NBER WORKING PAPER SERIES

PRODUCTIVITY VARIATION AND INPUT MISALLOCATION: EVIDENCE FROM HOSPITALS

Amitabh Chandra
Carrie H. Colla
Jonathan S. Skinner

Working Paper 31569
http://www.nber.org/papers/w31569

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

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 Amelia Bond, Daniel Gottlieb, Peter Hull, Peter Klenow, Edward Norton, Douglas Staiger, Chad Syverson, Taylor Watson, John Wennberg, and especially Joe Doyle for insightful comments. Seminar participants at Dartmouth, Chicago, Wisconsin, UCLA, the ASHE conference, and Stanford offered very helpful suggestions. Weiping Zhou and Ben Usadi provided extraordinary programming and data assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w31569

Productivity Variation and Input Misallocation: Evidence from Hospitals
Amitabh Chandra, Carrie H. Colla, and Jonathan S. Skinner
NBER Working Paper No. 31569
August 2023
JEL No. E23,I1,I10

ABSTRACT

There are widespread differences in total factor productivity across producers in the U.S. and around the world. To help explain these variations, we devise a general test for misallocation in input choices – the underuse of effective inputs and overuse of ineffective ones. Misallocation implies that conditional on total input use, the return to using a particular input is not zero (a positive return implies underuse, and a negative return implies overuse). We measure misallocation across hospitals, where inputs and outputs are better measured than in other industries. Applying our test to a sample of 1.6 million Medicare beneficiaries with heart attacks (of which 436 thousand were admitted by ambulance), we reject the hypothesis of productive efficiency; moving a patient from a 10th percentile to a 90th percentile hospital with respect to misallocation, holding spending constant, is predicted to increase survival by 3.1 percentage points. With misallocation accounting for as much as 25 percent of the variation in hospital productivity, our results suggest that how the money is spent, rather than how much money is spent, is central to understanding productivity differences both in health care, and in the rest of the economy.

Amitabh Chandra
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
amitabh_chandra@harvard.edu

Jonathan S. Skinner
Department of Economics
6106 Rockefeller Hall
Dartmouth College
Hanover, NH 03755
and NBER
jonathan.skinner@dartmouth.edu

Carrie H. Colla
The Dartmouth Institute for
Health Policy and Clinical Practice
Geisel School of Medicine
1 Medical Center Dr.
Lebanon, NH
carrie.h.colla@dartmouth.edu

I. Introduction

There are widespread differences across producers in total factor productivity (TFP), in the United States and around the world, in industries as disparate as manufacturing, banking, concrete and health care (Syverson, 2011). There are a variety of explanations for these differences including variation in the allocation of credit (Hsieh and Klenow, 2009), use of information technology (Lee et al., 2013) and the quality of management (Bloom et al., 2020, Tsai et al., 2015; Otero and Muñoz, 2022). Motivated by several classic studies from macroeconomics, we emphasize a different mechanism for productivity differences—the misallocation of inputs, through underuse of effective inputs, the overuse of ineffective inputs, or both. Hsieh and Klenow (2009) found that input misallocation 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.

We focus our attention on the hospital industry, because of its size, consequence, and the ability to measure outputs directly (Chandra et al., 2016a,b; Chandra and Staiger, 2020; Hull, 2020). In this setting, we measure the extent of misallocation in the health care sector using individual patient level data with accurately measured inputs and a meaningful output: 1-year survival following a heart attack (acute myocardial infarction, or AMI). Our test exploits simple economic intuition: an efficient producer would equate the ratio of marginal benefits to the marginal costs of every input, and one implication of this is that conditional on total input cost, an efficient producer has used every input to the point at which its marginal net benefit is zero. Rather than testing whether marginal returns (or markups) are equalized across firms or sectors of the economy, as in much of the macroeconomics literature on misallocation, we test for misallocation by examining whether the incremental health returns to inputs across or within hospitals are different from zero, conditional on total costs. Our test is general, and can be applied to any industry.

The idea of input misallocation—the overuse of ineffective medical treatments and underuse of effective ones-- has been a central topic in health care, for example in the "Choosing Wisely" campaigns to eliminate low-value health services (Colla et al., 2015a,b), the measurement of productive inefficiency (Hollingsworth, 2008), 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 distortions that may arise from informational asymmetries, inaccurate physician beliefs, inefficient reimbursement policies, or other distortions. 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 output (survival) *conditional on total expenditures*. That is, including specific input choices on the right-hand side of an equation for survival should result in zero coefficients (in the limit) on input choices because in a well-optimized delivery system (or firm), expenditure should be a sufficient statistic for all inputs.

Our orthogonality test is robust to a variety of hospital objective functions. For example, even if not-for-profit hospitals seek to maximize health outcomes regardless of costs, the first-order condition ensures that the incremental value of any additional input is zero (as is the incremental effect of spending more). The goal of minimizing production inefficiency – saving the most lives for a given level of costs – should be a relevant objective regardless of ownership status.[1] Another implication of the model is that in the presence of misallocation, conventional regressions seeking to estimate "the" association between expenditures and outcomes are problematic since one must know whether the incremental expenditures were used for effective or ineffective inputs.

We operationalize our test using the population of 1.6 million elderly (age 65+) fee-for-service Medicare enrollees with acute myocardial infarction (AMI), or heart attacks, during 2007-2017. We measure total inputs costs at the patient level using price-adjusted Medicare payments, and test the sensitivity of this measure to using operating expenses per admission from Medicare Cost Reports; these capture all costs (not just those allocated to Medicare patients).

In principle, there are many inputs that we could include for our test, so to reduce the dimensionality of the enterprise we focus on three types of input choices identified in the literature. Following Chandra and Skinner (2012), we consider cost-effective "Category I" inputs such as drug treatments effective for nearly all AMI patients and the use of same day percutaneous coronary interventions (PCI, or "stenting); "Category II" inputs are those with heterogeneous treatment effects – for example, the first MRI or physician visit is probably more valuable than the 47th. By contrast, "Category III" treatments are both costly and lack evidence of effectiveness, such as post-acute care (Doyle et al., 2017; McKnight, 2006; Einav et al., 2018) and treatments deemed by professional physician groups to be wasteful (Colla, et al., 2015).[2] We hypothesize that greater use of the often underused "Category I" treatments, and reduced use of the "Category III" treatments, will be associated with higher measured productivity, with intermediate effects for "Category II" treatments.

---

[1] If hospitals gain financially by inefficient input choices (e.g., supplier-induced demand), this would violate productive efficiency, as we demonstrate below.

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

The use of these categories to proxy for various types of inputs has several advantages: collapsing the dimensionality of the input space into clinically relevant groupings of inputs and reducing the need to interpret multiple coefficients on individual tests, diagnostics, procedures and treatments (e.g., Griliches, 2013). The categories that we use were defined over 10-years ago and so are pre-specified with respect to the current analysis. Of course, the inputs we consider may be surrogates for other inputs—for example, use of beta-blockers (a Category I treatment) may be correlated with unmeasured aspirin use (another Category I treatment)—and this is one reason that we also grouped them by category using principal component analysis. For some of our inputs, we use estimated treatment effects from randomized clinical trials to assess whether the regression estimates are capturing causal effects or whether the input is a surrogate for other unmeasured inputs.

The most important challenge for our estimation approach is inadequate risk adjustment; hospitals may appear to be productive when in fact they are treating healthier patients. We address this by considering detailed risk adjusters for one acute event – a heart attack – and thus avoid conflating different types of diseases in the same regression analysis. We include extensive information on individual health measures, including socioeconomic status and Medicaid dual-eligibility (which is associated with low income and poor health). Even with these controls, unmeasured health still may still bias estimates of productivity because of selection of patients to hospitals. We pursue several related approaches to address this concern. First, we note that Chandra and Staiger (2020) used chart-data for heart-attacks and found that the returns to a key input (reperfusion therapy after a heart-attack), are the same when estimated from models with simple controls such as age, gender, and race versus those with richer controls from chart-data.

Importantly, these OLS models replicate clinical trial results. That result is helpful for us, but may not apply to every category of input use.

We consider three model specifications. First, we include patient zip code fixed-effects (as in Garthwaite et al, 2019), which identifies misallocation through differences in the timing of input adoption across hospitals serving patients who live in the same zip code. Second, we identify misallocation using within-hospital temporal changes and across-hospital variation using random effects, and within-hospital variation alone using fixed-effects models.

We also consider ambulance-service-specific effects, which builds on an approach pioneered by Doyle et al. (2015, 2017) who use ambulance services as a natural randomization (instrument) for the hospital at which a patient is treated.[3] If, as in Doyle et al., ambulances take patients to preferred hospitals, then the use of ambulance-service fixed effects identifies input misallocation from the exogenous assignment of the patient to the service's preferred hospital, as well as any associated treatments provided (and charged) by the ambulance service prior to admission. Since we do not use ambulance-services as an instrument, we do not make additional assumptions about monotonicity and the exclusion restriction by which ambulances affect outcomes only through the choice of hospital (Chan et al., 2022b). Estimates of misallocation are robust across these different sources of identification.

Quantitatively, like Chandra et al. (2016a), we find wide variability across hospitals in productivity. Conditional on expenditures and detailed risk-adjusters, one-year risk-adjusted survival rates varied from 65.2 percent in the 10th percentile of hospitals to 74.4 percent in the

[3] That is, when there are multiple ambulance services in a region, with each service preferring a specific hospital, the arrival of a particular ambulance service generates quasi-random assignment of hospitals for patients because ambulances have a preferred hospital to which they transport patients.

90^{th} percentile of hospitals. Depending on the specification of the model, we find generally that higher expenditures are associated with better outcomes (e.g., a 10% increase in expenditures leading to a 0.2 - 0.5% increase in one-year survival), but these estimates are often attenuated after adjusting for hospital- or ambulance-specific effects and input choices. At least for AMI patients, we find little evidence of patient selection to high- or low-cost hospitals; the ambulance analysis yields estimates similar to those for the full sample.

Efficient production is strongly rejected by the data; in the full sample the $\chi^2(3)$ test for the null of efficient production is 771. Conditional on expenditures, hospitals more likely to use Category I treatments are associated with substantially higher survival, while Category II and III treatments are associated generally with lower survival. In a hypothetical risk adjusted counterfactual, moving hospitals from the 10^{th} to the 90^{th} percentile in terms of the degree of misallocation, holding expenditures constant, would decrease survival for those lower-performing hospitals by 3.1 percentage points, or 11 percent of average one-year mortality.

We also consider whether the 2017 "US News & World Report Top 25 Cardiovascular Hospitals" exhibit higher survival rates conditional on Medicare expenditures, and diminished misallocation, relative to those not in the top-25. This set of hospitals constitute a popular measure of marquee hospitals, and in our analysis exhibit risk-adjusted survival 5 percentage points higher than the non-Top-25 hospitals, with somewhat lower levels of misallocation.

Results were largely insensitive to a variety of alternative specifications; In addition to using different sources of identification, our results are similar using 30-day rather than one-year survival rates, and using hospital operating expenses from CMS cost reports to capture differences across hospitals in revenue and costs financed from sources other than Medicare. Coefficient estimates were consistent with existing clinical randomized trial estimates, reducing concerns about bias in production function estimates (e.g., Ackerberg, et al., 2015). Nor do we find evidence that teaching hospitals exhibit systematically different production functions from non-teaching hospitals.

Our estimates suggest that misallocation contributes to around 25 percent of overall variation in hospitals' total hospital factor productivity. Less well understood is why some hospitals systematically use too few effective treatments or overuse ineffective treatments. Unlike the macroeconomic literature, we cannot so easily appeal so easily to distortions such as taxation or capital controls to explain systematic underuse of effective treatments. Here, a rich literature from health economics on sources of inefficiency is highly informative, including the structure of reimbursements (Pauly, 1970), diagnostic skill (Abaluck et al., 2021; Chan, 2021); performance and ability to triage patient appropriately for procedures (Chandra and Staiger, 2021; Chan et al., 2022a; Mullainathan and Obermeyer, 2022; Doyle, 2020), and physician beliefs (Cutler et al., 2019). This literature and our empirical estimates imply a substantial degree of misallocation and productive inefficiency in U.S. health care, suggesting that what matters for health outcomes is less how much money is spent, but instead how the money is spent.

II. The Model

Conventional regression estimates of the marginal returns to expenditures (or factor inputs) seek to estimate the slope of the production function, as shown in Figure 1 by finding exogenous shifts in factor inputs that might plausibly affect survival. Depending on the coefficient estimated (typically the slope of the line between A and A′) the results are viewed as informative as to whether the U.S. is allocatively efficient, in the sense of spending too much or too little on health care. If the slope of the production function (e.g., the gains in survival per dollar of spending) exceeds some societal or cost-benefit threshold, then researchers conclude that more resources should be devoted to health care, and conversely.[4]

The problem arises when hospitals are not all on the same production function, or when productive inefficiency is present. To illustrate this, we consider a very simple model in which hospitals differ with respect to their total factor productivity (TFP) as in Chandra et al. (2016a). In this paper, we distinguish between two sources of TFP variation. The first is the conventional Hicks-neutral productivity difference (by which one hospital gets better outcomes than another for a given set of measured inputs but where the ratio of marginal products of inputs are unaffected), while the second arises from misallocation, or suboptimal input choices.

Assume two hospitals, A and B, shown in Figure 2 with expenditures on the horizontal axis and survival on the vertical axis; prices are the same for each hospital. At points A and B, expenditures are identical, but A yields better outcomes for the same spending level. Assuming we have accurately adjusted for differences in health risks between the two hospitals, there are two reasons why Hospital A yields better outcomes. The conventional explanation for productivity differences is that Hicks-neutral technological differences explain the entire gap between $S_A(X^*)$ and $S_B(X^*)$, perhaps because physicians, nurses, and support staff are of higher quality at Hospital A; these are shown by the hypothetical production functions that are drawn through A and B.

The second explanation for TFP differences is independent of conventional skill differences, but instead is the consequence solely of misallocation. 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

[4] The inverse of this slope – how much one must spend to save a life – is related to the cost-effectiveness ratio, which may range from $50,000 to $250,000 per life-year.

alternative "red" Category IfII 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, and would be detected as the positive return to the Category I technology times the differential use of the input at Hospital A relative to Hospital 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. The regression coefficient does not measure 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. (2015a) that regional use of low-value (Category III) care is positively correlated with regional spending overall. For these reasons, estimates of the slope of the production function require both hospital-specific productivity measures (as in Skinner and Staiger, 2015, Hull, 2020, or Chandra et al., 2016a) and differences across hospitals in input choices.

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

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

and the health outcome Y_{ih} is defined (for technical reasons) as the exponent of survival for 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 and assume diminishing returns ($\sum_k \beta_k < 1$).

In this model, the hospital system (including physicians, nurses, ancillary health employees, and administrators) is the relevant decision-making entity; the Hicks-neutral productivity measure 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. We also allow for systematic errors in decision-making of the institution.

Assume that the objective function of the hospital is:

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

[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 average survival.

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. 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 non-Medicare patients arising from commercial and Medicaid services. The first-order condition of the objective function in (2) is:

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

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.

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

By contrast, in a fee-for-service setting where Medicare is viewed as the major revenue source, $\phi_h = 1$; 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). At Point H, hospitals ignore prices (either efficient or distorted) or total expenditures – they simply maximize health regardless of costs.[7] In this special case, the incremental value of increasing any input (or overall expenditures) would be zero, so the F-test for the joint significance of input choices would fail to reject the null – and thus would correctly signal the absence of productive inefficiency.[8]

In general, however, we assume below that in the absence of misallocation, hospitals respond to prices and determine input allocation as if they were maximizing social welfare by setting the social value of the marginal product equal to input price. In this case, either an additional (say) PCI or additional ICU-day will contribute to survival in proportion to its input price, which in turn is equal to the coefficient on total expenditures; thus any small fluctuations in either PCI or ICU use will (by the envelope theorem) not affect outcomes since their beneficial survival effects are already captured by the commensurate increase in expenditures arising from the increased input use.

Now suppose that regulations, informational barriers, or internal resistance to using effective treatments exist 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] One justification for this— at least in fee-for-service Medicare – is that there is little supervision of hospital billing so even if input prices were wrong, they would be irrelevant and thus not affect outcomes.
[8] Because the estimated coefficient on inputs condition on total expenditures, we interpret their coefficients not as average treatment effects, but marginal effects relative to the optimum input mix.
[9] See Díaz-Hernández et al. (2008) for a similar "shadow price" approach.

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

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, 2021). 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 in hospital-level testing rates.[10]

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; in that case, there would be overuse of low-productivity high-profitability treatments.[11] For analytical purposes, it is easier to define the distortion as proportional to the price, so that

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

The proportional input distortions λ_{hk} are defined implicitly based on actions or decisions by the hospital, so they will differ across hospitals or over time.

We first derive our estimating equation in the special (nested) case where $\lambda_{hk} = \phi_h = 0$

[10] If the overall rate is optimal, but physicians over-test low-risk patients and under-test high-risk patients, this type of misallocation would be reflected in α.

[11] 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.

and $\pi_k = 1$ so there is no misallocation of factor inputs for informational or financial reasons (the complete derivation is in Appendix B). 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}:

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

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 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 face common input prices, 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, specific input choices should not predict outcomes conditional on m. 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:[12]

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

where $\eta^*_{hk} = \left[\frac{(\pi_k + \lambda_{hk} - \phi_h)\beta_k}{(\pi_1 + \lambda_{h1} - \phi_h) p_k \beta_1} \right]$.

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

[12] The Cobb-Douglas specification of the production function we use in (1) does not allow for negative coefficients on inputs; thus one may think of the log-linear specification as a first-order Taylor-series approximation of an arbitrary (log) production function.

overused. With prices equal to marginal cost, (π_k =1) the implicit shadow price for input 1 is positive, and for input 2 negative; 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 in the productivity of their inputs, for example in teaching hospitals relative to non-teaching hospitals (Burke et al., 2018) or for hospitals specializing in a specific input such as stenting (e.g., Chandra and Staiger, 2007). In the context of our production function, this corresponds to a larger β for effective treatments in hospitals with more skilled or specialized physicians. We test for this by stratifying our sample into teaching and non-teaching hospitals, and conducting an F-test for whether input coefficients differ systematically across the two groups of hospitals;[13] we also test whether hospitals specializing in PCI gain higher returns to PCI (Chandra and Staiger, 2007).

III. Data and Estimation

Medicare Claims. We use a cohort of patients hospitalized with acute myocardial infarction (AMI) in the fee-for-service Medicare population during 2007-2016, 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.

[13] In theory one could also compare the US News and World Report top-25 hospitals with other hospital production functions, but their small sample sizes lacked statistical power.

Measuring Costs and Inputs. In the primary analysis considering all AMI admissions, we measure inputs and average expenditures using the "leave-out" approach which excludes the individual's treatment (and costs) from the hospital-level average.

There are two approaches to measuring total expenditures. The primary measure comprises individual-level measures of Medicare reimbursements (Parts A and B, including physician fees, post-acute care, outlier payments, and outpatient care) with price adjustment as in Gottlieb et al. (2010). Medicare reimbursement rates have been characterized as roughly equal to average variable cost (MedPAC, 2019) and so we use these to proxy variable hospital costs for the AMI patients. Yet one shortcoming of these standardized Medicare reimbursements is that they do not capture differences across hospitals in revenue (and costs) financed from sources other than Medicare, most importantly from privately insured patients reimbursed at a much higher rate which in turn can provide support for high-quality physicians, nurses, and technicians. For this reason, we consider a second measure of costs: Per-admission operating expenses by hospital by year from Medicare Cost Reports; these capture all costs (not just those allocated to Medicare patients).[14] Two disadvantages of this secondary measure are that it may be subject to accounting strategies to shift cost allocation from (or towards) operating expenses, and it is not specific to AMI patients.

Risk Adjustment. The risk adjustment approach we use includes admission-level comorbidities such as cancer, diabetes, liver disease, peripheral vascular disease, congestive heart failure, the clinical location of the AMI (e.g., inferior, anterior, right-side, subendocardial), whether the patient has been dually-eligible for Medicaid within 6 months of the admission (a

[14] We are grateful to Adam Sacarny for posting these data; http://sacarny.com/data/. The operating cost measure has been adjusted for the labor portion of the wage index to account for geographic differences in labor rates (Sloan and Edmunds, 2012).

marker for poor health, low income, or both), 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).[15] Hierarchical Condition Categories (HCCs) count the number of different diagnoses that patients have received in the 6 months prior to the index admission, and weights them for severity.[16] We also adjust for the fraction of the hospital referral region enrolled in Medicare Advantage to capture the possibility that the fee-for-service population could exhibit greater unmeasured health deficits if healthier enrollees select into managed care.[17]

Zip Codes. Another approach is to sweep out all neighborhood variation by including zip code fixed effects; this will absorb common health behaviors, average socioeconomic status, environmental health effects, and any other neighborhood factor common to the zip code, although we must assume that hospital productivity is not correlated with unmeasured patient characteristics within the zip code. A deeper concern is that because hospital admissions are largely local, zip code fixed effects will also absorb the average (community) hospital quality. For example, if we were comparing health outcomes in City A served by three very high-quality hospitals with outcomes in distant City B with three very poor-quality hospitals, the zip code dummy variables will absorb any differences in average quality between City A and City B and capture only relative differences within cities. This would lead to incorrectly ranking City B's best hospital over City A's worst hospital. For this reason, we do not use zip code fixed effects when estimating hospital-level random effects at the national level.

---

[15] The Research Triangle Institute definitions are used for race and ethnicity. Unfortunately, Medicare data does not distinguish between ethnicity and race, so these definitions are mutually exclusive.

[16] The use of HCC measures can lead to biases because physicians who treat patients more intensively will tend to find (and code) more diseases (Song et al., 2010; Finkelstein et al., 2016).

[17] While managed care patients are absent from this sample because of lack of billing data, risk adjusters are specific to the fee-for-service population.

Ambulance Services. In pioneering research, Doyle et al. (2015, 2017) addressed selection by patients to hospitals based on unobservables by using the randomization of ambulance services loyal to one hospital over another. They argue that while it may be chance which ambulance makes it to the patient's address first, ambulance service loyalty to specific hospitals leads to quasi-randomization to hospitals for severely ill patients. Following their approach, we construct our measure of inputs at the hospital-level, and then calculate average hospital inputs and expenditures by ambulance service using a "leave out" condition to exclude the contribution of any given individual to the ambulance-service average.[18] In their instrumental variable model, Doyle et al. focused primarily on estimating the health return to one variable, Medicare expenditures. We relax their exclusion restriction – that survival is affected only through expenditures -- as well as the monotonicity condition (Chan et al., 2022b) to focus on reduced-form estimates with a variety of other inputs with clinically plausible causal effects on survival. Using the reduced form approach will tend to bias our estimates of misallocation towards zero.

Clinically Relevant Inputs: There are many treatments for AMI patients, both in the acute setting, and subsequently post-discharge. We consider a range of such treatments or procedures where the initial hypothesis of effectiveness is based on existing clinical evidence. To help organize the data analysis, we follow Chandra and Skinner (2012) by appealing to clinical evidence to collapse this broad array of treatment effectiveness into three groups. The first is "effective" or Category I inputs which are distinguished by their high cost-effectiveness and limited scope for overuse. Examples used in this study are beta blocker, statin, and ACE/ARB prescription fills for AMI patients during the 6 months after discharge from the hospital for AMI

[18] The National Provider Identifier (NPI) is used to identify ambulance services.

(Munson et al., 2013).[19] Nearly everyone should get such treatments, regardless of health status. We also include same-day PCI (stenting); a procedure highly effective in saving lives if administered for appropriate patients within 12 or 24 hours of a heart attack (Keeley et al., 2003).

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, 2020). Another example of potential Category II treatment occurs when many different physicians treat the same patient. As demonstrated by Becker and Murphy (1992), greater specialization within each physician can improve productivity (for example, having a cardiologist, as in Doyle, 2020), but at some point, there are diminishing returns owing to rapidly rising costs of coordinating care (Baicker and Chandra, 2004b); Additional examples are the number of intensive-care unit or cardiology-care unit days (ICU plus CCU), and MRI and CT scans; for these treatments, however, evidence on incremental effectiveness is sparse (Ahmed et al., 2013). 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, negative, or unknown, but that have a large effect on spending. For example, Doyle et al., (2017) using ambulance services as randomization, found adverse effects on survival of higher levels of post-acute care, while Einav et al. (2018) estimated long-term care hospitals (LTCH) to be wasteful; similarly, home health care exhibits fraudulent behavior in many regions (O'Malley et al., 2023). We therefore include three measures of post-acute care:

[19] These measures are based on the Medicare Part D data; because they represent just a 40 percent random sample, they comprise a smaller fraction of AMI patients.

average spending per AMI patient for home health care, average spending per AMI patient for skilled nursing facility (SNF) care, and the faction of AMI patients admitted to either a LTCH or an inpatient rehabilitation facility (IRF).[20] In addition, we use two "Choosing Wisely[21]" measures involving the use of "double CT" scans of the chest and abdomen, one with iodine contrast and the other without. They provide no additional clinical information and expose patients to double the radiation (Bogdanich and McGinty, 2011); we interpret these as markers for poor hospital quality control. 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.

Measurement Error and Minimum Cell Size. Measurement error creates challenges in estimating misallocation (Bils et al., 2021), particularly with highly skewed expenditures. We seek to measure inputs and expenditures at the hospital-year or ambulance-service-year level, but if the cell size is less than a minimum N* we assign the average value for the hospital or ambulance service over the entire period. If the cell size for the entire period is less than N*, the hospital or ambulance service is excluded. We adopt a minimum N* of 50, but consider both minimum sample sizes of 25 and 75.

Econometric Specification. To describe the basic patterns of the data, we begin by using Equations (1) and (10) to express survival for individual i in either hospital or ambulance service h, at time t, y_{iht}, in a linear probability model[22] as a function of Hicks-neutral productivity α_h, total (log) expenditures m_{ht}, risk adjusters z_{imt}, inputs x_{hkt}, and an error term ζ_{iht}:

[20] The first two measures average over all AMI patients admitted to the hospital, and thus reflects both the fraction receiving such care, and average spending conditional on receipt of care.
[21] For example, a list of procedures with no clinical benefit are here: http://www.choosingwisely.org/
[22] Fixed or random effects models in nonlinear models such as probit or logit are both difficult to interpret and computationally challenging given the number of zip codes and hospitals.

$$y_{iht} = \alpha_h + \beta' m_{ht} + \left[\sum_k \gamma_k x_{kht} \right] + \left[\sum_m \omega_m z_{imt} \right] + \zeta_{iht} \qquad (11)$$

where γ_k are the first-order approximation of changes in η^*_{hk} with respect to x_{kh} holding expenditures constant; in practice these are nearly identical to the coefficients β_k that do not hold expenditures constant. The null hypothesis for productive efficiency is that the inputs conditional on expenditures are jointly equal to zero: $\gamma_1 = \gamma_2 ... = \gamma_k = 0$.

A long-standing concern with the estimation of production function models is when the inputs x_{kht} are themselves correlated with the error term ζ_{iht}, for example when a firm hires more labor in response to a productivity shock not observed by the econometrician (e.g., Ackerberg et al., 2015; Levinsohn and Petrin, 2003). Our production function differs from these models in important ways, as it relates to inputs (and outcomes) per patient, rather than a total quantity produced using capital and labor.[23] That said, we are still concerned with the possibility that an unobservable secular change in α_h would be associated with a different mix of inputs, for example high-quality physicians instituting hospital-wide standards for improved use of post-discharge statins, or perhaps instituting a reduction in the use of Category III treatments. As a test for bias, we therefore allow for time-varying α_h, as well as comparing the regression coefficients (where possible) with estimates from randomized clinical trials.

The hospital-specific (or ambulance-service-specific) α_h can be estimated using either random- or fixed-effects models, although the statistical precision of fixed-effects models is limited given year-specific sample size considerations. In addition, we also estimate the model

[23] The analogy would be that a positive productivity shock in producing automobiles would likely lead to higher levels of labor, intermediate goods, and car sales, but might not directly affect the mix of inputs per car.

with conventional zip-code fixed effects regression models, with clustering at the level of the hospital referral region (HRR).

We consider fully specified models with all inputs entered separately, but then consider a simplified model that is easier to interpret using the principal components model. We use the estimated coefficients from the random-effect model, along with the observed values of X_{ht} (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 Hicks-neutral productivity difference.

Are Highly-Ranked Hospitals More Productive? We first test to see if our AMI data predict better outcomes at the U.S. News & World Report "25 Best Cardiovascular Hospitals" and then decompose the difference in outcomes into two parts: misallocation and Hicks-neutral productivity differences.

IV. Results

Table 1A presents summary statistics for the entire sample of 1,617,039 AMI patients age 65 or over enrolled in fee-for-service Medicare during 2007-17. The average age is 78.1, and on average 70.8 percent survive to one-year post-admission, and 86.3 percent to 30 days. Average Medicare expenditures in the first 30 days are $26,547 (with a standard deviation at the patient level of $21,331) while the hospital-wide operating expenditures per patient (as reported in the CMS cost reports, for all ages) are $21,462. There are high rates of comorbid diseases, with 27.3 percent having diabetes, 39.2 percent with congestive heart failure, and 17.6 percent with chronic lung disease.

We also stratify hospitals by quartiles of their 30-day Medicare expenditures; those admitted to hospitals in the highest-quartile of hospital Medicare spending (mean of $29,332) are slightly less likely to survive to one year (69.2%) while those in the lowest-quartile (mean of $23,788) are more likely to survive (71.4%). However, the Hierarchical Condition Category (HCC) score used for risk-adjustment is slightly higher (1.380) in the highest quartile hospitals, implying sicker patients, than for the lowest quartile (1.315); comorbidities are broadly similar.[24]

Table 1B considers the smaller subset of 436,519 patients admitted to the hospital by ambulance. The average age is slightly older (78.9 versus 78.1) with a higher HCC score (1.417 versus 1.339) compared to the overall sample in Table 1A. High- and low-expenditure quartiles exhibit similar one-year survival (67.5 percent versus 67.2 percent) and rates of comorbidities, although the HCC score is higher in the highest-quartile (1.484) compared to the lowest (1.344).

Table 2A presents regression estimates for the full sample that follow the standard conventions of placing log hospital expenditures on the right-hand side of the regression equation; all regressions include a full set of risk adjusters (see Appendix Table A.1). Model 1 reports coefficients of one-year survival on the log of patient-level 30-day average hospital Medicare expenditures; the coefficient is 0.0317 (s.e. 0.0067), indicating greater spending is associated with survival; shifting from the lowest to the highest quartile of expenditures is predicted to increase survival by 0.7 percentage points. Model 2 yields a similar coefficient of 0.0319 (s.e. 0.0025) when the log of hospital-level operating expenses (which is not limited to Medicare patients) is substituted for the AMI-specific Medicare expenditures. For 30-day Medicare expenditures, results are larger with zip-code fixed effects (0.0527, s.e. 0.0055), but considerably less with hospital random effects (0.0167, s.e. 0.0037) and smaller still for the fixed-effects model (0.0010, s.e. 0.0044). That the standard deviation of the hospital-specific

[24] Average operating expenditures (from the CMS cost reports) exhibit only a modest difference ($21,117 versus $22,257) because they are not strongly correlated with Medicare expenditures. This arises because of the lack of correlation between private insurance spending (which largely determines hospital revenue and hence expenses), and Medicare spending (Cooper et al., 2022).

random effect is 2.53 percentage points (Model 4) suggests an important role for variation in hospital productivity, as in Chandra et al. (2016a).

Figure 3 shows risk-adjusted measures of log expenses and survival (based on the risk-adjustment model in Appendix Table A.1) evaluated at the sample means of the risk-adjustment variables and averaged across all years. The sample is limited to hospitals with at least 500 AMI patients over 2007-17, so the wide range in both operating expenses and the range in survival is not the consequence of small sample variability. While a regression line yields a similar coefficient to those found in Table 2A, spending alone explains at most 2 percent of the variance in survival rates. We also highlight (in orange) hospitals included in the 2017 U.S. News & World Report "25 Best Cardiovascular Hospitals."[25] These hospitals exhibit (on average) risk-adjusted survival rates equal to 76.1 percent, substantially above the 71.1 percent observed in hospitals not included on the list.[26] We consider below the extent to which these differences are the consequence of misallocation and conventional TFP differences.

Table 2B presents estimates using the sample of patients who arrived at the hospital by ambulance. The coefficient in Model 1 is 0.0226 (s.e., 0.0197), although the coefficient for log operating expenditures is larger and precisely estimated (0.0513, s.e. 0.0040). With zip-code fixed effects, the estimate is 0.0539 (s.e. 0.0206), similar to the estimate in Table 2A for the entire sample.[27] Neither the random-effect or fixed-effect model coefficients on expenditures are significantly different from zero.

---

[25] A few "top 25" hospitals failed to meet the 500-patient minimum.

[26] CMS inpatient Medicare mortality rates comprise 30% of the ranking (U.S. News and World Report, https://health.usnews.com/health-care/best-hospitals/articles/faq-how-and-why-we-rank-and-rate-hospitals); so there may be a partially mechanical correlation between our estimates and theirs.

[27] Results are similar when restricting the estimates to at least two ambulance services and two hospitals served in each zip code, as in Doyle et al. (2015).

Table 3 describes the full model with each input entered individually; the first three columns present results for the full sample with zip code fixed effects, hospital random effects, and hospital fixed effects; the next three columns are for the ambulance-service sample with zip code fixed effects, and ambulance service random and fixed effects. The independent variables are expressed as fractions of the population (e.g., the fraction who filled a prescription for statins) except for home health and nursing home care where it measures average spending (in $1,000) per AMI patient. Most of the Category I inputs exhibit positive coefficients although many are not significant individually; however, early (same-day) stenting (PCI) however is a strong predictor of survival across all equations. For some Category II inputs, there is no evidence of diminishing returns; for the number of different physicians seen, for example, there is a consistent negative coefficient for the bottom quartile and positive coefficient for the top quartile. For the case of scans, the results are reversed; generally relative to the bottom quartile, more use is associated with worse outcomes. Category III treatments show generally negative associations between greater use and survival, although again, individual coefficients are not consistently significant. Depending on the model specification, the coefficients for IRF or LTCH admission, and home health care, are either not significantly different from zero, or negative and significant, the latter of which are consistent with Einav et al. (2018) and Doyle et al. (2017).

It is difficult to interpret the disparate coefficients in Table 3 (Griliches, 2013); the fixed-effects ambulance model (Column 6) sometimes exhibits implausible coefficients with very wide confidence intervals. We therefore turn in Table 4 to using first principal components for Categories I, II and III; see Appendix Table A.2 for further details on the principal components results. As in Table 3, the first three columns are for the full sample (zip code fixed effects, random effects, and hospital effects) while the second three columns are similarly for the ambulance sample. The estimates for the Category I first principal component are strongly positive across all models, ranging from 0.59 percentage points (s.e. 0.07 percentage points) per standard-deviation shift for the full-model fixed effect in Column 3 to 1.32 percentage points (s.e. 0.54) for the ambulance fixed effect model in Column 6; these estimates taken together show that, regardless of specification, greater use of effective treatments lead to higher survival conditional on expenditures. Estimates for Category II treatments are negative and smaller in magnitude ranging from -0.01 percentage points (s.e. 0.08) in Model 5 to -0.49 percentage points (s.e. 0.07) in Model 1. For Category III treatments, coefficient estimates are most strongly negative in the ambulance sample, with (for example) a coefficient of -0.81 percentage points (s.e. 0.30) in the fixed effects model (Column 6).[28]

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 from Model 2 in Table 3; 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, using hospitals with at least 500 AMIs during the period of analysis. Moving from the 10^{th} to 90^{th} percentile of estimated misallocation predicts a reduction of 3.1 percentage points in survival. Measuring the importance of misallocation in hospital-level survival variability by comparing the variance of hospital-level random effects with specific inputs (Table 3) and

[28] In Columns 3 and 6 in Table 4, we report the correlation coefficient between predicted survival and the hospital- or ambulance-specific fixed effects. That they are so small (0.035 and 0.060, respectively) is consistent with the assumption in random-effects models that the hospital- or ambulance-specific random effects are not correlated with either inputs or risk adjusters.

without inputs (Table 2A) implies that 25 percent of the variability is explained by misallocation; the equivalent for the ambulance-level estimates is 22 percent.[29]

What are 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 efficient input choices independent of Hicks-neutral productivity differences (Skinner and Staiger, 2015). This hypothesis appears to be consistent with the data; we illustrate the strong association between predicted survival based on the extent of misallocation ($\hat{S}_h$) and hospital volume by decile in Figure 5. The US News & World Report top-25 hospitals continue to demonstrate lower levels of misallocation even conditioning on volume; differences in measured misallocation between the 25 hospitals and other hospitals predict a difference in survival of 0.5 percentage points. Thus, most of the survival gap between the "top 25" and other hospitals is because of higher α_h rather than misallocation.

Robustness. One concern is that the results may be sensitive to the one-year survival rate if (for example) other unmeasured factors affected longer-term health outcomes. In Appendix Table A.3 we also consider the principal components models for 30-day survival; results are similar albeit with coefficients that are generally smaller in magnitude. Another concern is that results are sensitive to the minimum cell size. We therefore estimated the model using a cutoff point of 25 or 75; results are presented in Appendix Table A.4A (full sample) and A.4B (ambulance sample); results for the three categories are again broadly similar.

[29] For example, in the ambulance sample, with only log expenditures and risk adjusters (Table 2B, Model 4) the standard deviation of ambulance-level outcomes is 0.0258, while with all inputs the estimate is 0.0228 (Table 3 Model 2); the ratio of the squared values (e.g., the variances) is 0.78, so accounting for specific hospital inputs reduces the variance in hospital productivity by 22 percent, with a similar calculation for the full sample yielding a drop of 25 percent. A different approach is a hospital-level regression of the data in Figure 4 (not restricted to larger hospitals) which yields an R^2 of 28 percent.

Hospitals may reasonably choose different input combinations because they have different production functions, for example if highly trained physicians at teaching hospitals are more skilled with both diagnosis of appropriate patients and skill in the specific treatment (Burke et al., 2018). To consider this possibility, in our estimation we first allow for a completely different set of input measures γ_k for teaching hospitals compared to non-teaching hospitals. Table A.5 (Model 1) presents the coefficients for each input categories (and log expenditures), separately for teaching and non-teaching hospitals, but coefficients appear similar; the F-test for equality of the two sets of 4 coefficients is not significant (p = 0.22). More importantly, predicted hospital-level misallocation using the fully flexible random-effects model coefficients (as in Table 3) using (a) teaching hospital coefficients and (b) non-teaching hospital coefficients were highly correlated (ρ = 0.92).[30] Thus we do not find evidence of a different production function for inputs between teaching and non-teaching hospitals.

A more subtle concern is that α_k evolves over time, for example if a more productive manager replaces a less efficient one (Otero and Muñoz, 2022). In the fixed-effects models this could generate both changes in survival rates and changes in input choices, leading to a spurious correlation between survival and inputs. However, when we allow each hospital a different "early" α_{ke} (for AMIs occurring in 2007-11) and a "late" α_{kl} (for 2012-16), the estimated survival effects for our three categories are unaffected (Appendix Table A.5 Model 2 for the full sample and Model 3 for the ambulance sample). Nor did the use of the CMS operating cost per patient measure affect our estimates, also shown in Appendix A.5.

[30] Following Chandra and Staiger (2007), we also tested whether hospitals favoring the use of PCI experienced higher marginal returns. However, we did not find evidence for higher returns, perhaps because by the late 2000s interventional cardiologists had gained experience with the procedure.

Finally, the coefficient estimates for β_k (or γ_k) may be biased because of a correlation between the error term and the input measures, for example when the use of statins proxies for the use of other unmeasured inputs. We use randomized clinical trial estimates when available to compare with these estimated coefficients. For example, a randomized trial for older AMI patients estimated survival gains of 15 percentage points from same-day PCI (de Boer et al., 2002), somewhat larger than the median estimate of 0.082 (across all 6 specifications). In randomized trials, statin use is estimated to exhibit a mortality odds ratio of 0.83 (Josan et al., 2008), or 3.6 percentage points, close to the median coefficient estimate in Table 3 of 3.3 percentage points. For Beta Blockers, the median coefficient is 0.036 but with considerably more variability across the 6 models. This median estimate is close to the trial estimates prior to the general use of PCI, but higher than for more recent trials finding smaller incremental effects (Bangalore et al., 2014).

V. Conclusion

In this paper, we estimated the importance of misallocation of inputs (and productive efficiency more generally) in explaining the wide differences across hospitals in total factor productivity (TFP). Rather than testing whether returns to inputs are equalized across firms, as in the previous literature, we develop instead a new test for misallocation and productive inefficiency more generally: Whether the returns to inputs within firms or hospitals are different from zero, conditional on total operating costs. Using a sample of 1.6 million Medicare patients with acute myocardial infarction (AMI) between 2007 – 2017 (and a sample of 436,000 AMI ambulance patients), we demonstrated that the hypothesis of productive efficiency is strongly rejected; high-misallocation hospitals reduce one-year survival rates by 3.1 percentage points, and contribute to between 22 and 25 percent of overall variation in total hospital (or ambulance service) factor productivity. This estimate is larger than the estimate by Baqaee and Farhi (2020) who estimate a penalty of 15% of GDP arising from misallocation for the entire U.S. economy.

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 and accurately measured outcome – survival following an index event, a heart attack requiring immediate hospital treatment – and well measured inputs, allowing an estimate of misallocation with a minimum of assumptions about elasticities of outputs and inputs.

We recognize several key assumptions of the model. The first assumption is that prices of inputs are assumed constant across hospitals; otherwise, firms may optimally adjust inputs to minimize costs and thus falsely reject the null of no misallocation. While Medicare prices are regulated and adjusted for cost-of-living differences across regions, David et al. (2022) have shown that these adjustments can bias the selection of technology by changing the relative prices of labor inputs, which reflect local cost-of-living versus surgical devices priced at the national level. To the extent that relative prices of inputs are relatively stable *within* hospital markets over time, the ambulance and hospital fixed-effects models should not be affected by such price differences. A second key assumption is that hospitals are subject to the same production function, but in our comparison of teaching and non-teaching hospitals, we do not find systematic evidence that hospital production functions differ between the two groups.

We recognize the limitations of estimating production functions for health outcomes using observational data, and so 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 measuring intensity at the level of the ambulance service. And while the results are broadly similar for input choices, we acknowledge that the model does not predict why such input choices vary. Finally, the reduced-form regression coefficients must be interpreted cautiously, since the estimates are based on comparisons of marginal patients receiving treatment in Hospital A but not in Hospital B, and may themselves be correlated with other unobserved inputs or patient characteristics. For example, lowering inefficient hospital 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). However, when randomized trials are available (typically for effective Category 1 treatments), the causal treatment effects found in the trials are consistent with regression results.

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 (Garber and Skinner, 2008). Our research sidesteps this debate, because interpreting coefficients as “the” marginal value of greater spending implicitly assumes that all inputs are optimally determined, an assumption rejected by the data. That said, we find a generally positive association between spending and health outcomes, although the coefficient is sensitive to the model specification.

What are the implications of these results? As a first step, we suggest that identifying hospital quality based on input use can provide information about high- and low-performing hospitals (e.g., Ganguli et al., 2021), although identifying the structural forces behind input misallocation may be more challenging. Experiments seeking to alight payment incentives for hospitals with improving quality and reducing costs suggest that some basic Category I treatments can be improved (e.g., Colla et al., 2012). Yet many institutions face challenges with the overuse of expensive Category III treatments because of lack of knowledge, biases, litigation concerns, and budgetary pressure. Clinical practice patterns are slow to change, and these activities may indirectly fund other activities, or clinicians may fail to understand 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 proven effectiveness of their value (Cutler et al., 2019). Despite these caveats, policies focused on measuring and potentially reducing misallocation in health care could improve productivity in a sector comprising one-fifth of the U.S. economy.

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, and Arjun Venkatesh, 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." American Economic Review 106(12): 3730-64.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, 2015. "Identification Properties of Recent Production Function Estimators." Econometrica 83(6): 2411-2451.

Ahmed, Nadeem, David Carrick, Jamie Layland, Keith G. Oldroyd, and Colin Berry, 2013. "The Role of Cardiac Magnetic Resonance Imaging (MRI) in Acute Myocardial Infarction (AMI)." Heart, Lung and Circulation 22(4): 243-255.

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

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

Bangalore, Sripal, Harikrishna Makani, Martha Radford, Kamia Thakur, Bora Toklu, et al. "Clinical Outcomes with β-blockers for Myocardial Infarction: a Meta-Analysis of Randomized Trials." The American Journal of Medicine 127(10): 939-953.

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

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

Bils, M., Klenow, P. J., and Ruane, C. (2021) "Misallocation or mismeasurement?" Journal of Monetary Economics, 124 (Supplement), pp. S39-S56.

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.

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

Chan, David C., Matthew Gentzkow, and Chuan Yu, 2022a. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." Quarterly Journal of Economics.

Chan, David C., Jr, David Card, and Lowell Taylor, 2022b. "Is There a VA Advantage? Evidence from Dually Eligible Veterans." NBER Working Paper No. 29765.

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 Douglas O Staiger, Identifying Sources of Inefficiency in Healthcare, The Quarterly Journal of Economics, Volume 135, Issue 2, May 2020, Pages 785–843.

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

Colla, Carrie H., David E. Wennberg, Ellen Meara, Jonathan S. Skinner, Daniel Gottlieb, et al., 2012. "Spending Differences Associated with the Medicare Physician Group Practice Demonstration." JAMA 308(10): 1015-1023.

Colla CH, Sequist TD, Rosenthal MB, Schpero WL, Gottlieb DJ, Morden NE, 2015a. “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, 2015b. “Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States.” J General Internal Med. 30(2):221-8.

Cooper, Zack, Olivia Stiegman, Chima D. Ndumele, Becky Staiger, and Jonathan Skinner, 2022. “Geographical Variation in Health Spending Across the US Among Privately Insured Individuals and Enrollees in Medicaid and Medicare." JAMA Open 5(7): e2222138.

Cooper, Zack, Joseph J. Doyle Jr, John A. Graves, and Jonathan Gruber, 2022. “Do Higher-Priced Hospitals Deliver Higher-Quality Care?” WP 29809, National Bureau of Economic Research.

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, Guy, Candace Gunnarsson, Liisa Laine, Michael Ryan, Seth Clancy, Gunnar Gunnarsson, Kimberly Moore, and William Irish, 2022. "The Unintended Consequences of Medicare's Wage Index Adjustment on Device-Intensive Hospital Procedures." American Journal of Managed Care 28(3): e96-e102.

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.

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, 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.

Einav, Liran, Amy Finkelstein, and Neale Mahoney, 2018. "Long-Term Care Hospitals: A Case Study in Waste." The Review of Economics and Statistics: 1-57.

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, Elliott. S., David. E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. Lee 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, Elliott. S., David. E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. Lee 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.

Ganguli, Ishani, Nancy E. Morden, Ching-Wen Wendy Yang, Maia Crawford, and Carrie H. Colla, 2021. "Low-Value Care at the Actionable Level of Individual Health Systems." JAMA Internal Medicine 181(11): 1490-1500.

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:

Griliches, Zvi, 2013. “Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." In Price Indexes and Quality Change, Harvard University Press: 55-87.

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.

Hollingsworth, Bruce, 2008. "The Measurement of Efficiency and Productivity of Health Care Delivery." Health Economics 17(10): 1107-1128.

Hsieh, Chang-Tai, and Peter J. Klenow, 2009. "Misallocation and manufacturing TFP in China and India." 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.

Josan, Kiranbir, Sumit R. Majumdar, and Finlay A. McAlister, 2008. "The Efficacy and Safety of Intensive Statin Therapy: A Meta-Analysis of Randomized Trials." Canadian Medical Association Journal 178(5): 576-584.

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.

Levinsohn, James, and Amil Petrin, 2003. "Estimating Production Functions using Inputs to Control for Unobservables." The Review of Economic Studies 70(2): 317-341.

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.

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.

Munson J.C., Morden N.E., Goodman, D.C., Valle, L.F., Wennberg, J.E. The Dartmouth Atlas of Medicare Prescription Drug Use. October 15, 2013.

Otero, Cristóbal, and Pablo Muñoz, 2022. "Managers and Public Hospital Performance." University of California, Berkeley.

Pauly, Mark, 1970. "Efficiency, Incentives and Reimbursement for Health Care." Inquiry 7: 114-131.

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 29(8): 1523-31.

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

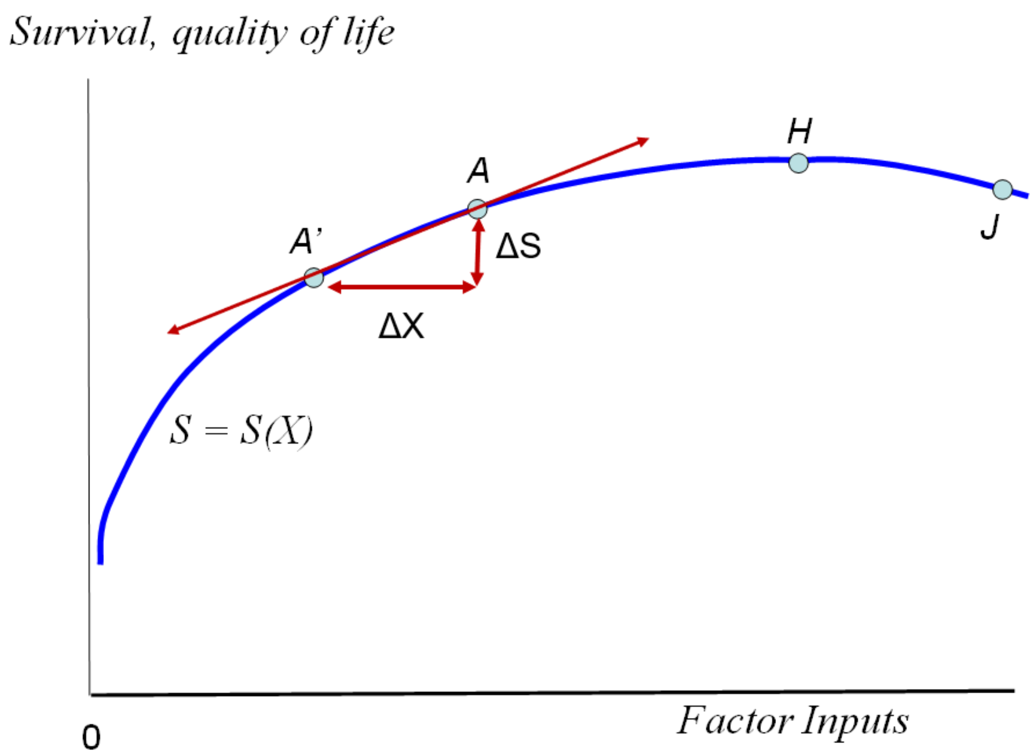


Figure 2: Sources of Differences in Outcomes: Total Factor Productivity and Input Misallocation

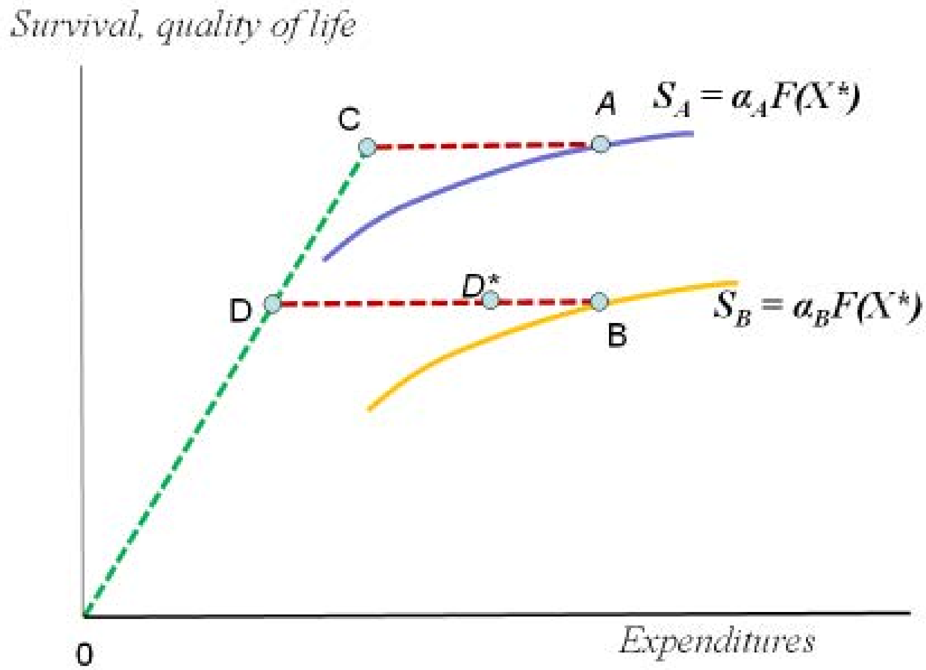


Figure 3: Association between Risk-Adjusted 30-Day Medicare Expenditures and One-Year Survival: 2007-2017

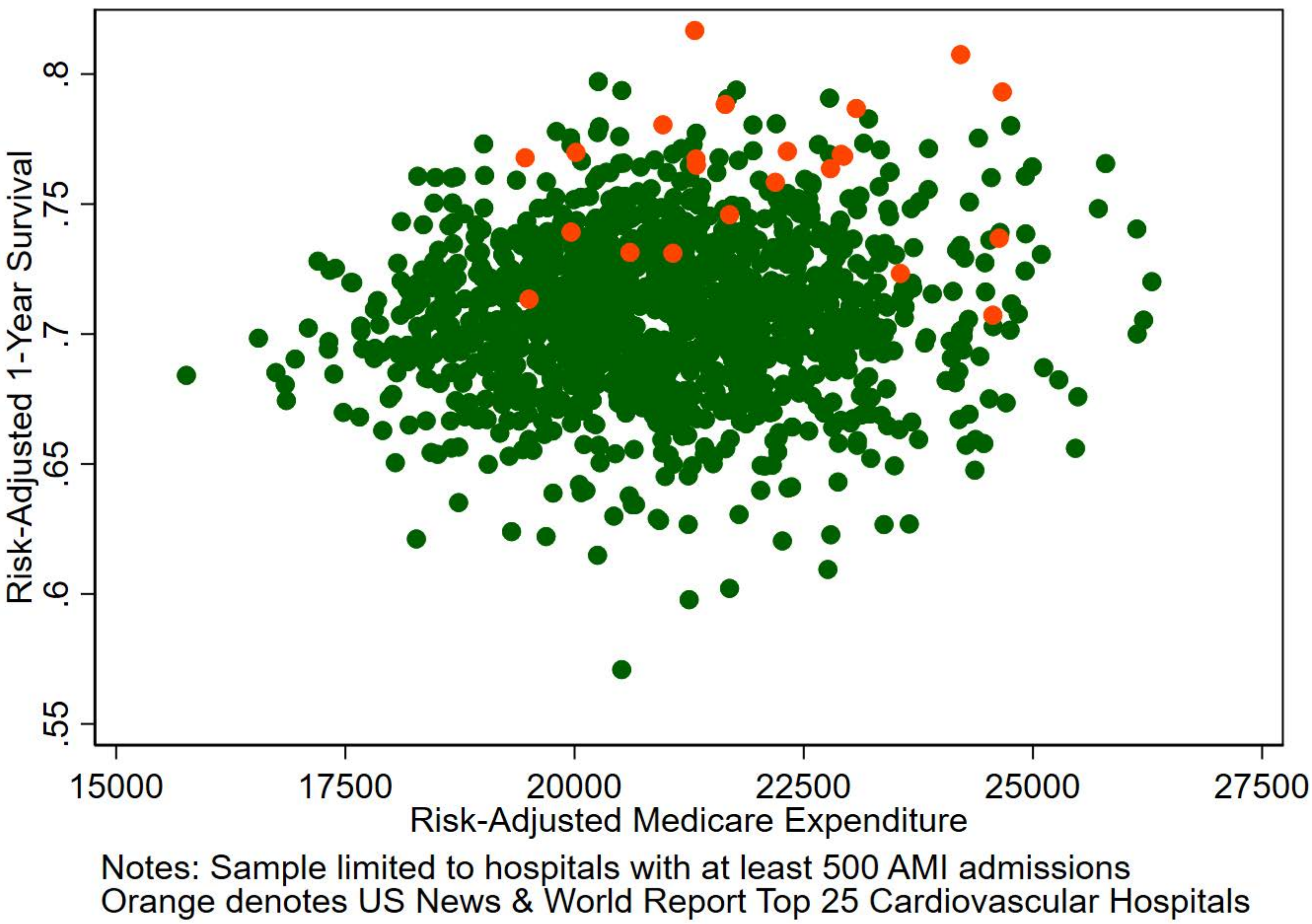
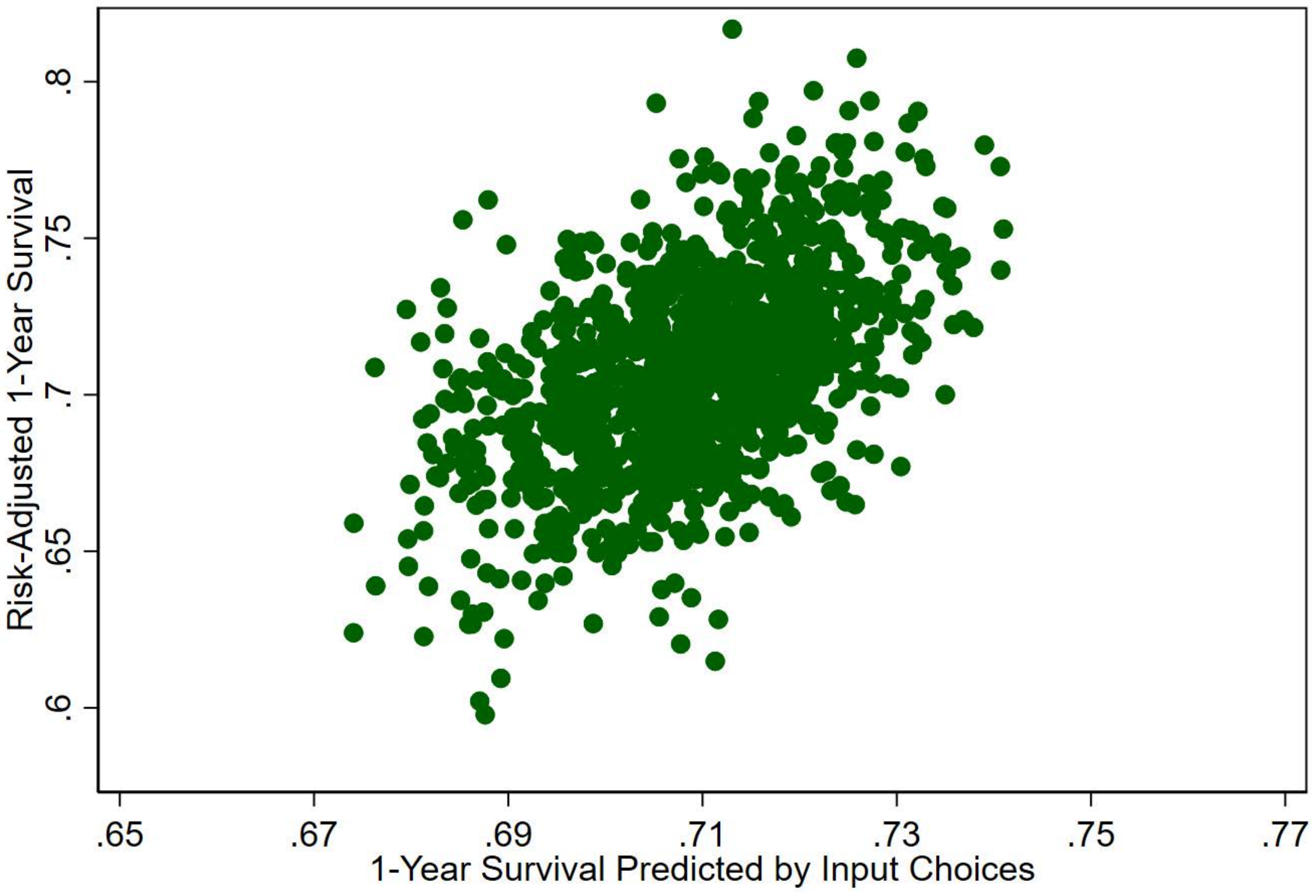


Figure 4: Association between Predicted 1-Year Survival Based on Misallocated Inputs and Actual 1-Year Risk-Adjusted Survival



Notes: Sample limited to hospitals with at least 500 AMI admissions

`

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

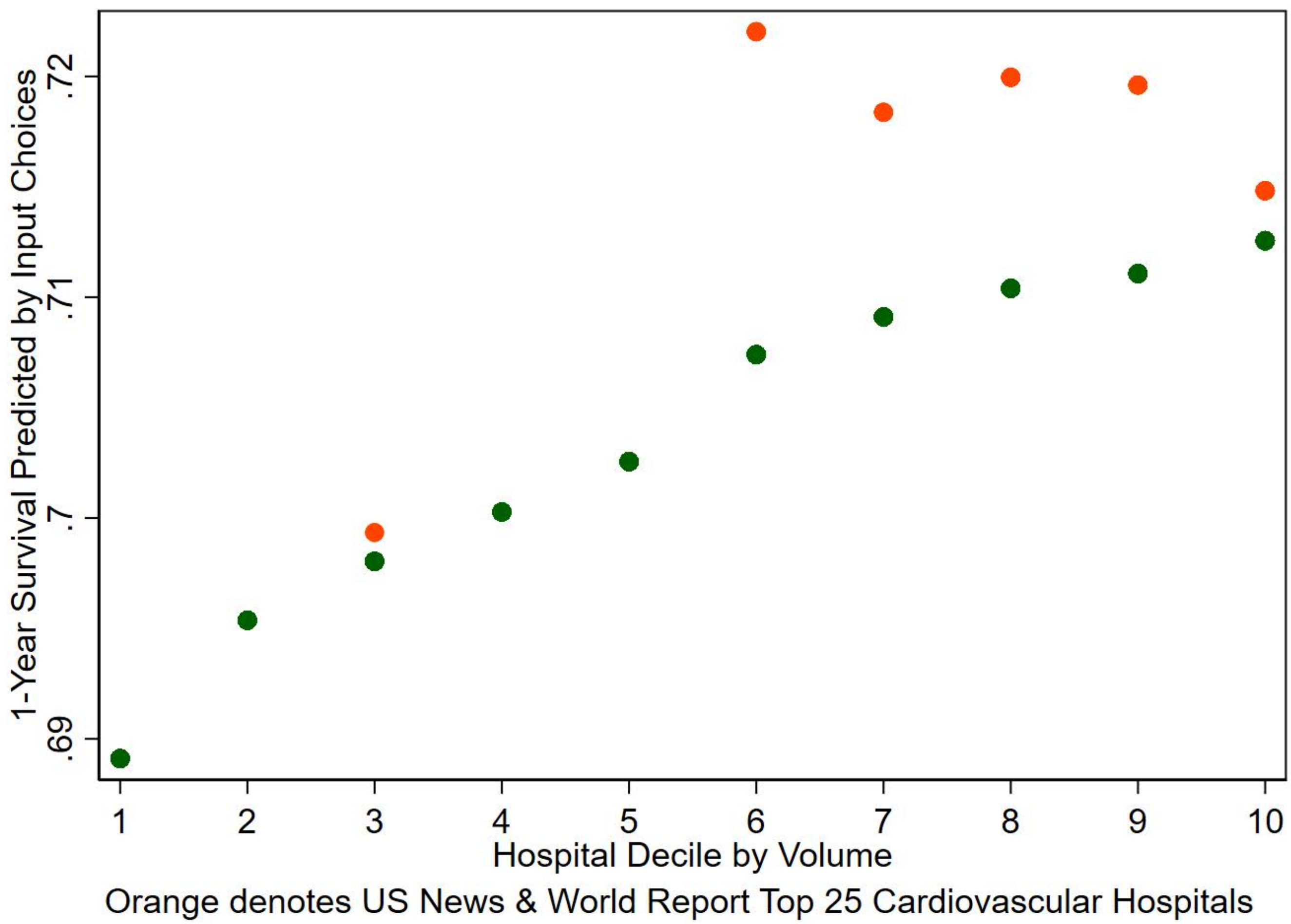


Table 1A: Summary Statistics for AMI Patients, 2007-17: Full Sample

	All	Lowest-Quartile of Medicare Expenditures	Highest-Quartile of Medicare Expenditures
	mean (s.d.)	mean (s.d.)	mean (s.d.)
Fraction Survive 1-Year	0.708 (0.455)	0.692 (0.462)	0.713 (0.452)
Fraction Survive 30-Day	0.863 (0.343)	0.855 (0.352)	0.866 (0.341)
30-Day Medicare Expenditures (000)	26.547 (21.331)	23.788 (18.508)	29.332 (24.149)
Average Operating Expenses (000)	21.462 (8.453)	21.117 (7.948)	22.257 (9.321)
Average Age	78.144 (8.320)	78.682 (8.448)	77.753 (8.221)
Fraction Female	0.480 (0.500)	0.495 (0.500)	0.475 (0.499)
Comorbidities			
Peripheral Vascular Disease	0.082 (0.275)	0.081 (0.273)	0.082 (0.275)
Chronic Lung Disease	0.176 (0.381)	0.186 (0.389)	0.168 (0.374)
Chronic Renal Failure	0.188 (0.391)	0.195 (0.396)	0.185 (0.389)
Congestive Heart Failure	0.392 (0.488)	0.397 (0.489)	0.395 (0.489)
Diabetes	0.273 (0.445)	0.278 (0.448)	0.273 (0.446)
HCC score*	1.339 (1.086)	1.315 (1.031)	1.380 (1.144)
Race/Ethnicity			
Asian	0.013 (0.111)	0.010 (0.102)	0.019 (0.136)
Black	0.074 (0.262)	0.073 (0.260)	0.081 (0.274)
Hispanic	0.015 (0.122)	0.009 (0.093)	0.024 (0.152)
Native American	0.005 (0.070)	0.005 (0.069)	0.005 (0.070)
Other	0.011 (0.102)	0.010 (0.097)	0.013 (0.113)
White	0.883 (0.322)	0.893 (0.309)	0.858 (0.349)
Subendocardial AMI	0.728 (0.445)	0.739 (0.439)	0.720 (0.449)
Zip Code Median Income	55869 (25672)	53465 (23213)	58008 (27367)
Medicare Advantage (% HRR)	0.257 (0.127)	0.259 (0.127)	0.259 (0.133)
Medicaid Enrollment**	0.163 (0.370)	0.164 (0.370)	0.175 (0.380)
Observations	1,617,039	363,193	418,382

Notes: * Hierarchical Category Condition score based on comorbidities 6 months prior to the Acute Myocardial Infarction (AMI). ** Medicaid enrollment during one of 6 months prior to the AMI. The standard deviation of Medicare expenditures is considerably greater than for CMS operating expenses because the Medicare expenditures are measured at the individual level, while the operating expenses are reported at the hospital level. The quartiles are by the number of hospitals, not the number of patients.

Table 1B: Summary Statistics for AMI Patients, 2007-17: Ambulance Admissions Sample

	All	Lowest-Quartile of Medicare Expenditures	Highest-Quartile of Medicare Expenditures
	mean (s.d.)	mean (s.d.)	mean (s.d.)
Fraction Survive 1-Year	0.673 (0.469)	0.675 (0.468)	0.672 (0.470)
Fraction Survive 30-Day	0.845 (0.362)	0.841 (0.365)	0.846 (0.361)
30-Day Medicare Expenditures (000)	27.249 (21.195)	24.831 (18.787)	29.804 (23961)
Average Operating Expenses (000)	21.519 (8.432)	21.707 (8.392)	22.019 (9.207)
Average Age	78.923 (8.421)	79.162 (8.463)	78.827 (8.396)
Fraction Female	0.505 (0.500)	0.505 (0.500)	0.501 (0.500)
Comorbidities			
Peripheral Vascular Disease	0.081 (0.273)	0.081 (0.273)	0.079 (0.270)
Chronic Lung Disease	0.181 (0.385)	0.182 (0.385)	0.177 (0.381)
Chronic Renal Failure	0.187 (0.390)	0.186 (0.389)	0.188 (0.391)
Congestive Heart Failure	0.413 (0.492)	0.400 (0.490)	0.420 (0.494)
Diabetes	0.265 (0.441)	0.260 (0.438)	0.269 (0.444)
HCC score*	1.417 (1.155)	1.344 (1.066)	1.484 (1.226)
Race/Ethnicity			
Asian	0.012 (0.107)	0.009 (0.094)	0.018 (0.135)
Black	0.073 (0.260)	0.058 (0.233)	0.072 (0.258)
Hispanic	0.015 (0.121)	0.007 (0.081)	0.027 (0.162)
Native American	0.004 (0.065)	0.006 (0.075)	0.004 (0.062)
Other	0.010 (0.098)	0.008 (0.090)	0.013 (0.111)
White	0.886 (0.317)	0.913 (0.282)	0.866 (0.340)
Subendocardial AMI	0.707 (0.455)	0.691 (0.462)	0.708 (0.455)
Zip Code Median Income	54367 (25013)	53485 (22172)	58291 (29024)
Medicare Advantage (% HRR)	0.265 (0.124)	0.255 (0.117)	0.277 (0.138)
Medicaid Enrollment**	0.187 (0.390)	0.160 (0.366)	0.209 (0.407)
Observations	436,519	108,481	105,866

Notes: * Hierarchical Category Condition score based on comorbidities 6 months prior to the Acute Myocardial Infarction (AMI). ** Medicaid enrollment during one of 6 months prior to the AMI. The standard deviation of Medicare expenditures is considerably greater than for CMS operating expenses because the Medicare expenditures are measured at the individual level, while the operating expenses are reported at the hospital level. The quartiles are by the number of hospitals, not the number of patients.

Table 2A: Association between Expenditures and One-Year Survival: Full Sample

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
Log Medicare Expenditures*	0.0317		0.0527	0.0167	0.0010
	(0.0067)		(0.0055)	(0.0037)	(0.0044)
Log Operating Expenses**		0.0319			
		(0.0025)			
Zip code fixed effects			X		
Hospital Random Effects				X	
Hospital Fixed Effects					X
R^2	0.179	0.180	0.199	0.179	0.179
σ_μ				0.0253	

Notes: All Regressions include full set of risk-adjustment covariates, see Appendix A.1. N = 1,617,039, including 2,024 hospitals; σ_μ denotes the standard error of the hospital-level random effect. * Average log hospital-level 30-day Medicare expenditures (with individual leave-out); ** Log per-admission operating expenditure, from CMS cost reports.

Table 2B: Association between Expenditures and One-Year Survival: Ambulance Sample

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5
Log Medicare Expenditures*	0.0226 (0.0197)		0.0539 (0.0206)	0.0179 (0.0107)	0.0162 (0.0206)
Log Operating Expenses**		0.0528 (0.0041)			
Zip code fixed effects			X		
Ambulance Random Effects				X	
Ambulance Fixed Effects					X
R^2	0.180	0.181	0.234	0.180	0.180
σ_μ				0.0258	

Notes: All Regressions include full set of risk-adjustment covariates, see Appendix A.1. N = 436,519, including 2,990 ambulance services; σ_μ denotes the standard error of the ambulance-level random effect. * Average log ambulance-level 30-day Medicare expenditures (with individual leave-out); ** Log per-admission operating expenditure, from CMS cost reports, aggregated at the ambulance service level (with individual leave-out).

Table 3: Regressions Estimates Explaining 1-Year Survival with Individual Measures of Misallocation: Full and Ambulance Samples

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Sample	Full Sample	Full Sample	Full Sample	Ambulance Sample	Ambulance Sample	Ambulance Sample
Log Medicare Expenditures *	0.0193	0.0099	0.0020	0.0500	0.0379	0.0233
	(0.0052)	(0.0043)	(0.0050)	(0.0222)	(0.0126)	(0.0233)
Cat I: Beta-blocker	0.0561	0.0162	-0.0483	0.0840	0.0782	-0.8452
	(0.0134)	(0.0110)	(0.0156)	(0.0550)	(0.0282)	(0.4600)
Cat I: Statins	0.0466	0.0666	0.0104	0.0198	0.0920	-0.2889
	(0.0119)	(0.0096)	(0.0138)	(0.0543)	(0.0248)	(0.4847)
Cat I: Ace Inhibitor/ARB	-0.0084	-0.0116	-0.0430	0.0169	0.0036	-0.8834
	(0.0124)	(0.0097)	(0.0147)	(0.0404)	(0.0239)	(0.3627)
Cat I: Early PCI	0.1090	0.0746	0.0241	0.1076	0.0814	0.0827
	(0.0067)	(0.0046)	(0.0067)	(0.0219)	(0.0133)	(0.0303)
Cat II: Late PCI (bottom quartile)	-0.0017	-0.0011	-0.0000	-0.0006	-0.0026	-0.0024
	(0.0010)	(0.0009)	(0.0010)	(0.0028)	(0.0019)	(0.0030)
Cat II: Late PCI (top quartile)	0.0020	0.0021	0.0014	0.0029	-0.0002	0.0003
	(0.0009)	(0.0009)	(0.0010)	(0.0022)	(0.0018)	(0.0028)
Cat II: # Doctors seen (bottom quartile)	-0.0027	-0.0033	-0.0006	-0.0043	-0.0085	-0.0034
	(0.0014)	(0.0012)	(0.0014)	(0.0034)	(0.0020)	(0.0036)
Cat II: # Doctors seen (top quartile)	0.0073	0.0059	0.0027	0.0076	0.0090	0.0034
	(0.0014)	(0.0011)	(0.0013)	(0.0033)	(0.0022)	(0.0041)
Cat II: # Scans (bottom quartile)	0.0026	0.0019	0.0005	0.0033	0.0022	0.0017
	(0.0011)	(0.0010)	(0.0011)	(0.0028)	(0.0021)	(0.0034)
Cat II: # Scans (top quartile)	-0.0053	-0.0024	-0.0018	-0.0062	-0.0062	-0.0010
	(0.0012)	(0.0010)	(0.0011)	(0.0027)	(0.0021)	(0.0030)
Cat II: ICU/CCU days (bottom quartile)	0.0046	0.0034	-0.0013	0.0050	0.0078	0.0011
	(0.0013)	(0.0011)	(0.0015)	(0.0034)	(0.0020)	(0.0049)
Cat II: ICU/CCU days (top quartile)	0.0002	0.0002	0.0000	0.0038	0.0007	-0.0015
	(0.0011)	(0.0011)	(0.0013)	(0.0032)	(0.0021)	(0.0041)
Cat III: IRH or LTCH admission**	-0.0156	-0.0623	-0.0593	-0.0015	0.0057	0.0018
	(0.0163)	(0.0142)	(0.0183)	(0.0022)	(0.0012)	(0.0025)
Cat III: SNF Days**	-0.0021	0.0013	-0.0000	-0.0122	-0.0043	-0.0205
	(0.0006)	(0.0005)	(0.0006)	(0.0155)	(0.0105)	(0.0207)
Cat III: Home Health Care	-0.0002	0.0037	0.0016	-0.1307	-0.1377	-0.1808
	(0.0056)	(0.0041)	(0.0050)	(0.0586)	(0.0359)	(0.0892)
Cat III: Abdomen CT measure***	0.0014	-0.0039	-0.0055	0.0201	0.0109	0.0141
	(0.0050)	(0.0038)	(0.0049)	(0.0230)	(0.0114)	(0.0940)
Cat III: Thorax CT measure***	-0.0170	-0.0201	-0.0042	-0.0514	-0.0587	-0.1887

	(0.0079)	(0.0067)	(0.0082)	(0.0440)	(0.0253)	(0.1829)
Zip Code Fixed Effects	X			X		
Hospital/Ambulance Random Effects		X			X	
Hospital/Ambulance Fixed Effects			X			X
Observations	1,617,039	1,617,039	1,617,039	436,519	436,519	436,519
R^2	0.200	0.181	0.181	0.235	0.168	0.168
σ_μ		0.0219			0.0228	

All regressions include risk-adjustment (see Appendix Table A.1. * Log 30-day Medicare Expenditures for hospitals (first two columns) or ambulance services (second two columns) with leave-out rule. ** Admission to either an inpatient rehabilitation hospital (IRH) or a long-term care hospital (LTCH); SNF denotes skilled nursing facility. *** Based on Hospital Compare data (for entire patient population, not just AMI patients).

Table 4: Regressions Estimates of Survival with Principal Component Estimates of Misallocation: Full and Ambulance Samples

VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Full Sample	Full Sample	Full Sample	Ambulance Sample	Ambulance Sample	Ambulance Sample
Log Hosp. Expenditures*	0.0515	0.0272	0.0173	0.0701	0.0500	0.0384
	(0.0058)	(0.0041)	(0.0049)	(0.0267)	(0.0123)	(0.0237)
PCA: Category I	0.0125	0.0088	0.0059	0.0092	0.0077	0.0132
	(0.0004)	(0.0004)	(0.0007)	(0.0013)	(0.0006)	(0.0054)
PCA: Category II	-0.0049	-0.0027	-0.0048	-0.0020	-0.0001	-0.0021
	(0.0007)	(0.0005)	(0.0006)	(0.0017)	(0.0008)	(0.0022)
PCA: Category III	-0.0006	-0.0026	-0.0015	-0.0027	-0.0057	-0.0081
	(0.0005)	(0.0004)	(0.0005)	(0.0015)	(0.0007)	(0.0030)
Zip Code Fixed Effects	X			X		
Hospital / Ambulance Random Effects		X			X	
Hospital / Ambulance Fixed Effects			X			X
Observations	1,617,039	1,617,039	1,617,039	436,519	436,519	436,519
R^2	0.200	0.180	0.180	0.234	0.168	0.168
σ_μ		0.0229			0.0247	
Correlation between predicted survival and hospital/ ambulance fixed-effects[a]			0.035			0.060

Notes: All Regressions include full set of risk-adjustment covariates; see Appendix A.1; σ_μ denotes the standard error of the ambulance-level random effect. * Average log hospital 30 day Medicare expenditures (Columns 1 and 2) and ambulance-level expenditures (Columns 3 and 4), all with individual leave-out). [a] This measures the correlation coefficient between the combined predicted impact on survival of both risk adjustment variables (Z) and input choice variables (X), and the hospital or ambulance-level fixed effects. That they are both so small in magnitude is why fixed-effect and random-effect estimates are similar.

Appendix A: Additional Regression Tables

Appendix Table A.1: Risk Adjustment Models for Log Medicare Expenditures and Survival

	Model 1	Model 2	Model 3
	Log 30-Day Medicare Expenditures	1 Year Survival	1 Year Survival
	Full Sample	Full Sample	Ambulance Sample
Peripheral Vascular Disease	-0.0280	0.0105	0.0101
	(0.0023)	(0.0013)	(0.0024)
Chronic Lung Disease	-0.0358	-0.0283	-0.0260
	(0.0018)	(0.0010)	(0.0019)
Chronic Renal Failure	0.0249	-0.0587	-0.0572
	(0.0017)	(0.0010)	(0.0019)
Cancers	-0.0533	-0.0694	-0.0747
	(0.0027)	(0.0018)	(0.0036)
Cancer: Metastatic	-0.0801	-0.3229	-0.304
	(0.0050)	(0.0036)	(0.0070)
Congestive Heart Failure	0.1962	-0.1198	-0.118
	(0.0020)	(0.0009)	(0.0016)
Liver Disease	-0.0726	-0.0864	-0.0714
	(0.0087)	(0.0052)	(0.0099)
Diabetes	-0.0449	0.0452	0.0481
	(0.0018)	(0.0008)	(0.0015)
Rheumatologic Disease	-0.0859	0.0277	0.0285
	(0.0041)	(0.0025)	(0.0049)
Dementia	-0.1712	-0.1410	-0.136
	(0.0032)	(0.0022)	(0.0038)
Female age 65-69	-0.0484	0.0088	0.0120
	(0.0031)	(0.0013)	(0.0027)
Male age 70-74	0.0297	-0.0209	-0.0248
	(0.0027)	(0.0011)	(0.0024)
Female age 70-74	-0.0301	-0.0140	-0.0130
	(0.0028)	(0.0013)	(0.0027)
Male age 75-79	0.0321	-0.0535	-0.0642
	(0.0027)	(0.0013)	(0.0027)
Female age 75-79	-0.0313	-0.0393	-0.0448
	(0.0028)	(0.0014)	(0.0028)
Male age 80-84	-0.0349	-0.1041	-0.118
	(0.0028)	(0.0016)	(0.0032)
Female age 80-84	-0.0894	-0.0826	-0.0892

	(0.0028)	(0.0015)	(0.0030)
Male age 85-89	-0.1467	-0.1795	-0.200
	(0.0032)	(0.0019)	(0.0036)
Female 85-89	-0.1905	-0.1533	-0.162
	(0.0030)	(0.0017)	(0.0032)
Male age 90+	-0.2953	-0.2988	-0.319
	(0.0037)	(0.0024)	(0.0044)
Female age 90+	-0.3214	-0.2782	-0.289
	(0.0033)	(0.0018)	(0.0036)
Native American	0.0643	-0.0023	-0.00704
	(0.0104)	(0.0049)	(0.0092)
Hispanic	-0.0324	0.0287	0.0266
	(0.0053)	(0.0029)	(0.0060)
Other Race/Ethnicity	0.0089	-0.0021	0.00267
	(0.0059)	(0.0033)	(0.0064)
Asian	0.0152	0.0135	0.0205
	(0.0058)	(0.0032)	(0.00680)
Black	-0.0530	-0.0008	-0.0011
	(0.0031)	(0.0016)	(0.0033)
Location of MI: anterior	0.1897	0.0948	0.0909
	(0.0036)	(0.0023)	(0.0041)
Location of MI: inferior1	0.1450	0.1346	0.1410
	(0.0036)	(0.0022)	(0.0039)
Location of MI: Right	0.1729	0.1117	0.1150
	(0.0060)	(0.0035)	(0.0063)
Location of MI: Subendocardial	0.0352	0.1648	0.1490
	(0.0035)	(0.0021)	(0.0035)
Location of MI: Other	0.1059	0.0912	0.0906
	(0.0053)	(0.0031)	(0.0058)
2nd Quintile Income*	-0.0016	0.0020	0.00200
	(0.0021)	(0.0011)	(0.0023)
3rd Quintile Income*	-0.0141	0.0029	0.00231
	(0.0023)	(0.0011)	(0.0023)
4th Quintile Income*	-0.0227	0.0021	0.0017
	(0.0024)	(0.0012)	(0.0024)
5th Quintile Income*	-0.0328	0.0073	0.0056
	(0.0026)	(0.0013)	(0.0026)
Medicaid Enrollment**	0.0132	-0.0300	-0.0337
	(0.0018)	(0.0012)	(0.0025)
Fraction HRR MA***	-0.0134	-0.0024	0.0034
	(0.0149)	(0.0069)	(0.0126)
Second Quintile HCC	0.0669	-0.0285	-0.0372
	(0.0021)	(0.0010)	(0.0020)
Third Quintile HCC	0.0827	-0.0609	-0.0760
	(0.0024)	(0.0011)	(0.0022)
Fourth Quintile HCC	0.0921	-0.1161	-0.134

	(0.0027)	(0.0012)	(0.0024)
Fifth Quintile HCC	0.1216	-0.2584	-0.269
	(0.0032)	(0.0014)	(0.0026)
Constant	9.8765	0.8203	0.820
	(0.0067)	(0.0034)	(0.0060)
Observations	1,614,071	1,617,039	436,519
R^2	0.0531	0.1726	0.168

Notes: * Income measured at the Zip Code level. ** Dual eligibility; Medicaid enrollment during one of 6 months prior to the AMI. *** Fraction of Medicare beneficiaries enrolled in a Medicare Advantage plan, by HRR. Includes year dummy variables.

Appendix Table A.2: Results from Principal Components Analysis: Proportion of Variance Explained, and Eigenvectors, of the First Component

	Full Sample	Ambulance Sample
Category 1 (Proportion of variance of first component)	0.683	0.698
Beta-blocker	0.534	0.538
Statins	0.543	0.546
Ace Inhibitor/ARB	0.484	0.486
Early PCI	0.431	0.420
Category 2 (Proportion of variance of first component)	0.400	0.462
Late PCI	0.410	0.286
# Doctors seen	0.523	0.577
# Scans	0.569	0.625
ICU/CCU days	0.484	0.441
Category 3 (Proportion of variance of first component)	0.310	0.363
IRH or LTCH admission*	0.342	0.348
SNF Days	-0.299	-0.222
Home Health Care	0.228	0.479
Abdomen CT measure**	0.611	0.554
Thorax CT measure**	0.607	0.541

* Admission to either an inpatient rehabilitation hospital (IRH) or a long-term care hospital (LTCH). ** Based on Hospital Compare data (for entire patient population, not just AMI patients). N = 21,944 hospital/years.

Appendix Table A.3: Regressions Estimates of 30-Day Survival with Principal Component Estimates of Misallocation: Full and Ambulance Samples

VARIABLES	Model 1	Model 2	Model 3	Model 4
	Full Sample	Full Sample	Ambulance Sample	Ambulance Sample
Log Hosp. Medicare Expenditures (30-day)	0.0146	0.0002	0.0528	0.0328
	(0.0040)	(0.0033)	(0.0154)	(0.0080)
PCA prediction for Category 1	0.0070	0.0051	0.0057	0.0052
	(0.0003)	(0.0003)	(0.0008)	(0.0004)
PCA prediction for Category 2	-0.0006	0.0014	0.0008	0.0030
	(0.0005)	(0.0004)	(0.0013)	(0.0005)
PCA prediction for Category 3	-0.0009	-0.0018	-0.0014	-0.0031
	(0.0004)	(0.0003)	(0.0012)	(0.0004)
Zip Code Fixed Effects	X		X	
Hospital Random Effects		X		
Ambulance Service Random Effects				X
Observations	1,617,039	1,617,039	436,519	436,519
R^2	0.104	0.083	0.1411	0.085
σ_μ		0.0173		0.000**

Notes: All Regressions include full set of risk-adjustment covariates; see Appendix A.1; σ_μ denotes the standard error of the ambulance-level random effect. * Average log hospital 30-day Medicare expenditures (Columns 1 and 2) and ambulance-level expenditures (Columns 3 and 4), all with individual leave-out. There are 2,024 hospitals and 2,990 ambulance services included in the analysis with minimum sample size of 50. ** An estimate for σ_μ of 0 means that there is not a statistically significant contribution of the random effects to the predictive value of the model.

Appendix Table A.4A: Principal Component Regression Results with Sensitivity Analysis for Minimum Cell Sizes (Full Sample)

VARIABLES	Model 1	Model 2	Model 3	Model 4
	Cell ≥25	Cell ≥25	Cell ≥75	Cell ≥75
Log Hosp. Medicare Expenditures	0.0615	0.0358	0.0512	0.0275
	(0.0049)	(0.0036)	(0.0066)	(0.0046)
PCA prediction for Category 1	0.0118	0.0080	0.0124	0.0088
	(0.0004)	(0.0003)	(0.0005)	(0.0004)
PCA prediction for Category 2	-0.0063	-0.0043	-0.0041	-0.0017
	(0.0007)	(0.0005)	(0.0007)	(0.0005)
PCA prediction for Category 3	-0.0009	-0.0022	-0.0006	-0.0028
	(0.0005)	(0.0004)	(0.0005)	(0.0004)
Zip Code Fixed Effects	X		X	
Hospital Random Effects		X		X
Observations	1,682,076	1,682,076	1,536,197	1,536,197
R^2	0.200	0.182	0.200	0.180
σ_μ		0.0267		0.0196

All regressions include risk-adjustment (see Appendix Table A.1. For minimum sample size of 25, there were 2,534 hospitals, for the minimum sample size of 75, 1,680.

Appendix Table A.4B: Principal Component Regression Results with Sensitivity Analysis for Minimum Cell Sizes (Ambulance Sample)

VARIABLES	Model 1	Model 2	Model 3	Model 4
	Cell ≥25	Cell ≥25	Cell ≥75	Cell ≥75
Log Hosp. Medicare Expenditures	0.0362	0.0243	0.0929	0.0715
	(0.0150)	(0.0098)	(0.0335)	(0.0117)
PCA prediction for Category 1	0.0098	0.0077	0.0099	0.0074
	(0.0010)	(0.0006)	(0.0016)	(0.0005)
PCA prediction for Category 2	-0.0024	-0.0010	-0.0029	0.0004
	(0.0013)	(0.0007)	(0.0020)	(0.0007)
PCA prediction for Category 3	0.0007	-0.0041	-0.0026	-0.0056
	(0.0012)	(0.0006)	(0.0019)	(0.0006)
Zip Code Fixed Effects	X		X	
Hospital Random Effects		X		X
Observations	540,218	540,218	321,463	321,463
R^2	0.200	0.182	0.243	0.181
σ_{μ}		0.0288		0.000

All regressions include risk-adjustment (see Appendix Table A.1. For minimum sample size of 25, there were 2,534 ambulance services, for the minimum sample size of 75, 2,068. ** An estimate for σ_{μ} of 0 means that there is not a statistically significant contribution of the random effects to the predictive value of the model.

Appendix Table A.5: Regressions Estimates of Survival with Principal Component Estimates of Misallocation: Further Robustness Tests

VARIABLES	Model 1		Model 2	Model 3	Model 4	Model 5
	Teaching Hospitals	Non-Teaching Hospitals	Early/Late α	Early/Late α	CMS Operating Costs	CMS Operating Costs
Log Hospital Expenditures [a]	0.0365	0.0253	0.0306	0.0499	0.0122	-0.0011
	(0.0108)	(0.0044)	(0.0042)	(0.0125)	(0.0017)	(0.0030)
PCA: Category I	0.0069	0.0086	0.0090	0.0076	0.0087	0.0060
	(0.0009)	(0.0004)	(0.0003)	(0.0007)	(0.0004)	(0.0007)
PCA: Category II	-0.0046	-0.0029	-0.0025	-0.0002	-0.0011	-0.0037
	(0.0011)	(0.0005)	(0.0005)	(0.0009)	(0.0004)	(0.0006)
PCA: Category III	-0.0045	-0.0023	-0.0033	-0.0057	-0.0020	-0.0014
	(0.0014)	(0.0004)	(0.0004)	(0.0007)	(0.0004)	(0.0005)
Hospital-Level Sample	X		X		X	X
Ambulance-Level Sample				X		
Hospital / Ambulance Random Effects	X		X	X	X	
Hospital Fixed Effects						X
Observations	1,617,039		1,617,039	436,519	1,616,480	1,616,480
R^2	0.181		0.181	0.181	0.181	0.180
σ_μ	0.0228		0.0272	0.0264	0.0229	

Notes: All Regressions include full set of risk-adjustment covariates; see Appendix A.1; σ_μ denotes the standard error of the ambulance-level or hospital-level random effect. *Average log hospital 30-day Medicare expenditures (Models 1 and 2) and ambulance-level expenditures (Model 3), with individual-leave-out means; per patient operating expenditures from CMS cost reports (Models 4 and 5.. For a few hospital/years, CMS operating costs are not available.

Appendix B: 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:

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

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

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

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

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:

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

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):

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

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

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):

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

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:

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

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 $\beta' = \sum_k \beta_k$.