NBER WORKING PAPER SERIES

IDENTIFYING SOURCES OF INEFFICIENCY IN HEALTH CARE

Amitabh Chandra
Douglas O. Staiger

Working Paper 24035
http://www.nber.org/papers/w24035

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

This research was funded by the National Institute of Aging (NIA) P01 AG19783-02. We thank Janet Currie, Joe Doyle, Mark Duggan, Amy Finkelstein, Peter Hull, Matt Notowidigdo, Jonathan Skinner, and Heidi Williams for comments that have greatly improved our paper. We obtained access to the proprietary data used in this paper through a data use agreement between the Centers for Medicare and Medicaid Services (CMS) and Dartmouth Medical School. Readers wishing to use these data must obtain them from CMS, but programs and output are available from the authors. 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 a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24035.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

Identifying Sources of Inefficiency in Health Care
Amitabh Chandra and Douglas O. Staiger
NBER Working Paper No. 24035
November 2017
JEL No. I1,I3

ABSTRACT

In medicine, the reasons for variation in treatment rates across hospitals serving similar patients are not well understood. Some interpret this variation as unwarranted, and push standardization of care as a way of reducing allocative inefficiency. However, an alternative interpretation is that hospitals with greater expertise in a treatment use it more because of their comparative advantage, suggesting that standardization is misguided. We develop a simple economic model that provides an empirical framework to separate these explanations. Estimating this model with data on treatments for heart attack patients, we find evidence of substantial variation across hospitals in both allocative inefficiency and comparative advantage, with most hospitals overusing treatment in part because of incorrect beliefs about their comparative advantage. A stylized welfare-calculation suggests that eliminating allocative inefficiency would increase the total benefits from this treatment by about a third.

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

Douglas O. Staiger
Dartmouth College
Department of Economics
HB6106, 301 Rockefeller Hall
Hanover, NH 03755-3514
and NBER
douglas.staiger@dartmouth.edu

A large and influential literature in economics and medicine has documented substantial variation in treatment rates and patient outcomes across hospitals even after carefully controlling for differences in patient risk [Skinner (2013); Institute of Medicine (2013)]. But this variation in treatment rates could arise from two different mechanisms. The conventional interpretation in the medical literature is that there is a correct amount of use, so that variation across providers in risk-adjusted treatment rates is evidence of allocative inefficiency: some providers are using too much care and others are using too little. This interpretation of variation has led to an emphasis on guidelines and developing and disseminating information on cost-effectiveness of care. An alternative interpretation argues that the ability to deliver treatment varies across providers, so that hospitals who can obtain higher benefits from a given treatment deliver more of that treatment because of their comparative advantage. This interpretation leads to an emphasis on understanding the sources of variation in hospital-specific skill and efforts to improve quality, instead of trying to standardize care.

We develop a simple economic framework that can distinguish between these explanations and shed light on the mechanisms behind them. Our framework builds on a generalized Roy model of treatment choice along the lines of Chandra and Staiger (2007), where treatment choice depends on the expected benefits of treatment relative to usual care. In this model, differences across hospitals in risk-adjusted treatment rates do not separately identify allocative inefficiency because they also capture differences in comparative advantage across hospital in providing the treatment. However, if treatment is being allocated efficiently to patients then any difference in the propensity to be treated, whether across patients or across hospitals, should solely reflect differences in the expected benefit from the treatment. Therefore, allocative inefficiency can be identified when the benefit of treatment is different across hospitals for patients with the same propensity to be treated. Furthermore, since low propensity patients are those least likely to benefit from treatment, overuse of the treatment in a given hospital can be identified when the treatment does harm among patients with the lowest propensity to receive treatment.

We apply this model to clinical data on heart-attacks and their treatment—reperfusion therapy. While our model can be applied to any healthcare setting, heart-attack treatments have several features that make them particularly suited for the analysis: outcomes are easily measured and agreed upon, and questions about overuse, underuse and comparative-advantage are central to treatment decisions. Estimating our model with data for elderly patients following a heart-attack, we find strong evidence of allocative inefficiency, with most hospitals overusing reperfusion therapy to the point that low propensity patients are harmed by the treatment. However, we also we find substantial variation in hospitals' ability to perform treatment (comparative advantage), with the variation across hospitals in the survival benefit from reperfusion being the same order of magnitude as the average treatment effect of reperfusion. Thus,

we find that both allocative inefficiency and comparative advantage contribute to variations in treatment-rates.

We use this framework to explore mechanisms that could lead to the allocative inefficiency that we observe in the data. One possibility, motivated by Currie and MacLeod (2017), is that allocative inefficiency would arise if hospital's had imperfect information and misperceived their ability to deliver treatment. In this mechanism, allocative inefficiency arises because hospitals base treatment decisions on their incorrect perception of the benefits of treatment in their patients, rather than on the true benefits of treatment. Given the general lack of systematic performance feedback and small samples of their own treated patients to observe, it is quite plausible that hospitals and physicians will have inaccurate beliefs about their own treatment effectiveness. We find evidence in favor of this mechanism, with smaller hospitals having particularly imprecise information about their own treatment effectiveness. Another explanation is that hospitals are optimizing something other than the survival of a given patient, e.g. over-treating for financial gain (particularly in for-profit hospitals) or because of benefits to future patients through learning-by-doing (particularly in teaching hospitals). This type of mechanism would suggest that allocative inefficiency would be related to hospital characteristics such as ownership, teaching status, etc. We find little *prima facie* evidence for this hypothesis—overuse is not correlated with a hospital's for-profit status or other characteristics such as being a teaching-hospital.

Our contribution connects the vast empirical literature on variations in medical care to the broader economics literature on productivity and technology adoption. Most of the literature on variations in medical care has ignored the role of productivity in driving variation in treatment, and instead debated whether finding variation in risk-adjusted treatment rates that is unrelated to patient outcomes points to allocative inefficiency or is simply due to inadequate risk-adjustment [Fisher et al (2003a, 2003b), Yasaitis (2009), Skinner (2011), Doyle (2011), Doyle, Graves, Gruber, Kleiner (2015), Finkelstein, Gentzkow, Williams (2016)]. In contrast, research influenced by the productivity literature (Syverson, 2011) has emphasized productivity differences in healthcare [Chandra and Staiger (2007), Chandra et al (2016), Skinner and Staiger (2015), Currie and MacLeod (2017)], but ignored the possibility of allocative inefficiency across hospitals. More specifically, evidence that comparative advantage drives variation across firms in technology adoption has been found in agriculture (Suri, 2011) and in health care (Chandra and Staiger, 2007), but these papers do not account for allocative inefficiency. Our paper is the first to separately identify variation due to comparative advantage from that due to allocative inefficiency. We build on the framework used in Chandra and Staiger (2010), but where that paper focused on differences in treatment rates across demographic groups this paper focusses on differences across hospitals. Our contribution is closest to that of Abaluck et al. (2016), who build on an earlier working-paper version of this paper. While they allow for allocative inefficiency and physician level expertise in

selecting patients for testing, we differ in allowing for comparative advantage and productivity differences across hospitals in addition to differences in allocative efficiency.

The paper proceeds as follows: Section I provides some background on heart attack biology and treatment, and describes the data. Section II develops the theoretical model underlying our analysis, and links it to our estimation strategy, paying particular attention to how allocative inefficiency will be identified separately from comparative advantage and productivity differences. Section III presents results: we start with simple graphical results and regressions which then motivate a more parametric approach. In Section IV we use our framework to understand mechanisms and estimate a simple bayesian learning model for learning comparative advantage and compare its fit, relative to simpler stories about learning and financial-incentives. We conclude by performing a highly stylized calculation of the welfare-loss from variation in treatment rates.

I. Heart-Attacks: Treatments, and Data

A. Treatments

Heart attacks (more precisely, acute myocardial infarction (AMI)) occur when the heart-muscle (the myocardium) does not receive sufficient oxygen, because of a blockage in one of the coronary arteries which supply blood to the heart. The blockage is typically caused by a blood clot that occurs because of coagulation induced by the rupture of atherosclerotic plaque inside the coronary arteries, and must be reperfused rapidly. There are two ways to give patients reperfusion (which is the treatment that we study): first, thrombolytics, also known as fibrinolytics, are administered intravenously and break down blood clots by pharmacological means (these drugs include tissue plasminogen activators, streptokinanse and urokinase). Not everyone is appropriate for thrombolytics—patients with strokes, peptic ulcers, head-trauma, dementia, advanced liver disease, and uncontrolled hypertension aren't appropriate for this treatment because of the risk of further bleeding induced by the treatment. In our data from the mid-1990s, over 90 percent of patients receiving reperfusion received thrombolytics. Reperfusion can also be performed through angioplasty (where a balloon on a catheter is inflated inside the blocked coronary artery to restore blood flow). Following the clinical literature, we define a patient to have received reperfusion if any of these therapies was provided within 12 hours of the heart attack.

We focus our empirical work on the treatment of AMI for a number of reasons. First, cardiovascular disease, of which heart attacks are the primary manifestation, is the leading cause of death in the US. A perusal of the leading medical journals would indicate that heart attack treatments are constantly being refined, and a large body of trial evidence points to significant therapeutic gains from many of these treatments. In this context, variation in treatments across hospitals may directly translate

into lost lives, and there is a rich tradition of studying variation across hospitals in treatments and outcomes after heart attacks.

Second, because of what is known about heart attack treatments from randomized controlled trials, we are able to assess whether our regression estimates of the benefits from reperfusion are comparable to those found in the medical literature, or whether they are confounded by selection-bias. We focus on reperfusion, where our use of chart data allows us to replicate the RCT evidence that is summarized by the Fibrinolytic Therapy Trialists' Collaborative Group (1994). Chart data provides comprehensive documentation on the patient's condition at the time that the treatment decision is made, and therefore minimizes the possibility that unobserved clinical factors related to a patient's survival are correlated with treatment.

Third, because mortality post-AMI is high (mortality rates at 30 days are nearly 20 percent), a well-defined endpoint is available to test the efficacy of heart attack treatments. Moreover during the acute phase of the heart attack the therapeutic emphasis is on maximizing survival, which is achieved by timely reperfusion, and hospital staff (not patients and their families) make treatment decisions. This would not be true if we focused on treatment variation for more chronic conditions such as diabetes, chronic obstructive pulmonary disease, or arthritis where because of the importance of quality-of-life there would be considerable disagreement on how to measure productivity.

Fourth, heart attacks are an acute condition for which virtually all patients are hospitalized at a nearby hospital and receive some medical care. This may not be true of more chronic conditions such as diabetes or heart-failure where many patients aren't diagnosed and selection into the sample confounds the analysis.

B. Data

Because acute myocardial infarction is both common and serious, it has been the topic of intense scientific and clinical interest. One effort to incorporate evidence-based practice guidelines into the care of heart attack patients, begun in 1992, is the Health Care Financing Administration's Health Care Quality Improvement Initiative Cooperative Cardiovascular Project (CCP). Information about patients admitted to hospitals for treatment of heart attacks in 1994/1995 was obtained from clinical records. The CCP is considerably superior to administrative/claims data of the type used by McClellan et al. (1994) as it collects chart data on the patients—detailed information is provided on laboratory tests, enzyme levels, the location of the myocardial infarction, and the condition of the patient at the time of admission. Detailed clinical data were abstracted from each patient's chart using a standard protocol. Further details about the CCP data are available in Marciniak et al. (1998), O'Connor et al. (1999), and in the appendix to this paper. The choice of sample and variables is identical to what we used and described in Barnato et

al. (2005) and Chandra and Staiger (2007, 2010).

In Table 1 we report some basic characteristics of our sample overall, and by whether the patient received reperfusion within 12 hours of admission to the hospital. In our sample, 19% of patients received reperfusion within 12 hours of admission for a heart attack. Overall, 81% of patients were still alive 30 days after admission, but survival was higher for patients receiving reperfusion (86%) than for patients who did not receive reperfusion (80%). However, much of the difference in survival between these two groups was due to differences in underlying health and pre-existing conditions, rather than the result of reperfusion. Patients receiving reperfusion were younger, and much less likely to have pre-existing conditions such as congestive heart failure, hypertension, diabetes, and dementia. Because of the selection of healthier patients into reperfusion, controlling for the detailed clinical variables available in the CCP is critical for estimating the effect of reperfusion on survival.

II. Theory and Estimation

A Roy model of patient treatment choice guides our empirical work. We assume that a hospital must choose between two treatment options for every patient: whether to offer reperfusion (treatment) or not (usual care). Treatment is provided to each patient whenever a patient's expected benefit from the treatment exceeds a minimal threshold. In our framework, there are two ways in which a patient's hospital could affect treatment. First, because of comparative-advantage, the benefit of treatment for a given patient may vary across hospitals, reflecting each hospital's expertise in providing the treatment. Second, because of allocative efficiency, the minimum threshold for receiving care may vary across hospitals. From the patient's point of view, treatment should be provided whenever the expected benefit from treatment exceeds zero. Therefore, there is underuse of the treatment in hospitals that set a minimum benefit threshold above zero, and overuse in hospitals that set a minimum threshold below zero.

A. Patient Outcomes

To formalize this, let Y_{ih}^1 represent the survival for patient *i* at hospital *h* if the patient receives the treatment (reperfusion) and let Y_{ih}^0 represent the survival if the patient does not receive the treatment, but otherwise receives usual medical care. We focus on the health benefits of the treatment, which in our setting is survival, but in other settings would include any reduction in mortality or morbidity that was expected from the treatment, e.g. the impact of the treatment on Quality Adjusted Live Years (QALYs).[1]

[1] For now, we abstract from the problem that hospitals should stop treatment prior to achieving zero marginal benefits—that is, that providers should maximize benefits net of costs. In practice, the cost of treating heart-attacks is small relative to the survival benefit but in other settings, such as oncology, this may not be true.

Treatment decisions are based on expected survival given the information available to the provider at the time of treatment. If receiving usual care, a patient's expected survival $E(Y_{ih}^0)$ depends on the hospital's general level of expertise α_h^0, observable patient characteristics X_{ih} such as age, medical history and lab results, and other unmeasured factors affecting baseline mortality v_{ih}^0 that are observed by the healthcare provider but not by the econometrician. If treated with reperfusion, a patient's expected survival $E(Y_{ih}^1)$ depends on a similar set of factors representing the hospital's expertise at providing the treatment α_h^1, patient characteristics (which may have a different relationship to survival when patients receive the treatment), and other unmeasured factors v_{ih}^1 that affect the expected benefits of reperfusion. The presence of two productivity parameters α_h^1 and α_h^0, allows us to model hospital specific benefits at both forms of medicine— usual care and reperfusion (treatment).

Actual (realized) survival if receiving usual care or reperfusion is equal to expected survival plus a random error term $(\varepsilon_{ih}^0, \varepsilon_{ih}^1)$, which yields survival equations of the following form:

(1a) $$Y_{ih}^0 = E(Y_{ih}^0) + \varepsilon_{ih}^0 = \alpha_h^0 + X_i\beta_h^0 \quad + v_{ih}^0 + \varepsilon_{ih}^0$$

(1b) $$Y_{ih}^1 = E(Y_{ih}^1) + \varepsilon_{ih}^1 = \alpha_h^1 + X_i\beta_h^1 \quad + v_{ih}^1 + \varepsilon_{ih}^1$$

The benefit, or gain, or return, from reperfusion treatment for patient *i* in hospital *h* is Y_{ih}^{Δ} given by:

(1c) $$Y_{ih}^{\Delta} = \alpha_h^{\Delta} + X_i\beta^{\Delta} + v_{ih}^{\Delta} + \varepsilon_{ih}^{\Delta},$$

where $\alpha_h^{\Delta} = \alpha_h^1 - \alpha_h^0$, $\beta_h^{\Delta} = \beta_h^1 - \beta_h^0$, $v_{ih}^{\Delta} = v_{ih}^1 - v_{ih}^0$ and $\varepsilon_{ih}^{\Delta} = \varepsilon_{ih}^1 - \varepsilon_{ih}^0$

And similarly the expected benefit from reperfusion at the time of choosing treatment is given by:

(1d) $$E(Y_{ih}^{\Delta}) = \alpha_h^{\Delta} + X_i\beta_h^{\Delta} + v_{ih}^{\Delta}$$

In Equation (1d), α_h^{Δ} represents the hospital-specific benefit in providing reperfusion. One could think of α_h^0 as representing a hospital's Total Factor Productivity (TFP)— because increases in it reflect improvements that are unrelated to specific treatments such as reperfusion or surgery [Garber and Skinner (2012) and Syverson (2001)]. Efforts to increase α_h^0 are efforts to increase productive efficiency--increasing the fraction receiving beta-blockers or improving patient safety are examples. The higher the α_h^0, the lower the benefit from reperfusion, for a fixed level of α_h^1. Because α_h^{Δ} represents the difference between the ability to perform reperfusion and usual care, we call it comparative-advantage at reperfusion. Hospitals may have comparative advantage in providing reperfusion because of either being particularly good at reperfusion treatment or being particularly bad at usual care for patients. In the above equations, we have also allowed for hospital-level variation in how patient characteristics affect outcomes through the β_h^{Δ} term, although in the empirical work we found these to be unnecessary and assumed $\beta_h^{\Delta} = \beta^{\Delta}$.

B. Treatment Choice

Each patient receives treatment if the expected benefit from treatment exceeds a minimal threshold τ_h, where the threshold may vary across hospitals due to incentives or information as discussed further below. Since $E(Y_{ih}^{\Delta})$ captures the total expected benefit to the patient of providing treatment, then the optimal decision from the patient's perspective would let τ_h=0 and provide treatment whenever the expected benefits to the patient exceed zero. There is underuse if $\tau_h > 0$, since patients with positive benefits are under the threshold and do not receive treatment. There is overuse if $\tau_h < 0$, since patients with negative benefits (who would do better without treatment) are above the threshold and receive treatment.

Figure 1A illustrates the intuition behind a Roy model of treatment at the hospital level. The two lines denote patient survival if a hospital treats a given patient with usual care (intercept is α_h^0) or using reperfusion (intercept is α_h^1) as a function of patient characteristics (i.e. patient X's) on the x-axis. To simplify exposition, we have suppressed the distribution of unobservables (v_{ih}^{Δ}). In reality, as well as in our model and empirical work, providers observe these unmeasured characteristics and use them to determine treatment. Expertise at usual care and reperfusion is captured by the intercepts α_h^0 and α_h^1 respectively, with comparative advantage being the difference between them. Allocative efficiency means that reperfusion should be performed to the point that the marginal patient receives zero benefit, (τ_h = 0), so that everyone to the right of the point of intersection should be treated and to the left should receive usual care.

First, consider the role of comparative-advantage in explaining treatment rates: *ceteris paribus*, a hospital that is better at reperfusion would have a higher intercept for reperfusion α_h^1, which would increase the fraction of patients receiving reperfusion at that hospital. A hospital may also have a relative advantage at reperfusion because it is worse at usual care. Either would increase $\alpha_h^{\Delta} = \alpha_h^1 - \alpha_h^0$ and also increase the fraction of all patients being reperfused. Next, consider allocative inefficiency by a hospital that over-treats patients with reperfusion therapy, and treats patients to the left of point of intersection. This harms patients and lowers the average benefit from reperfusion amongst all patients receiving reperfusion. Overuse is equivalent to setting $\tau_h < 0$, where some patients with negative benefits (harm) are treated. Underuse of reperfusion happens when patients who are appropriate for reperfusion (to the right of the intersection of the two lines) don't receive it—a possibility that increases the benefits of reperfusion amongst patients receiving it.

This figure provides four pieces of intuition. First, knowledge of comparative-advantage doesn't tell us where it originates from—it could arise from low α_h^0, a high α_h^1, or both. Second, allocative inefficiency may arise from overuse $\tau_h < 0$ (a willingness to perform reperfusion even if the benefit is

negative) or underuse $\tau_h > 0$ (an unwillingness to perform reperfusion even when the benefit is positive). Third, how a patient is treated depends on patient characteristics, the hospitals comparative advantage at delivering reperfusion (α_h^{Δ}), and the level of allocative efficiency at the hospital (τ_h): all three determine the propensity to be reperfused for a given patient at a given hospital. This brings us to the fourth insight: variation across hospitals in treatment rates does not imply anything about the presence of comparative-advantage versus allocative efficiency. Risk-adjusted hospital treatment rates capture both mechanisms—high risk-adjusted rates may arise because of high levels of hospital-specific benefits at performing the treatment or a very low threshold for performing the treatment-- and do not, by themselves, isolate the source of variation even with perfect risk-adjustment.

Allocative inefficiency ($\tau_h \neq 0$) could come from a variety of sources. Figure 1B illustrates how a hospital that misperceives its comparative-advantage from reperfusion and believes it to be higher than it is, through overconfidence or imperfect knowledge about its comparative advantage, would overuse reperfusion. It could also be that a hospital overuses reperfusion because it is maximizing something other than health. These are alternative mechanisms that that we explore in Section IV (we find evidence for the misperception mechanism). Regardless of the mechanism for allocative inefficiency, they cause a welfare loss whose magnitude is illustrated by the area of the triangle in the figure. The height of the triangle is the threshold, and the base is the threshold multiplied by how much the threshold increases the probability of receiving reperfusion. At the end of the paper, we aggregate the area of these triangles to estimate the welfare loss from allocative inefficiency.

We now specify our model of treatment choice more completely, paying particular attention to how one can identify the different sources of inefficiency. The probability of receiving treatment is the probability that expected benefits exceed the minimum threshold:

$$(2) \qquad \Pr(Treatment_{ih} = 1) = \Pr\left(E\left(Y_{ih}^{\Delta}\right) > \tau_h\right) = \Pr\left(\alpha_h^{\Delta} + X_i\beta^{\Delta} + v_{ih}^{\Delta} > \tau_h\right) = \Pr\left(-v_{ih}^{\Delta} < I_{ih}\right),$$
$$where\ I_{ih} = X_i\beta^{\Delta} + \theta_h \ \ and \ \ \theta_h = \alpha_h^{\Delta} - \tau_h$$

In the terminology of Heckman, Urzua and Vytlacil (2006), our model allows for *essential* heterogeneity where the decision to provide treatment to each patient is made with knowledge of their idiosyncratic response to treatment (v_{ih}^{Δ}). If we make the standard assumption that the distribution of patient-level idiosyncratic gains (v_{ih}^{Δ}) are i.i.d. (an assumption we return to below), then the parameters (β^{Δ}, θ_h) of Equation (2) can be estimated (up to scale) with a single index model such as a logit or OLS regression of treatment on patient characteristics and hospital effects.[2] The hospital-specific intercept (θ_h)

[2] Technically, logit models estimate $I_{ih} = \left(X_i\beta^{\Delta} + \theta_h\right)/\sigma_v$ where σ_v is the standard deviation of the patient-level idiosyncratic gains (v_{ih}^{Δ}). For now, we make the standard assumption that σ_v=1. Most of the

in this equation is commonly referred to as the hospital's risk-adjusted reperfusion rate, with higher values indicating a more aggressive hospital where identical patients are more likely to receive reperfusion. The hospital effect is $\theta_h = \alpha_h^{\Delta} - \tau_h$, which means that a hospital may be more likely to provide treatment because of greater comparative advantage at delivering treatment ($\alpha_h^{\Delta} > 0$), or because of using a lower benefit threshold for providing care $(\tau_h < 0)$ reflecting overuse. Even if treatment rates were the same across hospitals, there could still be overuse or underuse if, say, hospitals with greater comparative advantage set a correspondingly higher threshold for providing care. Thus, because variation in treatment rates across hospitals confounds variation in hospital comparative advantage with hospital treatment thresholds, such variation cannot by itself say anything about overuse or underuse.

C. Identifying Allocative Inefficiency

We now demonstrate that allocative efficiency can be identified separately from comparative-advantage if we can estimate the treatment effect for those patients receiving treatment. The treatment-on-the-treated parameter is the average gain from treatment amongst those who were given treatment, and can be obtained by conditioning the expression for Y_{ih}^{Δ} (equation 1c) on the condition for receiving treatment (equation 2):

(3) $$E\left(Y_{ih}^{\Delta} | Treatment_{ih} = 1\right) = E\left(Y_{ih}^{\Delta} | -v_{ih}^{\Delta} < I_{ih}\right) = X_i\beta^{\Delta} + \alpha_h^{\Delta} + E\left(v_{ih}^{\Delta} | -v_{ih}^{\Delta} < I_{ih}\right)$$

Noting that $X_i\beta^{\Delta} + \alpha_h^{\Delta} = I_{ih} + \tau_h$, we can rewrite Equation (3) as:

(4) $$E\left(Y_{ih}^{\Delta} | Treatment_{ih} = 1\right) = \tau_h + g(I_{ih})$$

where $g(I_{ih}) = I_{ih} + E\left(v_{ih}^{\Delta} | -v_{ih}^{\Delta} < I_{ih}\right)$

Here, *g(I)* is an unknown function of the propensity to receive treatment and the conditional expectation $E\left(v_{ih}^{\Delta} \middle| -v_{ih}^{\Delta} < I_{ih}\right)$. If we assume that the distribution of patient-level idiosyncratic gains (v_{ih}^{Δ}) are the same for all patients and hospitals, then this conditional expectation is only a function of the index, which means that g(I_{ih}) is only a function of I_{ih}. This is an important assumption because our empirical work relies heavily on the single-index property. The assumption would be violated, for example, if the variance or distribution of patient level idiosyncratic gains (v_{ih}^{Δ}) differed across hospitals. It is not possible to test the single-index assumption by looking at differences in treatment on the treated by type of hospital (for example, by hospital volume, size, non-profit status) because in our framework these differences, conditional on propensity, reflect differences in thresholds. Because this is a key assumption in our model, we will discuss supporting evidence for it in the empirical work. In particular, if the

---

results in the paper are invariant to the scale of I_{ih}, but we will return to this point in the results section when we try to recover estimates of α_h^{Δ}.

variance of idiosyncratic gains differed across hospitals then we would also expect differences across hospitals in the coefficients on patient characteristics in Equation 2 (predicting treatment), but we do not find this to be the case.

Equation (4) is the key result of the model that allows us to identify allocative inefficiency. Equation (4) states that after conditioning on patient propensity to receive treatment, differences across hospitals in the treatment effect on the treated are due solely to differences in the hospital's minimum threshold to deliver care (τ_h). Note that the propensity to receive treatment (I_{ih}) depends on the hospital effect (θ_h) and that includes both the presence of comparative advantage and allocative efficiency. By conditioning on this propensity and examining differences in benefit across hospitals, we can isolate differences in hospital thresholds.

Figure 2 illustrates why conditioning on the propensity (rather than, say, patient characteristics) allows us to identify differences across hospitals in τ_h. The first panel illustrates selection into treatment allowing for selection on gains that are observable to the physician but not to us. For patients in this hospital, Hospital A, the mean of the benefit distribution for a patient with characteristics *X* is $X\beta^{\Delta} + \alpha_h^{\Delta}$, with the distribution around this reflecting v^{Δ}, which is observed by the physician and used in determining treatment. Patients above τ_h are treated, so that the area under the curve to right of τ_h is the probability of treatment and the difference between $X\beta^{\Delta} + \alpha_h^{\Delta}$ and τ_h is our index ($I = X\beta^{\Delta} + \alpha_h^{\Delta} - \tau_h \equiv X\beta^{\Delta} + \theta_h$). The conditional mean to the right of τ_h represents the treatment effect on the treated (mean of shaded area in light blue). Holding patient characteristics fixed at $X\beta^{\Delta}$, the treatment on treated could be higher at one hospital versus another because that hospital has either a higher threshold (τ_h) or higher comparative advantage (α_h^{Δ})—which would increase the mean of the distribution upwards-- or both. Thus, differences across hospitals in the treatment on treated for identical patients do not help to identify allocative inefficiency (τ_h).

The second panel in the figure illustrates how conditioning on the propensity identifies variation in τ_h. In panel (b) which illustrates treatment decisions in Hospital B, we consider a patient who is treated at a hospital with the same comparative advantage (α_h^{Δ}) as hospital A but a higher threshold ($\tau_h + d$). Compared to hospital A, patients with the same *X*'s will have a lower propensity to be treated at hospital B because of the higher threshold, i.e. $\theta_h = \alpha_h^{\Delta} - \tau_h$ in hospital A and $\theta_h = \alpha_h^{\Delta} - \tau_h - d$ in hospital B. Thus, when we match on propensity, patients in hospital A will be matched to patients with higher X's in order to offset the lower θ_h in hospital A. More precisely, in order to hold the index (I) constant, a patient with characteristics *X* in hospital A will be matched to a patient with characteristics $\tilde{X}$ in hospital B such that $\tilde{X}\beta^{\Delta} = X\beta^{\Delta} + d$. Therefore, after matching on propensity, both the threshold ($\tau_h + d$) and the mean of the benefit distribution $\left(\tilde{X}\beta^{\Delta} + \alpha_h^{\Delta} = X\beta^{\Delta} + d + \alpha_h^{\Delta}\right)$ are shifted up by *d* in hospital B relative to

hospital A, and as a result the treatment on the treated is also shifted up by *d* in hospital B relative to hospital B. In other words, when we hold the propensity index constant, differences across hospitals in treatment on treated are *exactly* equal to the difference across the hospitals in τ_h (which in this case is *d*). Note that the same figure applies to a setting where both the threshold and comparative advantage increase by *d* in the second hospital. Here, θ_h would be the same in the two hospitals (since $\theta_h = \alpha_h^{\Delta} - \tau_h = \left(\alpha_h^{\Delta} + d\right) - (\tau_h + d)$), so that matching on the index would match patients with the same $X\beta^{\Delta}$. However, this would still imply a shift of *d* in both the threshold and the benefit distribution, resulting in a corresponding shift in treatment on treated. None of this would be possible if we matched patients on X's instead of the propensity to be treated, which includes X's, τ_h, and α_h^{Δ}. Finally, recall that an important assumption is that the distribution of v^{Δ} is the same across all patients with the same propensity (so that the distribution only depends on a single index, *I*). If this were not true, then the distribution in hospital B could differ from the distribution in hospital A, and would not be a simple shift to the right by *d*.

The third panel in Figure 2 illustrates how conditioning on the propensity controls for variation in comparative advantage across hospitals. In panel (c), we consider a patient who is treated at a hospital with the same threshold (τ_h) as hospital (a) but a lower comparative advantage ($\alpha_h^{\Delta} - d$). Compared to hospital A, patients with the same X's will have a lower propensity to be treated at hospital C because of the lower comparative advantage, i.e. $\theta_h = \alpha_h^{\Delta} - \tau_h$ in hospital A and $\theta_h = \alpha_h^{\Delta} - d - \tau_h$ in hospital C. The propensity to treat a patient with characteristics X is the same in hospital C as it was in hospital B, but in hospital C this is because of lower comparative advantage rather than a higher treatment threshold. Thus, when we match on propensity, patients in hospital A will be matched to patients with higher X's in hospital C just as they were in hospital B. Therefore, after matching on propensity, both the threshold (τ_h) and the mean of the benefit distribution $\left(\tilde{X}\beta^{\Delta} + \alpha_h^{\Delta} - d = X\beta^{\Delta} + d + \alpha_h^{\Delta} - d = X\beta^{\Delta} + \alpha_h^{\Delta}\right)$ are the same in hospital C as in hospital A, implying that treatment on the treated is also the same. In other words, when we hold the propensity index constant, differences across hospitals in treatment on the treated are *not affected* by differences across hospitals in comparative advantage.

Note that our model does not, by itself, uncover mechanisms for overuse or underuse—we will investigate these later. It is possible that overuse occurs because providers are worried about malpractice, because they're maximizing something other than health, because they incorrectly believe that they're better at offering the treatment, or because they inaccurately assess patients as more appropriate for treatment than they actually are (perceiving a rightward shift in the distribution of patient's Xs).

More insights from our model are illustrated in Figure 3. In this figure, we plot the treatment effect on the treated, $E\left(Y_{ih}^{\Delta} | Treatment_{ih} = 1\right) = E\left(Y_{ih}^{\Delta} | E\left(Y_{ih}^{\Delta}\right) > \tau_h\right)$, on the vertical axis, while the propensity of being treated (which is a function of I_{ih}) is given on the horizontal axis. The horizontal line

at zero indicates the efficient threshold, below which the expected benefit of treatment is negative (harm). The top curve in Figure 3 represents the treatment-on-the-treated effect for a patient with a given propensity that is treated in a hospital with a high minimum threshold for treatment, *i.e.* it represents $E\left(Y_{ih}^{\Delta}|E\left(Y_{ih}^{\Delta}\right) > \tau_h\right) = \tau_h + g(I_{ih})$. The lower curve represents the same thing for a hospital with a low minimum threshold. Treatment-on-the-treated approaches the minimum threshold (τ_{high} or τ_{low}) for a patient with a low propensity of being treated (small value of I_{ih}), since no patient is ever treated with a benefit below this threshold. For a patient with a high propensity of being treated (large value of I_{ih}), truncation becomes irrelevant and the treatment-on-the-treated effect asymptotes to the unconditional benefit of treatment. However, conditional on a patient's propensity, the treatment effect is always higher by exactly $\tau_{high} - \tau_{low}$ in the hospital with the higher threshold.

The graph illustrates two implications of the theoretical model. First, we can identify overuse and underuse by focusing on patients with the lowest probability of receiving treatment. In these patients, there is overuse when the treatment effect for the lowest propensity patients is negative, and underuse when the treatment effect for the lowest propensity patients remains positive. In particular, a hospital is over treating its patients ($\tau_h < 0$) whenever the treatment effect on the treated is negative (indicating harm) among low propensity patients.

Second, differences in comparative-advantage at performing reperfusion show up as a movement along the curves – higher comparative advantage at reperfusion (α_h^{Δ}) increases the propensity of patients to be treated, and therefore the treatment effect, but does not affect treatment effects conditional on propensity. Being treated at a hospital with a higher comparative advantage (α_h^{Δ}) is equivalent to having patient characteristics ($X_i\beta^{\Delta}$) that increase your benefits from treatment – both raise your expected benefit from treatment and therefore raise your propensity to be treated.

In summary, the key difference between identifying comparative advantage from allocative inefficiency is that differences in hospital comparative advantage have an impact on treatment effects by shifting the propensity to be treated, while differences in the minimum threshold have an impact on treatment effects conditional on the propensity to receive reperfusion.

D. Estimation

In a potential outcomes framework, the equation relating the level of survival to treatment is:

$$(5) \quad Y_{ih} = Y_{ih}^0 + Y_{ih}^{\Delta} Treatment_{ih}$$
$$= \alpha_h^0 + X_{ih}\beta^0 + Y_{ih}^{\Delta} Treatment_{ih} + \left(v_{ih}^0 + \varepsilon_{ih}^0\right)$$

Here, survival for patient *i* at hospital *h* depends on a hospital effect that captures the hospital's general level of expertise (or TFP) providing usual care α_h^0, patient risk adjusters X_{ih} and a patient-specific

treatment effect Y_{ih}^{Δ}. Regression estimates of this equation identify the treatment-on-the-treated effect, $E\left(Y_{ih}^{\Delta}\middle|Treatment = 1\right)$, if the receipt of treatment is uncorrelated with the unobservable characteristics of patients who were not reperfused ($v_{ih}^{0}\ and\ \varepsilon_{ih}^{0}$). This treatment-on-the-treated effect is the same as Equation 4—and will be used to identify overuse (if negative) and underuse (if positive).

It is important to see that we are not assuming that receipt of treatment is uncorrelated with the gain from treatment, which is the conventional assumption required to estimate average treatment effects. Indeed, we explicitly allow for 'selection on gains' where providers use information on the patient's idiosyncratic gain, $v_{ih}^{\Delta} = v_{ih}^{1} - v_{ih}^{0}$, to determine treatment [Wooldridge (2002, p.606)]. This is particularly likely in our setting because of idiosyncratic factors occurring in the hospital that will affect the benefits of reperfusion. For example, whether a patient gets treated with reperfusion or standard care depends on factors such as the experience of the particular doctor and team as well as the capacity of the hospital at the moment of the patient's arrival. Providers observe these factors, which are idiosyncratic to every patient situation, and act on them. This is a weaker set of identifying assumptions than the conventional 'selection on observables' model where one assumes that conditional on rich observable patient characteristics, patients receive treatments randomly. In terms of the analogy to random assignment, we are assuming that conditional on X's, patients are randomly assigned to hospitals, but within hospitals doctors triage them according to the benefit that they would receive from treatment.[3] In other words, the random assignment that we require is that treatment is random with respect to v_{ih}^{0} (baseline unobservables), which is plausible given the rich covariates that we have. We are *not* assuming that patients are randomly allocated to hospitals, and within hospitals that they are randomly assigned to treatment (conditional on X's). In the results section, we provide evidence supporting the case that we can estimate unbiased estimates of the treatment on treated effect, based on comparing our estimates to evidence from randomized trials.

Our test for allocative efficiency requires comparing the treatment on the treated parameter across hospitals, while holding the propensity to receive treatment constant. The index for the propensity to receive treatment, I_{ih}, was obtained from a random-effects logit model of treatment receipt on the patient risk adjusters (X_{ih}) and hospital-level random intercepts (θ_h) estimated using xtmelogit in Stata. Bayesian posterior estimates of the hospital random effects ($\hat{\theta}_h$), commonly referred to in the literature as shrinkage estimates, were used as estimates of θ_h. For more details see Appendix B.

[3] Hull (2017) clearly shows that patients are not randomly choosing hospitals, and that there is a 'selection on gains' in the hospital that they, or their ambulance driver, chose. This is consistent with our earlier work in Chandra et al. (2016b) which also showed that hospital choice is far from random, and aligned with some notion of 'market learning.' In this paper, we're relying on CCP data to assume away this selection and modeling what happens inside the hospital.

Using the fact that OLS estimates of Equation (5) estimate treatment on the treated, we can plug in our model's implication for treatment on treated from Equation (4) into Equation (5) to yield:

(6) $$Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \big(\tau_h + g(I_{ih})\big)Treatment_{ih} + v_{ih}^0 + \varepsilon_{ih}^0$$

To estimate Equation (6), we consider two approximations to the function $g(I)$: a linear approximation ($g(I) = \lambda_o + \lambda_1 I$), and a more flexible approximation using indicator variables for the 100 percentiles of that allows $g(I)$ to have any shape ($g(I) = \sum_{p=1}^{100} \delta_p 1\big(g_{p-1} < I < g_p\big)$). While the theory only predicts a monotonic relationship, we find that estimates from a simple linear specification are very similar to those that allow *g(I)* to have a completely flexible form. We will exploit the linear specification later in the paper, where we impose additional parametric structure to recover hospital measures of α_h^{Δ}.

Adding these approximations for *g(I)* into Equation (6) yields estimating equations:

(6a) $$Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \tau_h Treatment_{ih} + \lambda_0 Treatment_{ih} + \lambda_1 I_{ih} Treatment_{ih} + v_{ih}^0 + \varepsilon_{ih}^0$$

(6b) $$Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \tau_h Treatment_{ih} + \sum_{p=1}^{100} \delta_p 1\big(g_{p-1} < I_{ih} < g_p\big) * Treatment_{ih} + v_{ih}^0 + \varepsilon_{ih}^0$$

Here, the hospital-specific coefficient on treatment identifies differences across hospitals in τ_h. Note that separating the average level (as opposed to differences across hospitals) of τ_h from the intercept of $g(I)$ would require stronger parametric assumptions, so we focus on identifying differences between hospitals (which are indications of allocative inefficiency). In the linear specification, we demean I_{ih} to have a value of 0 for the average treated patient so that the coefficient λ_0 captures the average effect of reperfusion among the treated. The hospital-specific intercept in this regression identifies hospital TFP (α_h^0). The coefficient (λ_1) on the interaction $I_{ih}Treatment_{ih}$ or on the indicator variables for the percentiles of g(I) provide a test for whether the benefit of reperfusion therapy is increasing with the propensity to receive such treatment—as would be the case if a Roy Model of treatment allocation was at work, as opposed to model where providers select patients randomly or without regard to benefits.

Estimating Equations (6a) and (6b) involves estimating hospital-specific coefficients on treatment rates for thousands of hospitals. Rather than including hospital dummies interacted with treatment, which would yield imprecise estimates and suffer from small sample problems, we estimate hierarchical logit models for survival with hospital-level correlated random coefficients for the hospital-specific intercept and slope (α_h^0 and τ_h). We document that key results are similar using fixed effect models.

We also consider an alternative specification that allows for a relatively straightforward test for allocative inefficiency while avoiding the necessity of estimating hospital-specific coefficients. Recall that the risk-adjusted hospital reperfusion rate (the hospital intercept from Equation 2) is $\theta_h = \alpha_h^{\Delta} - \tau_h$, , and we obtain estimates $\hat{\theta}_h$ of this intercept (up to scale) from estimating the propensity equation. Under two extreme cases, we can say how treatment on the treated ($\tau_h + g(I_{ih})$) is related to θ_h. In the first case, if there is no allocative inefficiency ($\tau_h = 0$) then $\theta_h = \alpha_h^{\Delta}$ and variation in reperfusion across

hospitals is driven purely by comparative advantage. In this case, θ_h is unrelated to the treatment effect on the treated after conditioning on the propensity (since in this case $\tau_h + g(I_{ih}) = g(I_{ih})$). At the other extreme, if there is no variation in comparative advantage ($\alpha_h^{\Delta} = 0$) then $\theta_h = -\tau_h$ and variation in reperfusion across hospitals is driven purely by treatment thresholds. In this case, θ_h will be negatively related to the treatment effect on the treated after conditioning on the propensity (since in this case $\tau_h + g(I_{ih}) = -\theta_h + g(I_{ih})$).

These two extreme cases suggest including an interaction between $\hat{\theta}_h$and treatment in Equations (6), rather than estimating hospital-specific coefficients on treatment, as a simple test for allocative inefficiency. Therefore, we estimate specifications of the form:

(7a) $Y_{ih} = \alpha_h^0 + X_i\beta^0 + \lambda_2\hat{\theta}_h Treatment_{ih} + \lambda_0 Treatment_{ih} + \lambda_1 Treatment_{ih} * I_{ih} + v_{ih}^0 + \varepsilon_{ih}^0$

(7b) $Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \lambda_2\hat{\theta}_h Treatment_{ih} + \sum_{p=1}^{100}\delta_p 1\left(g_{p-1} < I_{ih} < g_p\right) * Treatment_{ih} + v_{ih}^0 + \varepsilon_{ih}^0$

If the coefficient on the interaction (λ_2) is zero, this suggests that variation in hospital-level reperfusion rates was entirely driven by comparative advantage in treatment (case 1 above). Alternatively, if the coefficient on the interaction is negative, this suggests that variation in hospital-level reperfusion is associated with allocative inefficiency. This simple specification provides an intuitive test of the key insights from our model: conditional on propensity, higher treatment rates due to comparative advantage will be unrelated to treatment effects, while higher treatment rates due to lower treatment thresholds will be negatively related to treatment effects.

III. Results

A. Identifying Allocative Inefficiency

Our model implies that if treatment is being allocated efficiently, then patients with a higher propensity to be treated for any reason should have higher expected benefit from the treatment, and two patients with the same propensity should have the same expected benefit from treatment. Allocative inefficiency can be identified when the benefit of treatment differs across hospitals for patients with the same propensity to be treated. In Figure 4, we evaluate these implications graphically by plotting the estimated survival benefit from reperfusion and 95% CI against a patient's treatment propensity index (I_{ih}) for patients treated in different hospitals. The graphs do not impose structure on the data and are designed to graphically illustrate the main findings of our paper using simple and transparent plots, that can be easily replicated by others.

The treatment benefit at each point in the distribution of the propensity to receive care was estimated flexibly using a local-linear version of equation 6a with a triangular kernel that included 30% of the sample on either side. The patient's treatment propensity was obtained from hierarchical logit estimation of equation (2), and is demeaned so that 0 is the propensity for an average patient receiving

reperfusion. We estimate separate panels for hospital's in the lowest tercile and highest tercile of the estimated hospital effect ($\hat{\theta}_h$), also estimated from the propensity equation (2), as described in Appendix B. These hospital effects are estimates of the risk-adjusted reperfusion rate at each hospital, so hospitals in the top tercile are those that treat patients more aggressively; the distribution of hospital-effects is also graphed in the appendix. As noted in the Appendix, we estimate hospital random-effects which allow for empirical Bayes shrinkage, but their correlation with hospital fixed-effects is over 0.999.

Both plots show a strong upward slope, with higher benefit from treatment for patients with a higher propensity to receive reperfusion—and exactly mirror the theoretical illustration in Figure 3. But at every propensity, the benefits of reperfusion are lower in the top-tercile hospitals, as would be expected if higher treatment rates were due to lower treatment thresholds. At the lowest propensity levels, the survival benefits from reperfusion are significantly negative for the top-tercile hospitals, suggesting that there is overuse among these hospitals. In the bottom-tercile hospitals, the estimated survival benefits from reperfusion for the lowest propensity patients are less negative and not significantly different from zero, which is consistent with appropriate use of reperfusion in these hospitals. Finally, we note that plots are also linear in log-odds despite the non-parametric nature of the estimation— this will allow us to use logit models that control for the propensity linearly as in Equations 6a and 7a.[4] The linearity in log-odds result was not implied by our model, but it will greatly simplify our empirical work.

Figure 5 is similar to Figure 4, but plots the estimated survival benefit from reperfusion and 95% CI against the hospital effect from the propensity equation ($\hat{\theta}_h$), controlling non-parametrically for the propensity index (a local linear estimate of Equation 7b). The hospital-effects are mean zero. The left-hand panel included all patients, while the right-hand panel was estimated only for low-propensity patients whose propensity index implied that they had below a 20% probability of receiving reperfusion. Both plots show a clear downward slope, with lower benefit from treatment for patients treated by hospitals with higher risk-adjusted reperfusion rates ($\hat{\theta}_h$). Again, this would be expected if higher treatment rates were due to lower treatment thresholds, and is evidence of allocative inefficiency. Among all patients (the left-hand plot), the estimated survival benefit from reperfusion is positive for all hospitals, although it is small and not significant in hospitals with the highest treatment rates (those 2 standard deviations above average, with $\hat{\theta}_h$=0.6). In contrast, among the lowest propensity patients (the right-hand plot), only hospital's with the lowest treatment rates are estimated to have survival benefits from

[4] Changes in log-odds (i.e. logit coefficients) can be approximately converted into absolute changes in probability by multiplying them by p x (1-p), where p is the probability of success. Table 1 reports 30-day survival as 80%, so a change of .3 in log-odds means a .3x.8.x2=4.8 percentage points increase in survival.

reperfusion that are near to zero. The estimated survival benefit from reperfusion is negative and significant in hospitals with average or higher treatment rates, suggesting that there is overuse in most hospitals, i.e., we were able to identify substantial subsets of low-propensity patients who were harmed by reperfusion treatment in most hospitals.

Table 2 reports regression estimates of equation 7a and 7b that are analogous to the results reported in the figures. The table reports estimates of the effect of reperfusion on 30-day survival allowing for interactions of reperfusion with the propensity index (I_{ih}) and the hospital effect from the propensity equation ($\hat{\theta}_h$). The regressions control for patient characteristics and include hospital fixed-effects, as the theory tells us to condition on them to control for hospital TFP (α_h^0). To help with interpretation, we have normed the propensity-index so that a value of 0 refers to the average patient receiving reperfusion. Thus, the coefficient on reperfusion is an estimate of the effect of reperfusion on an average patient receiving reperfusion. The first three columns report OLS estimates and the last two logit estimates where the coefficients are odds ratios. The second and third columns of Table 2 report estimation of Equations 7 using two different approaches to control for the propensity index. In the first (which corresponds to Equation 7a), we restrict the index to be a simple linear function, and interact this with reperfusion—an approach that was justified in the graphical analysis in Figure 4. In the second we control for the propensity index with 100 indicator variables also interacted with reperfusion (this corresponds to Equation 7b). The final two columns repeat these specifications using a logit model.

Column (1) does not include the interactions of reperfusion with $\hat{\theta}_h$, but it is included to demonstrate that the benefit of reperfusion is clearly increasing in the propensity to receive reperfusion, and consequently, that a Roy-model of triage describes provider decision making. The coefficient on the interaction of reperfusion with the propensity index is positive and highly significant, implying that the treatment effect of reperfusion on survival is increasing in the patient's propensity index ($\lambda_1 > 0$) as predicted by our model. The coefficient on this interaction implies that an increase in the propensity index of one (about one standard deviation of the propensity index in the treated population) is associated with roughly a doubling of the treatment effect. Thus, it appears that hospitals are choosing patients for treatment based on the benefit of the treatment, and the heterogeneity in the treatment effect is large relative to the average treatment effect.

In column 2, we add an interaction of reperfusion with the hospital effect from the propensity equation ($\hat{\theta}_h$). The coefficient on this interaction is negative and significant, meaning that conditional on a patient's propensity, more aggressive hospitals (those with a higher propensity to treat patients, $\hat{\theta}_h$) have lower returns to reperfusion. As noted earlier, this is evidence of allocative inefficiency – if the hospital's

high use of reperfusion was entirely due to comparative advantage in treatment, we would get a coefficient of 0 on this variable. The negative coefficient is consistent with what would be expected if the variation was due to differences in thresholds (τ), where more aggressive hospitals have lower minimum thresholds for treatment, treat more patients, and have lower benefits to treatment. The coefficient is similar in column 3, where we non-parametrically control for the interaction of reperfusion with a set of 100 dummies for each propensity percentile, suggesting that controlling for the linear interaction of propensity with reperfusion is a sufficient approximation to g(I). The last two columns of Table 2 are logit analogs to the earlier OLS regressions, and yield similar results in logodds terms. The estimated coefficients suggest that a one standard deviation increase in the hospital effect from the propensity equation (about 0.3) lowers the survival benefit of reperfusion by about 1 percentage point or lowers the odds of survival by about 7%.

In Table 3 we estimate the logit models for 7-day and 365-day survival to investigate the sensitivity of our results to alternative survival windows. The purpose of using 7-day survival was to examine whether the patterns noted above are evident soon after admission, and reflect decisions about how the heart-attack was initially treated. If they do not appear at 7-day survival, the concern would be that we are picking up the effect of later treatments—for example, the quality of post-discharge care. At 7 days relative to 30 days, we expect the effect of the treatment to be even more tightly linked to a patient's propensity to receive it and that is exactly what we find in Panel A. This relationship is half as strong for 1-year survival (Panel B) relative to 7 day survival, and represents the importance of post-discharge factors in affecting 1-year survival. In both panels, the benefits of reperfusion fall in hospitals that do more of it which is consistent with allocative inefficiency, as more aggressive hospitals work into less appropriate patients.

The regressions in Tables 2 and 3 identify allocative inefficiency indirectly, by estimating whether a particular hospital-level characteristic ($\hat{\theta}_h$) is associated with the survival benefit of reperfusion, after controlling for patient propensity. A more direct approach is to estimate how much the survival benefit of reperfusion varies across hospitals, after conditioning on patient propensity. In Table 4, we do this by estimating Equation 6a using hierarchical (mixed effects) logit models that treat the hospital-specific intercept and reperfusion coefficient as correlated random effects (these models side-step the challenges of fixed-effects estimation for small hospitals). The first column of Table 4 reports results from estimating a logit model that only includes a hospital-level random intercept to account for the hospital's general level of expertise (or TFP) at providing usual care (α_h^0). The random effects assumption may appear to be restrictive relative to the fixed effect models in Table 2, but the restrictions do not meaningfully change the estimated coefficients on reperfusion or reperfusion*propensity.

The second column of Table 4 estimates the logit with both a hospital-level random intercept and a hospital-level random coefficient on reperfusion, allowing us to estimate the variance and correlation in both the hospital-level TFP (α_h^0) and the hospital-level thresholds (τ_h). We find that the standard deviation of hospital thresholds is large (.313 in logodds) —and of the same magnitude as the effect of reperfusion for the average treated patient (.314). There is also considerable variation across hospitals in TFP, as seen by the standard deviation of α_h^0 estimated to be .198. This estimate implies that the standard deviation across hospitals in the risk-adjusted odds of survival for patients not receiving reperfusion is nearly 20 percent (or nearly 4 percentage points off a base survival rate of 81 percent). Finally, we estimate that the hospital-level coefficient on reperfusion is negatively correlated with the hospital-level intercept, meaning that hospitals with higher thresholds (conservative hospitals that do less) have worse outcomes for patients not receiving reperfusion. Later in the paper we find evidence that this stems from hospitals that are worse at caring for patients without reperfusion being unaware, especially if they're small, that the benefits from doing more reperfusion are actually high for them.

B. Evidence Supporting Key Assumptions

Our analysis relies on three key assumptions: (1) that hospitals triage patients according to a Roy-model; (2) the 'single-index' assumption, that the distribution of unobservables does not have a hospital specific component; and (3) that we are able to estimate a 'treatment on the treated' parameter. We have already presented evidence supporting the first assumption in Figure 4 and Table 2, which document that the benefit of reperfusion is clearly increasing in the propensity to receive reperfusion. Thus, it appears that hospitals are choosing patients for treatment based on the benefit of the treatment.

Evidence on the single-index assumption comes from noting that logit models implicitly estimate $\Pr(Treatment) = (X\beta + \theta)/\sigma_v$, where σ_v is the standard deviation of the unobservables. If hospitals vary in σ_v, then estimates of β from more aggressive or less aggressive hospitals will be different. But this is testable—we estimated separate propensity equations (as in equation 2) by aggressiveness of the hospital (above and below median on $\hat{\theta}_h$) and found that their predictions are nearly identical, with a correlation of 0.9987, suggesting that differences in variances are not a first-order concern. We acknowledge that we are not able to reject a world in which the variances vary across hospitals and where the coefficients scale up to perfectly offset the increased variance, but we're unaware of an economic story for why this should be the case.

To evaluate the plausibility of our third assumption we estimated a simple logit model for the impact of reperfusion on 30-day mortality, controlling for the rich patient risk-adjusters in the CCP data, and compared our estimates to those obtained from clinical trials. A summary of nine trials was published in the journal *Lancet* by the Fibrinolytic Therapy Trialists' Collaborative Group (FTTCG, 1994). This was

the same time-period as the CCP data and each trial evaluated reperfusion therapy in heart-attack patients. Across these nine trials, reperfusion within 12 hours reduced 35-day mortality from 11.5% to 9.6%, which implies that the treatment on the treated effect of reperfusion on the log-odds of mortality is 0.20. In our CCP data, a logit model controlling for the CCP risk-adjusters estimates an identical effect, with reperfusion reducing the logodds of mortality by 0.206 (S.E. = 0.023). We take this evidence as supporting the case that we can estimate unbiased estimates of the treatment on treated effect.[5]

C. Identifying Hospital Comparative Advantage

To summarize the evidence so far, we have shown that (i) patients with higher appropriateness receive higher benefits from treatment, (ii) this relationship is approximately linear, which is why simpler linear-controls for the propensity to receive care do as well as non-parametric controls for the propensity to receive care, (iii) less appropriate patients are harmed in high-reperfusion hospitals, which is consistent with overuse, (iv) more aggressive hospitals have lower average treatment benefits for patients at every propensity, which is consistent with these hospitals having a lower treatment threshold, and (v) after conditioning on patient propensity there remains substantial variation in the survival benefit of reperfusion across hospitals, which is consistent with allocative inefficiency. While these results identify the presence of allocative inefficiency, they say nothing directly about the presence or absence of comparative advantage. In order to simultaneously estimate variation in hospital thresholds and comparative advantage, we now turn to a more parametric framework to estimate both quantities.

In this section, we jointly estimate the treatment propensity and survival equations, yielding estimates of the joint distribution of the hospital-level parameters ($\theta_h, \alpha_h^0, \tau_h$). These estimates are then used to identify variation across hospitals in comparative advantage. This approach requires us to use the linear-approximation for $g(I) = \lambda_0 + \lambda_1 I$ instead of the non-parametric control. The linear assumption was justified by Figure 4 and in Table 2 where we showed very similar results with this restriction compared to the fully non-parametric approach.

To pursue this approach, recall that the treatment propensity was estimated using a random effect logit to estimate Equation 2 (see appendix B):

(2) $\Pr(Treatment_{ih} = 1) = \mathrm{F}(X_{ih}\beta + \theta_h)$

Note that since $I_{ih} = X_{ih}\beta + \theta_h$, equation 6a can we rewritten as:

(7c) $Y_{ih} = \alpha_h^0 + X_{ih}\beta^0 + \lambda_0 Treat_{ih} + (\lambda_1\,\theta_h + \tau_h)Treat_{ih} + \lambda_1\, Treat_{ih} * X_i\beta + v_{ih}^0 + \varepsilon_{ih}^0$

[5] A completely different approach would instrument for reperfusion. Key for our empirical approach is that the instrument should recover a treatment-on-the treated estimate that may or may not be the same as the Local Average Treatment Effect (LATE).

Equation 7c is a logit model with a hospital-level random intercept (α_h^0) and a hospital-level random coefficient on reperfusion ($\lambda_1\,\theta_h + \tau_h$). In estimating Tables 2-4, we used a 2-step approach that first estimated θ_h from Equation 2, and then plugged this estimate into the survival equation. We now estimate the treatment propensity equation (2) and the survival equation (7c) jointly, treating the hospital-effect in the propensity equation (θ_h) and the hospital-level intercept (α_h^0) and coefficient on reperfusion ($\lambda_1\,\theta_h + \tau_h$) in the survival equation as jointly normal correlated random coefficients. The remaining parameters determining the effect of reperfusion (λ_0, λ_1) were estimated along with the variance and covariance of the hospital-level random coefficients by maximum likelihood using xtmelogit in Stata.[6] Finally, knowledge of $(\lambda_1\,\theta_h + \tau_h)$, λ_1 and θ_h allows one to restate all of the estimates in terms of τ_h rather than $(\lambda_1\,\theta_h + \tau_h)$.

We report the results from joint estimation of equation 2 and 7c in Table 5. Reassuringly, the more complicated joint estimation procedure replicates the results and magnitudes from simpler models. For example the coefficient on reperfusion from the table 5 is 0.27 compared to 0.31 in the simpler logit model in Table 4 (both coefficients are in log-odds). The benefit of reperfusion increases with the index with similar magnitudes in both models—0.29 in Table 4 vs. 0.28 from Table 5. The threshold and general productivity are correlated -0.331 in the simpler model and are correlated -0.321 in the joint model, and both are estimated to have similar standard deviations using the simple and the joint model. In particular, the joint model continues to estimate considerable variation across hospitals in τ_h, the minimum threshold for treatment (Std. Dev. = 0.327). The consistency between the joint estimates and simpler approaches reassures us that the estimates are not a consequence of the structure that we have imposed. Consistent with the evidence presented in Table 2, there is a negative correlation (-0.341) between τ_h and the reperfusion intercept θ_h, suggesting that some of the variation in treatment rates across hospitals is associated with variation in the treatment threshold (mostly overuse, as suggested by figures 4-5), but that this correlation is far from -1 (as would be the case if there was no variation in comparative advantage) suggests that comparative advantage is also present.

The joint estimation in Table 5 yields estimates of the joint distribution of the hospital-level parameters ($\theta_h, \alpha_h^0, \tau_h$). However, our goal is to estimate the joint distribution of comparative advantage (α_h^Δ) and the treatment threshold (τ_h). Recall that $\theta_h = \left(\alpha_h^\Delta - \tau_h\right)/\sigma_v$, where σ_v is the standard deviation

[6] To simplify estimation, we first estimated the coefficients on all of the covariates (X) in equations 2 and 7c using simple logit models, and then used the estimated indices ($X_{ih}\hat{\beta}, X_{ih}\hat{\beta}^0$) rather than the individual covariates when estimating the random coefficient models. All the reported standard errors are conditional on the first-stage estimates $X_{ih}\hat{\beta}, X_{ih}\hat{\beta}^0$, but any adjustment for using these generated regressors is likely to be second-order because of the large samples used to estimate the patient-level coefficients.

of the patient-level idiosyncratic gains (v_{ih}^{Δ}), which we have so far ignored by assuming that it is one (all our earlier results were invariant to this scaling). This implies that $\alpha_h^{\Delta} = \tau_h + \sigma_v\theta_h$, so that the distribution of comparative advantage (and its correlation with the treatment threshold) depends on both σ_v and the joint distribution of (θ_h, τ_h). Therefore, it is important to know the scale factor in order to make statements about comparative advantage. The scale parameter represents the standard deviation of the unobservable factors determining expected benefit from treatment. While we cannot estimate it directly, we used a range of values for σ_v to calculate the standard deviation of α_h^{Δ} and its correlation with τ_h. These are presented in Figure 6. The left hand panel plots estimates of the standard deviation of comparative advantage (α_h^{Δ}) for values of σ_v from 0.01 to 3, while the right hand panel plots estimates of the correlation of α_h^{Δ} with τ_h for the same range of σ_v. Interestingly, the estimates from Figure 6 bound the standard deviation of α_h^{Δ} to be above 0.3. Thus, our estimates imply that the variation across hospitals in comparative advantage is at least as large as the variation across hospitals in the treatment threshold (SD=.327) and possibly much larger. For σ_vr < 1, corresponding to relatively less variation in idiosyncratic differences across patients in the expected benefits from treatment, our estimates imply similar amounts of variation in τ_h and α_h^{Δ}, and that the two are strongly positively correlated (between 0.4 and 1). In this case, a hospital's minimum treatment threshold τ_h is positively correlated with a hospital's comparative advantage α_h^{Δ}, meaning that hospitals with low comparative advantage tend to have low thresholds and overuse reperfusion.

To explore the relationships between τ_h and α_h^{Δ} further, we plotted the hospital-level relationship between the risk-adjusted treatment rates and in Figure 7a and between τ_h and α_h^{Δ} in Figure 7b. We did these using $\sigma_v = .25$, which is value of σ_v that minimizes the standard-deviation of α_h^{Δ}. This was chosen because the null-hypothesis in the literature is to assume that there is no comparative-advantage, and we wanted to examine the greatest scope for comparative-advantage under the null. The graphs report shrunken estimates of τ_h and α_h^{Δ} so they understate the variation in both. However, clear patterns are obvious—consistent with Figure 4 and Table 5, higher risk-adjusted treatment rates mean lower thresholds, and hospitals lower thresholds also have lower comparative advantage at delivering reperfusion. Interestingly, such a positive correlation (which was also noted above in Table 5) would arise if all hospital's *incorrectly* believed that they had high comparative advantage in performing the treatment, resulting in overuse among those hospitals that actually did not have a high comparative

advantage in performing the treatment (as we illustrated in Figure 1B). We examine this mechanism in the next section.

IV. Mechanisms

As noted earlier, there are two broad mechanisms that could lead to allocative inefficiency. First, hospitals may be over-treating for financial gain (particularly in for-profit hospitals) or because of benefits to future patients through learning-by-doing (particularly in teaching hospitals). This type of mechanism would suggest that allocative inefficiency (τ) would be related to hospital characteristics such as ownership, teaching status, etc. To investigate this hypothesis, we estimated the model from Table 4 adding interactions of treatment with a number of hospital-characteristics such as ownership, teaching status and size (Table 6). Overuse at for-profit hospitals or teaching hospitals, or at hospitals with other characteristics that are included in the table, would mean that the return to treatment would be lower at such facilities. The results demonstrate that there is no evidence that these characteristics are associated with the return to treatment, conditional on the patient's propensity to receive treatment. A joint-test on all the *Treatment*Hospital Characteristics* interactions can't reject the null-hypothesis that these variables are jointly zero (chi-squared statistic=2.96; p-value=0.96). Yet, since overtreatment is clearly in evidence from the earlier exhibits, we need another mechanism for why it happens. Note that these hospital characteristics do predict variation in the use of reperfusion: for-profit hospitals and high-volume hospitals were more likely to perform reperfusion, while teaching hospitals and hospitals that treated high poverty populations were less likely to perform more reperfusion. The evidence in Table 6 implies that differences across hospitals with these characteristics in reperfusion rates reflect differences in comparative-advantange rather than differences in treatment thresholds.

A second mechanism for allocative inefficiency is that hospitals have imperfect information and misperceive their comparative advantage, as we illustrated earlier in Figure 1B.[7] Given the general lack of systematic performance feedback and small samples of their own treated patients to observe, it is quite plausible that hospitals and physicians will have inaccurate beliefs about their own comparative advantage. Put differently, there is no reason to think that physicians or hospitals know their α_h^{Δ} perfectly—it's the difference of two parameters ($\alpha_h^0 \; and \; \alpha_h^1$) that are both measured with error. In this mechanism, θ represents a hospital's belief about their comparative advantage and τ represents a hospital's misperception (or prediction error) of their own comparative advantage.

More formally, we can reinterpret our empirical model in the following way. Suppose that a hospital does not know its comparative advantage, but instead has a belief about its comparative advantage which

[7] We are grateful to Janet Currie for suggesting this interpretation and alerting us to related work in Currie and MacLeod (2017).

is given by θ. Based on this belief, they treat patients if the expected benefit of treatment is positive. Thus, patients are treated based on beliefs (if $\theta_h + X_i\beta^{\Delta} + v_{ih}^{\Delta} > 0$) rather than based on actual comparative advantage (if $\alpha_h^{\Delta} + X_i\beta^{\Delta} + v_{ih}^{\Delta} > 0$). Let τ_h represent the difference between a hospital's actual comparative advantage and their beliefs, so that $\tau_h = \alpha_h^{\Delta} - \theta_h$ is the hospital's prediction error (and therefore $\theta_h = \alpha_h^{\Delta} - \tau_h$, as in our empirical model). Thus, this framework interprets τ_h as arising from an inaccurate belief about α_h^{Δ}, rather than assuming that hospitals know α_h^{Δ} and consciously set $\tau_h \neq 0$ to achieve other objectives. A negative τ_h implies that the hospital over-estimated their comparative advantage and, as a result, treated some patients who were in fact harmed by the treatment. In this reframing, the key question is how hospitals form their beliefs.

Suppose that each hospital receives a noisy signal of their comparative advantage (S), where $S=\alpha^{\Delta}+\omega$ and the noise (ω) is independent of α with variance σ_ω^2 (we have suppressed the subscripts to simplify notation). Based on this signal, the hospital forms a prediction of its comparative advantage (θ). If the hospital knew the reliability of the signal ($r=\sigma_\alpha^2/(\sigma_\alpha^2+\sigma_\omega^2)$, where σ_α^2 is the variance of α^{Δ} across hospitals), then the optimal prediction of α^{Δ} given S is the posterior mean, given by $E(\alpha^{\Delta}|S)=r*S$. More generally, we assume that hospitals may not know the reliability of the signal, and form their prediction using $\theta=w*S$, where $w\neq r$. Incorrectly weighting the signal generates additional variation in the prediction error (τ) which leads to greater allocative inefficiency. Even if hospital beliefs are optimal given S (i.e., w=r), there will be allocative inefficiency ($\tau\neq 0$) because hospitals have imperfect information, and this information only predicts a fraction (r) of the true variation in comparative advantage.

This simple framework delivers a number of strong empirical implications. First, because the error in the signal is assumed to be independent of the hospital's actual productivity and comparative advantage, this framework constrains the number of parameters in the structural model to six, which allows us to identify the scale parameter and the variance in comparative advantage. More specifically, our empirical model from Table 5 estimated 6 reduced-form moments (the variances and covariances of θ, τ, and α_0) which are a function of the 6 unknown structural parameters in this framework (the variance and covariance of α^{Δ} and α_0, the reliability of the signal r, the weight placed on the signal w, and the scale parameter from the logit σ_v).[8] Therefore, we can derive estimates of the unknown structural parameters for this model using minimum chi-squared estimation (Wooldridge 2010, p.442-446). Minimum chi-squared estimation chooses the structural parameters that provide a best fit of the reduced-form estimates (in a weighted least squares sense, using the standard errors & covariance of the reduced form estimates to form weights). In the just-identified case the resulting structural parameters fit the reduced form

[8] See Appendix-C for derivation of the equations stating the reduced-form moments in terms of the structural parameters.

moments exactly because they are one-to-one function of the reduced form parameters. Restrictions on the structural parameters can be tested based on how they affect the structural models ability to fit the reduced form estimates through a chi-squared goodness of fit statistic.

Just-identified estimates of the structural parameters for this model are provided in the first column of Table 7. There is substantial variation in comparative advantage (standard deviation of $\alpha = 0.317$), with the variation across hospitals being as large as the average treatment effect. The signal that hospital's receive about their comparative advantage is estimated to have very low reliability (r=.065), but hospital's place more weight on the signal than is optimal, with w=0.154.

If w=r in this framework then beliefs are optimal. When we impose w=r (column 2 of Table 7), we are over-identified (estimating 5 parameters from 6 moments) and can use the chi-squared goodness of fit statistic to test the restriction (Wooldridge, 2010, pp. 444-445). This statistic rejects the hypothesis that w=r (chi-squared with 1 df = 10.4, p=.001). In other words, the constrained model with w=r implies reduced-form variances and correlations of θ, τ, and α_0 that are significantly different from those estimated in Table 5. More specifically, if w=r then the hospital's prediction (θ) is optimal and, therefore, should be uncorrelated with the prediction error (τ). The fact that we estimated a significant negative correlation of -0.34 between θ and τ in table 5 implies that hospitals' predictions are not optimal and they are overweighting the noisy signal (w>r), i.e. they over-react to the signal. One might not expect hospitals to have the information necessary to form optimal weights – in particular, they may not know how much true variation in comparative advantage there is across hospitals, and are acting as if they are using an over-diffuse prior (placing too much weight on their own signal, and not shrinking enough to a prior mean).

Finally, if hospitals learn based on their experience with patients, then one would expect that low-volume hospitals would have less reliable signals of their comparative advantage than high-volume hospitals. In the third column of table 7 we fit our model to reduced-form moments estimated separately for low, medium and high-volume hospitals (6 moments for each group, for a total of 18 moments), allowing the reliability parameter to vary across the 3 groups but otherwise constraining the remaining model parameters to be equal across the 3 groups (8 parameters total to fit 18 moments). As expected, the reliability of the signal is estimated to be highest for the high-volume hospitals and lowest for the low-volume hospitals. Moreover, the goodness of fit statistic cannot reject our model (chi-squared with 10 df = 12.7, p=.24) suggesting that this simple model provides an adequate fit of the data. In other words, the model estimated in column 3 of Table 7 implies reduced-form variances and correlations of θ, τ, and α_0 that are not significantly different from the unconstrained reduced-form estimates for low, medium, and high-volume hospitals. Assuming that reliability of the signal is the same for high, medium and low-

volume hospitals (final column of table 7) is strongly rejected (chi-squared with 2 df = 49.7, p<.0001) and such a model is strongly rejected by the goodness-of-fit test (chi-squared with 12 df = 62.4, p<.0001).

V. Conclusion

Using a Roy model of treatment to motivate our empirical framework, we find significant evidence of allocative inefficiency across hospitals. We can use our results to construct a stylized estimate of the welfare loss generated by this allocative inefficiency, along the lines suggested by Phelps (2000). Returning to the intuition from Figure 1b, the effect of a non-zero τ generates a standard welfare loss triangle, where $Welfare\ Loss_h = \frac{1}{2}.\ \overbrace{(\tau_h)}^{height}\ \overbrace{(\tau_h)\ dPr(Treatmeant)/\ \mathrm{d}\tau}^{base}$. The welfare loss due to allocative inefficiency is the average reduction in (logodds) survival per patient.[9] This loss is given by $\frac{1}{2}.dPr(Treatment)/\ \mathrm{d}\tau_h E({\tau_h}^2)$

To get an estimate of $dPr(Treatment)/d\tau$, which is a change in the propensity to receive treatment for a small increase in τ, we took a tiny change of 0.01 in τ_h, divided it by our estimate of the scale factor (σ_v) of 0.44 to turn it into how much change that would create in the hospitals risk-adjusted treatment rate θ_h. This yielded an estimate of $dPr(Treatment)/d\tau$ =-0.26. To estimate $E({\tau_h}^2)$ note that $E({\tau_h}^2) = Var(\tau_h) + [E(\tau_h)]^2$. From the hierarchical-logit model in Table 5, we estimated $SD(\tau_h)$=0.33.[10] If we assume that there is no allocative inefficiency on average ($E(\tau_h) = 0$), then the welfare loss from variation in allocative inefficiency is (1/2)*(0.33^2)*-0.26=-0.014, i.e. the allocative inefficiency across hospitals results in an average reduction in the logodds of survival per patient of .014. The overall benefit from treatment is the benefit among the treated (0.20 in log-odds), who comprise 20 percent of the patient population for a total benefit of 0.04. This means that we could increase the effectiveness of treatment by about a third if we removed the allocative inefficiency across hospitals. There is additional welfare loss if the mean of τ is not equal to zero, e.g. if there is systematic overuse across all hospitals ($E(\tau_h) < 0$). This part of the welfare calculation is far more speculative, but a good guess about systematic overuse across all hospitals comes from the average treatment effect among low propensity patients, which is about -.1 from Figure 5. Thus, the additional welfare loss from systematic overuse would be (1/2)*(-0.1^2)*-0.26=-0.0013. This calculation suggests that systematic overuse is a minor concern relative to the welfare loss from the overall variation.

In addition to the welfare loss from allocative inefficiency, we also found evidence of substantial variation in comparative advantage across hospitals, with the benefits from treatment being much higher

[9] The welfare loss is measured in the same units as τ (logodds of survival in our estimates) and is the welfare loss per patient because we use the probability of treatment rather than total number treated.
[10] Table 4 yielded a very similar estimate of 0.31.

in some hospitals than others. This variation in the benefits from treatment implies that "one size fits all" policies such as strict treatment guidelines are incorrect, since hospitals with greater comparative advantage at a treatment should use it more among their patients. Moreover, our evidence suggests that much of the allocative inefficiency that we observe is due to hospitals having imperfect information and misperceiving their comparative advantage. This is a different mechanism than explaining variations by appealing to medical malpractice or financial entrepreneurship by providers (Gawande, 2009). Thus, rather than reducing treatment variation across hospitals, better information about treatment effect heterogeneity across hospitals is key to improving patient welfare. We don't know if these findings and conclusions generalize to settings beyond the treatment of heart attack patients, but our framework is general and can be applied to a variety of settings.

Finally, our work suggests new directions for research on productivity in healthcare. By uniting the literatures from economics and medicine on variations in medical care with insights from the productivity literature, we found that variation in comparative advantage (productive efficiency) plays an important role in generating treatment variation. Thus, future work should explore sources of variation in productive efficiency across hospitals and broaden the idea of productive efficiency beyond simple TFP (Garber and Skinner, 2008) to consider the reasons for comparative advantage in particular types of care. By separately identifying allocative inefficiency, we also found that lack of information about the variation across hospitals in comparative advantage generates substantial welfare loss. This finding is similar to Abaluck et al. (2016), who find that physician misperceptions about which patients benefit most from testing generate substantial welfare loss. Thus, future work should also explore how patients and providers learn about and respond to variation in productive efficiency. This would involve taking our framework for understanding how hospitals differ in efficiency, and combining it with the insights in Hull (2017), which finds evidence that patients select on gains in terms of choosing hospitals. More generally, our framework can also be applied to a range of related puzzles such as the presence of racial and gender disparities in treatment and the slow diffusion of new treatments that were proven effective in randomized trials [Chandra and Staiger, 2010)]. As with the variation we study across hospitals, our framework can identify the underlying source of these differences in treatment across populations. Our findings suggest that misperceptions and learning about the heterogeneous benefits of treatment across hospitals and patients may play a key role in understanding all of these puzzles.

References

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

Baicker Katherine, Amitabh Chandra, Jonathan Skinner. 2012. Saving Money or Just Saving Lives? Improving the Productivity of the U.S. Health Care Spending. *Annual Review of Economics* 2012;4: 33- 56.

Barnato, Amber E., F. Lee Lucas, Douglas Staiger, David E. Wennberg, and Amitabh Chandra. 2005. Hospital-Level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes. *Medical Care* 43.4; April: 308-319.

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

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

Chandra, Amitabh and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*. February.

Chandra, Amitabh and Douglas O. Staiger. 2010. "Identifying Provider Prejudice in Health care" *NBER Working Paper 16382.* September.

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

Currie, Janet and MacLeod, W. Bentley. 2017. Diagnosing Expertise: Human Capital, Decision Making and Performance Among Physicians. *Journal of Labor Economics*, forthcoming.

Doyle Jr, Joseph J. 2001. "Returns to local-area healthcare spending: evidence from health shocks to patients far from home." *American economic journal. Applied economics* 3.3: 221.

Doyle Jr, Joseph J.., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy* 123, no. 1 (February): 170-214.

Finkelstein, Amy and Gentzkow, Matthew and Williams, Heidi L. 2016. Sources of Geographic Variation in Health Care: Evidence from Patient Migration. *Quarterly Journal of Economics*:

Fibrinolytic Therapy Trialists' Collaborative Group. 1994. "Indications for Fibrinolytic Therapy in Suspected Acute Myocardial Infarction: Collaborative Overview of Early Mortality and Major Morbidity Results from all Randomized Trials of More Than 1000 patients." *Lancet* 343(8893): 311-22.

Fisher Elliott, et al. The Implications Of Regional Variations In Medicare Spending. Part 1: The Content, Quality, And Accessibility Of Care. *Annals of Internal Medicine*. 2003a February 18;138(4):273–87.

Fisher Elliott, et al. The implications of regional variations in Medicare spending. Part 2: Health outcomes and satisfaction with care. *Annals of Internal Medicine*. 2003b February 18;138(4):288–98.

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

Gawande, Atul. 2009. “The Cost Conundrum,” The New Yorker, June 1. Available at http://www.newyorker.com/magazine/2009/06/01/the-cost-conundrum

Heckman JH, S Urzua, E Vytlacil. 2006. “Understanding Instrumental Variables in Models with Essential Heterogeneity.” *The Review of Economics and Statistics* 88(3):389-432.

Hull, Peter D. 2017. “Estimating Hospital Quality with Quality-Experimental Data,” MIT Department of Economics, Unpublished. Available at http://about.peterhull.net/wp.

Marciniak, TA, EF Ellerbeck, MJ Radford, et al. 1998. “Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project.” *Journal of the American Medical Association* 279:1351-7.

McClellan, Mark, Barbara J. McNeil, Joseph P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272(11): 859-866.

O’Connor, GT, HB Quinton, ND Traven, et al. 1999. “Geographic Variation in the Treatment of Acute Myocardial Infarction: The Cooperative Cardiovascular Project.” *Journal of the American Medical Association* 281: 627-33.

Phelps, Charles E. 2000. Information Diffusion and Best Practice Adoption. *Handbook of Health Economics*, Volume 1, Elsevier: 223-264.

Skinner, Jonathan. 2011. “Causes and Consequences of Regional Variations in Health Care,” in *Handbook of Health Economics*, Volume 2, Elsevier: 45-94.

Suri, Tavnet. 2011. “Selection and Comparative Advantage in Technology Adoption,” *Econometrica*, 79: 159–209. doi:10.3982/ECTA7749

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326-65.

Yasaitis L, Fisher ES, Skinner JS, Amitabh Chandra. 2009. “Hospital quality and intensity of spending: is there an association?” *Health Affairs*,28(4): w566-w572.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: MIT Press.

Appendix-A

Construction of CCP Estimation Sample

The CCP used bills submitted by acute care hospitals (UB-92 claims form data) and contained in the Medicare National Claims History File to identify all Medicare discharges with an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal diagnosis of 410 (myocardial infarction), excluding those with a fifth digit of 2, which designates a subsequent episode of care. The study randomly sampled all Medicare beneficiaries with acute myocardial infarction in 50 states between February 1994 and July 1995, and in the remaining 5 states between August and November, 1995 (Alabama, Connecticut, Iowa, and Wisconsin) or April and November 1995 (Minnesota); for details see O'Connor et al. (1999). Among patients with multiple myocardial infarction (MIs) during the study period, only the first AMI was examined. The Claims History File does not reliably include bills for all of the approximately 12% of Medicare beneficiaries insured through managed care risk contracts, but the sample was representative of the Medicare fee-for-service (FFS) patient population in the United States in the mid-1990s. After sampling, the CCP collected hospital charts for each patient and sent these to a study center where trained chart abstracters abstracted clinical data. Abstracted information included elements of the medical history, physical examination, and data from laboratory and diagnostic testing, in addition to documentation of administered treatments. The CCP monitored the reliability of the data by monthly random reabstractions. Details of data collection and quality control have been reported previously in Marciniak et al. (1998). For our analyses, we delete patients who were transferred from another hospital, nursing home or emergency room since these patients may already have received care that would be unmeasured in the CCP. We transformed continuous physiologic variables into categorical variables (e.g., systolic BP < 100 mm Hg or ≥ 100 mm Hg, creatinine <1.5, 1.5-2.0 or >2.0 mg/dL) and included dummy variables for missing data.
Our choice of variables was based on those selected by Fisher et al. (2003a,b) and Barnato et al. (2005). With the exception of two variables that are both measured by blood-tests, albumin and bilirubin (where the rates of missing data were 24 percent), we do not have a lot of missing data (rates were less than 3 percent). Included in our model are the following risk-adjusters:

Age, Race, Sex (full interactions)
previous revascularization (1=y)
hx old mi (1=y)
hx chf (1=y)
history of dementia
hx diabetes (1=y)
hx hypertension (1=y)
hx leukemia (1=y)
hx ef <= 40 (1=y)
hx metastatic ca (1=y)
hx non-metastatic ca (1=y)
hx pvd (1=y)
hx copd (1=y)
hx angina (ref=no)
hx angina missing (ref=no)
hx terminal illness (1=y)
current smoker
atrial fibrillation on admission
cpr on presentation
indicator mi = anterior
indicator mi = inferior
indicator mi = other
heart block on admission
chf on presentation
hypotensive on admission
hypotensive missing
shock on presentation
peak ck missing
peak ck gt 1000
no-ambulatory (ref=independent)
ambulatory with assistance
ambulatory status missing
albumin low(ref>=3.0)
albumin missing(ref>=3.0)
bilirubin high(ref<1.2)
bilirubin missing(ref<1.2)
creat 1.5-<2.0(ref=<1.5)
creat >=2.0(ref=<1.5)
creat missing(ref=<1.5)
hematocrit low(ref=>30)
hematocrit missing(ref=>30)
ideal for CATH (ACC/AHA criteria)

Appendix-B

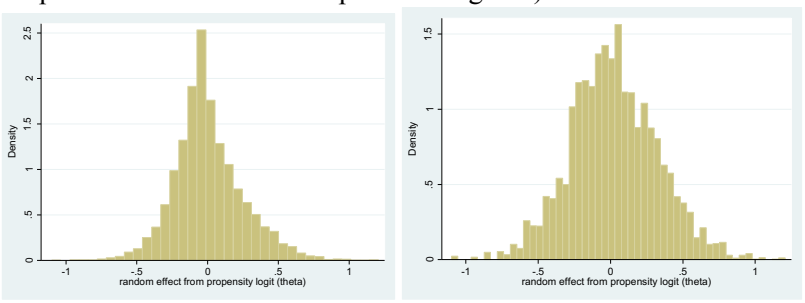
Estimation of Propensity to Receive Reperfusion

We compared results from fixed-effects and random-effects logits predicting reperfusion as a function of the full list of patient risk-adjusters and a random hospital-level intercept, and examined the sensitivity of different approaches to estimating the slope parameters for this equation, ranging from OLS with hospital fixed-effects, conditional logit with hospital fixed effecst, logit with random effects, and mixed-logit. If equation (2) is estimated using xtmelogit in Stata, then we obtain Bayesian posterior estimates of hospital random-effects, commonly referred to in the literature as shrinkage estimates or, in linear models, best linear unbiased predictions. As the table below notes, the choice between OLS and logit and between random effects and fixed effects is not significant for the estimation of hospital effects, but the use of shrinkage is because of the substantial number of small hospitals in our sample.

Model: Pr($\widehat{Treatment}_{ih}$)= $F(\hat{I}_{ih}) = F(X_i\hat{\beta} + \hat{\theta}_h)$	Correlation of Patient Characteristics ($X_i\hat{\beta}$)	Correlation of Hospital Effects ($\hat{\theta}_h$)
Fixed Effects OLS and Fixed-Effects Logit	0.9745	0.9997
Fixed Effects OLS and Random-Effects Logit (Unshrunk)	0.9997	0.9997
Fixed Effects OLS and Random-Effects (Shrunk)	0.9732	0.9998

The results from the shrunken random-effects were used to form posterior estimates of the hospital random effects $\hat{\theta}_h$ and an estimate of the propensity index for each patient $\hat{I}_{ih} = X_{ih}\hat{\beta} + \hat{\theta}_h$. The coefficients on the patient-level variables are consistent with the medical literature, with reperfusion being less likely among patients with pre-existing conditions and who are older, and also depending on the location and severity of the heart attack. The estimated standard deviation of the hospital effect is 0.44 (Std. Err. = 0.01), which implies that a one standard deviation in the hospital effect increases the logodds of receiving reperfusion by 0.44, which would increase an average patients probability of receiving reperfusion from 19% to 26%. Thus, there is sizable variation across hospitals in the rate at which they provide reperfusion to observationally similar patients. The model is able to predict much of the hospital-level variation, with the posterior prediction of each hospital's effect on reperfusion having a standard deviation of 0.30 in our data.

Also note that if hospitals vary in σ_v, then estimates of β from more aggressive or less aggressive hospitals will be different. This is testable—we estimated separate propensity equations (as in equation 2) by aggressiveness of the hospital (above and below median on $\hat{\theta}_h$) and found that their predictions are nearly identical, with a correlation of 0.9987, suggesting that differences in variances are not a first-order concern. Finally, histograms of hospital-risk adjusted treatment rates are below (Panel A is hospital level and Panel B is patient weighted)



Appendix-C

Equations used in minimum distance estimation

This appendix describes the equations used in minimum distance estimation that state the reduced-form estimates in terms of the parameters of the structural model. The mixed-logit model from Table 5 estimates 6 reduced-form moments (the variances and covariances of θ, τ, and α_0) and their associated variance covariance matrix. Call this 1x6 vector of reduced form parameters $\beta = \left(\sigma_\theta^2, \sigma_\tau^2, \sigma_{\alpha_0}^2, \sigma_{\theta\tau}, \sigma_{\theta\alpha_0}, \sigma_{\tau\alpha_0}\right)$ and let $\hat{\beta}$ be the vector of estimates of these parameters and $V = Var(\hat{\beta})$ be the associated 6x6 variance matrix of these estimates. Our structural model has 6 unknown structural parameters: the variance and covariance of α and α_0, the variance of the noise in the signal ω, the weight placed on the signal w, and the scale parameter from the logit σ_v. Call this 1x6 vector $\delta = \left(\sigma_{\alpha^\Delta}^2, \sigma_{\alpha_0}^2, \sigma_{\alpha^\Delta\alpha_0}, \sigma_\omega^2, w, \sigma_v\right)$. We can state the reduced form parameters as a function of the structural parameters (as shown below) so that $\beta = f(\delta)$. Then minimum distance estimates of δ minimize the objective function $\left(\hat{\beta} - f(\delta)\right) V^{-1} \left(\hat{\beta} - f(\delta)\right)'$. In the just-identified case the resulting structural parameters fit the reduced form moments exactly ($\hat{\beta} = f(\hat{\delta})$). Restrictions on the structural parameters can be tested based on how they affect the structural models ability to fit the reduced form estimates, using the fact that the objective function has a chi-squared distribution with degrees of freedom equal to the degree of over-identification (the difference between the dimension of $\hat{\beta}$ and the dimension of δ. Fitting the model to reduced form estimates from low, medium and high-volume hospitals is done similarly, where $\hat{\beta}$ stacks the estimates from the three samples into a 1x18 vector and V is 18x18 with the variance covariance matrix for estimates from each of the samples along the diagonal and zeros everywhere else.

The structural model interprets our original model parameters (θ, τ, and α_0) as follows:

1. α_0 is unchanged
2. $\theta^{**} = wS = w\left(\alpha^\Delta + \omega\right)$ which is the hospital's prediction of its comparative advantage given its signal. In the treatment propensity logit we estimate $\theta = \frac{\theta^*}{\sigma_v} = \frac{1}{\sigma_v} w\left(\alpha^\Delta + \omega\right)$
3. $\tau = \alpha^\Delta - \theta^* = (1-w)\alpha^\Delta - w\omega$

Using these definitions, it is straightforward to derive the following relationships between the reduced form estimates and the parameters of the structural model:

i. $\sigma_\theta^2 = Var\left[\frac{1}{\sigma} w\left(\alpha^\Delta + \omega\right)\right] = \frac{1}{\sigma_v^2} w^2\left(\sigma_{\alpha^\Delta}^2 + \sigma_\omega^2\right)$

ii. $\sigma_\tau^2 = Var\left[(1-w)\alpha^\Delta - w\omega\right] = (1-w)^2\sigma_{\alpha^\Delta}^2 + w^2\sigma_\omega^2$

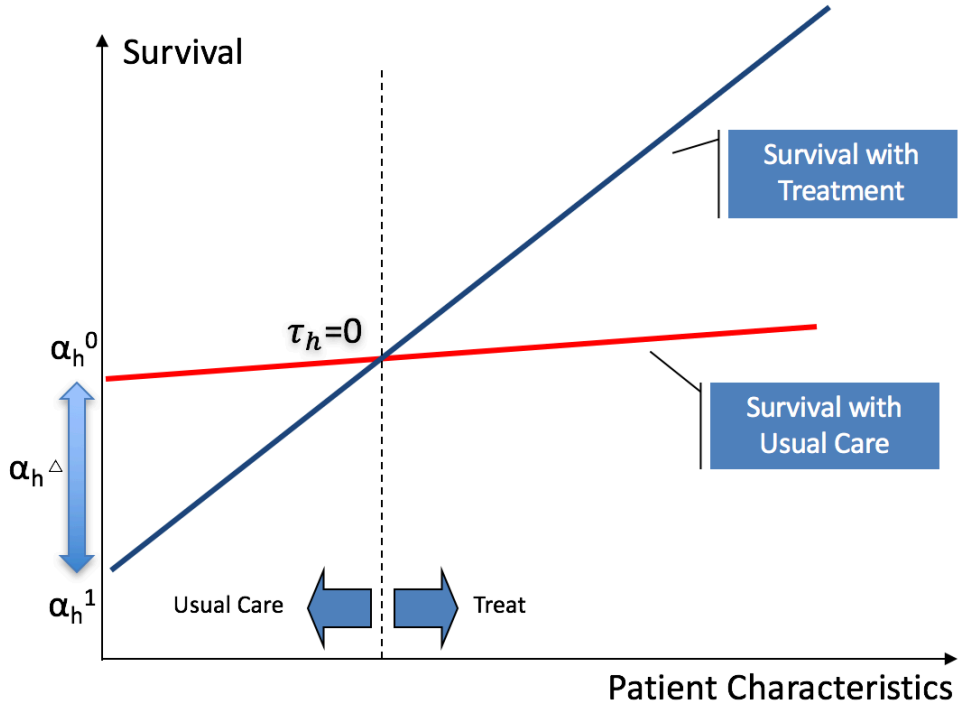
iii. $\sigma_{\alpha_0}^2 = Var[\alpha_0] = \sigma_{\alpha_0}^2$

iv. $\sigma_{\theta\tau} = Cov\left[\frac{1}{\sigma_v} w\left(\alpha^\Delta + \omega\right), (1-w)\alpha^\Delta - w\omega\right] = \frac{1}{\sigma_v} w(1-w)\sigma_{\alpha^\Delta}^2 - \frac{1}{\sigma_v} w^2\sigma_\omega^2$

v. $\sigma_{\theta\alpha_0} = Cov\left[\frac{1}{\sigma_v} w\left(\alpha^\Delta + \omega\right), \alpha_0\right] = \frac{1}{\sigma_v} w\sigma_{\alpha^\Delta\alpha_0}$

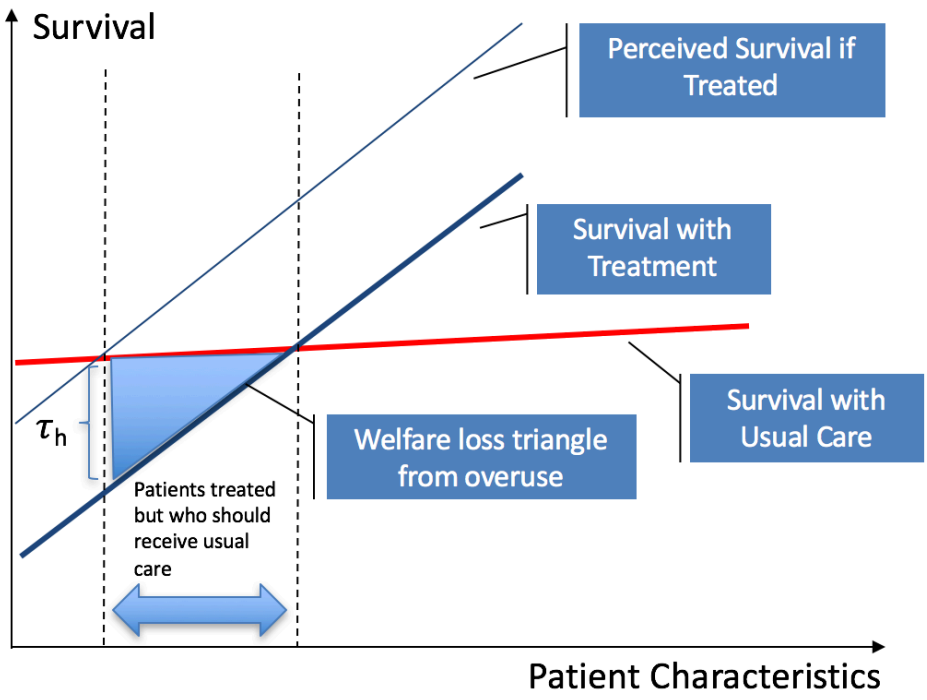
vi. $\sigma_{\tau\alpha_0} = Cov\left[(1-w)\alpha^\Delta - w\omega, \alpha_0\right] = (1-w)\sigma_{\alpha^\Delta\alpha_0}$

Figure 1A: A Roy model of Treatment at the Hospital level



The two lines denote patient survival if a hospital treats a given patient with usual care (intercept is α_h^0) or with reperfusion treatment (intercept is α_h^1) as a function of patient characteristics (i.e. patient X's) on the x-axis. We have suppressed the distribution of unobservables that come out of the plane. Expertise at usual care and reperfusion is captured by the intercepts α_h^0 and α_h^1 respectively, with comparative advantage being the difference between them. Allocative efficiency means that reperfusion should be performed to the point that the marginal patient receiving it receives zero benefit ($\tau_h = 0$).

Figure 1B: A Roy model of Treatment at the Hospital level with Allocative Inefficiency



The figure illustrates the presence of allocative inefficiency. Here, perceptions about comparative advantage at delivering the treatment result in more patients treated than is optimal. It is also possible that some hospitals overuse treatment because of maximizing something other than survival, rather than because of misperceptions about comparative advantage. As drawn, the hospital overuses treatment and uses a negative threshold (τ_h). A welfare loss triangle emerges and is illustrated in blue.

Figure 2. Why conditioning on the Propensity to Receive Treatment identifies differences in Allocative efficiency

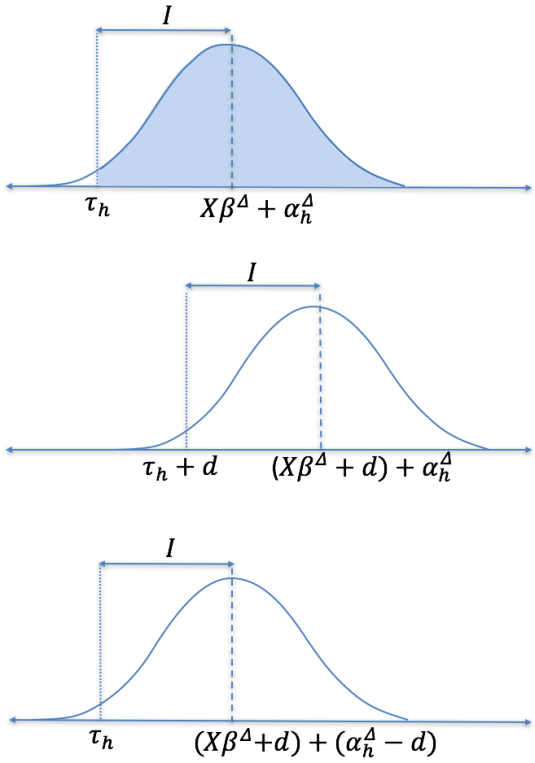
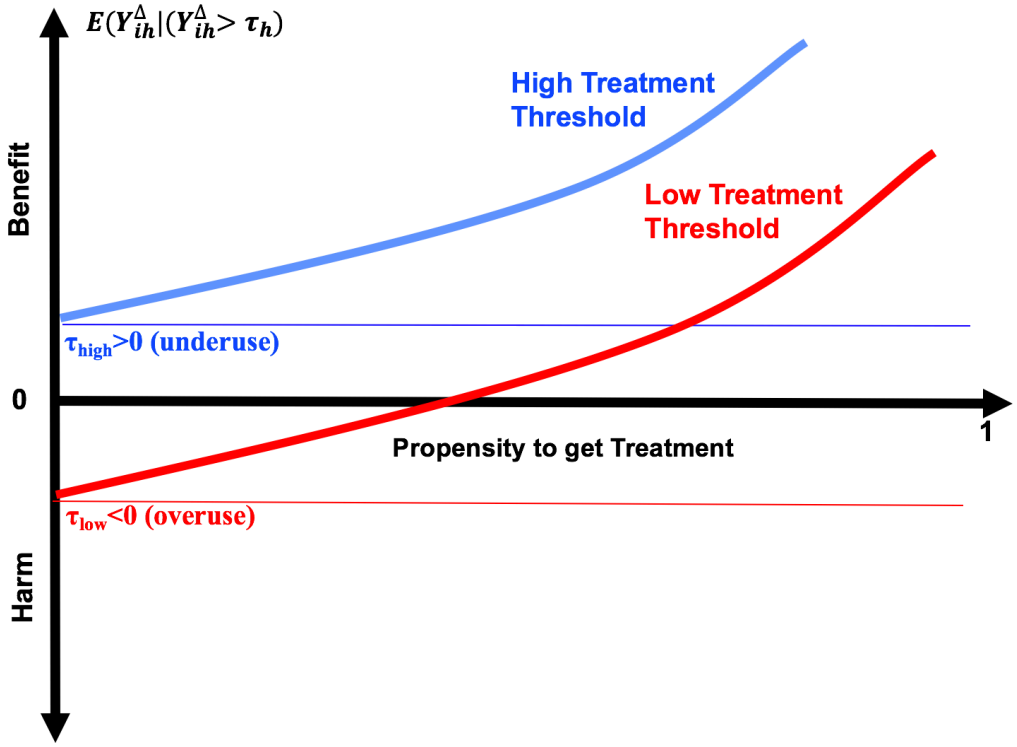


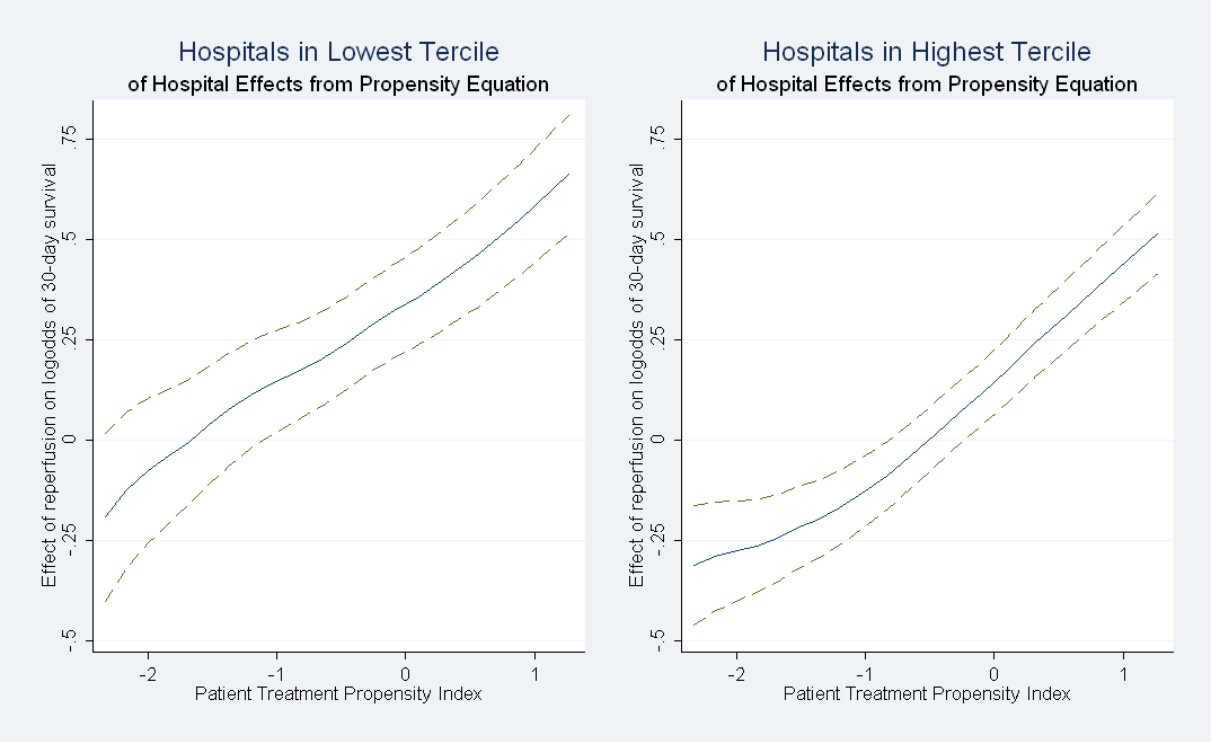
Figure show the distribution of expected benefits, $E(Y_h^{\Delta})$. Panel A illustrates selection into treatment allowing for selection on gains. Expected benefits for the *i*th patient is centered on $X\beta^{\Delta} + \alpha_h^{\Delta}$, with the distribution around this reflecting v_{ih}^{Δ}, which is used by the physician to determine treatment. Patients above τ_h are treated, so that the difference between $X_i\beta^{\Delta} + \alpha_h^{\Delta}$ and τ_h is the propensity index I . The conditional mean to the right of τ_h represents the treatment effect for the treated. Panel B illustrates how matching on the propensity isolates a higher threshold $(\tau_h + d)$ in Hospital B, that has the same comparative advantage (α_h^{Δ}) as hospital A. Compared to hospital A, patients with the same *X*'s will have a lower propensity to be treated at hospital B because of the higher threshold. However, after matching on propensity, both the threshold $(\tau_h + d)$ and the mean of the benefit distribution $\left(\tilde{X}\beta^{\Delta} + \alpha_h^{\Delta} = X\beta^{\Delta} + d + \alpha_h^{\Delta}\right)$ are shifted up by *d* in hospital B relative to hospital A, and as a result the treatment on the treated is also shifted up by *d* in hospital B relative to hospital A. In panel C we illustrate hospital C, with the same threshold (τ_h) as hospital A but lower comparative advantage $(\alpha_h^{\Delta} - d)$. When we match on propensity, patients in hospital A will be matched to patients with higher X's in hospital C. Therefore, after matching on propensity, both the threshold (τ_h) and the mean of the benefit distribution are the same in hospital C as in hospital A, implying that treatment on the treated is also the same, which implies no difference in thresholds. Note that the differences in thresholds and comparative-advantage would not be captured, if we only matched on X's.

Figure 3: Distinguishing underuse and overuse using the propensity to receive treatment.



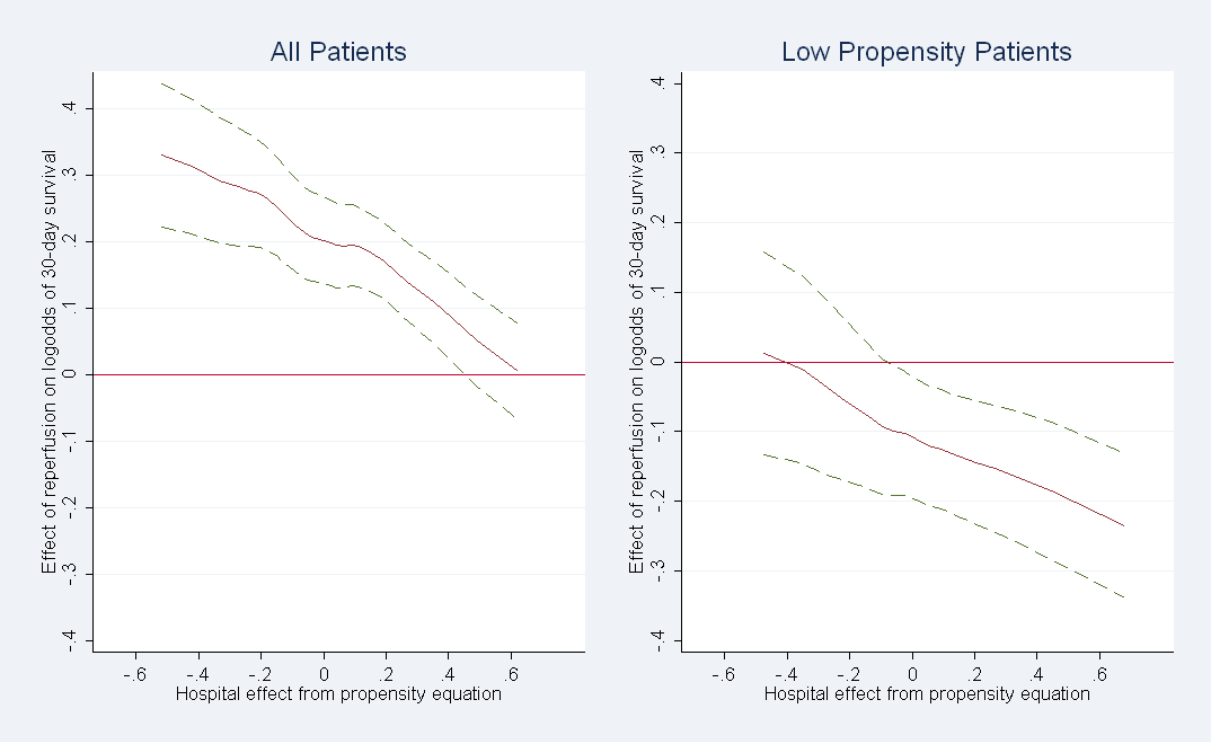
The figure illustrates the relationship between the expected benefit from treatment, $E\left(Y_{ih}^{\Delta} \mid Y_{ih}^{\Delta} > \tau_h\right)$, on the vertical axis, and the propensity index I on the horizontal axis. The propensity to receive treatment depends on patient characteristics and a hospitals assessment of its hospital specific benefit from treatment. The curves represent the treatment-on-the-treated effect for a patient with index I, and approach the minimum threshold (τ) for a patient with a low propensity of being treated. The top curve represents a hospital with a high treatment threshold (underuse) and the bottom curve represents a hospital with a low treatment threshold (overuse).

Figure 4: Survival Benefit from Reperfusion According to Patient's Treatment Propensity, Low-Treatment-Rate (Left) and High-Treatment-Rate (Right) Hospitals.



The figures plot the estimated survival benefit (and 95% confidence intervals) from reperfusion against a patient's treatment propensity index for hospital's in the lowest (left-hand side) and highest (right-hand side) terciles of the estimated hospital effect from the propensity equation. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix-B. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. All models include all CCP risk-adjusters.

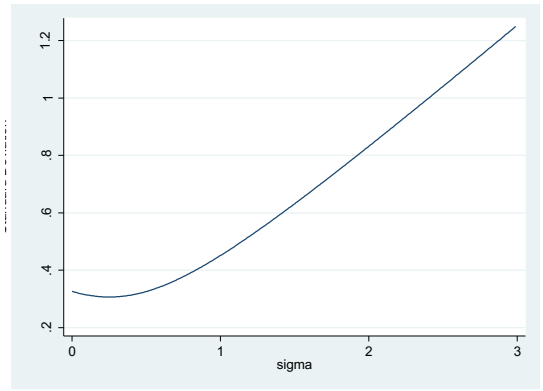
Figure 5: Survival Benefit from Reperfusion by to Hospital Effect from Treatment Propensity, All patients (Left) and Low-propensity patients (Right).



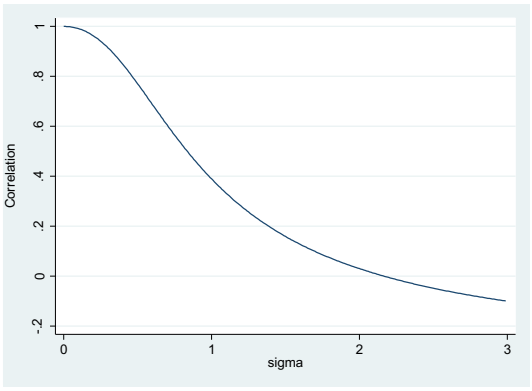
The left-hand panel plots the estimated survival benefit from reperfusion (and 95% confidence interval) against the hospital effect from the propensity equation using a locally-weighted logit model to estimate the reperfusion effect (controlling non-parametrically for the propensity index as was done in column 3 of Table 2). The right-hand panel is the analogous plot estimated only for low-propensity patients whose propensity index implied that they had below a 20% probability of receiving reperfusion. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix-B. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. All models include all CCP risk-adjusters.

Figure 6: Estimates of the Standard Deviation of Comparative Advantage (α_h^{Δ}) and its Correlation with a Hospital's Treatment Threshold (τ_h) for a Range of Values of the Scale Parameter (σ_v).

Standard Deviation of Comparative Advantage (α_h^{Δ})

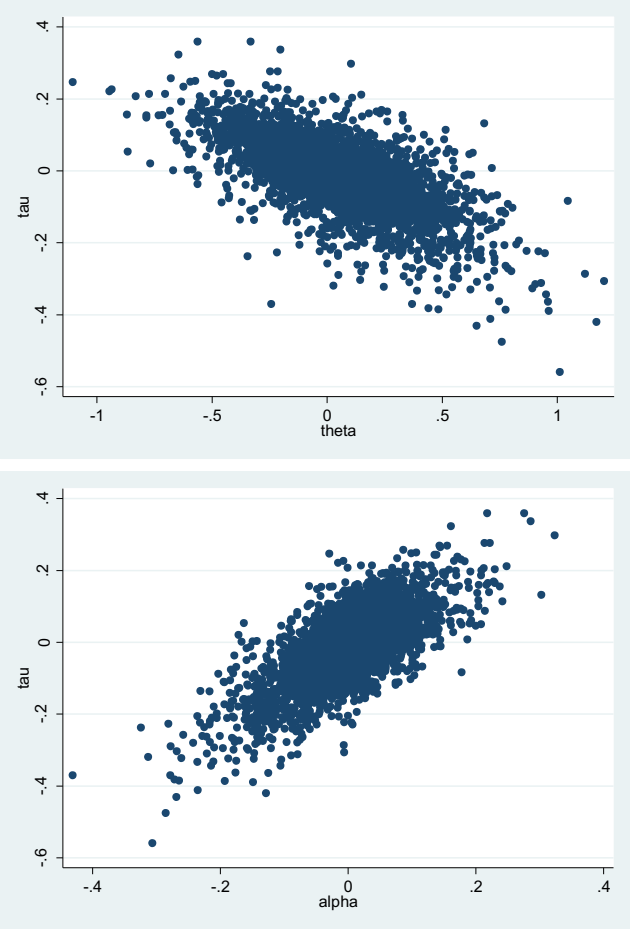


Correlation of Treatment Threshold (τ_h) With Comparative Advantage (α_h^{Δ})



We used the estimates from Table 5 to calculate estimates of the standard deviation of α_h^{Δ} and its correlation with τ_h for a range of values for σ_v. The left hand panel plots estimates of the standard deviation of comparative advantage α_h^{Δ} for values of σ_v from 0.01 to 3, while the right hand panel plots estimates of the correlation of α_h^{Δ} with τ_h for the same range of σ_v.

Figure 7: Hospital level correlation between Risk-Adjusted Treatment rates, Thresholds, and Comparative Advantage



We used the estimates from Table 5 to graph the correlation between risk-adjusted treatment rates (θ_h) and thresholds (τ_h), as well as the correlations between comparative-advantage (α_h^{Δ}) and τ_h, while setting σ_v to 0.25, the value that minimizes the role of comparative advantage across hospitals. All parameters are demeaned to zero for the average hospital. The positive correlation in the second panel would arise if hospital's *incorrectly* believed that they had high comparative advantage in performing the treatment, resulting in overuse among those hospitals that actually did not have a high comparative advantage in performing the treatment.

Table 1: Patient Characteristics, Full Sample and by Reperfusion

Variable	Full Sample	Received Reperfusion w/in 12 hours	No Reperfusion w/in 12 hours
Survival 30 days post-AMI	81%	86%	80%
Reperfusion within 12 hours	19%	100%	0%
Age	77	73	77
Previous diagnoses:			
Congestive Heart Failure	22%	7%	25%
Hypertension	62%	56%	63%
Diabetes	30%	23%	32%
Dementia	6%	2%	7%
Number of observations	138,957	25,876	113,081

Note: Full-list of variables is in Appendix-A.

Table 2: Effect of Reperfusion on 30-day Survival, OLS and Logit Estimates

	OLS (1)	OLS (2)	OLS (3)	Logit (4)	Logit (5)
Reperfusion	0.039	0.043	non-parametric	0.328	non-parametric
	(0.003)	(0.003)		(0.027)	
Reperfusion*Propensity index	0.040	0.042	non-parametric	0.291	non-parametric
	(0.002)	(0.002)		(0.018)	
Reperfusion* Hospital Treatment Rate (θ)		-0.031	-0.037	-0.211	-0.254
		(0.009)	(0.009)	(0.076)	(0.077)
Hospital Fixed-Effects	Yes	Yes	Yes	Yes	Yes
Control for Propensity Index	None	Linear	Non-Parametric	Linear	Non-Parametric

Note: Dependent variable is the whether patient survived to 30 days. Reperfusion measures receipt of reperfusion therapy within 12 hours of admission. OLS coefficients are percentage-point changes in survial and logit coefficients are log-odds. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix-B. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. All models include all CCP risk-adjusters. Columns 2 and 4 include linear controls for propensity-index. Columns 3 and 5 includes 100 percentiles of propensity-index interacted with the receipt of Reperfusion.

Table 3: Effect of Reperfusion on 7-day and 365-day Survival, Logit Estimates

	Conditional on Propensity	Conditional on Propensity
	(1)	(2)
Panel A: 7 Day Survival		
Reperfusion	0.233 (0.031)	non-parametric
Reperfusion * propensity index	0.362 (0.020)	non-parametric
Reperfusion * Hospital Treatment Rate (θ)	-0.368 (0.084)	-0.511 (0.087)
Control for Propensity Index	Linear	Non-Parametric
Panel B: 365 Day Survival		
Reperfusion	0.403 (0.024)	non-parametric
Reperfusion * propensity index	0.181 (0.016)	non-parametric
Reperfusion * Hospital Treatment Rate (θ)	-0.177 (0.066)	-0.351 (0.068)
Control for Propensity Index	Linear	Non-Parametric

Note: Coefficients are log-odds. Table is analogous to Table 2. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. Column 3 reports equation 7c and Column 4 reports equation 7b. All models include all CCP risk-adjusters. Sample-size in every regression is 138,957.

Table 4: Effect of Reperfusion on 30-day Survival, Mixed-Logit Estimates

	(1)	(2)
Reperfusion	0.297	0.314
	(0.0218)	(0.0243)
Reperfusion*Propensity index	0.289	0.292
	(0.0169)	(0.0171)
Std dev of hospital intercept (α^0)	0.188	0.198
	(0.0151)	(0.0168)
Hospital Level Random-Intercept (α^0)	Yes	Yes
Hospital Level Random Coefficient on Reperfusion (τ)	No	Yes
Std dev of hospital coefficient on reperfusion (identifies τ; hospital level thresholds)		0.313
		(0.0557)
corr(hospital level intercept, coefficient on reperfusion) (identifies corr (α^0, τ))		-0.331
		(0.154)
Number of Hospitals	4,690	4,690
Control for Propensity Index	Linear	Linear

Note: Coefficients are log-odds. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. All models include all CCP risk-adjusters. Columns 2 and 4 include linear controls for propensity-index as in equation 7a. Sample-size in every regression is 138,957.

Table 5: Effect of Reperfusion on 30-day Survival, Hierarchical-Logit Estimates

Reperfusion Equation:	
Std. Dev. Of Hospital Reperfusion Rate (θ)	0.442
	(0.013)
30-day Survival Equation:	
Reperfusion	0.265
	(0.026)
Reperfusion * Propensity Index	0.276
	(0.018)
Hospital-level intercept (α_0; general productivity)	
Standard Deviation	0.199
	(0.017)
Correlation with Hospital Reperfusion Rate (θ)	-0.100
	(0.073)
Hospital minimum treatment threshold (τ)	
Standard deviation	0.327
	(0.055)
Correlation with Hospital Reperfusion Rate (θ)	-0.341
	-(0.106)
Correlation with General Productivity (α^0)	-0.321
	(0.150)
Number of Hospitals	4,690
Control for Propensity Index	Linear

Note: Coefficients are log-odds. Table reports estimates from hierarchical logit, where the propensity to receive treatment is estimated simultaneously with the survival equation. See text and discussion of equation 7c for details.

Table 6: Effect of Reperfusion on 30-day Survival, By Type of Hospital

Reperfusion	0.205
	(0.195)
Reperfusion*Propensity Index	0.289
	(0.0171)
Reperfusion*Church Operated Hospital	-0.0119
	(0.0596)
Reperfusion*For-Profit Hospital	0.0341
	(0.0750)
Reperfusion*Government Hospital	0.0230
	(0.0719)
Reperfusion*ln (Discharge Volume)	0.0214
	(0.0429)
Reperfusion*Major Teaching Hospital	0.0352
	(0.0816)
Reperfusion*Minor Teaching Hospital	-0.00622
	(0.0675)
Reperfusion*Percent of DSH Patients	-0.156
	(0.200)
Reperfusion*ln (Beds)	0.00949
	(0.0516)
Reperfusion*Resident to Bed Ratio	-0.230
	(0.296)
Constant	0.0984
	(0.0757)
Propensity	Linear
Hospital Random Effects	Yes
Observations	138,957
Number of Hospitals	4,690

Note: Coefficients are log-odds, from mixed-logits that allow for random coefficients and are analogous to Table 4. Omitted characteristics is a non-profit hospital. 30-day survival is regressed on hospital-characteristics and hospital characteristics interacted with treatment. The table reports the interaction effects. A test of joint-significance on these interactions yielded a chi-square statistic of 2.96, p=.097.

Table 7: Minimum Chi-Squared Estimates of Structural Parameters

	Pooling All Hospitals		Separately by Hospital Volume	
	Just-identified	Constrain w=r	Different Reliability by Hospital Volume	Same Reliability by Hospital Volume
Std. Dev (α_0)	0.198	0.204	0.200	0.200
	(0.0167)	(0.017)	(0.016)	(0.016)
Std. Dev (α^{Δ})	0.317	0.407	0.337	0.336
	(0.058)	(0.059)	(0.057)	(0.057)
Corr(α^{Δ}, α_0)	-0.390	-0.438	-0.457	-0.438
	(0.145)	(0.130)	(0.148)	(0.154)
σ_v	0.435	0.367	0.431	0.441
	(0.152)	(0.218)	(0.117	(0.135)
w (weight)	0.154	constrained	0.119	0.114
	(0.169)		(0.107)	(0.127)
r (reliability)	0.065	0.155		0.040
	(0.106)	(0.162)		(0.069)
r (big Hospitals)			0.069	
			(0.093)	
r(Medium Hospitals)			0.047	
			(0.063)	
r(Small Hospitals)			0.019	
			(0.026)	
# moments being fit	6	6	18	18
Degrees of freedom	0	1	10	12
Chi-Squared statistic	NA	10.4	12.7	62.4
(p-value)		(p=.001)	(p=.24)	(p<.001)

The first two columns fit 6 reduced-form moments estimated from our empirical model (the variances and covariances of θ, τ, and α^0) as a function of the unknown structural parameters in this framework (the variance and covariance of α^{Δ} and α^0, the reliability of the signal r, the weight placed on the signal w, and the scale parameter from the logit σ_v). The reduced-form moments were estimated pooling all hospitals. The unknown structural parameters were estimated using minimum chi-squared methods. The last two columns fit our model to reduced-form moments estimated separately for low (20 or fewer patients), medium (21-80 patients) and high (81 or more patients) volume hospitals - 6 moments for each group, for a total of 18 moments.