variation, but are not. 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 address this by considering detailed risk adjusters for just one acute event – a heart attack – and thus avoid conflating different types of diseases in the same regression analysis. We also are able to include extensive information on socioeconomic status, such as whether patient is eligible for low-income drug subsidies or Medicaid (as in Lewis et al., 2019) , the zip code poverty rate, 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, more importantly there may be unobservable differences in health behaviors such as smoking that vary by region. We address this potential problem in two ways; we first include zip-code fixed effects (as in 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 occur when ambulance services are loyal to a given hospital as in Doyle et al., 2015; 2017). Second, following Doyle (2011), we consider the subset of patients whose heart attack occurred far from home, and are therefore likely to be assigned to hospitals in a plausibly random fashion.

Quantitatively, we find considerable variability with regard to both total operating expenditures and 30-day and 1-year survival. For example, even after adjusting for zip-code fixed-effects. 30-day risk-adjusted survival rates across hospitals ranges from 80.4 percent in the 10^{th} percentile of hospitals to 89.0 percent in the 90^{th} percentile of hospitals; for the same patients, annual risk- and wage index-adjusted operating expenses per discharge range from $14,348 in the 10^{th} percentile to $37,897 in the

90th percentile. Furthermore, we find that hospital-level misallocation has a large impact on survival, conditional on operating expenses. In a hypothetical counterfactual, moving all hospitals to at least the 90th percentile with regard to efficient use of inputs alone – without changing overall productivity -- would lead to a 1.4 percentage point improvement in 30-day survival with little impact on overall expenses. Furthermore, the degree of misallocation is much lower for high-volume hospitals, suggesting one mechanism by which higher-volume hospitals are of better quality.

Less well understood is why some hospitals systematically use too little effective treatments, or overuse ineffective treatments; unlike the macroeconomic literature, we cannot appeal to distortions such as taxation or crony capitalism (although regulations could still play a role). Further research is needed to link measures of poor management or provider diagnostic skills to systematic misallocation as opposed to other sources of total factor productivity. Still, these empirical estimates suggest that what matters for health outcomes is not so much how much money is spent for treatment, but how it's spent.

II. The Model

Conventional regression estimates of the marginal returns to expenditures seek to estimate "the" slope of the production function, as shown in Figure 1; depending on the coefficient (and the corresponding calculated cost-effectiveness ratio), the results are viewed as informative as to whether the U.S. is spending too much or too little on health care. The problem arises when hospitals are not all on the same production function, or when productive inefficiency is present. To illustrate this, we allow for two scenarios in the model, one where there are differences in total factor productivity (as in Chandra et al., 2016a), and second, where there is misallocation.

Assume two hospitals, A and B, but where we consider operating expenditures (on the horizontal axis) and survival 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 (for example) 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 (and might be interpreted as being on 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, which researchers may interpret as implying that less 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, Hull, 2020, or Chandra et al., 2016a) or allow for the presence of misallocation, or both.

To formalize the intuition, we follow Chandra et al. (2016) by defining a hospital-level production function for patient i in hospital h as

$$Y_{ih} = \tilde{Y}_{ih} e^{\varepsilon_i}$$
$$where\ \tilde{Y}_{ih} = A_h \left[\prod_w Z_{iw}^{\omega_w} \right] \left[\prod_k X_{ik}^{\beta_k} \right] \qquad (1)$$

and the health outcome Y_{ih} is defined (for technical reasons) as the exponent of survival for (e.g.) 30 days (we also consider 1-year survival). 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.[5] For the moment, we also assume that the production parameters β_k are constant across hospitals; we expect diminishing returns to scale ($\sum_k \beta_k < 1$).

[5] 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.

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

Assume that the objective function is:

$$\Omega_h = \varphi_h \tilde{Y}_h + \phi_h \sum_k p_k X_{hk} - \left[\sum_k \pi_k p_k X_{hk} + F_h \right] \qquad (2)$$

where Ω_h is the objective function for hospital h.[6] The first term is the value to the hospital of the outcome, where φ_h translates survival into a dollar value (e.g., the social value of health improvements). The second term captures potential benefits in a fee-for-service system that occurs when Medicare billing rises; p_k is the reimbursement rate paid by Medicare for input k, which in turn is weighted by ϕ_h, the extent to which hospitals value the revenue. Finally, hospital costs are given by the third term, which expresses marginal costs as a ratio π_k of the Medicare reimbursement rate; generally hospitals view Medicare reimbursement rates as being equal to marginal cost (or average variable cost); as a simplification we can consider $\pi_k = 1$ for all k (although we relax that assumption below). Fixed costs F_h are in general positive, but one can also view this objective function as for Medicare alone, in which case "fixed" costs can represent revenue minus cost for commercial and Medicaid services; thus F_h could be negative. The first-order conditions of the objective function in (2) is:

$$\varphi_h \frac{\partial \tilde{Y}_h}{\partial X_k} - (\pi_k - \phi_h) p_k = 0 \qquad (3)$$

[6] Because we are focusing here on hospital-level averages, we drop the i subscripts.

where the first term in (3) is the marginal value to the hospital of an improvement in survival because of an increase in the kth input, while the second term captures both the marginal cost and the financial incentives arising from doing more. When φ_h is equal to the (allocatively efficient) value of survival to society and $\phi_h = 0$, the hospital attains both allocative and productive efficiency by spending up to the point where the marginal social value of the health benefit is equal to marginal cost across all k inputs.[7]

By contrast, in a fee-for-service setting where Medicare is viewed as the major revenue source, then $\phi_h = 1$; when Medicare prices are equal to marginal cost (so that $\pi_h = 1$), the hospital would simply maximize outcomes $\tilde{Y}$, placing it at the top of the production function (point H in Figure 1); varying ϕ_h from 0 to 1 would effectively trace out the production function between the allocatively efficient point and the peak of the production function. When the price exceeds marginal cost and φ_h is near zero, or when hospitals are pure profit-maximizers so that $\varphi_h = 0$, the point on the production curve could correspond to a point where spending is harmful to health (Point J in Figure 1).[8]

Suppose that regulations, informational barriers, or internal resistance to using effective treatments exists across hospitals; this is summarized by a "shadow" price μ_{hk} for the kth input; in the absence of pre-existing distortions (e.g., when $\phi_h = 0$) the shadow price can be either positive, leading to underuse, or negative, leading to overuse: [9]

[7] This takes the level of hospital productivity as given.

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

[9] See Díaz-Hernández et al. (2008) for a similar "shadow price" approach. More generally, in the presence of pre-existing distortions, shadow prices may nudge the hospital towards social efficiency, for example in the case of medical professionalism (McWilliams, 2020).

$$\varphi_h \frac{\partial \tilde{Y}_h}{\partial X_k} - (\pi_k - \phi_h) p_k = \mu_{hk} \qquad (4)$$

Why might these inefficiencies occur? For underuse, information about effective treatments may have been scarce because of high search costs (e.g., Skinner and Staiger, 2015), poor organizational or management structure (e.g., Bloom et al., 2014; McConnell et al., 2013; Tsai et al., 2015) the lack of leaders championing their use (Bradley et al., 2005), or systematic differences in training environments (Chan, 2020). More recently, Mullainathan and Obermeyer (2022), Abaluck et al. (2016; 2021), Chan et al. (2022) and others have identified substantial inefficiency arising from both underuse and overuse of diagnostic tools; if physicians were either consistently under-using or over-using, these inefficiencies would be captured in our measures of misallocation of hospital-level testing rates. These studies often find that physicians both overtest low-risk patients and under-test high-risk patients leading to aggregate rates that could be optimal; in this case, the inefficiency would be reflected in the overall productivity measure rather than in measures of misallocation.

Suppose there were strong financial incentives to "overuse" some (but not all) treatments because of incorrect physician beliefs about the marginal value of the input (e.g., Cutler et al., 2019); in that case, there would be overuse of the low-productivity high-profitability treatment.[10] For analytical purposes, it is easier to define the distortion as proportional to the price, so that

$$\lambda_{hk} \equiv \frac{\mu_{hk}}{p_k}$$

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

[10] Allocative inefficiency across hospitals might also occur if profit margins are positive for some treatments (e.g., cardiac surgery) but not for others (e.g., primary care visits), and these in turn are interacted with profit motives of hospitals.

assume temporal stability).

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

$$X_{hk} = X_{h1}\left[\frac{\beta_k}{p_k \beta_1}\right] \tag{5}$$

so that total expenditures M_h can be written solely as a function of X_{h1}:

$$M_h = \sum_k p_k X_{hk} = X_{h1}\left(\sum_k \left[\frac{\beta_k}{\beta_1}\right]\right) \tag{6}$$

Using the same first-order condition in (5), we can similarly express output solely as a function of the numeraire input X_{h1}

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

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:

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

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 note that all of these terms that depend solely on (fixed) 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, a test for the null of no misallocation.

More generally, Equation (9) can be rewritten to account for misallocation, differences in financial motivation, and 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[11] $\eta^*_{hk} = \left[\frac{(\pi_k + \lambda_{hk} - \phi_h)\beta_k}{(\pi_1 + \lambda_{h1} - \phi_h) p_k \beta_1} \right]$.

For example, recall that Hospital B in Figure 2 is assumed to experience lower output because Input 1 (the green technology) is under-used and Input 2 (the red technology) is overused. With prices equal to marginal cost, (π_k=1) the implicit shadow price for input 1 is positive, and for input 2 negative;

[11] We rule out cases in which financial incentives are such that $\pi_1 + \lambda_{hk} - \phi_h \leq 0$; given the Cobb-Douglas structure where firms are fully compensated to increase production, they will do so without limit.

the ratio $(1+\lambda_{h2}-\phi_h)/(1+\lambda_{h1}-\phi_h) < 0$. 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.

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, Doyle (2020) where treatment by cardiologists produces better outcomes, or in Chan et al. (2019) where some radiologists are more productively efficient in diagnosing their patients. In the context of our production function, this could correspond to a larger β for effective treatments in hospitals with more skilled or specialized physicians. For example, 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 the impact might be of varying values of β across inputs. Holding the output elasticity or the sum β^* (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.[12] We test for bias by stratifying hospitals by measured TFP, and test whether the high-TFP hospitals gain a higher marginal return in survival for the three input categories (or to $\beta*$ more generally).

III. Data

[12] 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 0.125 and 0.075 (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 less than a 1 percent increase in outcomes, to 2.080.

Medicare claims. We use 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 did not detect coding-induced changes 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.

Measuring Costs. There are two approaches to measuring total expenditures. The first simply adds up price-adjusted Medicare billing and represents individual-level measures of Medicare reimbursements. While Medicare reimbursement rates appear to be quite similar to average variable cost (MedPAC, 2019), they may not capture differences across hospitals in actual costs, for example financial support for high-quality physicians, nurses and allied health workers; this could be reflected by net subsidies ($F_h < 0$) and a cost structure that is subsidized by higher prices paid by commercial insurance. For this reason we use per-admission operating expenses by hospital by year from the Medicare Cost Reports because operating expenses capture all costs (and not just those allocated to Medicare patients).[13] The downside of these measures is that they are noisy and subject to accounting strategies designed to shift cost allocation from (or towards) operating expenses. For the regression analysis, cost measures have been adjusted for the labor portion of the wage index to account for geographic differences in labor rates (Sloan and Edmunds, 2012).

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

[13] We are grateful to Adam Sacarny for posting these data in a consistent way; http://sacarny.com/data/

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 (Black, Hispanic, Asian, Native American).[14] 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 a hospital referral region (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.[15]

Zip Codes. Another approach is to sweep out all neighborhood variation by including zip (Zone Improvement Plan) code fixed effects; this will of course absorb common health behaviors, average socioeconomic status, environmental health effects, and any other neighborhood factor. A deeper concern is that because hospital admissions are largely local, zip code fixed effects will also absorb the

[14] We use the Research Triangle Institute definitions for race and ethnicity. Unfortunately, Medicare data does not distinguish between ethnicity and race, so these definitions are mutually exclusive.
[15] There are 306 HRRs in the U.S. defined by the Dartmouth Atlas Project.

average (community) hospital quality. For example, if the fortunate residents of Fredonia are served by three very high quality hospitals, the zip code dummy variables will absorb those benefits, so the third-ranked hospital will appear to be below-average relative to other areas of the US. Thus the zip code-adjusted measures provide only a locally relative performance.

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. To avoid biases that might occur because of differences in underlying mortality rates across hospitals – those who die early are unlikely to be treated – we use utilization measures for those who remained alive for the 30-day period; these are then applied at the hospital level in individual regression analysis.[16]

To help organize the data analysis, we follow 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 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. 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.[17]

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

[17] 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.

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; Keeley et al., 2003; De Boer et al., 2002).

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 large 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 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 allow for diminishing returns we include dummy variables for hospitals if they are in the (weighted) top or bottom quartiles of each Category II treatment.

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.[18] As noted above,

[18] 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/

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*.[19]

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 two "Choosing Wisely" measures involving the use of "double CT" scans of the chest and abdomen, one with iodine contrast and the other without. These provide no additional clinical information and expose patients to double the radiation (Bogdanich and McGinty, 2011).[20] To aid in the interpretation of the coefficients, we also perform a principal components analysis for each of the 3 categories, with just the first principal component for each category as explanatory variables.

Econometric Specifications. To describe the basic patterns of the data, we begin by estimating conventional model specifications consistent with the previous literature where 30-day and 1-year survival are the dependent variables, with individual patient covariates, year effects, and (log) per-admission operating expenses from the Medicare Cost Reports are on the right-hand side. All standard errors are clustered at the hospital level.

We then allow for hospital-level random effects which, under the assumption of no misallocation, capture Hicks-neutral productivity differences (as in Chandra et al., 2016a). These

[19] 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."

[20] A CT scan of the chest involves roughly 70 times the radiation dose of an X-ray of the chest.

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), a result that was also found using chart-data by Chandra and Staiger (2020).

We next turn to the fully specified model that includes the hospital-specific random effects and hospital-specific Category I through Category III inputs, both the principal components of each category (to aid interpretation) as well as a regression model with all specific measures included separately. We can then use the estimated coefficients along with the observed values of X (evaluated at the mean of Z to abstract from patient comorbidities and year effects) to estimate what fraction of the total difference in productivity across hospitals is the consequence of input misallocation, and what fraction is the traditional productivity difference a_h.

IV. Results

Table 1A presents summary statistics for the entire sample of 1.7 million individuals with AMI, while Table 1B considers only the 115,282 out-of-HRR tourist sample. In the highest-cost quintiles, patients were slightly younger, less likely to be Hispanic, a slightly lower zip code income, and more likely to survive 30 days (86.9 percent versus 84.9 percent in the lowest quintile) and one year (72.1 percent versus 67.3 percent), but were otherwise quite similar with regard to patient characteristics. It is interesting to note that price-adjusted Medicare spending was higher ($53,821 versus $49,466) in the *lowest* operating cost quartile compared to the highest-cost quartile; more generally the correlation between per-patient Medicare expenditures and operating cost per patient measure is -0.13 ($p < .001$).

As shown in Table 1B, tourists were slightly younger, less likely to be Black or Hispanic, less likely to present with subendocardial (or less serious) AMIs, yet were somewhat healthier based on comorbidities at admission. Differences between the highest and lowest quintiles with regard to characteristics of patients were quite modest, as were differences in survival rates for both 30-day and 1-year horizons.

Table 2A presents regression 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 survival on the log of operating expenses per admission; with a coefficient of 0.0105 (s.e. 0.0016), it implies a positive association between spending and survival; the estimates are smaller in magnitude than Doyle et al. (2015) but similar to estimates in Hull (2020) using a similar dataset; the coefficient for one-year survival (0.0232, s.e. 0.0024) is larger in magnitude. While hospital-level Medicare expenditures is an entirely independent measure of inputs, the simple regression using such expenditures is again quite similar in magnitude, with a coefficient of .0092 (s.e. 0.0027). Estimates adjusting for zip code yield an estimate of .0099 (s.e. 0.0015) for 30-day mortality and 0.0211 (s.e. 0.0023) for 1-year mortality, Finally, Models 6 (30-day survival) and 7 (1-Year Survival) estimate a random-effects model along the lines of Chandra et al. (2016a); these estimate a very important role for differences across hospitals in TFP (standard deviation of 0.025 and 0.031, respectively) but the coefficients on log operating costs are no longer significant, suggesting that, conditional on hospital TFP, spending more does not lead to better health outcomes.

Table 2B replicates regressions from Table 2A for tourists, although the smaller sample size precludes the use of zip-code fixed effects. In Model 1 explaining 30-day survival, the coefficient on operating expenses (0.0038, s.e. 0.0033) is not significantly different from zero, although for 1-year

survival it is larger (0.0173, s.e. 0.0048); otherwise estimates are modest except for 1-year survival for the random-effects model (0.153, standard error 0.005).

In Figure 3A, we create risk-adjusted measures of log expenses and survival (based on the risk-adjustment model in Appendix Table A.1) to illustrate hospital-specific measures of expenses and survival, evaluated at the sample means of the risk-adjustment variables.[21] The sample is limited to hospitals with at least 400 AMI patients over 2007-17, so the wide range in both operating expenses and the range in survival of at least 10 percentage points is not the consequence of small sample variability.

The previous literature has focused on whether the slope of the regression line is positive or negative, but that hypothesis seems less relevant given the much larger differences in productivity; hospitals in the Northwest corner of the graph are achieving above-average outcomes at low costs, while those in the Southeast corner spend considerably more without commensurate improvements in outcomes. Figure 3B presents a similar graph for the tourist sample (given the smaller overall sample, we include all hospitals with at least 50 tourist AMIs during the period) using a similar risk-adjustment model (Appendix Table A.1). Like Figure 3A, there is wide variability in both spending and outcomes, with a very weak correlation between the two.

In Table 3A, we turn to our key results and examine for the allocation of inputs within the hospital using the first principal components for Categories I, II and III, alongside the risk adjustment tools described above, as well as the log of operating expenses per discharge adjusted for cost-of-living. Models 1 and 2 are random-effects, while Models 3 and 4 use zip-code fixed effects; all models reject the null that input choices are irrelevant to survival conditional on expenditures. The principal component score on Category I services is significant and positive across specifications, as hypothesized; the coefficient estimates predict as much as a 0.8 percentage point improvement in

[21] The weighted slope of the line in Figure 3A thus corresponds to the positive coefficient in Table 2A, Model 4.

survival based on a 2-standard-deviation increase in the Category I principal component; similarly a 2-standard-deviation increase in Category III is predicted to reduce survival by as much as 0.4 percentage points, holding expenses constant. The estimated linear impact of the Category II principal component is more sensitive to model specification and survival horizon.

We repeat the analyses on our sample of tourists (Table 3B), beginning with Models 1 and 2 using the full sample but only for the subset of tourist hospitals with sufficient sample size. 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 (Table 4). The top quartile of late PCI remains positive and significant in the tourist sample. Similar to the analyses on the full sample, the coefficient on the Category III principal component score is consistently negative and significant (Table 3B). In sum, the coefficients on the three principal components are remarkably stable across different specifications, for both tourists and the entire sample.

Table 4 shows similar regressions to Table 3, but with each input entered individually; the first two columns are for the full sample while the second two columns are for tourists only. Each of the Category I inputs remain positive or essentially zero (e.g., several have negative coefficients but are not significant at conventional levels); the results are dominated by use of primary PCI, with large effects on 30-day (0.113, s.e. .0032) and 1-year (0.155, s.e. .0042) survival; estimates are attenuated (0.089 for 30 day, 0.122 for 1-year) among tourists. While the estimated effect seems large relative to average survival, at least one earlier randomized trial for older AMI patients estimated survival gains of 0.15 (de Boer et al., 2002).

Recall that we include dummy variables if the hospital is in the bottom quartile, or the top quartile, of Category II inputs; a negative coefficient in the first quartile indicates underuse for this

group, while a negative coefficient in the fourth quartile indicates overuse. There is weak evidence for both over- and under-use of MRI and CT scans, although estimates are not precise. For 1-year survival (Model 2), there appears to be a generally negative association between ICU/CCU days and survival, with the converse true for late (after the admission date) PCI. None of the Category III measures are significant at conventional levels individually, but the factor model (Table 3A) demonstrates that they are jointly predictive.

How important is misallocation? To address this question, 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 18 percent of the variance. Moving all hospitals to the 90th percentile hospital with regard to misallocation is predicted to raise survival by 1.4 percentage points.

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 learn about more efficient care (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.

We next consider several sensitivity analyses. First, we restrict the sample solely to those admitted to hospital through the Emergency Room (Appendix Tables A.2A and A.2B); results are similar although standard errors are generally wider. Second, we consider only the more serious ST-

Elevated myocardial infarctions (STEMI) by excluding the 73 percent of the sample comprising sub-endocardial MIs which are unlikely to be STEMI infarctions. As shown in Appendix Tables A.3A and A.3B the regression estimates again are roughly similar in magnitude although again less likely to be significantly different from zero.

Another concern is that 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) have shown using data from the mid-1990s, hospitals tended to favor inputs in which they enjoyed a comparative advantage, thus potentially explaining why input choices were associated with outcomes. However, we do not find substantive differences in coefficients for the principal components of Category I, II, or III when the sample is split into terciles based on their estimated TFP productivity.

V. Conclusions

In this paper, we have estimated the importance of *misallocation* of inputs in explaining the sharp differences across hospitals in total factor productivity (TFP). Using a sample of 1.7 million Medicare patients with acute myocardial infarction (AMI) we have shown that there is both considerable variability in TFP across hospitals and a high degree of input misallocation. Rather than testing whether returns to inputs are equalized across firms, as in the previous literature, we develop instead a new test for misallocation: whether the returns to inputs within firms or hospitals are different from zero, conditional on total operating costs. We find that the combination of overusing ineffective or potentially harmful treatments, coupled with the underuse of effective care, accounts for nearly one-fifth of overall variation in total factor productivity (TFP) across hospitals, an estimate consistent with Baqaee and Farhi (2020) who estimate a 15 percent reduction in productivity arising from misallocation for the U.S.

as a whole. Reducing misallocation in the U.S. to the 90th percentile hospital is predicted to raise average 1-year survival rates by 1.4 percentage points.

Industry studies of misallocation have often harnessed assumptions about elasticities or other relevant parameters to make inferences about the degree of misallocation (e.g., Hsieh and Klenow, 2009; Haltiwanger et al., 2018). Our use of a well-defined outcome – survival following an index event, a heart attack requiring immediate hospital treatment – and well measured inputs with clinical evidence on effectiveness allows the estimation of misallocation with a minimum of assumptions about elasticities of outputs and inputs. That said, we also recognize several key assumptions of the model. One is that prices of inputs are constant across hospitals; otherwise firms may optimally adjust inputs to minimize costs. Yet Medicare prices are regulated and adjusted for cost-of-living differences across regions, and while health care providers (e.g., physicians, nurses, other health workers) may be more expensive in some areas than others, the opportunity for substitution is limited. Another key assumption is that hospitals are subject to the same Cobb-Douglas production function; however we have shown that the differences across hospitals in the production function would need to be large to affect our results.

The challenges for empirical work specific to health outcomes are twofold; when unobservable patient health is correlated with hospital admission, and when there is selection into hospitals. We have adopted several approaches to addressing the problem of unmeasured patient characteristics, including highly detailed individual covariates, the use of zip-code fixed effects, and a sample comprising tourists, but estimates of misallocation and the extent of TFP are similar across these different approaches. Furthermore, selection into hospitals in a Roy Model context (e.g., Chandra and Staiger, 2020) may be more closely related to skill conditional on inputs rather than input choice.

What are the research and policy implications on this work? Our estimated parameters are not causal, and probably emanate from a variety of hospital and physician characteristics. For example,

lowering inefficient CT scans may not have a direct impact on hospital productivity if such scans are a symptom of poor organizational structure or a lack of information about efficient practice (e.g., David et al., 2016) rather than a causal measure *per se*. Understanding the structural forces behind input misallocation through randomized trials or natural experiments would be useful for revealing incentive, informational or design levers that would reduce misallocation.[22]

Much of the research 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 estimating the slope of the production function for health care, thus addressing the question of whether U.S. health care is allocatively efficient or inefficient (Garber and Skinner, 2008). Our research cautions againt this line of inquiry—for it implicitly assumes that inputs are efficiciently utilized. If inputs are missallocated, the marginal return to highly effective care could be high even if the association between overall spending and outcomes is weak at best.

On the policy dimension, 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 with regard to a reduction in misallocation 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 that indirectly fund other activities, or with regard to understanding how specific treatments affect their financial bottom line (Kaplan and Witkowski, 2014). 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

[22] Restuccia and Rogerson (2017) distinguish between direct approaches to estimating misallocation by focusing on specific plausibly exogenous policy changes such as a tax hike, and indirect approaches which instead attempts to measure, as we do, the extent of misallocation without identifying the specific sources.

proven effectiveness of their value. Despite these caveats, a policy focus on reducing misallocation in health care appears promising given the potential for improving survival at little or no additional cost.

REFERENCES

Abaluck, Jason, Leila Agha, David C. Chan Jr, Daniel Singer, and Diana Zhu, 2021. "Fixing Misallocation with Guidelines: Awareness vs. Adherence." NBER Working Paper No. 27467.

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, ans Arjun Venkatesh, 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106(12): 3730-64.

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.

Baqaee, David R. and Emmanuel Farhi, 2020. "Productivity and Misallocation in General Equilibrium" *Quarterly Journal of Economics*, 135(1): 105–163.

Bau, Natalie, 2019. "Estimating an equilibrium model of horizontal competition in education." Working Paper, UCLA.

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.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen, 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals," *The Review of Economic Studies*, 82, 2, 457–489.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen, 2020. "Healthy Business? Managerial Education and Management in Health Care." *Review of Economics and Statistics* 102(3): 506-517.

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

Bolhuis, Marijn A., Swapnika R. Rachapalli, and Diego Restuccia, 2021. "Misallocation in Indian Agriculture." NBER Working Paper No. 29363.

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, 2021. "Influence and Information in Team Decisions: Evidence from Medical Residency," *American Economic Journal: Economic Policy* 13(1): 106-37

Chan Jr, David C., Matthew Gentzkow, and Chuan Yu, 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *Quarterly Journal of Economics* (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 (Papers and Proceedings)* 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.

Currie, Janet, and W. Bentley MacLeod, 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics* 35(1): 1-43.

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 Boer, Menko-Jan, Jan-Paul Ottervanger, Arnoud WJ van't Hof, Jan CA Hoorntje, Harry Suryapranata, et al., 2002. "Reperfusion Therapy in Elderly Patients with Acute Myocardial Infarction: A Randomized Comparison of Primary Angioplasty and Thrombolytic Therapy." *Journal of the American College of Cardiology* 39(11): 1723-1728.

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): 1928-37.

Deryugina, Tatyana, and David Molitor, 2020. "Does When you Die Depend on Where You Live? Evidence from Hurricane Katrina" *American Economic Review* 110(11): 3602-33.

Doyle Jr, Joseph J. 2020. "Physician Characteristics and Patient Survival: Evidence from Physician Availability.". National Bureau of Economic Research Working Paper No. 27458.

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.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review* 111(8): 2697-2735.

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.

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*,” Brookings Papers on Economic Activity*, Fall 2019:

Gollin, Douglas, and Christopher Udry, 2021. "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture." *Journal of Political Economy* 129(1): 1-80.

Hartwell, Debbie, J. Colquitt, Emma Loveman, Andrew J. Clegg, Håkan Brodin, Norman Waugh, Pamela Royle, Peter Davidson, Luke Vale, and Lynda MacKenzie, 2005. "Clinical effectiveness and cost-effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation." In NIHR Health Technology Assessment programme: Executive Summaries. NIHR Journals Library.

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, 2010. Relationship between early physician follow-up and 30-day readmission among Medicare beneficiaries hospitalized for heart failure. *JAMA* 303(17):1716-22.

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

Hull, Peter, 2020. "Estimating hospital quality with quasi-experimental data." Working Paper, Brown University.

Kaplan, Robert S., and Mary L. Witkowski, 2014. "Better Accounting Transforms Health Care Delivery." *Accounting Horizons* 28, 2: 365-383.

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

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.

Lee, Jinhyung, Jeffrey S. McCullough, and Robert J. Town, 2013. "The impact of health information technology on hospital productivity." *The RAND Journal of Economics* 44.3: 545-568.

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.

McWilliams, J. Michael, 2020. "Professionalism Revealed: Rethinking Quality Improvement in the Wake of a Pandemic." *NEJM Catalyst Innovations in Care Delivery* 1, no. 5.

Medicare Payment Advisory Commission (MedPAC), 2019. Report to the Congress: Medicare Payment Policy, Washington, D.C. http://www.medpac.gov/docs/default-source/reports/mar19_medpac_entirereport_sec.pdf

Mullainathan, Sendhil and Ziad Obermeyer, 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *Quarterly Journal of Economics* (forthcoming).

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, 2010. Little evidence of correlation between growth in health care spending and reduced mortality. *Health Affairs* (Millwood) 29(8): 1523-31.

Sharma G, Kuo YF, Freeman JL, Zhang DD, Goodwin JS, 2010. 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, and Douglas Staiger, 2015. “Technology Diffusion and Productivity Growth in Health Care.” *Review of Economics and Statistics*.

Sloan, Frank A., and Margaret Edmunds, eds. 2012. Geographic Adjustment in Medicare Payment: Phase I: Improving Accuracy. National Academies Press.

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.

Tsai, Thomas C., Ashish K. Jha, Atul A. Gawande, Robert S. Huckman, Nicholas Bloom, and Raffaella Sadun, 2015. "Hospital Board and Management Practices Are Strongly Related to Hospital Performance on Clinical Quality Metrics." *Health Affairs* 34, 8, 1304–1311.

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

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

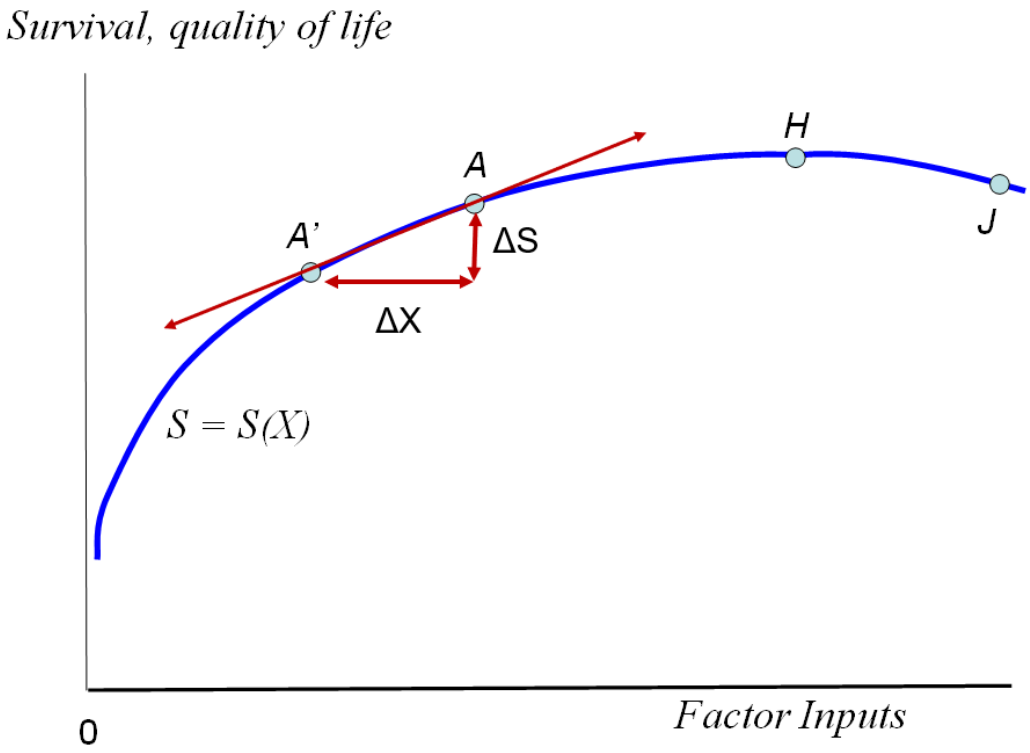


Figure 2: Sources of Differences in Outcomes: Total Factor Productivity and Input Misallocation

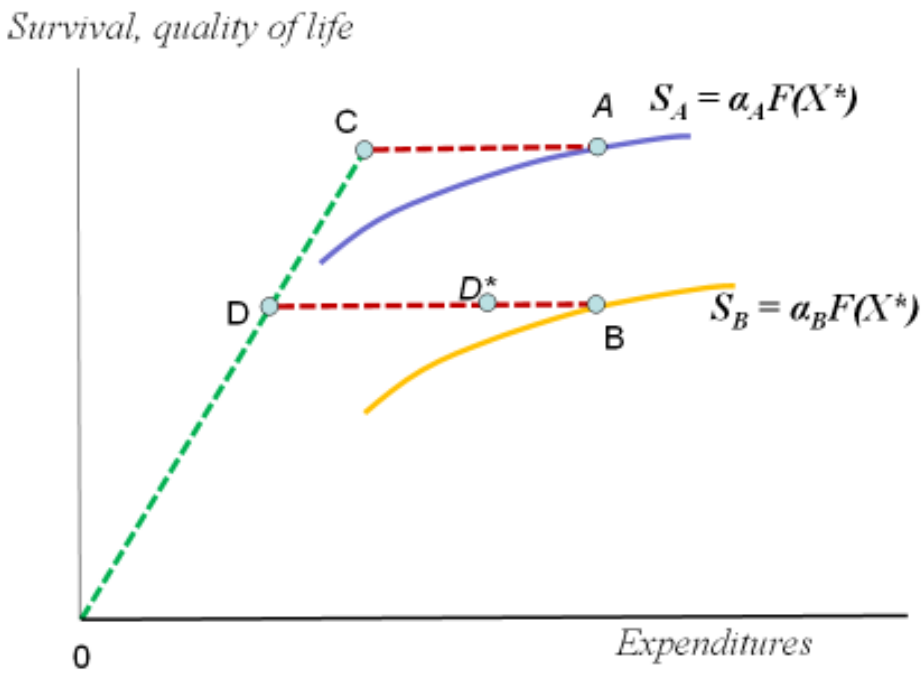


Figure 3a: Association between Risk-Adjusted Operating Expenses and Survival: 2007-2017

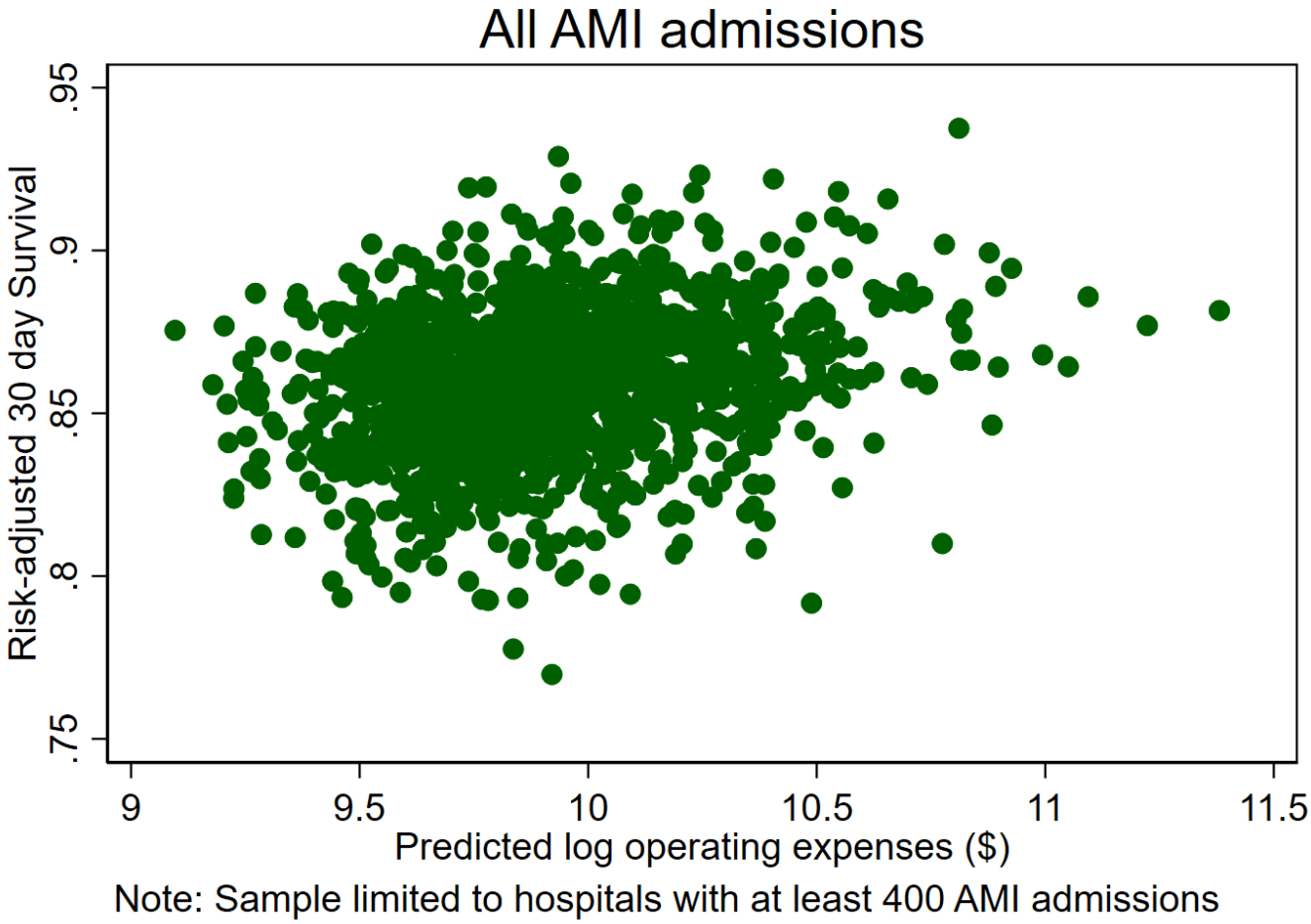
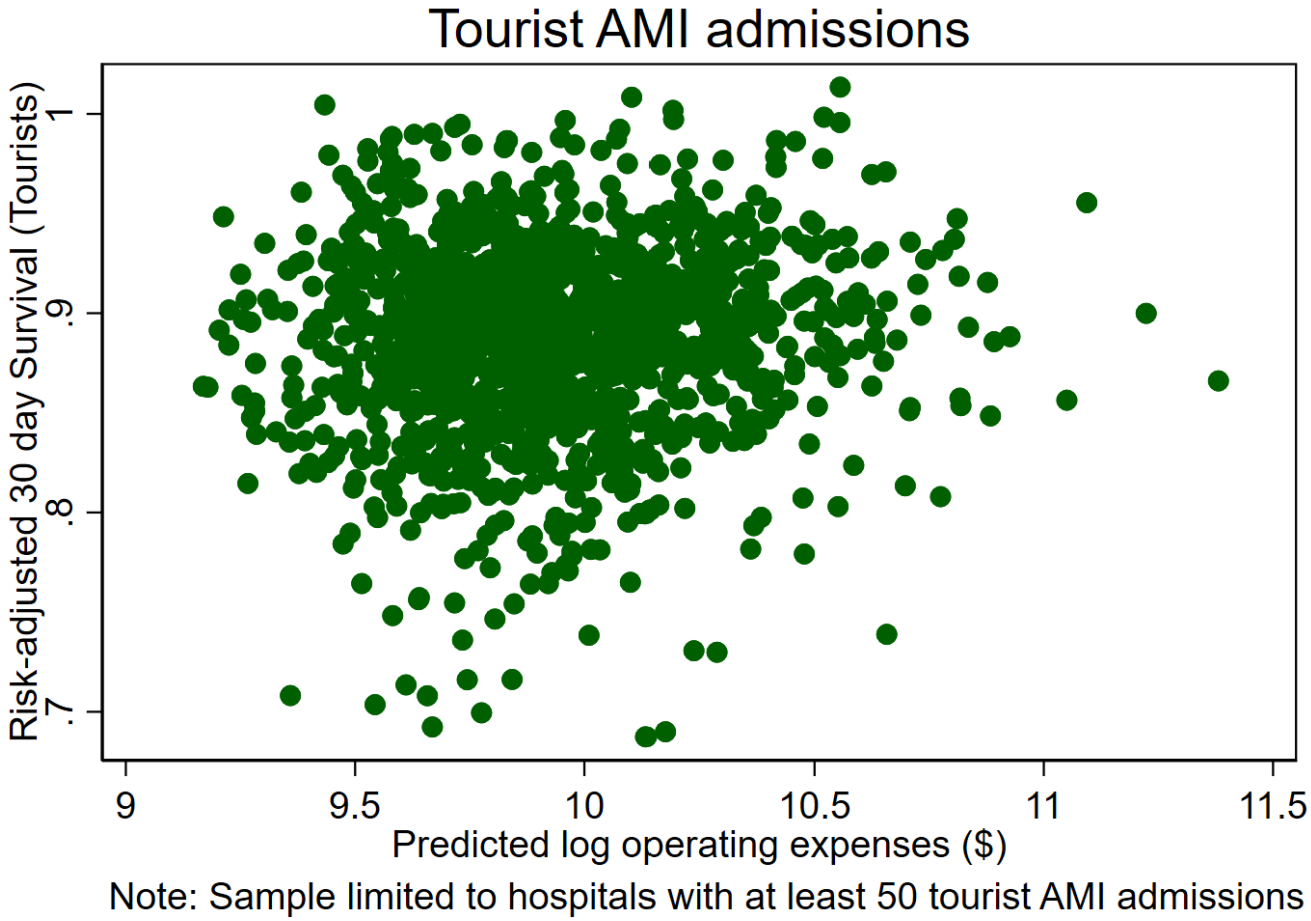
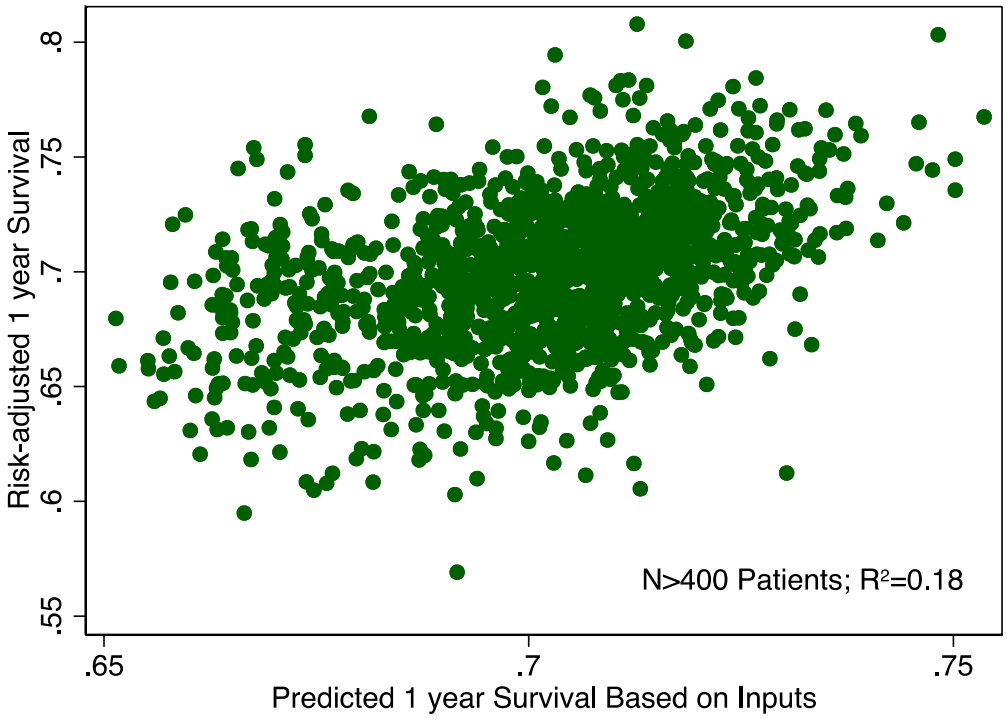


Figure 3b: Association between Risk-Adjusted Operating Expenses and Mortality for Tourists: 2007-2017



Note: Sample limited to hospitals with at least 50 tourist AMI admissions

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



b

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

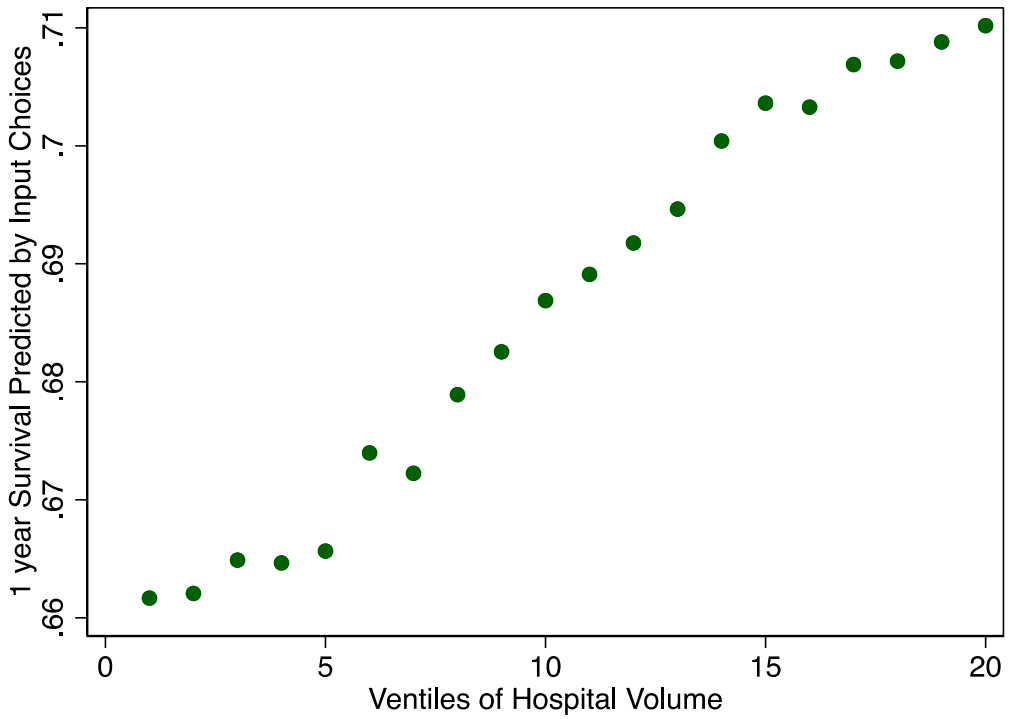


Table 1A: Descriptive Characteristics of Medicare Beneficiaries with AMI: 2007-17: Entire Sample

	FULL:ALL	FULL: TOP COST HOSPITAL QUINTILE	FULL: BOTTOM COST HOSPITAL QUINTILE
	mean (sd)	mean (sd)	mean (sd)
Survival			
1 year	0.701 (0.458)	0.721 (0.449)	0.673 (0.469)
30 day	0.861 (0.346)	0.869 (0.338)	0.849 (0.358)
Log operating expenses per discharge	9.928 (0.338)	10.403 (0.270)	9.548 (0.197)
Price Adjusted Spending			
1 year	50872 (47272)	49466(46926)	53821 (49765)
30 day	26368 (21351)	26408 (22060)	26595 (21002)
age at index admitted date	78.254 (8.368)	77.861 (8.310)	78.764 (8.439)
female	0.483 (0.500)	0.466 (0.499)	0.499 (0.500)
Comorbidities			
Peripheral Vascular Disease	0.081 (0.273)	0.082 (0.274)	0.079 (0.269)
Chronic Non-Asthmatic Lung Disease	0.178 (0.383)	0.168 (0.374)	0.188 (0.390)
dementia	0.039 (0.193)	0.034 (0.182)	0.045 (0.207)
Chronic Renal Failure	0.189 (0.392)	0.195 (0.396)	0.187 (0.390)
Cancers (various)	0.050 (0.218)	0.054 (0.226)	0.050 (0.217)
Metastatic Solid Tumor	0.013 (0.111)	0.013 (0.115)	0.013 (0.113)
Congestive Heart Failure	0.398 (0.489)	0.391 (0.488)	0.410 (0.492)
HIV	0.001 (0.023)	0.001 (0.028)	0.001 (0.024)
Hemi/Paraplegia	0.003 (0.052)	0.003 (0.053)	0.003 (0.051)
Liver Disease	0.005 (0.070)	0.005 (0.073)	0.005 (0.071)
Diabetes	0.274 (0.446)	0.272 (0.445)	0.279 (0.449)
Peptic Ulcer Disease	0.008 (0.090)	0.008 (0.091)	0.008 (0.091)
Rheumatologic Disease	0.017 (0.130)	0.018 (0.134)	0.016 (0.127)

HCC score 6m before index admission	1.344 (1.091)	1.311 (1.061)	1.435 (1.166)
Race			
Native American	0.005 (0.070)	0.007 (0.082)	0.002 (0.046)
Hispanic	0.015 (0.121)	0.009 (0.094)	0.030 (0.171)
Other Race/Ethnicity	0.011 (0.102)	0.012 (0.109)	0.011 (0.103)
Asian	0.012 (0.111)	0.014 (0.116)	0.015 (0.123)
Black	0.073 (0.261)	0.078 (0.268)	0.077 (0.266)
White	0.884 (0.320)	0.881 (0.324)	0.865 (0.342)
AMI subtype			
Anterior	0.085 (0.279)	0.090 (0.286)	0.074 (0.262)
Inferior	0.103 (0.304)	0.109 (0.311)	0.089 (0.284)
Right-Heart	0.012 (0.109)	0.015 (0.123)	0.009 (0.094)
Subendocardial	0.731 (0.443)	0.719 (0.449)	0.752 (0.432)
Other sites	0.019 (0.135)	0.018 (0.134)	0.017 (0.130)
Unspecified	0.050 (0.218)	0.049 (0.215)	0.059 (0.236)
Zip Code Median Income	55431(25357)	54906 (25124)	56112 (26098)
Medicare Advtg proportion (0 to 1) (for zipcode of bene residence)	0.261 (0.128)	0.253 (0.132)	0.283 (0.128)
any month enrolled in Medicaid 6m prior to admission	0.166 (0.372)	0.164 (0.371)	0.192 (0.394)
Observations	1,730,535	278,138	323,827

Table 1B: Descriptive Characteristics of Medicare Beneficiaries with AMI: 2007-17: Tourist Sample

	TOURISTS: ALL	TOURISTS: TOP COST HOSPITAL QUINTILE	TOURISTS: BOTTOM COST HOSPITAL QUINTILE
	Mean(sd)	Mean(sd)	Mean(sd)
Survival			
1 year	0.783 (0.412)	0.799 (0.401)	0.771 (0.420)
30 day	0.891 (0.311)	0.897 (0.304)	0.889 (0.314)
Log operating expenses per discharge	9.961 (0.351)	10.398 (0.276)	9.560 (0.197)
Price Adjusted Spending			
1 year	50951(47939)	50972 (49138)	52309 (49274)
30 day	29085 (23923)	29836(25185)	28720 (23467)
age at index admitted date	76.042 (7.844)	75.550 (7.599)	76.664 (8.021)
female	0.413 (0.492)	0.397 (0.489)	0.414 (0.493)
Comorbidities			
Peripheral Vascular Disease	0.074 (0.262)	0.080 (0.271)	0.066 (0.249)
Chronic Non-Asthmatic Lung Disease	0.150 (0.358)	0.145 (0.353)	0.152 (0.359)
dementia	0.025 (0.156)	0.021 (0.145)	0.031 (0.173)
Chronic Renal Failure	0.141 (0.348)	0.141 (0.348)	0.136 (0.342)
Cancers (various)	0.040 (0.197)	0.043 (0.202)	0.042 (0.201)
Metastatic Solid Tumor	0.008 (0.088)	0.008 (0.090)	0.008 (0.091)
Congestive Heart Failure	0.341 (0.474)	0.338 (0.473)	0.348 (0.476)
HIV	0.000 (0.020)	*	*
Hemi/Paraplegia	0.003 (0.051)	0.003 (0.055)	0.002 (0.048)
Liver Disease	0.004 (0.064)	0.005 (0.070)	0.005 (0.068)
Diabetes	0.255 (0.436)	0.252 (0.434)	0.260 (0.439)
Peptic Ulcer Disease	0.008 (0.089)	0.009 (0.094)	0.008 (0.090)
Rheumatologic Disease	0.014 (0.117)	0.014 (0.119)	0.012 (0.110)
HCC score 6m before index admission	1.082 (0.929)	1.099 (0.954)	1.105 (0.953)

Race			
Native American	0.008 (0.090)	0.008 (0.090)	0.005 (0.067)
Hispanic	0.011 (0.106)	0.007 (0.085)	0.022 (0.146)
Other Race/Ethnicity	0.011 (0.105)	0.011 (0.105)	0.012 (0.110)
Asian	0.010 (0.102)	0.011 (0.106)	0.013 (0.112)
Black	0.055 (0.227)	0.042 (0.200)	0.059 (0.236)
White	0.904 (0.294)	0.920 (0.271)	0.889 (0.314)
AMI subtype			
Anterior	0.109 (0.312)	0.112 (0.315)	0.097 (0.297)
Inferior	0.137 (0.343)	0.137 (0.344)	0.123 (0.328)
Right-Heart	0.016 (0.126)	0.021 (0.142)	0.011 (0.105)
Subendocardial	0.679 (0.467)	0.678 (0.467)	0.703 (0.457)
Other sites	0.020 (0.140)	0.018 (0.134)	0.019 (0.136)
Unspecified	0.039 (0.194)	0.034 (0.181)	0.047 (0.211)
Zip Code Median Income	54926 (26171)	54331 (25766)	59187 (27943)
Medicare Advtg proportion (0 to 1) (for zipcode of bene residence)	0.262 (0.131)	0.260 (0.129)	0.265 (0.132)
any month enrolled in Medicaid 6m prior to admission	0.128 (0.334)	0.118 (0.323)	0.125 (0.331)
Observations	115,282	25,357	20,993

* Suppressed (cell size < 11)

Table 2A: Association between Operating Expenses or Expenditures and Survival: Full Sample

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Log Operating Expenses per Admission	0.0105	0.0232		0.0099	0.0211	0.0009	0.0077
	(0.00161)	(0.00245)		(0.00150)	(0.00230)	(0.00148)	(0.00206)
Log 30-Day Medicare Spending			0.00924				
			(0.00273)				
Outcome	30-Day Survival	1-Year Survival	30-Day Survival	30-Day Survival	1-Year Survival	30-Day Survival	1-Year Survival
Zip code fixed effects				X	X		
Hospital Random Effects						X	X
Observations	1,729,854	1,652,183	1,730,535	1,729,854	1,652,183	1,729,854	1,652,183
N of Hospitals	2,912	2,901	2,912	2,912	2,901	2,912	2,901
R-squared	0.085	0.182	0.085	0.105	0.202	NA	NA
Sigma_u	NA	NA	NA	NA	NA	0.0250	0.0314

Note: All Regressions include full set of risk-adjustment covariates.

Table 2B: Association between Operating Expenses or Expenditures and Survival: Tourists

Variables	Model 1 Tourists	Model 2 Tourists	Model 3 Tourists	Model 4 Tourists	Model 5 Tourists
Log Operating Expenses per Admission	0.0038	0.0173		0.0030	0.0153
	(0.00335)	(0.00476)		(0.00324)	(0.00471)
Log 30-Day Medicare Spending			0.00240		
			(0.00811)		
Outcome	30-Day Survival	1-Year Survival	30-Day Survival	30-Day Survival	1-Year Survival
Hospital Random Effects				X	X
Observations	115,263	109,514	115,282	115,263	109,514
N of Hospitals	1,290	1,289	1,290	1,290	1,289
R-squared	0.070	0.150	0.070	NA	NA
Sigma_u	NA	NA	NA	0.0194	0.0276

Note: Hospitals limited to those with at least 20 tourists during 2007-17. Regressions include full set of risk-adjustment covariates. We do not include zip Code fixed effects in tourist estimates because of lack of sample size in the tourist cohort.

Table 3A: Regressions Estimates Survival Allowing for Misallocation: Entire Sample

VARIABLES	Model 1 30-Day Survival	Model 2 1-Year Survival	Model 3 30-Day Survival	Model 4 1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	0.00106	0.00399	0.00867	0.0176
	(0.00149)	(0.00209)	(0.00150)	(0.00230)
PCA prediction for input1 component1	0.000934	0.00277	0.00189	0.00472
	(0.000194)	(0.000243)	(0.000205)	(0.000272)
PCA prediction for input2 component1	0.00145	-0.00445	0.000189	-0.00479
	(0.000458)	(0.000598)	(0.000512)	(0.000735)
PCA prediction for input3 component1	-0.000893	-0.00222	-0.000868	-0.00218
	(0.000405)	(0.000478)	(0.000437)	(0.000570)
Hospital Random Effects	X	X		
Zip Code Fixed Effects			X	X
Observations	1,729,854	1,652,183	1,729,854	1,652,183
R-squared			0.105	0.202
Sigma_u	0.0249	0.0311		

All regressions incude the full set of risk-adjustment covariates. Adjusted for wage index.

Table 3B: Regressions Estimates for Survival Allowing for Misallocation: Tourist Hospitals and Tourists

VARIABLES	Tourist Hospitals Model 1 30-Day Survival	Tourist Hospitals Model 2 1-Year Survival	Tourists Model 3 30-Day Survival	Tourists Model 4 1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	0.00952	0.0179	0.00123	0.00976
	(0.00166)	(0.00233)	(0.00329)	(0.00481)
PCA prediction for input1 component1	0.000858	0.00237	0.00135	0.00356
	(0.000209)	(0.000264)	(0.000604)	(0.000792)
PCA prediction for input2 component1	0.00108	-0.00470	-0.000185	-0.00333
	(0.000555)	(0.000746)	(0.00122)	(0.00168)
PCA prediction for input3 component1	-0.000717	-0.00189	-0.00195	-0.00352
	(0.000432)	(0.000526)	(0.00106)	(0.00134)
Hospital Random Effects	X	X	X	X
Observations	1,354,078	1,291,346	115,263	109,514
Sigma_u	0.0154	0.0205	0.0190	0.0264

Restricted to hospitals with >20 tourist AMI admissions.
All regressions incude the full set of risk-adjustment covariates. Adjusted for wage index.

Table 4: Regressions Estimates for Survival Allowing for Misallocation with a Full Set of Variables: Full Sample and Tourists

	Model 1 Full	Model 2 Full	Model 3 Tourists	Model 4 Tourists
Outcome	30 day survival	1 year survival	30 day survival	1 Year survival
Variables				
Log Hospital Operating Expenses	0.00111	0.00468	-0.00373	0.00228
	(0.00142)	(0.00190)	(0.00338)	(0.00476)
Cat1: Beta-blocker (hospital)	0.00127	0.00941	-0.00907	0.00724
	(0.00304)	(0.00396)	(0.0106)	(0.0135)
Cat1: Statins (hospital)	-0.000114	0.00623	0.0131	0.0291
	(0.00285)	(0.00376)	(0.00982)	(0.0132)
Cat1: Ace Inhibitor/ARB (hosp.)	-0.00525	-0.00354	-0.00520	-0.0140
	(0.00298)	(0.00391)	(0.00961)	(0.0126)
Cat1: Early PCI (hospital)	0.113	0.155	0.0890	0.122
	(0.00319)	(0.00419)	(0.0109)	(0.0143)
Cat1: Visits within 14 days of discharge	-0.00167	-0.00212	-0.00902	0.00509
	(0.00518)	(0.00641)	(0.0139)	(0.0186)
Cat2: Late PCI (bottom quartile)	-0.00762	-0.00980	-0.00507	-0.00557
	(0.000738)	(0.000914)	(0.00250)	(0.00326)
Cat2: Late PCI (top quartile)	0.00779	0.0103	0.00745	0.00956
	(0.000707)	(0.000919)	(0.00235)	(0.00304)
Cat2: # Doctors seen (bottom quartile)	-0.000752	-0.00248	0.000217	-0.00428
	(0.000871)	(0.00115)	(0.00259)	(0.00338)
Cat2: # Doctors seen (top quartile)	0.000520	0.00161	0.00413	0.00321
	(0.000879)	(0.00117)	(0.00269)	(0.00372)
Cat2: # Scans (bottom quartile)	-0.00315	0.00136	-0.000422	0.00295
	(0.000734)	(0.000933)	(0.00246)	(0.00313)
Cat2: # Scans (top quartile)	-0.000585	-0.00751	0.000395	-0.00400
	(0.000822)	(0.00104)	(0.00255)	(0.00343)
Cat2: SNF Days (bottom quartile)	-0.00578	4.89e-05	-0.000164	0.00310
	(0.000740)	(0.000929)	(0.00226)	(0.00302)
Cat2: SNF Days (top quartile)	0.00663	-0.000185	-0.000351	-0.00489
	(0.000813)	(0.00106)	(0.00284)	(0.00378)
Cat2: ICU/CCU days (bottom quartile)	0.000517	0.00332	0.000304	0.00324
	(0.000921)	(0.00116)	(0.00268)	(0.00352)
Cat2: ICU/CCU days (top quartile)	-0.000523	-0.00517	-0.00277	-0.00316
	(0.000871)	(0.00114)	(0.00264)	(0.00337)

Cat3: Fraction of 30 day Home Health spending outliers	-0.00334	-0.00126	0.00157	-0.0137
	(0.00496)	(0.00629)	(0.0164)	(0.0204)
Cat3: Abdomen CT measure from Hospital Compare	-4.29e-06	-1.55e-05	-3.90e-05	-0.000121
	(3.21e-05)	(3.91e-05)	(8.23e-05)	(0.000110)
Cat3: Thorax CT measure from Hospital Compare	-4.15e-05	-0.000199	-0.000202	-0.000239
	(5.71e-05)	(6.53e-05)	(0.000164)	(0.000193)
Hospital Random Effects	X	X	X	X
Observations	1,729,854	1,652,183	115,263	109,514
Sigma_u	0.0233	0.0280	0.0190	0.0258

All regressions incude hospital random effects; adjusted for wage index.

Appendix Table A.1: Risk Adjustment Models for Log Operating Expenses and 1-Year Outcomes

	Model 1	Model 2	Model 3	Model 4	Model 5
Outcome	Log Operating Expenses	30 day Survival	1 Year Survival	30 day Survival	1 Year Survival
Sample	Full	Full	Full	Tourists	Tourists
Peripheral Vascular Disease	(0.00238)	0.0184	0.0112	0.0150	-0.000987
	2.371	(0.000931)	(0.00133)	(0.00322)	(0.00468)
	-0.0122	19.74	8.387	4.652	-0.211
Chronic Non-Asthmatic Lung Disease	(0.00215)	0.00913	-0.0325	0.00866	-0.0285
	-5.683	(0.000749)	(0.00101)	(0.00257)	(0.00377)
	0.0118	12.20	-32.18	3.369	-7.567
Chronic Renal Failure	(0.00176)	-0.00692	-0.0579	-0.00710	-0.0560
	6.704	(0.000783)	(0.00104)	(0.00280)	(0.00370)
	0.0103	-8.828	-55.53	-2.537	-15.12
Cancers	(0.00178)	-0.0128	-0.0698	0.00421	-0.0485
	5.812	(0.00135)	(0.00177)	(0.00472)	(0.00647)
	0.00965	-9.451	-39.33	0.891	-7.487
Cancer: Metastatic Solid Tumor	(0.00293)	-0.175	-0.346	-0.180	-0.352
	3.294	(0.00340)	(0.00334)	(0.0161)	(0.0175)
	-0.000617	-51.60	-103.5	-11.17	-20.12
Congestive Heart Failure	(0.00147)	-0.0476	-0.120	-0.0411	-0.103
	-0.420	(0.000725)	(0.000911)	(0.00226)	(0.00292)
	0.0448	-65.62	-131.9	-18.15	-35.24
HIV	(0.0143)	0.0261	0.00784	0.0401	0.0270
	3.140	(0.00905)	(0.0143)	(0.0399)	(0.0573)
	0.00687	2.883	0.550	1.005	0.472
Hemi/Paraplegia	(0.00494)	-0.119	-0.170	-0.136	-0.194
	1.391	(0.00634)	(0.00717)	(0.0245)	(0.0290)
	0.00982	-18.79	-23.71	-5.549	-6.711
Llver Disease	(0.00454)	-0.0244	-0.0868	-0.0259	-0.0796
	2.165	(0.00373)	(0.00518)	(0.0139)	(0.0201)
	-0.00337	-6.546	-16.76	-1.858	-3.955
Diabetes	(0.00126)	0.0395	0.0429	0.0334	0.0347
	-2.682	(0.000611)	(0.000845)	(0.00207)	(0.00290)
	0.000941	64.63	50.79	16.15	11.95
Peptic Ulcer Disease	(0.00376)	0.0176	0.000248	0.00397	-0.0122
	0.250	(0.00267)	(0.00375)	(0.0101)	(0.0135)
	0.00947	6.612	0.0660	0.391	-0.902
Rheumatologic Disease	(0.00217)	0.0334	0.0266	0.0280	0.0233
	4.369	(0.00176)	(0.00250)	(0.00645)	(0.00955)
	-0.00600	18.91	10.66	4.343	2.436
female age 65-69	(0.00141)	0.00150	0.00793	0.00110	0.00734
	-4.269	(0.000969)	(0.00134)	(0.00304)	(0.00395)
	-0.00389	1.546	5.934	0.363	1.856

male age 70-74	(0.00126)	-0.0121	-0.0206	-0.0124	-0.0199
	-3.099	(0.000843)	(0.00112)	(0.00257)	(0.00342)
	-0.00985	-14.35	-18.38	-4.806	-5.809
female age 70-74	(0.00163)	-0.0109	-0.0151	-0.0144	-0.0236
	-6.059	(0.000978)	(0.00135)	(0.00339)	(0.00467)
	-0.00686	-11.10	-11.19	-4.234	-5.055
male age 75-79	(0.00179)	-0.0298	-0.0542	-0.0238	-0.0447
	-3.824	(0.000972)	(0.00135)	(0.00325)	(0.00416)
	-0.0111	-30.68	-40.23	-7.328	-10.75
female age 75-79	(0.00204)	-0.0246	-0.0414	-0.0250	-0.0369
	-5.435	(0.00103)	(0.00136)	(0.00340)	(0.00454)
	-0.00765	-23.81	-30.38	-7.339	-8.141
male age 80-84	(0.00257)	-0.0563	-0.106	-0.0520	-0.0935
	-2.980	(0.00121)	(0.00160)	(0.00421)	(0.00567)
	-0.0145	-46.59	-66.40	-12.36	-16.48
female age 80-84	(0.00260)	-0.0478	-0.0855	-0.0418	-0.0797
	-5.555	(0.00112)	(0.00148)	(0.00427)	(0.00595)
	-0.0102	-42.83	-57.85	-9.801	-13.41
male age 85-89	(0.00326)	-0.0993	-0.183	-0.0884	-0.176
	-3.136	(0.00149)	(0.00189)	(0.00585)	(0.00741)
	-0.0164	-66.76	-96.95	-15.13	-23.76
female 85-89	(0.00351)	-0.0838	-0.157	-0.0813	-0.148
	-4.681	(0.00129)	(0.00166)	(0.00509)	(0.00681)
	-0.0141	-64.82	-94.77	-15.96	-21.73
male age 90+	(0.00459)	-0.168	-0.303	-0.169	-0.307
	-3.076	(0.00207)	(0.00234)	(0.00944)	(0.0109)
	-0.0205	-81.08	-129.8	-17.95	-28.20
female age 90+	(0.00462)	-0.157	-0.284	-0.147	-0.283
	-4.432	(0.00153)	(0.00180)	(0.00695)	(0.00867)
	0.0635	-102.5	-157.8	-21.09	-32.68
Native American	(0.0184)	0.00228	0.00472	-0.0111	0.00114
	3.454	(0.00354)	(0.00499)	(0.00966)	(0.0155)
	-0.120	0.644	0.946	-1.144	0.0739
Hispanic	(0.0172)	0.0140	0.0264	0.00107	0.0140
	-6.961	(0.00251)	(0.00326)	(0.00877)	(0.0115)
	0.0138	5.586	8.087	0.122	1.225
Other Race/Ethnicity	(0.0129)	-0.00478	0.000259	-0.0172	-0.00847
	1.071	(0.00244)	(0.00351)	(0.00923)	(0.0104)
	0.00442	-1.961	0.0736	-1.859	-0.814
Asian	(0.0169)	0.000856	0.0180	-0.00574	-0.00900
	0.261	(0.00285)	(0.00359)	(0.00955)	(0.0133)
	-0.00466	0.301	5.006	-0.601	-0.675
Black	(0.0116)	0.00824	-0.00697	0.000407	-0.0237
	-0.403	(0.00122)	(0.00174)	(0.00419)	(0.00602)
	0.0276	6.745	-4.003	0.0970	-3.944
Location of MI: anterior	(0.00511)	0.105	0.101	0.0923	0.0880
	5.406	(0.00229)	(0.00225)	(0.00722)	(0.00802)
	0.0265	45.79	45.04	12.78	10.96
Location of MI: inferior1	(0.00525)	0.142	0.141	0.126	0.122

	5.059	(0.00223)	(0.00215)	(0.00722)	(0.00786)
	0.0667	63.64	65.88	17.48	15.50
Location of MI: Right	(0.00883)	0.126	0.123	0.0975	0.0948
	7.558	(0.00323)	(0.00352)	(0.0105)	(0.0123)
	0.0132	39.09	34.80	9.302	7.705
Location of MI:subendocardial	(0.00456)	0.204	0.168	0.180	0.157
	2.896	(0.00210)	(0.00201)	(0.00680)	(0.00737)
	0.0198	96.91	83.68	26.45	21.33
Location of MI: Other	(0.00500)	0.100	0.0962	0.0770	0.0802
	3.966	(0.00301)	(0.00307)	(0.00991)	(0.0107)
	0.0192	33.37	31.40	7.763	7.511
2nd Quintile zip Income	(0.00589)	0.00207	0.00490	0.00533	0.00812
	3.263	(0.000882)	(0.00118)	(0.00275)	(0.00364)
	0.0249	2.345	4.142	1.937	2.229
3rd Quintile zip Income	(0.00686)	0.00464	0.00843	0.00465	0.0116
	3.632	(0.000895)	(0.00126)	(0.00285)	(0.00397)
	0.00839	5.185	6.705	1.628	2.920
4th Quintile zip Income	(0.00816)	0.00676	0.0106	0.0106	0.0198
	1.028	(0.000987)	(0.00139)	(0.00305)	(0.00412)
	-0.0168	6.846	7.657	3.474	4.803
5th Quintile zip Income	(0.0116)	0.0151	0.0225	0.0232	0.0373
	-1.445	(0.00111)	(0.00167)	(0.00310)	(0.00433)
	-0.449	13.61	13.51	7.475	8.618
year2007	(0.00947)	-0.0180	-0.0355	-0.0142	-0.0225
	-47.36	(0.00241)	(0.00300)	(0.00873)	(0.0102)
	-0.388	-7.450	-11.85	-1.630	-2.203
year2008	(0.00890)	-0.0195	-0.0370	-0.0158	-0.0274
	-43.63	(0.00242)	(0.00300)	(0.00833)	(0.0100)
	-0.332	-8.071	-12.35	-1.902	-2.737
year2009	(0.00784)	-0.0175	-0.0331	-0.0179	-0.0239
	-42.29	(0.00241)	(0.00302)	(0.00834)	(0.0101)
	-0.286	-7.247	-10.96	-2.149	-2.369
year2010	(0.00734)	-0.0150	-0.0289	-0.00868	-0.0181
	-38.96	(0.00239)	(0.00296)	(0.00834)	(0.0101)
	-0.247	-6.270	-9.763	-1.041	-1.795
year2011	(0.00667)	-0.0176	-0.0253	-0.0178	-0.0189
	-37.08	(0.00238)	(0.00298)	(0.00842)	(0.0102)
	-0.210	-7.411	-8.483	-2.115	-1.858
year2012	(0.00586)	-0.0166	-0.0260	-0.0155	-0.0186
	-35.78	(0.00239)	(0.00296)	(0.00840)	(0.0102)
	-0.155	-6.945	-8.755	-1.851	-1.829
year2013	(0.00511)	-0.0157	-0.0207	-0.0166	-0.0168
	-30.44	(0.00237)	(0.00296)	(0.00843)	(0.0105)
	-0.109	-6.641	-6.984	-1.971	-1.604
year2014	(0.00456)	-0.0183	-0.0256	-0.0163	-0.0169
	-23.94	(0.00238)	(0.00296)	(0.00822)	(0.0101)
	-0.0719	-7.683	-8.656	-1.987	-1.677
year2015	(0.00388)	-0.0146	-0.0175	-0.0195	-0.0139

	-18.50	(0.00202)	(0.00241)	(0.00692)	(0.00809)
	-0.0328	-7.244	-7.268	-2.814	-1.721
year2016	(0.00310)	-0.00676		-0.00809	
	-10.57	(0.00135)		(0.00494)	
	-0.000575	-4.992		-1.637	
icd10 indicator (adjust for coding shift)	(0.00212)	-0.00405	-0.00531	-0.00897	-0.00559
	-0.272	(0.00193)	(0.00254)	(0.00647)	(0.00876)
	-0.0172	-2.103	-2.090	-1.386	-0.638
Dementia	(0.00299)	-0.0693	-0.150	-0.0532	-0.133
	-5.771	(0.00218)	(0.00247)	(0.00945)	(0.0119)
	-0.00213	-31.85	-60.88	-5.630	-11.18
Dementia ICD10 Indicator interacted	(0.00384)	0.0177	0.00854	0.000665	-0.0172
	-0.555	(0.00348)	(0.00446)	(0.0150)	(0.0210)
	(0.00238)	5.084	1.915	0.0442	-0.823
Any month enrolled in Medicaid prior 6 months	-0.00430	-0.0153	-0.0317	-0.0132	-0.0294
	(0.00464)	(0.00101)	(0.00144)	(0.00334)	(0.00442)
	-0.928	-15.22	-22.13	-3.958	-6.651
Medicare Advtg proportion (0 to 1)	-0.141	0.00453	-0.00198	0.0321	0.0283
	(0.0542)	(0.00380)	(0.00572)	(0.00799)	(0.0104)
	-2.598	1.194	-0.346	4.022	2.712
Second Quintile HCC	0.00363	-0.0149	-0.0289	-0.0139	-0.0232
	(0.00129)	(0.000731)	(0.000959)	(0.00234)	(0.00299)
	2.826	-20.32	-30.10	-5.951	-7.753
Third Quintile HCC	0.000755	-0.0292	-0.0617	-0.0263	-0.0500
	(0.00172)	(0.000850)	(0.00113)	(0.00314)	(0.00409)
	0.439	-34.34	-54.73	-8.362	-12.21
Fourth Quintile HCC	-0.00467	-0.0510	-0.117	-0.0473	-0.0956
	(0.00230)	(0.000892)	(0.00124)	(0.00342)	(0.00448)
	-2.031	-57.23	-94.36	-13.82	-21.35
Fifth Quintile HCC	-0.0123	-0.112	-0.259	-0.103	-0.231
	(0.00315)	(0.00110)	(0.00144)	(0.00442)	(0.00577)
	-3.900	-102.1	-179.6	-23.29	-40.09
Constant	10.18	0.795	0.819	0.808	0.814
	(0.0195)	(0.00348)	(0.00395)	(0.0110)	(0.0129)
	521.6	228.4	207.0	73.14	62.91
Observations	1,729,854	1,730,535	1,652,864	115,282	109,533
R-squared	0.167	0.085	0.182	0.070	0.150

Robust standard errors in parentheses

p<0.001, p<0.01, p<0.05

Appendix Table A.2A: Regressions Estimates Survival Allowing for Misallocation: Full Sample Admitted through the Emergency Room

VARIABLES	Model 1 30-Day Survival	Model 2 1-Year Survival	Model 3 30-Day Survival	Model 4 1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	0.000160	0.00360	0.00762	0.0115
	(0.00167)	(0.00220)	(0.00163)	(0.00229)
PCA prediction for input1 component1	0.000728	0.00261	0.00120	0.00346
	(0.000219)	(0.000270)	(0.000225)	(0.000285)
PCA prediction for input2 component1	0.00206	-0.00393	0.000463	-0.00594
	(0.000504)	(0.000626)	(0.000557)	(0.000729)
PCA prediction for input3 component1	-0.000715	-0.00187	-0.000342	-0.00107
	(0.000443)	(0.000521)	(0.000447)	(0.000564)
Zip Code Fixed Effects			X	X
Hospital Random Effects	X	X		
Observations	1,349,049	1,287,109	1,349,049	1,287,109
Sigma_u	0.0309	0.0311		
R-squared			0.112	0.210

All regressions incude the full set of risk-adjustment covariates. Adjusted for wage index.

Appendix Table A.2A: Regressions Estimates for Survival Allowing for Misallocation: Tourist Hospitals and Tourists Admitted through ER

VARIABLES	Tourist Hospitals Model 1	Tourist Hospitals Model 2	Tourists Model 3	Tourists Model 4
	30-Day Survival	1-Year Survival	30-Day Survival	1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	0.00836	0.0131	0.00169	0.00332
	(0.00183)	(0.00249)	(0.00444)	(0.00598)
PCA prediction for input1 component1	0.000717	0.00225	0.000967	0.00359
	(0.000235)	(0.000300)	(0.000778)	(0.00104)
PCA prediction for input2 component1	0.00212	-0.00351	0.00146	-0.00236
	(0.000594)	(0.000786)	(0.00153)	(0.00205)
PCA prediction for input3 component1	-0.000499	-0.00129	-0.00112	-0.00277
	(0.000463)	(0.000570)	(0.00139)	(0.00176)
Zip Code Fixed Effects				
Hospital Random Effects	X	X	X	X
Observations	1,020,697	971,859	67,648	64,053
Sigma_u	0.0160	0.0219	0.0213	0.0319

Full sample except restricted to hospitals with >20 tourist AMI admissions.
All regressions incude the full set of risk-adjustment covariates. Adjusted for wage index.

Appendix Table 3A_STEMI: Regressions Estimates Survival Allowing for Misallocation: Full Sample; STEMI type AMI

VARIABLES	Model 1 30-Day Survival	Model 2 1-Year Survival	Model 3 30-Day Survival	Model 4 1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	-0.00404	-0.00435	-0.00121	-0.00153
	(0.00317)	(0.00323)	(0.00276)	(0.00311)
PCA prediction for input1 component1	0.00222	0.00422	0.00278	0.00520
	(0.000407)	(0.000440)	(0.000420)	(0.000470)
PCA prediction for input2 component1	-0.000594	-0.00419	-0.00214	-0.00644
	(0.000969)	(0.000975)	(0.00102)	(0.00114)
PCA prediction for input3 component1	-0.00134	-0.00262	-0.00201	-0.00272
	(0.000786)	(0.000787)	(0.000804)	(0.000879)
Zip Code Fixed Effects			X	X
Hospital Random Effects	X	X		
Observations	465,482	449,019	465,482	449,019
R-Squared			0.171	0.261
Sigma_u	0.0576	0.0481		

All regressions incude the full set of risk-adjustment covariates. Adjusted for wage index.

Table 3B_STEMI: Regressions Estimates for Survival Allowing for Misallocation: Tourist Hospitals and Tourists; STEMI type AMI

VARIABLES	Tourist Hospitals Model 1 30-Day Survival	Tourist Hospitals Model 2 1-Year Survival	Tourists Model 3 30-Day Survival	Tourists Model 4 1-Year Survival
log of hospital oper exp per discharge (from CMS cost reports)	0.00427	0.00638	-0.0117	-0.0123
	(0.00308)	(0.00338)	(0.00676)	(0.00712)
PCA prediction for input1 component1	0.00208	0.00340	0.00303	0.00447
	(0.000423)	(0.000463)	(0.00112)	(0.00127)
PCA prediction for input2 component1	-0.00149	-0.00557	-0.00448	-0.00598
	(0.000997)	(0.00111)	(0.00222)	(0.00248)
PCA prediction for input3 component1	-0.00143	-0.00181	-0.00209	-0.00453
	(0.000801)	(0.000857)	(0.00184)	(0.00207)
Zip Code Fixed Effects				
Hospital Random Effects	X	X	X	X
Observations	375,245	361,348	37,004	35,557
Sigma_u	0.0278	0.0262	0.0360	0.0191

Full sample except restricted to hospitals with >20 tourist AMI admissions.
All regressions

Complete Derivation of Estimation Equation

While Section II provides an overview of how the estimation equation is derived, we provide here a more thorough derivation for completeness. To begin, we start with the special (nested) case of the hospital objective function (2), where $\lambda_{hk} = \phi_h = 0$ and $\pi_k = 1$. The first step is to take the first order conditions of this function, as demonstrated in (3); specifically, we take it with regards to X_1:

$$\Omega_h = \varphi_h \tilde{Y}_h + \phi_h \sum_k p_k X_{hk} - \left[\sum_k \pi_k p_k X_{hk} + F_h \right] \qquad (2)$$

$$\frac{\partial \Omega_h}{\partial X_1} = \varphi_h \frac{\partial \tilde{Y}_h}{\partial X_1} - (\pi_1 - \phi_h) p_1 \qquad \text{(A1)}$$

Under the above conditions of $\pi_k = 1$ and $\phi_h = 0$, this equation simplifies to:

$$\varphi_h \frac{\partial \tilde{Y}_h}{\partial X_1} = p_1 \qquad \text{(A2)}$$

Repeating this process for a representative X_k, we can then set the marginal rate of substitution between X_1 and X_k equal to the marginal rate of transformation in optimality:

$$\frac{\varphi_h \frac{\partial \tilde{Y}_h}{\partial X_k}}{\varphi_h \frac{\partial \tilde{Y}_h}{\partial X_1}} = \frac{p_k}{p_1} \qquad \text{(A3)}$$

To simplify further, we need expressions for the partial derivatives. We return to (1) and take the partial derivatives with respect to X_1 and X_k; bear in mind that $\tilde{Y}_h$ here is expressed for an average patient (whose risk-adjustment product $\left[\prod_w Z_{iw}^{\omega_w} \right]$ is normalized to 1):

$$\frac{\partial \tilde{Y}_h}{\partial X_1} = A_h \left[\prod_k X_{hk}^{\beta_k} \right] \left[\beta_1 X_{h1}^{\beta_1 - 1} \right] \qquad \text{(A4)}$$

$$\frac{\partial \tilde{Y}_h}{\partial X_k} = A_h \left[\prod_{-k} X_{h-k}^{\beta_{-k}} \right] \left[\beta_k X_{hk}^{\beta_k - 1} \right] \qquad \text{(A5)}$$

Substituting these partial derivatives into (A3) results in significant simplification:

$$\frac{\varphi_h A_h \left[\prod_{-k} X_{h-k}^{\beta_{-k}}\right]\left[\beta_k X_{hk}^{\beta_k - 1}\right]}{\varphi_h A_h \left[\prod_{k} X_{hk}^{\beta_k}\right]\left[\beta_1 X_{h1}^{\beta_1 - 1}\right]} = \frac{p_k}{p_1} \qquad \text{(A6)}$$

$$\frac{\beta_k X_{h1}}{\beta_1 X_{hk}} = \frac{p_k}{p_1} \qquad \text{(A7)}$$

Letting $p_1 = 1$ as the numeraire, we can express the quantity of each input X_k as a function of X_{h1} as in (5):

$$X_{hk} = X_{h1}\left[\frac{\beta_k}{p_k \beta_1}\right] \qquad (5)$$

Next, we use (5) to develop an expression for total expenditures, M_h, again as a function of only X_{h1}. We begin by expanding the summation form of M_h into an expanded form including all inputs and prices:

$$M_h = \sum_k p_k X_{hk} = p_1 X_{h1} + p_2 X_{h2} ... + p_k X_{hk} \qquad \text{(A8)}$$

Each quantity factor can then be replaced in terms of X_{h1} using (5):

$$M_h = p_1 X_{h1} + p_2 X_{h1}\left[\frac{\beta_2}{p_2 \beta_1}\right] ... + p_k X_{h1}\left[\frac{\beta_k}{p_k \beta_1}\right] \qquad \text{(A9)}$$

Simplifying terms and factoring out the X_{h1} results in the simplified form of M_h, as seen in (6) in Section II:

$$M_h = X_{h1}\left(\sum_k \left[\frac{\beta_k}{\beta_1}\right]\right) \qquad (6)$$

We now turn to the main hospital output function, and will also express this as a function of X_{h1} as in (7). We begin by using (5) to simplify the product:

$$\tilde{Y}_h = A_h \left[\prod_k \left[X_{h1}\frac{\beta_k}{\beta_1 p_k}\right]^{\beta_k}\right] \qquad \text{(A10)}$$

The X_{h1} term can then be removed from the product. This allows us to derive an explicit term for total expenditure, via multiplication by a unit term, $\left(\frac{\sum_k \left[\frac{\beta_k}{\beta_1} \right]}{\sum_k \left[\frac{\beta_k}{\beta_1} \right]} \right)^{\sum_k \beta_k}$

$$\tilde{Y}_h = A_h X_{h1}^{\sum_k \beta_k} \left[\prod_k \left[\frac{\beta_k}{\beta_1 p_k} \right]^{\beta_k} \right] \quad \text{(A11)}$$

$$\tilde{Y}_h = \frac{A_h \left[X_{h1} \sum_k \left[\frac{\beta_k}{\beta_1} \right] \right]^{\sum_k \beta_k} \left[\prod_k \left[\frac{\beta_k}{\beta_1 p_k} \right]^{\beta_k} \right]}{\left[\sum_k \left[\frac{\beta_k}{\beta_1} \right] \right]^{\sum_k \beta_k}} \quad \text{(A12)}$$

$$\tilde{Y}_h = \frac{A_h \left[M_h \right]^{\sum_k \beta_k} \left[\prod_k \left[\frac{\beta_k}{\beta_1 p_k} \right]^{\beta_k} \right]}{\left[\sum_k \left[\frac{\beta_k}{\beta_1} \right] \right]^{\sum_k \beta_k}} \quad \text{(A13)}$$

We can now take the log of each side and represent logged values by lower-case letters:

$$y_h = \alpha_h + \beta' m_h + \left[\sum_k \beta' \ln \left(\frac{\beta_k}{\beta_1 p_k} \right) - \beta' \ln \left(\sum_k \frac{\beta_k}{\beta_1} \right) \right] + \varepsilon_h \quad \text{(A14)}$$

Where $\beta' = \sum_k \beta_k$.