NBER WORKING PAPER SERIES

PRODUCTIVITY VARIATION AND INPUT MISALLOCATION: EVIDENCE FROM HOSPITALS

Amitabh Chandra
Carrie Colla
Jonathan Skinner

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

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023, Revised July 2026

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, Jack Marshall, Edward Norton, Douglas Staiger, Chad Syverson, Taylor Watson, John Wennberg, and especially Joe Doyle for insightful comments. Seminar participants at Dartmouth, Chicago, Houston, Toulouse, Tulane, Wisconsin, NBER Productivity, UCLA, Stanford, and the HEAL and ASHEcon conferences 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 Colla, and Jonathan Skinner
NBER Working Paper No. 31569
August 2023, Revised July 2026
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 explain these variations, we use simple economic insights to test the extent to which U.S. hospitals misallocate inputs by examining if the marginal return to specific inputs, conditional on total cost, differs from zero. This test does not require knowledge of the production function or specific input prices. We find large and systematic deviations from efficiency: Holding spending fixed, misallocation is estimated to account for as much as 3.3 percentage points (11% of mortality) across hospitals, and explains roughly 25% of hospital productivity variation. These results suggest large gains from improving how—not how much—hospitals spend.

Amitabh Chandra
Harvard University
Harvard Business School,
and Harvard Kennedy School
and NBER
amitabh_chandra@harvard.edu

Carrie Colla
Dartmouth College
carrie.h.colla@dartmouth.edu

Jonathan Skinner
Dartmouth College
Department of Economics
and NBER
jonathan.skinner@dartmouth.edu

1 Introduction

A central theme in modern macroeconomics is that misallocation of inputs—caused by frictions such as distorted prices, financial constraints, or regulatory barriers—can substantially depress aggregate productivity. Seminal contributions by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008) showed that the reallocation of resources from low-productivity to high-productivity firms could, in theory, raise total factor productivity (TFP) by 30–60% in countries such as China and India. More recently, Baqaee and Farhi (2020) estimated that eliminating misallocation in the U.S. could increase aggregate TFP by up to 15%. These insights have stimulated a broader literature quantifying misallocation across a range of settings, from oil extraction (Asker et al. 2019), to finance (Banerjee and Moll 2010), to manufacturing (Haltiwanger et al. 2018).

While this literature has transformed macroeconomic understanding of productivity dispersion, it has rarely been applied to health care—a sector that accounts for nearly 20% of U.S. GDP and is high-stakes given the implications for health.[1] Hospitals, which are firms that deliver acute health care, exhibit wide variation in measured productivity (Chandra et al. 2016a,b; Hull 2020; Chandra and Staiger 2020), with prior studies noting that productivity differences may reflect the quality of management (Bloom et al. 2020; Tsai et al. 2015; Muñoz and Otero 2025; McConnell et al. 2013), adoption of information technology (Lee et al. 2013), physician beliefs (Cutler et al. 2019; Chandra and Staiger 2020), diagnostic and triage capabilities (Abaluck et al. 2021; Chan et al. 2022; Mullainathan and Obermeyer 2022) and capital allocation (Banerjee and Moll 2010). But inefficient allocation of inputs—e.g., underuse of effective treatments and overuse of ineffective ones—offers a complementary explanation that aligns naturally with theories of misallocation, where good management and operational efficiency can potentially reduce input misallocation and increase aggregate TFP.

[1] An exception is Raja (2025) who considers the impact of minimum nursing staffing regulations on input misallocation.

In this paper, we measure the extent of input misallocation in the U.S. hospital sector and use this misallocation to understand what share of productivity variation is caused by input misallocation. We use data on Medicare patients with acute myocardial infarction (AMI)—a condition with standardized treatment guidelines and a well-defined outcome of survival. Our empirical test for misallocation uses simple microeconomic intuition: Under efficient production, the marginal benefit of any input, conditional on cost, should be zero. That is, once the efficient hospital has spent a given amount, at the margin reallocating that spending across inputs should not improve outcomes. This approach allows us to detect misallocation of inputs without requiring strong assumptions about the production function or input prices. The test captures differences across hospitals in returns to a given input (as in the macroeconomic literature which considers capital and labor distortions) as well as variations in returns per dollar spent for different inputs *within* hospitals.[2]

We implement this test using the universe of 1.6 million Medicare fee-for-service patients aged 65 and over hospitalized with AMI from 2007–2017. We measure inputs and reimbursements at the patient level and aggregate them to the firm level, supplementing with hospital-level operating costs from Medicare Cost Reports to capture the cost constraint. Outputs are measured using 1-year survival, a meaningful and measurable outcome that has been validated as causal, in the sense of not being contaminated by unmeasurable characteristics of patients (Chandra et al. 2023). The measurement of inputs is key for our test, but complicated by the fact that there are many inputs. To make this feasible, we adopt a clinically grounded and pre-specified taxonomy of medical inputs introduced by Chandra and Skinner (2012), which classifies treatments into three categories based on their expected productivity:

- Category I inputs are consistently high-productivity, with almost all patients receiving

[2] A large literature in economics and medicine measures input misallocation at the level of specific inputs (for example, Abaluck et al. (2021) for diagnostic testing, and Chandra and Staiger (2020) for a specific treatment, reperfusion, in the context of treating heart-attacks. We examine a related question by proposing an economy-wide test for misallocation across many distinct inputs.

the therapy benefiting from it. Examples include early same-day PCI (percutaneous coronary interventions, typically stents), and beta blockers and statins at discharge.

- Category II inputs exhibit heterogeneous treatment effects and diminishing returns. For example, PCI more than 24 hours after admission has less consistent benefits, while a rising number of different physicians may exhibit decreasing returns because of communication costs.(Becker and Murphy 1992; Baicker and Chandra 2004)

- Category III inputs are costly and low-productivity, with little or no evidence of effectiveness. These include extended stays in post-acute care or repeated imaging (e.g., serial MRIs) and high rates of home health care use with unclear clinical benefit (McKnight 2006; Doyle et al. 2017; Einav et al. 2018, 2026; O'Malley et al. 2023).

This taxonomy was developed more than a decade ago and provides a clinically meaningful and productivity-relevant classification of inputs.[3] It allows us to collapse the large dimensionality of the treatment space into a tractable set of variables, and in some specifications further reduce dimensionality using principal components analysis. When available, we benchmark regression coefficients against treatment effects from clinical trials to assess whether observed associations are plausibly causal.

A key empirical challenge in this setting is risk adjustment. If some hospitals appear more productive merely because they treat healthier patients, our estimates would conflate selection with misallocation. To address this, we follow several strategies. First, risk-adjusted measures of hospital mortality for conditions like heart attacks have been demonstrated to be forecast-unbiased, meaning they represent causal effects and are not confounded by patients' unobservable characteristics (Chandra et al. 2023). Second, expecting within-hospital or within-zipcode variation in input use to have less selection bias, we estimate both random and fixed-effects models, of which the latter exploits only within-hospital variation over time. Third, we follow Doyle et al. (2015, 2017) in exploiting ambulance service preferences as a

[3] See also Wennberg (2010) for a similar but distinct categorization.

quasi-random assignment mechanism: when patients are transported by different ambulance services with distinct hospital preferences, assignment to hospitals is plausibly exogenous. We use ambulance-service random and fixed effects models to isolate variation in input use and outcomes not driven by patient choice.

Across these identification strategies, a formal test for the null of efficient input choice is strongly rejected. Holding total expenditures constant, hospitals that use more Category I inputs have significantly higher survival, while those that use more Category II or III treatments tend to perform worse. Consistent with Chandra et al. (2023), we find little evidence of favorable selection into high-spending hospitals: the ambulance-based estimates are similar in magnitude to our full-sample regressions.

We next connect our results on input misallocation to productivity variation (measured as survival conditional on inputs) across hospitals (Chandra et al. 2016a). Like earlier work, we find large evidence of productivity variation in the hospital industry: conditional on expenditures and risk adjustment, one-year survival rates (which measure TFP) vary from 66.7% (10th percentile) to 74.9% (90th percentile) across hospitals. We further show that input misallocation explains one-quarter of these overall variations, with a 3.3 percentage point gap (11% of mortality) between the 10th and the 90th percentile explained by input choice alone. These results suggest that whatever the organizational or behavioral mechanism for higher productivity, that they are likely to reduce input misallocation as well. In other words, our paper identifies one novel internal process potentially distinct from diagnostic ability and procedural skill – the more efficient choice of treatment inputs – by which management and providers potentially improves outcomes without raising costs.

One reasonable question is whether variations in input choices might be explained by some hospitals exhibiting different production function parameters or facing different prices, in which case input variation conditional on costs would be optimal. We first show formally that in the Cobb-Douglas framework, the hospital fixed-effect absorbs hospital-level permanent differences in production functions and prices; that our results are so similar to the random-effects model suggests that such heterogeneity is not first-order. Furthermore, the magnitude of price differences required in a Cobb-Douglas framework to explain the more than three-fold variation in (risk-adjusted) input choice appears implausible.

Why do some hospitals exhibit such large differences in input choices? Previous work has suggested that variation in physician beliefs about the effectiveness of treatments may explain variations in Medicare expenditures at the regional level. We find evidence that these physician beliefs, as measured using vignettes for hypothetical patients in Cutler et al. (2019) are predictive of hospital input choices. Those in regions with a higher fraction of evidence-based physicians are predicted to use Category I highly effective treatments, while hospitals in regions with a larger fraction of more aggressive "cowboys" are predicted to use "Category II" treatments with uncertain clinical benefits; Category III treatments are less closely associated with physician beliefs.

Our findings have clear policy implications. They suggest that improving how hospitals spend money – by shifting resources toward high-value care and away from low-value treatments – can yield substantial health gains without increasing total spending (Colla 2014). Syverson (2011) emphasizes that large productivity dispersion in services often signals inefficiency rather than hidden quality differences. More broadly, the implication is that misallocation, a concept developed in macroeconomics to explain productivity gaps in manufacturing and trade (Hsieh and Klenow 2009; Restuccia and Rogerson 2008, 2017), applies similarly for inputs in service sectors like health care, where providers vary widely in skill, information, and incentives. These results suggest that insights from the broader productivity literature extend naturally to health care, and that straightforward economic tests of how hospitals allocate inputs can reveal meaningful productivity differences in a sector too often overlooked in aggregate productivity research.

2 Productivity and Input Misallocation

Why might health outcomes vary so much across hospitals? In Figure 1 we show a hypothetical production function in which average survival or quality of life (S) of patients admitted to the hospital is a function of two inputs x_1 and x_2, and a measure A of Hicks-neutral productivity; $S = Af(x_1, x_2)$. Average total cost per patient is $Cost = p_1x_1 + p_2x_2 + \Phi$ where p_k measures prices and Φ average fixed costs; these are shown on the horizontal axis of the Figure. Note that this measures the average costs and average outcomes per patient; thus the vertical axis measures the probability that a representative (risk-adjusted) patient survives. Point V corresponds to "flat of the curve" health care where spending more yields no benefit.

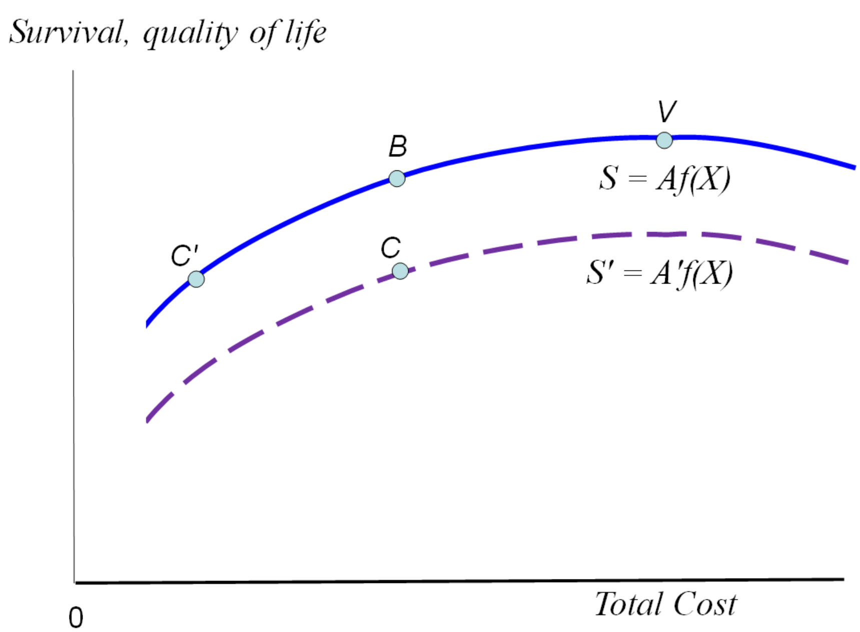

Figure 1: Productivity Differences in Health Care

One source of variation in outcomes across hospitals could come from differences in firm-level total factor productivity (TFP), A. For example, the dashed purple line S' assumes the identical production function $f(X)$ where X is the vector of inputs, but as drawn $A' < A$; this is the model estimated in Chandra et al. (2016a). In Figure 1, Hospital B and Hospital

C exhibit the same costs, but differ with regard to survival.[4]

A second possibility exists that could also lead to differences in productivity without appealing to variation in A: Misallocation, in which aggregate capital and labor (or in our case, total costs) are held constant, but the distribution of inputs across firms is suboptimal (Hsieh and Klenow 2009; Restuccia and Rogerson 2017). In the case of misallocation in the macroeconomic literature, the underlying cause might be distortionary sectoral taxes or tariffs, regulations, or "crony-capitalism" that results in resources being restricted in firms with higher marginal returns (for example, Hospital C' compared to Hospital B or C in Figure 1). Thus by shifting a unit of capital from the low-productivity firm to the high productivity firm, aggregate output can be increased holding aggregate capital and labor constant. Conditional on these sectoral distortions, however, firms are generally assumed to minimize costs. In our microeconomic approach, we focus on the misallocation of inputs *within* firms; even when two hospitals exhibit identical firm-level TFP and equal input prices, we observe hospitals at points B and C with equal costs but large differences in outcomes owing to suboptimal input choice.

2.1 Identifying Input Misallocation

Begin with the case in which all h hospitals exhibit efficient input choices across the k inputs. Then a natural question arises: What happens to survival (output) in hospital h, when we marginally increase one input, say x_{h1}, while holding *total costs* constant? In the case of a hospital with two inputs where $S_h = f(x_{h1}, x_{h2})$ that is operating efficiently at point (x_{h1}, x_{h2}), the effect of increasing x_{h1} while maintaining constant cost by reducing x_{h2} is:

[4] Productivity differences across hospitals create difficulties for researchers seeking to estimate "the" returns to health care spending; for example if hospitals cluster around C' and C in Figure 1 the regression coefficient will be essentially zero. One might be tempted to interpret the regression coefficient as suggesting that the health care system is at Point V (or "flat of the curve" health care spending) when in fact the true return to spending for each type of hospital is positive.

$$\left.\frac{dS_h}{dx_1}\right|_{\text{cost}} = f_1 - f_2\frac{p_1}{p_2} = p_1\left(\frac{f_1}{p_1} - \frac{f_2}{p_2}\right) \equiv \mu = 0 \tag{1}$$

At the optimum, this condition (summarized by μ) is equal to 0; this reflects reflects the fundamental principle that at this point, marginal cost-neutral reallocations yield no output improvement, and any infinitesimal movement along the isocost line results in offsetting gains and losses that exactly cancel out.[5] This is the hallmark of efficient resource allocation. To see this graphically, consider Figure 2 again featuring Hospital B and C with equal total costs. Hospital B uses inputs efficiently and chooses a level of output where the isocost line and survival isoquant are tangent to each other (point B). Any small change in the mix of x_1 and x_2 holding total costs constant will have no measurable impact on survival. Also illustrated is hospital C with an over-reliance on the less productive input 2; it sits on the same isocost line but produces a lower level of survival, thus replicating the greater apparent productivity of Hospital B compared to Hospital C shown in Figure 1. Thus the graph illustrates the key intuition: only for hospital C (and not for B), it is possible to improve outcomes by μ – holding total costs constant – by increasing input x_1 by $1/p_1$ and reducing input x_2 by $1/p_2$.

The difference between input use in the inefficient hospital (C) and an optimizing one (B) can be decomposed into two economically distinct steps, each addressing a different source of inefficiency. The first correction is a reallocation from C to C' along the same isoquant curve, shifting expenditure toward the input with higher output per dollar until the input mix lies on the optimal expansion path with a lower total cost but equal outcome or survival rate (This also corresponds to a shift from C to C' in Figure 1.) This step removes the inefficiency arising from misallocation, since each hospital is now using the cost-minimizing set of inputs; the output level cannot be further improved by shifting spending across inputs.[6]

[5] The above condition is easily seen by totally differentiating the cost constraint and substituting the result into the total differential of the production function. The former yields $p_1dx_1 + p_2dx_2 = 0 \Rightarrow dx_2 = -\frac{p_1}{p_2}dx_1$. The latter is $dS = f_1dx_1 + f_2dx_2 = \left(f_1 - f_2\frac{p_1}{p_2}\right)dx_1$

[6] We label this source of inefficiency as misallocation rather than productive inefficiency because of the

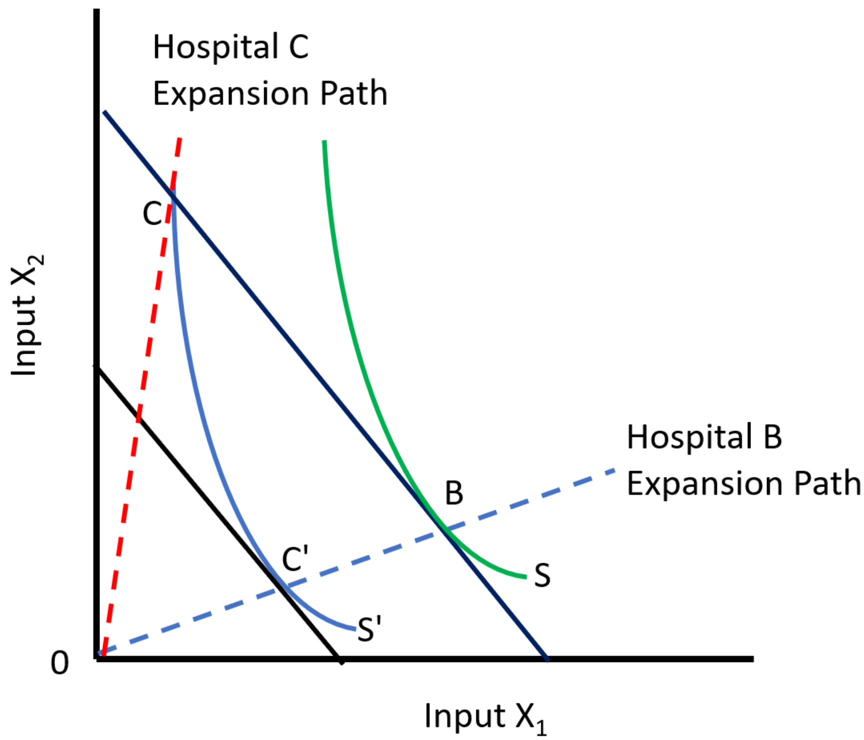

Figure 2: Misallocation in Factor Inputs

The second correction moves from $C' \to B$ by proportional scaling of the now-correct input mix (i.e., a move along the expansion path) until the bundle just touches the frontier at the isocost tangency. This second step removes the remaining inefficiency, delivering the cost-constrained optimum for both hospitals B and C at point B (which is the equivalent of moving from C' to B in Figure 1), where both firms operate on the frontier and use the cost-minimizing input mix given prices and the budget.

This simple framework also clarifies why efficiency analysis requires careful attention to constraints faced by decision-makers. The same firm might appear technically efficient (optimal input ratios) while remaining allocatively inefficient (suboptimal scale), or conversely. As noted earlier, the macroeconomics literature has shown that input misallocation causes sizable dispersion in marginal productions and large counterfactual productivity gains. However, identifying misallocation in the macroeconomic literature has proven to be challenging

inefficient allocation of the health care inputs; for a given level of resources devoted to health care, the economy is producing too much of input x_2 but not enough x_1.

given the presence of mismeasurement in inputs and outputs, adjustment costs, and other issues that could bias estimates of aggregate misallocation (Bils et al. 2021). Because of the unique data on survival and inputs at the individual and hospital level, we are able to use a novel empirical test for misallocation that can be used either within or across firms.[7] There is nothing about this setup that is restricted to the two input case; in Appendix A.1 we derive the same comparative static for the n-input case.

One may equivalently view misallocation as reflecting a "shadow price" encouraging or discouraging the use of an input because of lack of information, financial incentives that reward overusing an ineffective treatment (and conversely), or regulatory constraints. Thus one may write Equation 1 as

$$\left.\frac{dS_h}{dx_1}\right|_{\text{cost}} = p_1\left(\frac{f_1}{p_1} - \frac{f_2}{\lambda_2 p_2}\right) = 0 \tag{2}$$

where λ_k reflects the implicit price that must be faced by the hospital in order to justify their under- ($\lambda_k > 1$) or over-use ($\lambda_k < 1$) of the specific input.[8] For example, the conventional case of "supplier-induced demand" would correspond to a λ that is less than one.

The magnitude of $\left.\frac{dS_h}{dx_1}\right|_{\text{cost}}$ measures both the direction and intensity of beneficial reallocation.[9] A simple test of input misallocation is therefore to regress across hospitals h (or over time for a given hospital) output (in our case, patient survival) on the two inputs and average total cost per patient.

[7]The macroeconomics definition of misallocation focuses on the suboptimal distribution of capital and labor across the economy, holding total capital and labor constant. In our model, we focus on intermediate inputs which are in turn produced by capital and labor; thus misallocation corresponds both to moving inputs from low-productivity to high-productivity hospitals, but also to a reallocation from the production of the less productive x_2 towards the more productive x_1, holding total costs constant.

[8]We denote λ_1 as the numeraire for simplicity.

[9]This can also be viewed as a mathematical characterization of Leibenstein's X-inefficiency (Leibenstein 1966) where firms operating with the same technology and facing identical prices achieve different performance; his empirical finding could be generated by variation in input allocation decisions.

$$S_h = \alpha_h + \beta_1 x_{1h} + \beta_2 x_{2h} + \phi ln(Cost_h) + \epsilon_h \quad (3)$$

Here α_h is the first-order approximation of A_h, and β_k is an approximation of μ_k since total costs are held constant in the regression.[10] The test for input misallocation is a test of whether $\beta_1 = \beta_2 = 0$. Note that we include costs in log terms, with a coefficient ϕ, although we later test for further nonlinearity in cost and input coefficients. This equation therefore allows us to measure the distribution of α_h which captures firm-level TFP, while the β_k coefficients capture the extent and severity of misallocation at the margin.

2.2 The Cobb-Douglas Production Function

Assuming a Cobb-Douglas production function (as in Syverson (2011) and many other productivity studies) allows us to provide a more precise interpretation of the coefficients β_k in the linear regression described in Equation 3. Consider the simplest form of the two-input Cobb-Douglas production function:

$$exp(S) = A{x_1}^{\gamma_1} x_2^{\gamma_2} e^{\epsilon} \quad (4)$$

where γ_1 and γ_2 are the productivity coefficients in the Cobb-Douglas specification and output is the exponent of the survival rate as in Chandra et al. (2016a) to ensure that the dependent variable in the log-linear specification is binary (e.g., whether the individual survives or not). There are two ways by which the Cobb-Douglas specification aids in the interpretation of the model coefficients and sensitivity analysis; these are detailed in Appendix A.2. First, in a cross-section analysis, if the production function parameters γ_{kh} or prices p_{kh} differ across hospitals (or ambulance services), then our test of misallocation may be biased; some hospitals may optimally use one procedure over another. However,

[10] In the k-input case where only a subset of inputs are measured, β_k corresponds to the change in survival, holding other measured inputs constant, while varying unmeasured inputs to increase on average by p_k, thus holding costs constant.

as we show in Appendix A, in a Cobb-Douglas framework, this bias does not hold in a fixed-effects model where permanent hospital-specific prices and production parameters γ_k are captured in the fixed effect. Thus the test for misallocation is valid whether prices or production functions differ across hospitals.

Second, the first-order conditions for input choice can be rewritten to show that, for two hospitals h and h', the ratio of the kth input is written:

$$\frac{x_{kh}}{x_{kh'}} = \frac{s_{kh}}{s_{kh'}} \frac{\gamma_{kh}}{\gamma_{kh'}} \frac{p_{kh'}}{p_{kh}} \frac{\lambda_{kh'}}{\lambda_{kh}} \tag{5}$$

were as noted above, λ_{kh} is the shadow price of the kth input in hospital h. Given that survival s varies by roughly 10 percentage points across hospitals, this equation places bounds on the extent to which prices (p_{kh}) and production function differences (γ_{kh}), can explain the 3-fold per-patient differences in inputs such as PCI discussed below.

3 Data, Measurement and Empirical Approach

We examined a cohort of patients, aged 65 and above, who were hospitalized for acute myocardial infarction (AMI) or heart attacks, within the fee-for-service Medicare group from 2007 to 2016, and tracked follow-up data until December 31, 2017. The identification of AMI relies on the initial diagnosis code: 410.x0 or 410.x1, excluding 410.x2, in ICD9 coding (before October 2015) and I21.x in ICD10 coding starting from October 1, 2015; we found no evidence of coding-related changes in 2015-16. We employed comparable data in related studies (Chandra et al. 2023, 2016a).

3.1 Measuring Inputs

A variety of treatments exist for patients with AMI, applicable during both the acute phase and the post-discharge period. Our focus is on evaluating treatments or procedures whose initial effectiveness hypothesis is supported by clinical evidence and where data are available from Medicare claims. Following the approach by Chandra and Skinner (2012), we classify these treatments into three groups based on clinical evidence:

Category I inputs include those with high cost-effectiveness and low potential for overuse. Category I inputs are prescriptions for beta blockers, statins, and angiotensin-converting enzyme (ACE) inhibitors and angiotensin II receptor blockers (ARBs) for AMI patients during the six months following hospital discharge (Munson et al. 2013). These treatments should be widely administered, irrespective of patient health status. Also included is same-day percutaneous coronary intervention (PCI, stenting), an intervention shown to be very effective if given to suitable patients within 12 to 24 hours post-heart attack (Keeley et al. 2003).

Category II inputs show variable benefits depending on patient type– in other words, they exhibit diminishing marginal returns when used appropriately. While the advantages of same-day PCI are well-documented, performing a PCI for AMI after the first 24 hours offers diminished benefits, especially among some subgroups of the population.(Hochman et al. 2006; Chandra and Staiger 2020) A further instance of a Category II treatment arises when numerous doctors administer care to a single patient; greater specialization among physicians can boost productivity (Doyle 2020), but coordination costs among specialists can escalate, resulting in rapidly diminishing returns.(Becker and Murphy 1992; Baicker and Chandra 2004). Other Category II inputs include the number of days in ICU or cardiology care, and the utilization of MRI and CT scans, where evidence of incremental benefits is scant (Ahmed et al. 2013).

Category III inputs are characterized by marginal benefits that are minor, negative, or unclear. For instance, Doyle et al. (2017), using ambulance services for randomization, identified negative survival impacts at higher levels of post-acute care, while Einav et al. (2018) found long-term care hospitals (LTCHs) to be inefficient. Similarly, home health care often involves fraudulent activities in many areas (O'Malley et al. 2023; Einav et al.

2026). In AMI patient treatment, the application of post-acute care isn't governed by specific guidelines and is frequently influenced by non-clinical factors like service availability, payment generosity, or post-acute service ownership by the hospital.(Buntin et al. 2009). We evaluate three post-acute care measures: average spending per AMI patient on home health care, average spending per AMI patient on skilled nursing facility (SNF) care, and the proportion of AMI patients admitted to either a LTCH or inpatient rehabilitation facility (IRF).[11] Additionally, we include two "Choosing Wisely" indicators related to the use of "double CT" scans on the chest and abdomen, one with and one without iodine contrast, offering no additional clinical value while subjecting patients to excess radiation (Bogdanich and McGinty 2011).[12] These Category III measures are interpreted as indicators of poor hospital quality.

One shortcoming of including detailed input measures as independent variables is the difficulty in interpreting specific coefficients, particularly when measured variables may proxy for unmeasured ones (Griliches 2013) We therefore also conduct a principal components analysis for each category to provide clarity in the interpretation of the coefficients, using only the first principal component as an explanatory variable.[13]

3.2 Measuring Costs

Two primary methods exist for assessing costs. The first, direct measurement, involves individual-level data on Medicare reimbursements (specifically Parts A and B, encompassing physician fees, diagnostic scans, post-acute care, outlier payments, and outpatient services) adjusted for geographic price variations Gottlieb et al. (2010). A large portion of the cost is the lump-sum payment (known as a DRG payment) that Medicare makes for the hospital stay which is closely linked to the patient's diagnosis and treatment method (such as the use of a

[11]The first two metrics encompass all AMI patients hospitalized, considering both the portion receiving such care and average spending given receipt.

[12]These latter two measures are not limited to AMI patients.

[13]To assess technology complementarity and substitution, we also computed interaction terms between the principal components (Athey and Stern 1998), but our findings remained quantitatively similar.

stent versus medical management). We utilize the "leave-out" technique to calculate inputs and average expenditures, removing the individual's treatment costs from the hospital's average.[14]

A limitation of Medicare reimbursements is their inability to reflect disparities among hospitals concerning non-Medicare revenue, especially from privately insured patients who pay higher prices, potentially enabling higher quality inputs to be delivered. Consequently, we introduce a secondary cost measure: per-admission operating expenses by hospital and year, derived from Medicare Cost Reports, which encompass all costs, not merely those assigned to Medicare patients.[15]

3.3 Adjusting for Patient Risk

One of the principal empirical difficulties in estimating the regression in (2) is appropriately adjusting for mortality risk. If risk adjustment is inadequate, hospitals could seem more efficient solely because they manage healthier patients. To tackle this issue, we adopt several methods. 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 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).[16] We include Hierarchical Condition Categories (HCCs) that count the number of different diagnoses that

[14]While average Medicare reimbursements capture costs to the federal government (rather than costs to the hospital), the MedPAC March 2020 report (p.10) noted that Medicare reimbursements for inpatient services were just 8% higher than variable cost – in other words, Medicare reimbursements are were roughly equivalent to average variable cost.

[15]We are grateful to Adam Sacarny for providing these data; https://github.com/asacarny/hospital-cost-reports

[16]Medicare data does not distinguish between ethnicity and race, so these definitions are mutually exclusive.

patients have received in the 6 months prior to the index admission, and weight them for severity using well-established methods. Finally, the fraction of the hospital referral region enrolled in Medicare Advantage is included to capture the possibility that the fee-for-service population could exhibit greater unmeasured health deficits if healthier enrollees select into managed care.

Zip Code, Random, and Fixed Effect Estimation Approaches. We follow several approaches to ensure that unmeasured confounding does not bias our results. First, we rely on the work of Chandra et al. (2023) who note that risk-adjustment using claims data results in quality measures that are 'causal' (in the sense of representing the mortality outcomes of a randomly admitted patient; the bias is less than 7%). Still, in our first set of regressions, we include zip code fixed effects, so the variation is generated by differences in outcomes and expenditures when patients in a given zip code seek care in different hospitals (Garthwaite et al. 2019). Second, we estimate the hospital-specific TFP parameter α_h along with the input-specific parameters γ_k in a random-effects model that harnesses both within- and across-hospital variation. Our final approach is with a hospital fixed-effects model that uses only variation in within-hospital inputs and outcomes to estimate regression coefficients; thus in a Cobb-Douglas framework, any permanent differences across hospitals in prices or production functions are absorbed in the hospital fixed effect.

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.[17] In their

[17] The National Provider Identifier (NPI) is used to identify ambulance firms.

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) to focus on reduced-form estimates.[18]

Estimation Equation. In its most general form, the linear estimation equation is:

$$S_{iht} = \alpha_h + \beta^* m_{ht} + \left[\sum_{k=1}^{K'} \beta_k(x_{kht})\right] + \left[\sum_{j=1}^{J} \omega_j z_{ijt}\right] + \zeta_{iht} \tag{6}$$

where S_{iht} is a binary value of whether individual i survives to one year (or 30 days), m_{ht} is the log of total expenditures or costs for hospital (or ambulance service) h in year t, x_{kht} is the kth input for hospital or ambulance service h in year t, β_k the coefficient on the kth input, and ζ_{iht} is the error term. In this simple form we specify the model as a linear approximation to inputs and log expenditures or costs; in sensitivity analysis, we also consider an estimation model with a more flexible specification that allows for decreasing returns (or increasing returns) to factor inputs and overall expenditures. This latter model captures misallocation across hospitals because of diminishing returns to inputs.

The risk adjustment measures are summarized by z_{ijt} of which we observe J distinct measures; these are specific to individual i except when zip code fixed effects are included which are assigned to each individual living within the zip code; Finally, α_h is (as before) the hospital- or ambulance-service specific total factor productivity measure. In the above estimating equation, the null hypothesis for misallocation is that the inputs conditional on expenditures are jointly equal to zero: $\beta_1 = \beta_2 ... = \beta_k = 0$.

Note that we only observe a subset K' of the total K possible inputs to the production function; these have been chosen because of clinical or other studies suggesting an important role in the production function for health, as well as whether they appear in the claims data

[18] In both the ambulance and overall sample, we restrict our attention to hospital-year cell sizes with at least 50 AMI admissions so as to minimize biases arising from small samples – particularly with regard to measuring average expenditures. If the hospital-year cell size is less than 50, we measure inputs and outputs over the entire period of analysis; any hospital with fewer than 50 AMI patients over the 10-year period 2007-2016 is excluded.

or are available from other sources. Thus the interpretation of β_k is the value of increasing input k by one unit, balanced by a general reduction in costs of other inputs not measured in the regression equation, of which some are unmeasured but reflected in total costs per patient.[19]

3.4 Identification

The key empirical challenge is ensuring that the use of inputs and overall costs are not themselves determined by unobservable factors that also influence survival and outcomes. There are several long-standing concerns with the estimation of production functions. One is that the input vector X_{ht} may be correlated with the error term ζ, for example when a firm hires more labor in response to a productivity shock not observed by the econometrician (Ackerberg et al. 2015; Levinsohn and Petrin 2003). Our production function differs from these models however, as it relates to inputs (and outcomes) *per patient*, rather than the total quantity produced using capital and labor. 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.

Another concern, as noted by Ackerberg et al. (2015), is that under the null hypothesis of efficient input allocation, the β_k coefficients are not identified without exogenous variation in x_{it}.[20] Earlier studies have focused on productivity differences because of management quality, adoption of information technology, and diagnostic and triage capacity, but these are difficult to measure accurately at a national scale. However, we hypothesize that physician beliefs play a key role in predicting variation across hospitals in their input choice (Cutler et al. 2019; Chandra and Staiger 2020). As we demonstrate below, how physicians respond to vignettes regarding the clinical management of patients is strongly correlated at the regional

[19] Thus even if Category III treatments have no adverse impact on health, their coefficient may still be negative if, holding spending constant, they crowd out beneficial inputs.

[20] Even if the model is not identified, the F-test is unlikely to reject. However, as Jack Marshall pointed out, if all hospitals are following a linear decision rule that is inefficient (so that misallocation is present) the F-test would falsely fail to reject the null; thus we still require exogenous variation in x_{it}

level with whether hospitals are more or less likely to use effective Category I treatments and less effective Category II treatments.[21]

4 Results

4.1 Descriptive Statistics

Table 1 (Column 1) 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 as well as the smaller subset of 436,519 patients (27%) admitted to the hospital by ambulance. The average age is 78.1, and on average 86.3% survive to 30 days and 70.8% survive to one-year post-admission. 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 patients) are $21,462. There are high rates of comorbid diseases, with 27.3% having diabetes and 39.2% with congestive heart failure. Rates of Category I treatments are modest; just 51.4% receive beta blockers on discharge, and 31.1% same-day PCI. Rates of post-acute care (3.5%) and spending on home health care ($173) are both low on average across all AMI patients.For the subset of 436,519 in the ambulance sample (Column 2), average age is slightly older (78.9 versus 78.1 years), and survival rates somewhat lower (67.4% surviving one year compared to 70.8% for the full sample). Treatment rates are marginally higher, but are otherwise quite similar to the full sample.

[21] In the context of the fixed-effect models, factors that may vary over time within a hospital include when physicians moving from one hospital to another but retain their earlier treatment patterns (Badinski et al. 2023), the uneven pace of diffusion in new technologies (Skinner and Staiger 2015), or institutional changes, such as when PCI facilities are either opened or closed (Shen et al. 2023).

4.2 Regression Results

Table 2a presents regression estimates for the full sample that follow the standard convention 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 B.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 substitutes the log of hospital-level operating expenses (which is not limited to Medicare patients) for AMI-specific Medicare reimbursements; results are similar with a coefficient of 0.0319 (s.e. 0.0025). For Medicare expenditures, results are larger with zip-code fixed effects (0.0527, s.e. 0.0055, in Column 3), but considerably less with hospital random effects (0.0167, s.e. 0.0037, Column 4) and smaller still for the fixed-effects model (0.0010, s.e. 0.0044, Column 5). That the standard deviation of the hospital-specific random effect is 2.5 percentage points (Model 4) suggests an important role for variation in hospital productivity, as in Chandra et al. (2016a).

In Table 2b, which uses the ambulance sample, the coefficient in Column 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. Neither the random-effect or fixed-effect model coefficients on expenditures are significantly different from zero.

Figure 3 shows risk-adjusted measures of log expenses and one-year survival for the full sample (based on the risk-adjustment models in Appendix Table B.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 just 2% 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. These hospitals exhibit (on average) risk-adjusted survival rates equal to 76.1%, substantially above the 70.7% observed in hospitals not included on the list.

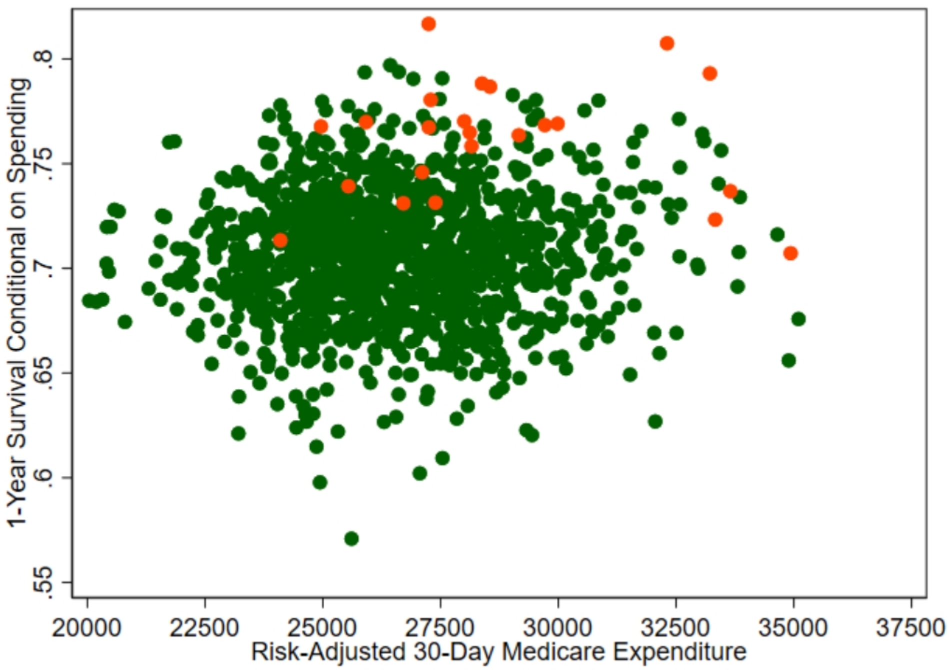

Figure 3: Medicare Expenditures and One-Year Survival Rates, by Hospital, 2007-17 Survival and Expenditures are risk-adjusted using model coefficients in Appendix Table B.1. The sample is limited to hospitals with at least 500 AMI admissions during the period 2007-2016. The orange dots denote US News & World Report Top 25 Cardiovascular Hospitals in 2017 (a few do not meet the 500-patient minimum). The R^2 of the regression is 0.02.

Table 3 and Table 4 report our key results from estimating Equation 5. 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.

Some independent variables are expressed as fractions of the population (e.g., the fraction who filled a prescription for statins), counts (e.g., the number of different physicians, or number of nursing home days), and in dollar terms (home health spending per $1,000) averaged across AMI patients in the hospital/year. Most of the Category I inputs exhibit positive coefficients although not all are significant individually; however, early (same-day) stenting (PCI) is a strong predictor of survival across all equations. For Category II inputs, there is no consistent association with outcomes; indeed the negative coefficient on the number of scans may capture lower-productivity physicians as in Doyle Jr et al. (2010). Category III treatments show generally negative associations between greater use and survival, although again, individual coefficients are not consistently significant across regression specifications.

It is difficult to interpret the disparate coefficients in Table 3, particularly as they may also reflect inputs missing from the regression. We therefore consider a principal components approach in which we categorize hospitals as high or low for Categories I, II and III using the first principal component of the combined inputs (see Appendix Table B.3). As shown in Table 4, 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 greater use of effective treatments are strongly associated with 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, per standard deviation shift. For Category III treatments, coefficient estimates are most strongly negative in the ambulance sample, with a coefficient of -0.81 percentage points (s.e. 0.30) in the fixed effects model (Column 6).[22]

Estimating the Fraction of Productivity Variation Attributable to Misallocation We use the estimated coefficients from the random-effect model (Table 3, Model 2), along with the observed values of x_{kh} (evaluated across years and for the representative patient to abstract

[22]For Columns 3 and 6 in Table 4, the correlation coefficient between predicted survival and the hospital- or ambulance-specific fixed effects are very small – 0.035 and 0.060, respectively. This is consistent with the assumption (inherent in random effects models) that the hospital- or ambulance-specific random effects are not closely correlated with either inputs or risk adjusters.

from patient comorbidities) to estimate $M_h = -[\sum_k \widehat{\beta}_k \tilde{x}_{kh}]$ where $\tilde{x}_{kh}$ is the hospital specific input with the mean removed, so that M_h, in units of survival percentage points, also has mean zero across hospitals, with higher values representing a greater degree of misallocation. Thus a higher misallocation measure represents greater mortality risk (or lower survival risk) based solely on hospital h's input choices. Values of M_h are shown on the horizontal axis in Figure 4, along with risk-adjusted survival rates conditional on spending (e.g., our measure of firm-level TFP) on the vertical axis, for hospitals with volumes of at least 500 patients during the period of analysis.

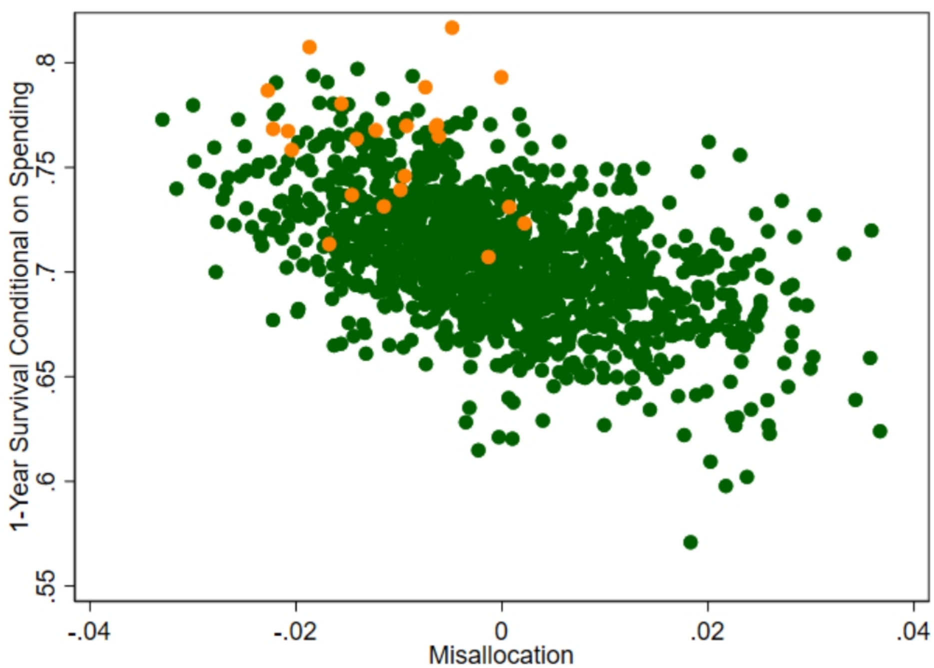

Figure 4: Association between Misallocation (M_h) and 1-Year Survival Conditional on Spending This graph shows the contribution of misallocation (with higher values denoting more misallocation) to explain risk- and spending-adjusted survival. The graph is limited to hospitals with at least 500 admissions during the period of analysis. The R^2 of the bivariate regression is 0.26.

Moving from the (patient-weighted) 10^{th} to 90^{th} percentile of estimated misallocation predicts a reduction of 3.3 percentage points in survival. A simple bivariate regression shows that our misallocation measure explains 26 percent of the overall variation in risk- and spending-adjusted survival (firm-level TFP).

The importance of misallocation in hospital-level survival variability can also be estimated by comparing the variance of hospital-level random effects with specific inputs (Table 3) and without inputs (Table 2A). That is, the standard error of the random effects estimates (without inputs) of hospital productivity is 0.0253 (Table 2A), or a change of 2.53 percentage points in mortality with regard to a one-standard deviation shift in hospital productivity. The standard error of the random effect *with* inputs included falls to 0.0219 (Table 3). The ratio of $.0219^2$ to $.0253^2$ is .75, so including input measures reduces the variance by 25 percent; the corresponding calculation for the ambulance sample is 24 percent.

4.3 Sensitivity Analysis

Heterogeneous Treatment Effects. In the macroeconomic misallocation literature, economic distortions such as subsidies to specific sectors or firms lead to different marginal rates of return for a given input; thus reallocating a given supply of resources to sectors with higher rates of return (and away from low-return sectors) would yield a higher aggregate output. As well, in our first-order approximation to an arbitrary production function, linearized inputs not be a valid simplification (as in the Cobb-Douglas specification which exhibit diminishing returns). We test for systematic differences in returns by creating dummy variables for the middle 50 percent of hospitals and the top quartile; the baseline is the bottom 25 percent. Note that hospitals may simultaneously be in the top quartile of one input and the bottom quartile of another.

The results are presented in Appendix Table B.2. There is little evidence of consistent diminishing or increasing returns; a combined F-test for these semi-parametric input measures compared to conventional linear inputs yields a p-value of only 0.07. An exception, however, is that there appears to be diminishing returns for PCI, with the incremental benefits among hospitals in the top quartile of utilization is just 60 percent of those in the middle two quartiles.[23]

Prices and Production Functions that Differ Across Hospitals. If prices or production function parameters γ_{hk} differ systematically across hospitals, then they should optimally choose different input combinations, thus potentially leading to bias in our F-test for input misallocation. We address this in several ways. First, and most simply, with fixed hospital or ambulance effects in a Cobb-Douglas framework, any differences in prices or production function parameters will be absorbed in the constant term, and while the F-tests are not as large as in the random effects models (e.g., an F-test of 4.96 ($p < 0.001$) in the full sample), they still reject the null of efficient input allocation. Second, we do not find evidence that coefficient estimates differ systematically between teaching and non-teaching hospitals.

We use the properties of the Cobb-Douglas model to estimate the potential impact of differences in prices or production parameters on input use across hospitals – as noted earlier, in Cobb-Douglas production functions, a 20% difference in prices or in the γ coefficients will induce a roughly 20% difference in the ratio of optimal inputs per output unit. For example, in our sample the early (same-day) PCI stenting rate for hospitals in the top decile is 0.401, while in the bottom decile the rate is 0.117, or a ratio of 3.43. Given diminishing returns to PCI shown above, this gap is unlikely to be explained by a difference in γ between the top and bottom deciles, nor is it likely that the bottom decile hospital faces costs of providing PCI that exceeds the top decile by a factor of 3.43. Far more likely is that variation in the shadow price λ_{hk} capturing inefficient variation in inputs.

Short-Run Outcome Measures. One concern is that the results may be sensitive to the

[23] The incremental survival benefits moving from the bottom quartile to the middle two quartiles is equal to 0.162 = .0202/.1248 where .1248 is the risk-adjusted change in the rate of early PCI from the bottom quartile to the middle two quartiles, while moving to the highest quartile yields an incremental benefit of 0.089 = (.0281-.0202)/0.0981, where 0.0981 is the corresponding incremental change in PCI rates from the middle quartiles to the highest quartile. Our results differ therefore from Chandra and Staiger (2007) who find increasing returns to PCI use; that our results differ is likely because in the period of the Chandra and Staiger analysis, the then emerging technology was not recommended for the vast majority of hospitals with inadequate experience in the use of PCI. For example, the 1996 American Heart Association guidelines states that primary PCI was indicated "...only if performed in a timely fashion by individuals skilled in the procedure and supported by experienced personnel in high-volume centers." (Ryan et al. 1996) By 2007-17, however, the use of PCI had become commonplace across nearly all hospitals.

one-year survival rate if unmeasured factors affected longer-term health outcomes. In Appendix Table B.3 we also consider the corresponding regression models for 30-day survival; results are similar albeit with coefficients that are generally smaller in magnitude.

Comparing Coefficient Estimates to Randomized Trial Treatment Effects. We used 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 8 percentage points. 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 estimate of 3.4 percentage points in Table 3. For β-Blockers, our median coefficient is 0.035; similar to the trial estimates prior to the general use of PCI, but larger than for more recent studies showing smaller beneficial effects of β-blockers.(Bangalore et al. 2014)

5 What Explains Input Differences Across Hospitals?

What are characteristics of hospitals that are associated with low or high levels of input misallocation? While we cannot establish causation, we can ask what might predict the degree of misallocation across hospitals that may be associated with the principal component estimates of Categories I, II, and III treatment inputs. We use vignette survey data of physician beliefs in the effectiveness of treatments with (or without) strong clinical support. (Chandra and Staiger 2020) Our survey data from Cutler et al. (2019) characterized physician-respondents as cowboys (those supporting more aggressive care without clinical evidence of effectiveness) and comforters (those choosing less intensive care with evidence-based support), as well those favoring high- or low-followup depending on their recommendation for the frequency of follow-up visits outside of professional guidelines. Because we could not match physicians to hospitals, we instead calculated the average of the vignette responses using hospital referral regions (HRRs); we then test whether survey responses predict meaningful differences in the use of Category I, II, and III components of care for hospitals in those regions.[24] At the hospital level, we also use teaching hospital status and the volume of heart attack patients during the period of analysis. Recall that the dependent variables have a standard deviation of 1.0 by construction.

The regressions are estimated at the hospital level, with weights corresponding to log AMI volume, and are clustered by HRR. As shown in Table 5 (Column 1), hospitals with a higher fraction of physicians recommending high follow-up visits, and a low fraction of low-follow-up physicians (and a higher fraction of cowboy physicians, albeit with marginal significance) were much less likely to use Category I effective treatments. In contrast, the same characteristics of physicians (cowboy physicians strongly favoring frequent follow-up visits) were far more likely to recommend the Category II treatments with diminishing returns to use. Finally, a a higher fraction of comforters was predictive of a significantly lower use of Category III treatments.

We might expect teaching hospitals to exhibit a more rapid diffusion of new technologies; we find in Table 5 we find that teaching hospital status is predicted to increase the use of Category I treatments by 0.49 standard deviations, and (surprisingly) Category II by 0.41 standard deviations, with no impact on Category III use. Finally, hospital volume is associated with a substantially higher use of Category I treatment rates, with no impact on Category II or Category III treatment rates. The R^2 for the Category III inputs is only 0.025; it may be that physician beliefs have little impact on Category III post-acute patient placements for nursing homes and home health care, which are often arranged by non-physician staff.

[24] The mean response rate was 9.6 physicians per HRR for the 138 larger HRRs with at least 4 physician respondents; these HRRs accounted for 72 percent of total hospitalizations.

6 Discussion and Conclusion

In this paper, we estimated the importance of misallocation of inputs in explaining the wide differences across hospitals in aggregate total factor productivity. Rather than testing whether returns to inputs are equalized across firms, we develop a new test: whether the returns to various inputs within and across firms are different from zero, conditional on total costs. Using a sample of 1.6 million Medicare patients and a subsample of 436,000 AMI ambulance patients with AMI between 2007 – 2017, we demonstrated that the hypothesis of productive efficiency is strongly rejected; high-misallocation hospitals reduce one-year survival rates by 3.3 percentage points based on input choice alone, and misallocation contributes roughly one-quarter of overall variation in total hospital (or ambulance service) factor productivity. This is larger than the estimate by Baqaee and Farhi (2020) who find 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 (Hsieh and Klenow 2009; Haltiwanger et al. 2018; Restuccia and Rogerson 2017). Our use of a well-defined, accurately measured primary outcome – survival following a heart attack requiring immediate hospital treatment – and well measured inputs, allow an estimate of misallocation with minimum assumptions about elasticities of outputs and inputs.

There are several potential limitations. First, we are concerned about the problem of unmeasured patient characteristics and hospital selection, and so have adopted several strategies including highly detailed individual covariates, the use of zip-code fixed effects and hospital random and fixed effects, as well as measuring intensity at the level of the delivering ambulance service. Our results are consistent across estimation approaches. Second, the reduced-form regression coefficients must be interpreted cautiously, since the estimates are based on comparisons of marginal patients receiving treatment in Hospital B but not Hospital C, 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 (David et al. 2016). Our goal is not to measure returns to a particular input per se, but instead to examine whether inputs matter for outcomes conditional on total spending. However, when evidence from randomized trials is available, the causal treatment effects found in the trials are consistent with regression results.

Third, a key assumption is that the rejection of the null of no misallocation is not the consequence purely of differences across hospitals in production functions and regional prices. In our comparison of teaching and non-teaching hospitals, we do not find evidence that hospital production functions differ. Most importantly, the results from the fixed-effects models, which adjust for differences across hospitals in prices and production parameters, are similar to the zip-code fixed-effect and random effects models.

It is important to consider other factors beyond physician opinion that may contribute to misallocation. Unlike the macroeconomics literature, we cannot as easily appeal to distortionary taxes or a lack of credit financing to explain why hospitals differ so much with regard to their input choices. One possibility is that the lack of population density and fragmented markets means that specialization will be limited by the extent of the market (Dingel et al. 2023). This view is supported by the very strong association between hospital volume and the use of Category I treatments. Another possibility is the quality of hospital management and clinical leadership, for example in the careful evaluation of new technologies;(McConnell et al. 2013). That the "Top 25" hospitals exhibit lower misallocation conditional on volume is consistent with this view.

Finally, we cannot identify the sources of the hospital-specific TFP, α_h. Management quality is again likely to play a role, for example through process standardization and improvement, coordination and communication, and reliance on data for decision making (McConnell et al. 2013). Indeed, Bloom et al. (2020) found that both high-quality management and clinical skills (rather than either in isolation) are essential to improving hospital-level performance (also see Tsai et al. (2015)). And while some of the difference between "Top-25" hospitals and other hospitals is because of lower rates of misallocation, most of the survival advantage is the consequence of systematic differences in α_h.

What are the implications of these results? Abstracting from health care, we have introduced a new way to assess the role of misallocated inputs that could be used across a wide range of industries, helping to understand the links between management, governance, and productivity gaps. In the health care sector, experiments seeking to align payment incentives for hospitals with improving quality and reducing costs suggest that some treatments can be increased (Colla et al. 2012). At the same time, many institutions face challenges with overusing low-value care. Increased diffusion of payment models designed to encourage use of high-value care, and reduce the use of low-value care may improve input allocation in US hospitals. Yet clinical practice patterns are slow to change, and clinicians may fail to understand how specific treatments affect their financial bottom line (Kaplan and Witkowski 2014). More worrisome is the difficulty in improving productivity if the firm's decision-makers hold strong beliefs about the use of specific inputs. Despite these caveats, policies focused on measuring and potentially reducing misallocation in input choice could improve productivity in a sector comprising one-fifth of the U.S. economy.

References

Abaluck, J., L. Agha, D. C. Chan, D. Singer, and D. Zhu (2021). Fixing misallocation with guidelines: Awareness vs. adherence. Working Paper 27467, National Bureau of Economic Research.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Ahmed, N., D. Carrick, J. Layland, K. G. Oldroyd, and C. Berry (2013). The role of cardiac magnetic resonance imaging (mri) in acute myocardial infarction (ami). *Heart, Lung and Circulation 22*(4), 243–255.

Asker, J., A. Collard-Wexler, and J. De Loecker (2019). (mis) allocation, market power, and global oil extraction. *American Economic Review 109*(4), 1568–1615.

Athey, S. and S. Stern (1998). An empirical framework for testing theories about complementarity in organizational design. Working Paper 6600, National Bureau of Economic Research.

Badinski, I., A. Finkelstein, M. Gentzkow, and P. Hull (2023). Geographic variation in healthcare utilization: The role of physicians. Working Paper 31749, National Bureau of Economic Research.

Baicker, K. and A. Chandra (2004). The productivity of physician specialization: Evidence from the medicare program. *American Economic Review 94*(2), 357–361.

Banerjee, A. V. and B. Moll (2010). Why does misallocation persist? *American Economic Journal: Macroeconomics 2*(1), 189–206.

Bangalore, S., H. Makani, M. Radford, K. Thakur, B. Toklu, S. D. Katz, J. J. DiNicolantonio, P. Devereaux, K. P. Alexander, J. Wetterslev, et al. (2014). Clinical outcomes with β-blockers for myocardial infarction: a meta-analysis of randomized trials. *The American journal of medicine 127*(10), 939–953.

Baqaee, D. R. and E. Farhi (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics 135*(1), 105–163.

Becker, G. S. and K. M. Murphy (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics 107*(4), 1137–1160.

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

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2020). Healthy business? managerial education and management in health care. *Review of Economics and Statistics 102*(3), 506–517.

Bogdanich, W. and J. C. McGinty (2011, June). Medicare claims show overuse for ct scanning. *The New York Times*.

Buntin, M. B., C. H. Colla, and J. J. Escarce (2009). Effects of payment changes on trends in post-acute care. *Health services research 44*(4), 1188–1210.

Chan, D. C., M. Gentzkow, and C. Yu (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics 137*(2), 729–783.

Chandra, A., M. Dalton, and D. O. Staiger (2023). Are hospital quality indicators causal? Technical Report 31789, National Bureau of Economic Research.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016a). Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review 106*(8), 2110–2144.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016b). Productivity dispersion in medicine and manufacturing. *American Economic Review 106*(5), 99–103.

Chandra, A. and J. Skinner (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature 50*(3), 645–680.

Chandra, A. and D. Staiger (2007). Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. *Journal of Political Economy 115*(1), 103–140.

Chandra, A. and D. O. Staiger (2020). Identifying sources of inefficiency in healthcare. *The quarterly journal of economics 135*(2), 785–843.

Colla, C. H. (2014). Swimming against the current—what might work to reduce low-value care? *The New England Journal of Medicine 371*(14), 1280–1281.

Colla, C. H., D. E. Wennberg, E. Meara, J. S. Skinner, and D. J. Gottlieb (2012). Spending differences associated with the medicare physician group practice demonstration. *JAMA 308*(10), 1015–1023.

Cutler, D., J. S. Skinner, A. D. Stern, and D. 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, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, misallocation, and aggregate productivity. *Quarterly Journal of Economics 131*(2), 943–1005.

de Boer, M.-J., J.-P. Ottervanger, A. W. J. van't Hof, J. C. A. Hoorntje, H. 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.

Dingel, J. I., J. D. Gottlieb, M. Lozinski, and P. Mourot (2023). Market size and trade in medical services. Technical Report 31030, National Bureau of Economic Research.

Doyle, J., J. Graves, J. Gruber, and S. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*(1), 170–214.

Doyle, J. J. (2020). Physician characteristics and patient survival: Evidence from physician availability. (27458).

Doyle, J. J., J. A. Graves, and J. Gruber (2017). Uncovering waste in us healthcare: Evidence from ambulance referral patterns. *Journal of Health Economics 54*, 25–39.

Doyle Jr, J. J., S. M. Ewer, and T. H. Wagner (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of health economics 29*(6), 866–882.

Díaz-Hernández, J. J., E. Martinez-Budría, and S. 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*, 321–329.

Einav, L., A. Finkelstein, Y. Ji, N. Mahoney, and G. Moore (2026). Medicare home health fraud: How much, where, and who? Working Paper 35280, National Bureau of Economic Research.

Einav, L., A. Finkelstein, and N. Mahoney (2018). Long-term care hospitals: A case study in waste. *The Review of Economics and Statistics*, 1–57.

Garthwaite, C., J. A. Graves, T. Gross, Z. Karaca, V. R. Marone, and M. J. Notowidigdo (2019). All medicaid expansions are not created equal: The geography and targeting of the affordable care act. *Brookings Papers on Economic Activity*.

Gottlieb, J. D., W. Zhou, Y. Song, K. G. Andrews, J. S. Skinner, and J. M. Sutherland (2010, March-April). Prices don't drive regional medicare spending variations. *Health Affairs (Millwood) 29*(3), 537–543.

Griliches, Z. (2013). Hedonic price indexes for automobiles: An econometric analysis of quality change. pp. 55–87.

Haltiwanger, J., R. Kulick, and C. Syverson (2018). Misallocation measures: The distortion that ate the residual. Technical Report 24199, National Bureau of Economic Research.

Hochman, J. S., G. A. Lamas, C. E. Buller, V. Dzavik, H. R. Reynolds, S. J. Abramsky, S. Forman, W. Ruzyllo, A. P. Maggioni, H. White, et al. (2006). Coronary intervention for persistent occlusion after myocardial infarction. *New England Journal of Medicine 355*(23), 2395–2407.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics 124*(4), 1403–1448.

Hull, P. (2020). Estimating hospital quality with quasi-experimental data. *Working Paper*.

Josan, K., S. R. Majumdar, and F. 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.

Kaplan, R. S. and M. L. Witkowski (2014). Better accounting transforms health care delivery. *Accounting Horizons 28*(2), 365–383.

Keeley, E. C., J. A. Boura, and C. 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.

Lee, J., J. S. McCullough, and R. J. Town (2013). The impact of health information technology on hospital productivity. *The RAND Journal of Economics 44*, 545–568.

Leibenstein, H. (1966). Allocative efficiency vs." x-efficiency". *The American economic review 56*(3), 392–415.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*, 317–341.

McConnell, K. J., R. C. Lindrooth, D. R. Wholey, T. M. Maddox, and N. Bloom (2013). Management practices and the quality of care in cardiac units. *JAMA internal medicine 173*(8), 684–692.

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

Mullainathan, S. and Z. Obermeyer (2022). Diagnosing physician error: A machine learning approach to low-value health care. *Quarterly Journal of Economics*.

Muñoz, P. and C. Otero (2025). Managers and public hospital performance. *American Economic Review 115*(11), 4040–4074.

Munson, J., N. Morden, D. Goodman, L. Valle, and J. Wennberg (2013). The dartmouth atlas of medicare prescription drug use. October 15.

O'Malley, A. J., T. A. Bubolz, and J. S. Skinner (2023, June). The diffusion of health care fraud: A bipartite network analysis. *Social Science and Medicine 327*, 115927.

Raja, C. (2025). *Regulating the Quality of Hospitals: Input Regulation, Factor Market Structure, and Misallocation in Healthcare*. Ph. D. thesis, University of California, Los Angeles.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Restuccia, D. and R. Rogerson (2017). The causes and costs of misallocation. *Journal of Economic Perspectives 31*, 151–174.

Ryan, T. J., J. L. Anderson, E. M. Antman, B. A. Braniff, and N. H. Brooks (1996). Acc/aha guidelines for the management of patients with acute myocardial infarction: A report of the american college of cardiology/american heart association task force on practice guidelines. *Journal of the American College of Cardiology*, 1328–1419.

Shen, Y.-C., H. M. Krumholz, and R. Y. Hsia (2023). Do pci facility openings and closures affect ami outcomes differently in high-vs average-capacity markets? *Cardiovascular Interventions 16*(10), 1129–1140.

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

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

Tsai, T. C., A. K. Jha, A. A. Gawande, R. S. Huckman, N. Bloom, and R. Sadun (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs 34*(8), 1304–1311.

Wennberg, J. E. (2010). *Tracking Medicine: A Researcher's Quest to Understanding Health Care*. Oxford University Press.

Table 1: Summary Statistics for AMI Patients, 2007-17: Full and Ambulance Samples

	Full Sample	*Ambulance Sample*
	Mean (sd)	Mean (sd)
N	1,617,039	436,519
Age	78.144 (8.320)	78.909 (8.416)
Fraction Female	0.480 (0.500)	0.505 (0.500)
Fraction Asian	0.013 (0.111)	0.012 (0.108)
Fraction Black	0.074 (0.262)	0.073 (0.261)
Fraction Hispanic	0.015 (0.122)	0.015 (0.121)
Fraction Native American	0.005 (0.070)	0.004 (0.065)
Fraction Heart Failure	0.392 (0.488)	0.413 (0.492)
Fraction Diabetes	0.273 (0.445)	0.265 (0.441)
Fraction Survive 1 Year	0.708 (0.455)	0.674 (0.469)
Fraction Survive 30 Day	0.863 (0.343)	0.845 (0.362)
Medicare Spending ($1000)	26.547 (21.33)	27.261 (21.18)
Operating costs per discharge ($1000)	21.462 (8.453)	21.492 (8.407)
Cat I: Beta Blockers	0.514 (0.069)	0.497 (0.059)
Cat I: Statins	0.443 (0.080)	0.432 (0.070)
Cat I: ACE/ARB Drugs	0.346 (0.065)	0.332 (0.051)
Cat I: Same-Day PCI	0.311 (0.130)	0.319 (0.083)
Cat II: Later PCI	0.093 (0.038)	0.088 (0.025)
Cat II: N of Doctors	11.625 (2.892)	13.280 (2.021)
Cat II: N of Scans	0.579 (0.200)	0.632 (0.151)
Cat II: ICU/CCU Days	4.350 (1.988)	4.491 (1.617)
Cat III: Post-Acute Care	0.035 (0.035)	0.039 (0.028)
Cat III: Nursing Home Days	2.460 (1.085)	2.939 (0.976)
Cat III: Home Health ($1000)	0.173 (0.106)	0.199 (0.096)
Cat III: Abdomen CT Measure	0.118 (0.140)	0.112 (0.089)
Cat III: Thorax CT Measure	0.035 (0.071)	0.033 (0.038)

Notes: 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.

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

	Model 1	Model 2	Model 3	Model 4	Model 5
Log Expenditures	$0.0317^{\dagger}$		$0.0527^{\dagger}$	$0.0167^{\dagger}$	0.0010
	(0.0067)		(0.0055)	(0.0037)	(0.0044)
Log Operating Expense		$0.0319^{\dagger}$			
		(0.0025)			
Intercept	$0.4907^{\dagger}$	$0.4928^{\dagger}$	$0.2851^{\dagger}$	$0.6401^{\dagger}$	$0.8105^{\dagger}$
	(0.0691)	(0.0256)	(0.0568)	(0.0377)	(0.0447)
Zip Code Fixed Effects			X		
Hospital Random Effects				X	
Hospital Fixed Effects					X
Number of observations	1,617,039	1,616,480	1,617,039	1,617,039	1,617,039
Adjusted R-squared	0.179	0.180	0.181		0.172
Panel-level standard deviation				0.0253	0.038

Notes: All Regressions include full set of risk-adjustment covariates, see Appendix B.1. N = 1,617,039, including 2,024 hospitals; σ_μ denotes the standard deviation 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. † p<.05

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

	Model 1	Model 2	Model 3	Model 4	Model 5
Log Medicare Spending	0.0226		$0.0539^{\dagger}$	0.0179	0.0162
	(0.0197)		(0.0206)	(0.0107)	(0.0206)
Log Operating Costs		$0.0528^{\dagger}$			
		(0.0041)			
Intercept	$0.5843^{\dagger}$	$0.6527^{\dagger}$	0.2777	$0.6363^{\dagger}$	$0.6558^{\dagger}$
	(0.2015)	(0.0141)	(0.2122)	(0.1090)	(0.2109)
Zip Code Fixed Effects			X		
Ambulance Random Effects				X	
Ambulance Fixed Effects					X
Number of observations	436,519	436,519	436,519	436,519	436,519
Adjusted R-squared	0.1801	0.1810	0.1827		0.1624
Panel-level standard deviation				0.0258	0.089

Notes: All Regressions include full set of risk-adjustment covariates, see Appendix B.1. N = 436,519, including 2,990 ambulance services; σ_μ denotes the standard deviation 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). † p<.05

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

	Full Sample			*Ambulance Sample*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Log Expenditures	0.0169†	0.0080	-0.0002	0.0489†	0.0398†	0.0449
	(0.0055)	(0.0045)	(0.0054)	(0.0238)	(0.0130)	(0.0263)
Cat I: Beta Blockers	0.0542†	0.0158	-0.0488†	0.0736	0.0647†	-0.8507
	(0.0134)	(0.0110)	(0.0156)	(0.0541)	(0.0282)	(0.4589)
Cat I: Statins	0.0445†	0.0657†	0.0114	0.0232	0.0973†	-0.2877
	(0.0117)	(0.0096)	(0.0138)	(0.0537)	(0.0247)	(0.4836)
Cat I: ACE/ARB Drugs	-0.0047	-0.0093	-0.0432†	0.0206	0.0061	-0.8962†
	(0.0123)	(0.0097)	(0.0147)	(0.0404)	(0.0238)	(0.3604)
Cat I: Same-Day PCI	0.1068†	0.0745†	0.0245†	0.1035†	0.0781†	0.0748†
	(0.0066)	(0.0047)	(0.0068)	(0.0219)	(0.0135)	(0.0310)
Cat II: Later PCI	0.0370†	0.0323†	0.0104	0.0856	0.0777†	0.0966
	(0.0128)	(0.0110)	(0.0126)	(0.0556)	(0.0331)	(0.0588)
Cat II: N of Doctors	0.0020†	0.0016†	0.0004	0.0021†	0.0033†	-0.0001
	(0.0003)	(0.0002)	(0.0003)	(0.0010)	(0.0006)	(0.0015)
Cat II: Number of Scans	-0.0207†	-0.0105†	-0.0044	-0.0282†	-0.0254†	-0.0166
	(0.0032)	(0.0027)	(0.0032)	(0.0112)	(0.0076)	(0.0124)
Cat II: ICU/CCU Days	-0.0008†	-0.0006	0.0007	-0.0000	-0.0024†	-0.0030
	(0.0003)	(0.0003)	(0.0005)	(0.0011)	(0.0006)	(0.0021)
Cat III: Post-Acute Care	-0.0130	-0.0613†	-0.0590†	-0.1261†	-0.1369†	-0.1800†
	(0.0164)	(0.0142)	(0.0184)	(0.0587)	(0.0357)	(0.0893)
Cat III: Nursing Home Days	-0.0020†	0.0013†	-0.0000	-0.0016	0.0055†	0.0020
	(0.0006)	(0.0005)	(0.0006)	(0.0022)	(0.0012)	(0.0025)
Cat III: Home Health ($1000)	-0.0002	0.0036	0.0015	-0.0116	0.0017	-0.0207
	(0.0053)	(0.0041)	(0.0050)	(0.0155)	(0.0105)	(0.0206)
Cat III: Abdomen CT Measure	0.0013	-0.0038	-0.0057	0.0201	0.0106	0.0259
	(0.0049)	(0.0038)	(0.0049)	(0.0230)	(0.0114)	(0.0943)
Cat III: Thorax CT Measure	-0.0165†	-0.0199†	-0.0040	-0.0510	-0.0572†	-0.2088
	(0.0080)	(0.0067)	(0.0082)	(0.0442)	(0.0252)	(0.1822)
Zip Code Fixed Effects	X			X		
Hospital / Ambulance Random Effects		X			X	
Hospital / Ambulance Fixed Effects			X			X
Number of observations	1,617,039	1,617,039	1,617,039	436,519	436,519	436,519
Adjusted R-squared	0.182		0.172	0.183		0.162
Panel-level std. deviation		0.0219	0.038		0.0225	0.1457

Notes: All regressions include risk-adjustment (see Appendix Table B.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). † p<.05

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

	Full Sample			*Ambulance Sample*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Log Expenditures	$0.0515^{\dagger}$	$0.0272^{\dagger}$	$0.0173^{\dagger}$	$0.0701^{\dagger}$	$0.0500^{\dagger}$	0.0384
	(0.0058)	(0.0041)	(0.0049)	(0.0267)	(0.0123)	(0.0237)
PCA prediction for input1 component1	$0.0125^{\dagger}$	$0.0088^{\dagger}$	$0.0059^{\dagger}$	$0.0092^{\dagger}$	$0.0077^{\dagger}$	$0.0132^{\dagger}$
	(0.0004)	(0.0004)	(0.0007)	(0.0013)	(0.0006)	(0.0054)
PCA prediction for input2 component1	$-0.0049^{\dagger}$	$-0.0027^{\dagger}$	$-0.0048^{\dagger}$	-0.0020	-0.0001	-0.0021
	(0.0007)	(0.0005)	(0.0006)	(0.0017)	(0.0008)	(0.0022)
PCA prediction for input3 component1	-0.0006	$-0.0026^{\dagger}$	$-0.0015^{\dagger}$	-0.0027	$-0.0057^{\dagger}$	$-0.0081^{\dagger}$
	(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
Number of observations	1,617,039	1,617,039	1,617,039	436,519	436,519	436,519
Adjusted R-squared	0.182		0.172	0.183		0.162
Panel-level std. deviation		0.023	0.034		0.0247	0.0882

Notes: All Regressions include full set of risk-adjustment covariates; see Appendix B.1; σ_{μ} denotes the standard deviation 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. † p<.05

Table 5: Regressions Estimates of the First Principal Components for Categories I, II, and III: Hospital Level Analysis

	PC: Category I	PC: Category II	PC: Category III
Fraction Cowboy	-0.519	1.213†	0.169
	(0.367)	(0.361)	(0.270)
Fraction Comforter	0.081	-0.181	-0.581†
	(0.393)	(0.308)	(0.265)
Fraction High Follow-Up	-1.543†	3.196†	0.232
	(0.741)	(0.565)	(0.494)
Fraction Low Follow-Up	2.178†	-2.256†	0.569
	(0.781)	(0.741)	(0.701)
Teaching Hospital	0.495†	0.410†	-0.068
	(0.124)	(0.104)	(0.080)
Log AMI Volume	0.517†	0.031	-0.041
	(0.078)	(0.061)	(0.046)
Intercept	-3.286†	-0.479	0.377
	(0.548)	(0.452)	(0.383)
Number of observations	1,490	1,490	1,490
R-squared	0.142	0.219	0.025

Notes: The dependent variable for each of these three regressions is the hospital-specific first principal component (PC) for Category I, II, and III AMI inputs as defined in the text. The measures of cowboys, comforters, low-follow-up and high-follow-up come from responses by physicians to vignettes about how they would treat a hypothetical patient, aggregated up to the HRR level; see Cutler et al. (2019) for details. We restrict the number of HRRs to those with at least 4 physicians responding to the vignettes, with 1,490 hospitals in total accounting for 72 percent of the entire sample; results were not sensitive to inclusion rules. We also include a dummy variable for teaching hospital and a measure of the log AMI volume for the entire period 2007-2016. Regressions weighted by log(volume) and clustered by HRR. † $p<.05$

A Appendix: Additional Derivations

A.1 Multi-Input Comparative Statics Analysis for a General Production Function

In this section we generalize Equation (1) by deriving the comparative statics for the n-input case. Consider a hospital production function with n inputs $f(x_1, x_2, \ldots, x_n)$ subject to a cost constraint $\sum_{i=1}^{n} p_i x_i = C$

If we increase input x_k slightly while adjusting the other inputs to maintain total cost:

$$\sum_{j \neq k} p_j \, dx_j = -p_k \, dx_k \tag{A.1}$$

The test requires understanding how output (survival) changes with changing the input mix while holding total costs constant. The total differential of output is:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \, dx_i = \frac{\partial f}{\partial x_k} dx_k + \sum_{j \neq k} \frac{\partial f}{\partial x_j} dx_j \tag{A.2}$$

Using the cost constraint adjustment:

$$df = \frac{\partial f}{\partial x_k} dx_k + \sum_{j \neq k} \frac{\partial f}{\partial x_j} \left(-\frac{p_k}{p_j} \cdot dx_k \right) \tag{A.3}$$

In a hospital that optimally allocates inputs:

$$\frac{\partial f / \partial x_i}{p_i} = \psi \quad \text{for all } i \quad \Rightarrow \quad \frac{\partial f}{\partial x_i} = \psi p_i \tag{A.4}$$

Substituting:

$$df = \psi \sum_{i=1}^{n} p_i dx_i = \psi \cdot 0 = 0$$

At the optimum:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = 0 \quad (\text{if } \sum_{i=1}^{n} p_i dx_i = 0) \tag{A.5}$$

If the firm is misallocating inputs, then $df \neq 0$, and reallocating cost toward inputs with higher marginal product per dollar will increase output.

A.2 The Cobb-Douglas Production Function

In this Appendix, we demonstrate how a Cobb-Douglas production model with multiple inputs yields an estimation equation of the form in Equation 6 Suppose output is written

$$\begin{aligned} & Y_{ih} = \tilde{Y}_{ih} e^{\varepsilon_i} \\ & where\ \tilde{Y}_{ih} = A_h \left[\prod_k X_{ik}^{\gamma_k}\right] \end{aligned} \tag{A.6}$$

We write "produced" survival $\tilde{Y}_{ih}$ as depending on hospital-specific productivity A_h, healthcare inputs X_{ik}, $k = 1, ..., K$; the actual hospital-specific survival rate is equal to $\tilde{Y}_{ih}$times the (exponential) error term ε_i.[25] Not implausibly we assume diminishing returns, so that $\sum_k \gamma_k < 1$; we ignore issues of risk adjustment here. In this model, the hospital system (including physicians, nurses, ancillary health employees, and administrators) is the relevant decision-making entity; Hicks-neutral TFP 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 consider the optimization problem of a hospital that seeks to maximize the hospital-specific value of survival minus costs.[26] That is, hospitals are assumed to maximize:

$$\Omega_h = \varphi_h \tilde{Y}_h - \left[\sum_{k=1}^{K} p_k x_{hk} + F_h\right] \tag{A.7}$$

where $\Omega_h$is the objective function for hospital h. The first term is the value to the hospital of the outcome, where $\varphi_h$translates survival into a dollar value (e.g., the social value of health improvements); it may differ across hospitals but corresponds to the slope of the production function S in Figure 1. Hospital costs are summed over all measured inputs 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 (A.7) is:

$$\varphi_h \frac{\partial \tilde{Y}_h}{\partial x_k} - p_k = 0 \tag{A.8}$$

where the first term in (A.8) 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 variable cost. Now suppose that regulations, informational barriers, or internal resistance to using effective treatments exist across hospitals; this is summarized by $\mu_{hk}$for the kth input as in the text; in the absence of pre-existing distortions (e.g., when $\phi_h = 0$) the shadow price can

[25]as noted previously, we take the exponent of survival as "produced" health so the log-linear specification allows for binary outcomes of survival.

[26]This approach ignores the complicated incentives provided to hospitals to maximize inputs for which the reimbursement rate exceeds marginal costs; this type of "supplier-induced demand" however would still lead to a rejection of the null.

be either positive, leading to underuse, or negative, leading to overuse: [27]

$$\varphi_h \frac{\partial \tilde{Y}_h}{\partial x_k} - p_k = \mu_{hk} \tag{A.9}$$

For analytical purposes, it is easier to define the distortion as a shadow price proportional to the true price, so that

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

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

$$x_{hk} = x_{h1} \left[\frac{\gamma_k}{p_k \gamma_1} \right] \tag{A.10}$$

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

$$M_h = \sum_k p_k x_{hk} = x_{h1} (\sum_k \left[\frac{\gamma_k}{\gamma_1} \right]) \tag{A.11}$$

so that total expenditures M_h can be written solely as a function of X_{h1} (e.g., the dual).

Using the same first-order condition in (A.10), we can similarly express output solely as a function of the numeraire input X$_{h1}$

$$\tilde{Y}_h = A_h \left[X_{h1} \prod_k \left[\frac{\gamma_k}{p_k \gamma_1} \right] \right] \tag{A.12}$$

where $\tilde{Y}_h$is again "produced" health by the hospital for an average patient. This means that normalized output can be written:

$$\tilde{Y}_h = A_h \frac{\left[{M_h}^{\gamma'} \prod_k \left[\frac{\eta_k}{p_k} \right]^{\gamma_k} \right]}{\sum_k [\eta_k]} \tag{A.13}$$

where $\eta_k = \gamma_k / \gamma_1$ and $\gamma' = \sum_k \gamma_k$.

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

$$y_h = \alpha_h + \gamma' m_h + \left[\sum_k \beta_k \ln \left(\frac{\eta_k}{p_k} \right) - \gamma' \ln(\sum_k \eta_k) \right] + \varepsilon_h \tag{A.14}$$

We are unlikely to separately identify the terms in the brackets but note that all these

[27]See Díaz-Hernández et al. (2008) for a similar "shadow price" approach.

terms that depend 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 input 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 $\gamma_k$coefficients are the same across hospitals, specific input choices should not predict outcomes conditional on m_h. That is, under the null of efficient input choices, survival can be written simply as

$$y_h = C + \alpha_h^* + \gamma' m_h + \varepsilon_h \tag{A.15}$$

where α^* denotes the hospital-specific productivity measures adjusted by the constant term in brackets in Equation A.14, and C is the constant term in brackets. Note that if *all* hospitals exhibit the same vector of distortions $\lambda_{hk} = \lambda_k \neq 1$ then we could falsely conclude a lack of distortions (since C would be constant across hospitals), as noted earlier in the text. However, there is sufficient variation in input choices across hospitals so as to make this scenario unlikely.

Suppose that the vector Γ of the Cobb-Douglas production function differed across hospitals. This would imply that each hospital should choose a different bundle of inputs. Thus when pooling data across hospitals, one would incorrectly reject the null hypothesis that outcomes are affected by input choices. That said, we show below that our test is valid in fixed-effects regression analysis using longitudinal data. We therefore rewrite Equation (A.14) to account for misallocation and differences in TFP across hospitals with hospital-specific γ_{hk} and hospital-specific prices p_{hk}.

$$y_h = \alpha_h^* + \gamma' m_h + \left[\sum_k \ln \eta_{hk}^* - \ln(\sum_k \eta_{hk}^*) \right] + \varepsilon_h \tag{A.16}$$

where

$$\eta_{hk}^* = \left[\frac{\lambda_{hk}\gamma_{hk}}{\lambda_{h1} p_{hk} \gamma_{h1}} \right]. \tag{A.17}$$

Once again, under the null hypothesis of efficient input use within the hospital, the regression specification would be

$$y_h = C_h + \alpha_h^* + \gamma_h' m_h + \varepsilon_h \tag{A.18}$$

where the constant term C_h is now specific to hospital h. Note also that the coefficient on total expenditures, $gamma_h'$ may vary to the extent that the sum of the $gamma_{hk}$ parameters by hospital, but empirically these estimates are modest and show little variation across hospital types. As before, any deviations from the optimal hospital-specific mix will appear as

We can also use the property of the Cobb-Douglas model to consider why the use of the kth input may differ across hospitals and over time. If we rewrite Equation A.10 to compare input choice (on average, per patient) across hospitals, then in its most general form,

$$\frac{x_{hk}}{x_{h'k}} = \frac{\tilde{Y}_h \lambda_{h'k} p_{h'k} \gamma_{h'k}}{\tilde{Y}_{h'} \lambda_{hk} p_{hk} \gamma_{hk}} \tag{A.19}$$

That is, the ratio of the kth input in hospital h versus hospital h' depends on (a) the ratio of output or (the exponent) of survival rates, (b) the ratio of the shadow price distortion, (c) the ratio of prices faced by hospitals – for example, if the price of an interventional cardiologist is disproportionately higher at hospital h' (relative to the cost-of-living reimbursements provided by Medicare), and (d) the ratio of the Cobb-Douglas production parameters γ_k. Thus a 20% difference in prices, output, or production parameters would each imply a 20% difference in input choices (given that the elasticity of substitution is 1 in Cobb-Douglas production functions) but variation in the ratio of inputs (for example in the use of PCI or home health care) on the order of 3 to 1 or more is unlikely to be generated by (a) through (c).

B Appendix: Additional Regression Results

Table B1: Risk Adjustment Coefficient Estimates

	Full Sample	*Ambulance Sample*
Vascular Disease	0.0111†	0.0108†
	(0.0014)	(0.0026)
Pulmonary Disease	-0.0286†	-0.0277†
	(0.0011)	(0.0019)
Renal Disease	-0.0578†	-0.0576†
	(0.0011)	(0.0020)
Non-metastatic Cancer	-0.0695†	-0.0733†
	(0.0020)	(0.0037)
Metastatic Cancer	-0.3230†	-0.3065†
	(0.0032)	(0.0072)
Congestive Heart Failure	-0.1187†	-0.1189†
	(0.0012)	(0.0018)
Liver Disease	-0.0864†	-0.0734†
	(0.0053)	(0.0111)
Diabetes	0.0450†	0.0479†
	(0.0010)	(0.0016)
Rheumatologic Disease	0.0268†	0.0299†
	(0.0026)	(0.0047)
Dementia	-0.1416†	-0.1437†
	(0.0024)	(0.0038)
Fraction Native American	-0.0044	0.0037
	(0.0044)	(0.0087)
Fraction Hispanic	0.0290†	0.0233†
	(0.0026)	(0.0060)
Fraction Other Race/Ethnicity	-0.0005	0.0082
	(0.0034)	(0.0065)
Fraction Asian	0.0142†	0.0246†
	(0.0033)	(0.0063)
Fraction Black	0.0019	-0.0091†
	(0.0019)	(0.0033)
Anterior MI	0.0955†	0.0944†
	(0.0028)	(0.0046)
Inferior MI	0.1355†	0.1434†
	(0.0027)	(0.0041)
Right MI	0.1127†	0.1212†
	(0.0038)	(0.0066)
Subendocardial MI	0.1658†	0.1536†
	(0.0027)	(0.0044)

Other MI	0.0914†	0.0936†
	(0.0035)	(0.0066)
Zip Code Income Q2	0.0022	0.0063†
	(0.0018)	(0.0024)
Zip Code Income Q3	0.0009	0.0074†
	(0.0021)	(0.0025)
Zip Code Income Q4	-0.0005	0.0067†
	(0.0020)	(0.0025)
Zip Code Income Q5	0.0015	0.0182†
	(0.0024)	(0.0033)
Patient Dual Eligible	-0.0312†	-0.0346†
	(0.0021)	(0.0042)
Proportion MA in HRR	-0.0014	0.0137
	(0.0168)	(0.0126)
HCC Score Q2	-0.0280†	-0.0369†
	(0.0010)	(0.0021)
HCC Score Q3	-0.0604†	-0.0760†
	(0.0011)	(0.0022)
HCC Score Q4	-0.1154†	-0.1349†
	(0.0014)	(0.0026)
HCC Score Q5	-0.2577†	-0.2728†
	(0.0016)	(0.0028)
Number of observations	1,617,039	436,519
Adjusted R-squared	0.182	0.180

Notes: † p<.05. The dependent variable is one-year survival. Column 1 is the least squares regression on the full sample, and Column 2 is the least squares regression for the ambulance sample.

Table B2: Regression Estimates of 1-Year Survival with Flexible Coefficients: Full Sample

	Full Sample		
	Model 1	Model 2	Model 3
Log Spending Q23	0.0051†	0.0035†	0.0011
	(0.0010)	(0.0010)	(0.0011)
Log Spending Q4	0.0089†	0.0056†	0.0021
	(0.0014)	(0.0013)	(0.0015)
Beta Blockers Q23	0.0080†	0.0052†	-0.0034
	(0.0020)	(0.0014)	(0.0021)
Beta Blockers Q4	0.0119†	0.0071†	-0.0068†
	(0.0026)	(0.0019)	(0.0027)
Statins Q23	0.0033	0.0051†	-0.0006
	(0.0019)	(0.0014)	(0.0021)
Statins Q4	0.0104†	0.0121†	-0.0027
	(0.0024)	(0.0019)	(0.0028)
ACE/ARB Drugs Q23	0.0027	-0.0003	-0.0038
	(0.0017)	(0.0013)	(0.0020)
ACE/ARB Drugs Q4	0.0017	-0.0017	-0.0075†
	(0.0021)	(0.0017)	(0.0025)
Same Day PCI Q23	0.0202†	0.0120†	0.0019
	(0.0017)	(0.0011)	(0.0014)
Same Day PCI Q4	0.0281†	0.0181†	0.0036†
	(0.0020)	(0.0015)	(0.0018)
Later PCI Q23	0.0013	0.0005	-0.0002
	(0.0010)	(0.0009)	(0.0010)
Later PCI Q4	0.0028†	0.0020	0.0010
	(0.0013)	(0.0011)	(0.0013)
N of Doctors Q23	0.0024	0.0029†	0.0002
	(0.0014)	(0.0012)	(0.0014)
N of Doctors Q4	0.0102†	0.0087†	0.0028
	(0.0020)	(0.0016)	(0.0019)
N of Scans Q23	-0.0046†	-0.0030†	-0.0008
	(0.0012)	(0.0010)	(0.0011)
N of Scans Q4	-0.0112†	-0.0062†	-0.0031†
	(0.0016)	(0.0014)	(0.0015)
ICU/CCU Days Q23	-0.0051†	-0.0035†	0.0008
	(0.0013)	(0.0011)	(0.0015)
ICU/CCU Days Q4	-0.0053†	-0.0037†	0.0003
	(0.0016)	(0.0015)	(0.0020)
Post-Acute Care Q23	0.0025†	-0.0011	-0.0011
	(0.0011)	(0.0010)	(0.0011)

Post-Acute Care Q4	0.0021	-0.0038†	-0.0027
	(0.0015)	(0.0013)	(0.0016)
Nursing Home Days Q23	-0.0045†	-0.0010	-0.0027†
	(0.0011)	(0.0010)	(0.0011)
Nursing Home Days Q4	-0.0092†	0.0000	-0.0010
	(0.0016)	(0.0013)	(0.0015)
Home Health Q23	0.0007	0.0004	0.0006
	(0.0011)	(0.0009)	(0.0010)
Home Health Q4	-0.0001	0.0007	0.0009
	(0.0014)	(0.0012)	(0.0013)
Abdomen CT Q23	0.0003	0.0000	-0.0005
	(0.0013)	(0.0011)	(0.0013)
Abdomen CT Q4	-0.0016	-0.0024	-0.0015
	(0.0016)	(0.0014)	(0.0018)
Thorax CT Q23	0.0011	-0.0007	0.0002
	(0.0013)	(0.0010)	(0.0011)
Thorax CT Q4	-0.0014	-0.0035†	0.0000
	(0.0017)	(0.0014)	(0.0016)
Number of observations	1,617,039	1,617,039	1,617,039
Adjusted R-squared	0.182		0.172
Panel-level standard deviation		0.0224	0.0390

Notes: This regression allows for heterogeneity in the coefficients for each input. The reference group in each case is the bottom quartile of hospital-level input utilization; Q23 denotes quartiles 2 and 3 (the middle 50 percentile), while Q4 denotes the top quartile of utilization. † $p<.05$

Table B3: Principal Components Regression Results

	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

Notes: Results from Principal Components Analysis: Proportion of Variance Explained, and Eigenvectors, of the First Component.

Table B4: Regressions Estimates of 30-Day Survival with Principal Component Estimates of Misallocation: Full and Ambulance Samples

	Full Sample			*Ambulance Sample*		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Log Expenditures	$0.0146^{\dagger}$	0.0002	$-0.0112^{\dagger}$	$0.0528^{\dagger}$	$0.0328^{\dagger}$	$0.0457^{\dagger}$
	(0.0040)	(0.0033)	(0.0039)	(0.0154)	(0.0080)	(0.0194)
PCA prediction for input1 component1	$0.0070^{\dagger}$	$0.0050^{\dagger}$	$0.0044^{\dagger}$	$0.0057^{\dagger}$	$0.0052^{\dagger}$	$0.0117^{\dagger}$
	(0.0003)	(0.0003)	(0.0005)	(0.0008)	(0.0004)	(0.0044)
PCA prediction for input2 component1	-0.0006	$0.0014^{\dagger}$	$0.0017^{\dagger}$	0.0008	$0.0030^{\dagger}$	0.0019
	(0.0005)	(0.0004)	(0.0005)	(0.0013)	(0.0005)	(0.0018)
PCA prediction for input3 component1	$-0.0009^{\dagger}$	$-0.0018^{\dagger}$	$-0.0015^{\dagger}$	-0.0014	$-0.0031^{\dagger}$	$-0.0062^{\dagger}$
	(0.0004)	(0.0003)	(0.0004)	(0.0012)	(0.0004)	(0.0024)
Number of observations	1,617,039	1,617,039	1,617,039	436,519	436,519	436,519
Adjusted R-squared	0.083		0.079	0.083		0.072
Panel-level standard deviation (σ_{μ})		0.0180	0.0254		0.0000	0.0664

Notes: All Regressions include full set of risk-adjustment covariates; see Appendix B.1; σ_{μ} denotes the standard deviation 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. † p<.05