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, Revised December 2019

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 and the editors and referees 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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.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, Revised December 2019
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. 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. A simple economic model provides an empirical framework to separate these explanations. Estimating this model with data 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 the treatment that we study by 44%.

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 (2011); 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.[1] A different interpretation is 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 use rich clinical data from the Cooperative Cardiovascular Project for elderly patients experiencing heart-attacks in the mid-1990s and being treated with reperfusion therapy to estimate variation across patients and hospitals in both the propensity to be treated and treatment effects. Because of the richness of the clinical variables in our data, we are able to credibly estimate treatment effects,

[1] There are many sources of guidelines and recommendations in healthcare—ranging from screening guidelines from the US Preventative Task Force (USPTF), to treatment guidelines from the American College of Cardiology (O'Gara et al, 2013) for heart-attacks, to treatment guidelines for very specific cancers from the American Society of Clinical Oncologists (Burstein et al, 2018).

replicating estimates from randomized controlled trials and demonstrating robustness of our estimates to controlling for increasingly detailed covariates. We find that aggressive hospitals (with high risk-adjusted treatment rates) are much more likely to treat patients with negative treatment effects, suggesting overuse of treatment in these hospitals.

To interpret these results, we estimate the model implied by our economic framework and use it to separately identify variation across hospitals in allocative inefficiency and comparative advantage. 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, 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 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 hospitals 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, describes the data, and provides some motivating facts documenting how simple estimates of treatment rates and treatment effects vary across hospitals and patients. Section II develops a theoretical model for understanding these facts, 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 graphical results and regressions suggested by our theoretical model 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 some stylized calculations of the welfare-loss from variation in treatment rates. All appendix material can be found in an online appendix.

## I. HEART-ATTACKS: TREATMENTS, DATA, AND SOME MOTIVATING FACTS

### *I.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. 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. In our data from the mid-1990s, over 90 percent of patients receiving reperfusion received thrombolytics.

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 confounds the analysis.

Fifth, within hospitals, heart-attacks are treated by teams comprising doctors and nurses in the emergency room, in the hospital, in cardiac-care units and in post-acute facilities. This makes heart-attacks treatments more hospital-oriented than other areas of medicine where an individual physician may have a primary role. This motivates our focus on hospitals over individual physicians.[2] Most relevant to our paper, we don't have physician or team identifiers in our data. Nor do the data that we have, or most other claims data, identify the *first* physician who saw the patient in the emergency-room (this is key for attribution because all subsequent physicians are endogenous to this first physician). Even if we had these identifiers, we'd need assignment to subsequent physicians to be random. While this is possibly true for emergency-room physicians, it's unlikely to be true for other downstream physicians.

### *I.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). The CCP samples all Medicare beneficiaries who had an heart-attack between February 1994 and July 1995 (45 states), between August and November, 1995 for Alabama, Connecticut, Iowa, and Wisconsin, and between April and November 1995 (Minnesota). The CCP used bills submitted by acute care hospitals and contained in the Medicare claims to identify all Medicare discharges with a principal diagnosis of myocardial infarction. These data were matched to hospital 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

[2] We also note that the medical literature around measuring individual physician quality, does not examine the quality of physicians for heart-attack treatments. The literature on physician-level measures typically looks at high-mortality and common conditions such as carotid endarterectomy, coronary artery bypass grafting, valve replacement, abdominal aortic aneurysm repair, lung resection, cystectomy, pancreatic resection, and esophagectomy. Heart-attacks are common and high-mortality, so our sense for why the medical literature has not looked at physician quality in this setting is that it is hard to do and may not be key relative to hospital-level factors.

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 Appendix I. 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).

One concern with using the CCP is that data from the mid-1990s may be less relevant today: perhaps variations have fallen, or perhaps variations today aren't reflective of variations from many years ago.[3] To explore whether variation from the 1990s is similar to variation in more recent years, we used Medicare fee-for-service data and examined a sample of 2,970 hospitals that treated heart-attacks in Medicare in 1992-5 and in 2012-15. We dropped hospitals that closed or merged or opened in this window. Since reperfusion is not recorded in claims data, we instead used cardiac-catheterization as a proxy for intensive management (Chandra and Staiger, 2007).[4] Over this time period, cardiac-catheterization rates for heart-attack patients increased from 35% to 52%. Over twenty years, the correlation between hospital-level catheterization rates is 0.69 (unweighted) and 0.73 (patient weighted), showing that treatment intensity is highly correlated over time. In both periods, the standard-deviation of these hospital-level treatment rates was about the same: 0.23 and 0.25 respectively, suggesting that hospital variations continue to be a persistent phenomenon.

We report some basic characteristics of our sample in Table I. 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 may be due to differences in underlying health and pre-existing conditions, rather than the result of reperfusion. Overall, patients in the sample are elderly (average age of 76.7) and have high rates of chronic conditions such as hyptertension (62%), diabetes (30%) and congestive heart failure

[3] Chandra and Skinner (2012) argue that the scope of allocative inefficiency has increased over time because new medical technologies have fewer side-effects (e.g. stents and many pharmaceutical treatments) and that increases the scope for overusing them, in the sense of using a technology whose marginal benefit is less than its social cost. In the past, technologies were more invasive or had larger side-effects (for example cardiac bypass), and that reduced the willingness to overuse them because death may have been an immediate side-effect of using invasive procedures in marginally less appropriate patients. We are less sure about whether the scope for comparative advantage differences is increasing over time, but to the extent that more complicated medicine requires more expertise in how different inputs are combined, then the answer to this may be affirmative as well.

[4] In the CCP data, reperfusion and catheterization are correlated 0.35 at the hospital level.

(22%). However, patients receiving reperfusion were relatively younger, and much less likely to have these pre-existing conditions.

We have 4690 hospitals in the sample. Patients receiving reperfusion were treated in similar hospitals in terms of ownership, teaching status, and size of the hospital, reflecting the fact that reperfusion is a procedure that does not require special hospital capabilities and is done in nearly all hospitals. The average hospital in our sample has 30 patients, with 10th percentile hospital serving fewer than 11 patients and 90th percentile hospital having 73 patients. The average patient is in a hospital with 67 patients and the 10th percentile patient is in a hospital with 15 patients and the 90th percentile patient is in a hospital with 133 patients.

### *I.C. Motivating Facts*

We begin by documenting some key empirical facts about how treatment rates and treatment effects vary across hospitals and across patients. This model-free summary of key facts motivates the theoretical and empirical model that we develop below and use to interpret this variation. To describe variation in treatment rates we use the CCP sample (Table I) to estimate a random-effect logit model of whether a patient received reperfusion within 12 hours regressed on a rich set of covariates derived from patient charts (see Appendix I for a full list) and a hospital-level random intercept. We use empirical Bayes (shrinkage) estimates of the random logit intercepts to capture variation in treatment rates at the hospital level for observationally similar patients, and use the estimated treatment propensity index based on patient covariates (without the hospital intercept) to capture variation in treatment rates at the patient level. Empirical Bayes estimates account for estimation error due to small samples of patients in each hospital, and tend to understate the true amount of variation in the hospital-level parameters. To describe variation in the effect of reperfusion on survival we estimate a random-coefficient logit model of whether a patient survived 30 days after their heart attack regressed on whether the patient received reperfusion within 12 hours, controlling for the full set of patient covariates, and allowing for a hospital-level random intercept and (possibly correlated) random coefficient on reperfusion. We again use empirical Bayes estimates of the random coefficient on reperfusion to capture variation in the treatment effect at the hospital level. As we discuss in more detail below, this survival logit estimates an average effect of reperfusion that is very similar to estimates from randomized controlled trials, suggesting that the CCP control variables are sufficient to yield unbiased estimates of the treatment effect.

Panel A of Figure I illustrates how treatment rates and treatment effects vary across hospitals. For each of our 4690 hospitals, we plot the hospital intercept estimated from the reperfusion logit (x axis) versus the hospital treatment effect estimated from the survival logit (y axis). This figure illustrates

three key facts. First, there is large variation across hospitals in risk-adjusted treatment rates (the reperfusion intercepts), ranging from -1 to 1 (in logodds, where an average hospital is normed to 0). This fact is at the heart of the variations literature in economics and medicine, and has led some to argue that uniform treatment guidelines are needed to reduce unwarranted variation. Second, there is also substantial variation across hospitals in the benefit from treatment (the treatment effect from the survival logit), ranging from slightly negative (indicating harm) to a positive effect on the logodds of survival of over 0.4. This fact is fairly novel,[5] and challenges the notion of uniform treatment guidelines which implicitly assume that there is no hospital-level variation in the benefit from treatment. The third fact is that these hospital level treatment rates and treatment benefits are uncorrelated (correlation = -.02). This fact raises a puzzle—why are hospitals with a low benefit from treatment treating patients at the same rate as hospitals with a high benefit from treatment? One answer may be overuse.

Panel B of Figure I illustrates how treatment rates vary across patients. In Panel B, we non-parametrically plot the probability of treatment as a function of the treatment propensity index from the reperfusion logit based on patient covariates (without the hospital intercept), estimating separate lines for the probability of treatment in hospitals in the highest (most aggressive) and lowest (most conservative) terciles of risk-adjusted treatment rates (the hospital-level reperfusion intercept). This figure illustrates two additional facts. First, treatment increases with the propensity index for both types of hospitals. This suggests that both aggressive and conservative hospitals use patient characteristics similarly to sort patients into more or less appropriate for treatment. Second, the line for hospitals in the highest tercile is both higher and extends further to the left. This underscores the point that hospitals with higher risk-adjusted treatment rates are going much deeper into the distribution of observable patient characteristics.

Panel C of Figure I illustrates how the survival benefits of reperfusion (the treatment effects) vary across patients. Similarly to Panel B, Panel C non-parametrically plots the treatment effect as a function of the treatment propensity index based on patient characteristics, separately for hospitals in the highest (most aggressive) and lowest (most conservative) terciles of risk-adjusted treatment rates. The survival benefit at each point in the distribution of the propensity index was estimated flexibly using a local-logistic regression of 30-day survival on reperfusion that controlled for the detailed risk adjusters available in the CCP. This figure illustrates three final facts. First, for both types of hospitals, the benefit associated with reperfusion is increasing with the propensity index. This suggests that patients with

[5] Hospital-level variation in patient outcomes (and their relationship to treatment rates) is well documented in the literature, but few studies have tried to estimate hospital-level variation in the return to treatment (Skinner, 2011).

greater benefits from treatment are more likely to receive treatment and, conversely, that the treatment propensity serves as a strong indicator of a patient's expected benefit from treatment. Thus, while treatment rates are uncorrelated with treatment effects at the hospital level, they are strongly correlated at the patient level. The second fact is that reperfusion is associated with worse survival (indicating harm) for patients with low treatment propensity, particularly for patients treated in aggressive hospitals with high risk-adjusted treatment rates. The aggressive hospitals treat between 10 and 30 percent of patients for whom we estimate reperfusion is associated with worse survival, while the conservative hospitals treat fewer than 10 percent of these patients. This evidence suggests that aggressive hospitals are over-treating some patients. Finally, it is interesting to note that the benefit associated with reperfusion for patients with the same characteristics are about the same despite the 30 point higher probability of being treated in the higher tercile hospitals. If top tercile hospitals treated more patients because of comparative advantage (greater benefit from reperfusion in these hospitals), we would expect higher benefits associated with reperfusion in these hospitals. If anything, we observe lower returns in the aggressive hospitals among low propensity patients. This fact, along with observing negative benefits among the lower propensity patients, suggests the existence of allocative inefficiency.

## 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.

### *II.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.[6] 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).[7] 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 $\alpha_h^0$, 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 $\alpha_h^1$, 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 $\alpha_h^1$ and $\alpha_h^0$, 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 $Y_{ih}^{\Delta}$ 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 $\varepsilon_{ih}^{\Delta} = \varepsilon_{ih}^1 - \varepsilon_{ih}^0$

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), $\alpha_h^{\Delta}$ represents the hospital-specific benefit in providing reperfusion. One could think of $\alpha_h^0$ 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

[6] Throughout this section we treat survival, Y, as a continuous outcome measure. In the empirical work, we use binary outcomes indicating whether the patient survived beyond 30 days (or other thresholds). Thus, Y is the latent variable in the survival logits that we estimate.

[7] For now, we abstract from the problem that hospitals should also think about costs and stop treatment prior to achieving zero marginal benefits—that is, that maximize benefits net of cost. We visit this issue in Section III.B. The cost of treating heart-attacks is small relative to the survival benefit of the treatments, but this may not be true for other medical conditions.

Skinner (2012) and Syverson (2001)]. Efforts to increase $\alpha_h^0$ are efforts to increase productive efficiency-- increasing the fraction receiving beta-blockers or improving patient safety are examples. The higher the $\alpha_h^0$, the lower the benefit from reperfusion, for a fixed level of $\alpha_h^1$. Because $\alpha_h^\Delta$ 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 $\beta_h^\Delta$ term, although in the empirical work we found these to be unnecessary and assumed $\beta_h^\Delta = \beta^\Delta$.

***II.B. Treatment Choice***

Each patient receives treatment if the expected benefit from treatment exceeds a minimal threshold $\tau_h$, where the threshold may vary across hospitals due to incentives or information as discussed further below. Since $E\left(Y_{ih}^\Delta\right)$ captures the total expected benefit to the patient of providing treatment, then the optimal decision from the patient's perspective would let $\tau_h$=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 2A 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 $\alpha_h^0$) or using reperfusion (intercept is $\alpha_h^1$) 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 ($v_{ih}^\Delta$). 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 $\alpha_h^0$ and $\alpha_h^1$ 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, ($\tau_h$ = 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 $\alpha_h^1$, 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 $\alpha_h^0$, a high $\alpha_h^1$, 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 hospital's comparative advantage at delivering reperfusion ($\alpha_h^{\Delta}$), and the level of allocative efficiency at the hospital ($\tau_h$): 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 2B 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 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) $\Pr(Treatment_{ih}=1)=\Pr\left(E\left(Y_{ih}^{\Delta}\right)>\tau_h\right)=\Pr\left(\alpha_h^{\Delta}+X_i\beta^{\Delta}+v_{ih}^{\Delta}>\tau_h\right)=\Pr\left(-v_{ih}^{\Delta}<I_{ih}\right)$,
$where\ I_{ih}=X_i\beta^{\Delta}+\theta_h\ \ and\ \ \theta_h=\alpha_h^{\Delta}-\tau_h$

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 ($v_{ih}^{\Delta}$). If we make the standard assumption that the distribution of patient-level idiosyncratic gains ($v_{ih}^{\Delta}$) are i.i.d. (an assumption we return to below), then the parameters ($\beta^{\Delta},\theta_h$) 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. [8] The hospital-specific intercept ($\theta_h$) 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 ($\boldsymbol{\alpha_h^{\Delta}>0}$), or because of using a lower benefit threshold for providing care $\left(\tau_h<0\right)$ 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.

### *II.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 $\boldsymbol{Y_{ih}^{\Delta}}$ (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) $E\left(Y_{ih}^{\Delta}|Treatment_{ih}=1\right)=\tau_h+g(I_{ih})$

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

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 ($v_{ih}^{\Delta}$) 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 ($v_{ih}^{\Delta}$) 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 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). We will explore this possibility later, but do not find this to be the case. The conditional mean $E\left(v_{ih}^{\Delta} \middle| -v_{ih}^{\Delta} < I_{ih}\right)$ can also be modeled parametrically as a Mills-Ratio and we find support for this assumption as well.

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 ($\tau_h$). Note that the propensity to receive treatment ($I_{ih}$) depends on the hospital effect ($\theta_h$) 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. The intuition for this result is straightforward. By conditioning on the propensity to receive treatment ($I_{ih} = X_i\beta^{\Delta} + \alpha_h^{\Delta} - \tau_h$), we hold the difference between the mean of the benefit distribution ($X_i\beta^{\Delta} + \alpha_h^{\Delta}$) and the truncation threshold ($\tau_h$) fixed. Thus, holding $I_{ih}$ constant, any difference in $\tau_h$ shifts both the mean of the benefit distribution and the truncation point by the same amount (keeping the difference fixed), and therefore the truncated mean also increases by that same amount. Had we not conditioned on the propensity (or just conditioned on $X_i\beta^{\Delta}$ as we did for the motivating facts in Figure I), differences across hospitals in the treatment effect on the treated would be difficult to interpret because they would depend on both comparative advantage ($\alpha_h^{\Delta}$) and allocative inefficiency ($\tau_h$).

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 III. 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 III 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 ($\tau_{high}$ or $\tau_{low}$) 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 ($\alpha_h^{\Delta}$) 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 ($\alpha_h^{\Delta}$) 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.

***II.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, latent 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 $\alpha_h^0$, patient risk adjusters $X_{ih}$ and a patient-specific treatment effect $Y_{ih}^{\Delta}$. 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 relying on the conventional `selection on observables' assumption that both $v_{ih}^0$ and $v_{ih}^1$ are conditionally uncorrelated with treatment, which would imply that receipt of treatment is uncorrelated with the gain from treatment. 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)]. In our model, conditional on our control variables, we assume that treatment is uncorrelated with omitted factors in a patient's baseline survival ($v_{ih}^0$), but allow for selection on gains where treatment is correlated with omitted factors in the patient's unobserved gain ($v_{ih}^{\Delta} = v_{ih}^1 - v_{ih}^0$). Thus, instead of assuming that treatment is uncorrelated with $v_{ih}^{\Delta}$ we must assume that $v_{ih}^0$and $v_{ih}^{\Delta}$are conditionally uncorrelated. We think these assumptions are plausible given our rich data and the nature of the reperfusion treatment decision.

The assumption that treatment is uncorrelated with $v_{ih}^0$ is plausible given the rich covariates that we have. The chart data collected in the CCP was focused on patient factors that predict outcomes, so there is some reason to think that the CCP included all data from the patient's chart that physicians would use to predict baseline mortality ($v_0$). While the assumption that $v_{ih}^0$and $v_{ih}^{\Delta}$are uncorrelated is a strong one, we think there are a number of reasons that it is more likely in our setting than assuming treatment is uncorrelated with $v_{ih}^{\Delta}$. First, because reperfusion is a treatment that must be done quickly after admission, the physician is not able to wait and observe the patient's outcome without treatment ($Y_{ih}^0$). Thus, the physician's belief about the benefit from reperfusion ($v_{ih}^{\Delta}$) is unlikely to be driven by knowledge of a patient's baseline outcome ($Y_{ih}^0$). This would not be the case in many settings in which people get to see their actual outcome in the untreated state (e.g. workers may enroll in training because their current job is going poorly, and heart attack patients may not receive bypass because they died before being stabilized).

Second, unlike many surgical treatments, the benefits of reperfusion are not closely related to baseline mortality risk. The key question is whether gains from treatment (and selection into treatment) are driven by baseline mortality risk. For some treatments the answer is yes – e.g. invasive surgical treatments, or treatments that occur weeks after index admission, are much less successful if the patient is at high risk of death (patient is too frail to absorb the physical trauma of the procedure, or does not survive/stabilize to get eventual treatment). For reperfusion, this is less likely to be true because it happens immediately, and it is less invasive, so that success depends less on how frail the patient is and more on specific patient conditions that put them at risk for complications such as bleeding. Thus, the benefits of the treatment are specific to reperfusion, rather than a direct function of baseline mortality risk.

Third, many idiosyncratic factors occurring in the hospital that are unrelated to a patient's baseline mortality may affect the benefits of reperfusion. For example, whether a patient gets treated with reperfusion 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 $v_{ih}^{1}$) , that are idiosyncratic to every patient situation, and act on them.

While we think the selection on gains assumption is defensible on *a priori* grounds, it is still a strong assumption. In the results section, we provide evidence supporting the case that we can estimate the treatment on treated effect, including 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 ($\theta_h$) 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 $\theta_h$. For more details see Appendix II.

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 $\alpha_h^{\Delta}$.

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

Here, the hospital-specific coefficient on treatment identifies differences across hospitals in $\tau_h$. Note that separating the average level (as opposed to differences across hospitals) of $\tau_h$ 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 $\lambda_0$ captures the average effect of reperfusion among the treated. The hospital-specific intercept in this regression identifies hospital TFP ($\alpha_h^0$). The coefficient ($\lambda_1$) 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 ($\alpha_h^0$ and $\tau_h$). 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 $\theta_h$. 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, $\theta_h$ 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, $\theta_h$ 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 ($\lambda_2$) 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

### *III.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 IV, 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. These graphs are similar to Figure I, only now we include the hospital effect $\hat{\theta}_h$ in the propensity, as suggested by our model.

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 hospitals 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

II. 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 there, we estimate hospital random-effects which allow for empirical Bayes shrinkage. Empirical Bayes reduces prediction error in the estimates of the hospital effects by shrinking standard hospital fixed-effect estimates back toward zero, particularly for small hospitals with noisy estimates. In the appendix, we show that OLS and logit models with random and fixed-effects generate very similar (correlation>.97) estimates of the impact of patient characteristics ($X_i\hat{\beta}$) and unshrunk estimates of hospital effects ($\hat{\theta}_h$). However, the use of shrinkage does matter for the estimates of hospital effects because of the substantial number of small hospitals in our sample: The correlation between the shrunk and unshrunk random effect from the logit model is only 0.88.

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.[9] The linearity in log-odds result was not implied by our model, but it will greatly simplify our empirical work.

Figure V is similar to Figure IV, 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.

[9] 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.

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 hospitals with the lowest treatment rates are estimated to have survival benefits from 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 II 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 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 four columns report OLS estimates and the last three report logit estimates where the coefficients are odds ratios.

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 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, and consistent with the simpler graphical evidence

presented before, this is evidence of allocative inefficiency – if a hospital's high use of reperfusion was entirely due to comparative advantage, 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 ($\tau$), where more aggressive hospitals have lower minimum thresholds for treatment, treat more patients, and have lower benefits to treatment. In column 3 we control for the propensity-index through a mills ratio and note that this type of control is quite similar to controlling for the index linearly.[10] The coefficient is similar in column 4, 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 three columns of Table II 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 III we estimate the logit models for 7-day and 360-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 360-day survival (Panel B) relative to 7-day survival, and represents the importance of post-discharge factors in affecting 360-day 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 II and III 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 IV, we do this by estimating Equation 6a using hierarchical (mixed effects) logit models that treat

[10] In the columns with a mills-ratio control for the propensity index, we took the propensity Pr(Reperfusion=1|X) from the treatment logit, put it through an inverse normal to get the equivalent probit index, then created the mills-ratio using that probit index. We did this for ease of comparability across columns and avoiding a comparison of probit coefficients to logit and OLS coefficients.

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 IV 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 ($\alpha_h^0$). The random effects assumption may appear to be restrictive relative to the fixed effect models in Table II, but the restrictions do not meaningfully change the estimated coefficients on reperfusion or reperfusion*propensity.

The second column of Table IV 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 ($\alpha_h^0$) and the hospital-level thresholds ($\tau_h$). 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 $\alpha_h^0$ 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.

### *III.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 examine each of these in turn. Finally, we discuss other relevant issues including our decision to ignore costs and the possibility that comparative advantage and allocating inefficiency are not just hospital-level objects, but vary with patient characteristics.

*Roy Selection*: We have presented evidence supporting the first assumption in Figure IV and Table II, 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. This evidence is not consistent with other possible selection rules that would imply no

relationship or even a negative relationship between the propensity and benefit. For example, we would expect no relationship if selection were based on non-medical factors such as ability to pay or availability of needed personnel at the time of admission. Alternatively, we would expect a negative relationship between propensity and benefit if hospitals were more likely to treat patients who were about to die in a futile, last-ditch effort.

*Single Index*: If some hospitals have a wider distribution of unobservable gains from treatment ($v^{\Delta}$), because of patient selection or superior diagnostics, it will increase the return to treatment at these hospitals, but the mechanism would be patients' unobserved characteristics, not allocative inefficiency in the form of higher hurdles. This can be tested by examining whether the relationship between treatment effect and propensity-index varies by selected hospital characteristics: hospitals with more variation in $v^{\Delta}$ (gains in treatment unobserved to us but observable to the hospital, say through better diagnostics) will have a flatter relationship between propensity and the return to treatment relative to other hospitals. There are two reasons for this: (i) the propensity estimates $\beta/\sigma_v$ so as $\sigma_v$ increases, the index captures less about benefit relative to the increased variance of $v^{\Delta}$ and (ii) an increase in $\sigma_v$ would push relatively more low-propensity patients over the hurdle and this would flatten the relationship between returns and the propensity.

This test is implemented by adding a triple interaction (Reperfusion*Index*Hospital Characteristics) between hospital characteristics like (risk-adjusted) hospital treatment rates, volume, and major-teaching hospital and seeing whether the effect of Reperfusion*Index on survival varies by these characteristics. If these triple interactions are significant then it is evidence *against* our hypothesis that conditional on the X's, the distribution of $v^{\Delta}$ is identical across hospitals. This would be different than a parallel shift in the returns to treatment that would be generated by a model of allocative differences (as in Figure II).

We test this hypothesis in Table V. If we thought that hospitals with high treatment rates had access to superior diagnostic technology or somehow received patients with better unobservable characteristics, then they'd have a higher variance of $v^{\Delta}$, not higher tau, and the triple-interaction would be negative to reflect the flatter relationship. We do not see this; the coefficient on the triple interaction in the first column is essentially zero. But if hospitals with higher risk-adjusted treatment rates have lower hurdles (negative tau), then the coefficient on Reperfusion *Hospital-Effect should be negative because this is a story about hurdles ($\tau$) and implies a parallel shift down conditional on propensity—this is exactly what we find. In the remaining columns we estimated triple interactions with other hospital characteristics like volume and major-teaching. Here too, we found that the triple interactions

were insignificant, and interpreted that as being reassuring for the credibility of our empirical work (OLS models gave very similar results).

*Estimating Treatment-on-the-Treated*: To evaluate the plausibility of this 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. Clinical trials in medicine have highly specific and strict exclusion criteria based on contraindications and prior histories and therefore estimate the effect of the treatment on patients who will likely receive the treatment. For this reason, we think that the CCP data and the RCTs should recover similar treatment-on-treated effects. A summary of nine trials was published in the journal *Lancet* by the Fibrinolytic Therapy Trialists' Collaborative Group (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 survival is 0.20. In our CCP data, a logit model controlling for the CCP risk-adjusters estimates an identical effect, with reperfusion increasing the log-odds of survival by 0.207 (S.E. = 0.025). We take this evidence as supporting the case that we can estimate unbiased estimates of the treatment on treated effect.

Because of the centrality of the treatment-on-treated parameter to our model, we further explored the plausibility of the model's identifying assumptions. In our model, conditional on our control variables, we assume that treatment is uncorrelated with omitted factors in a patient's baseline survival ($v_{ih}^0$), but allow for selection on gains where treatment is correlated with omitted factors in the patient's unobserved gain ($v_{ih}^{\Delta} = v_{ih}^1 - v_{ih}^0$). Thus, we must assume that $v_{ih}^0$and $v_{ih}^{\Delta}$are conditionally uncorrelated. While we cannot test this assumption directly, we can test for a similar pattern in observables: as we add more detailed information about the patient to our controls, the additional controls may predict treatment (correlated with $v_{ih}^{\Delta}$) and predict mortality (correlated with $v_{ih}^0$) but should not change the estimated effect of treatment on survival (because $v_{ih}^{\Delta}$ uncorrelated with $v_{ih}^0$).

Table VI reports results from logit models of survival on reperfusion with increasingly detailed controls. The first column includes no controls, and estimates an impact of reperfusion on the logodds of survival of 0.424, well above the trial estimates and implying that some level of controls are needed for our assumptions to hold. The second column adds age-sex-race controls, and estimates a treatment effect of 0.192, in line with trial estimates. The third column adds all of the CCP risk adjusters, and yields a similar estimate of the treatment effect of 0.207. Thus, while the additional CCP risk-adjusters

are very strong predictors of both treatment (p<.001) and survival (p<.001), controlling for them has very little impact on the estimated treatment effect, which is consistent with our identifying assumptions.

In the final column we add two types of additional controls that were not on the patient's chart (so were not included in our baseline specification) but were added to the data later and are highly predictive (p<.001) of both survival and treatment: patient zip-code characteristics (average income and fraction with a high school and college degree); and American Hospital Association and American College of Cardiology (AHA/ACC) criteria for reperfusion and cardiac catheterization that were created for each patient by expert reviewers using the chart data (where ideal patients for a treatment are those for whom the treatment would almost always be indicated, and less-than-ideal candidates are patients for whom the therapy would be controversial). Controlling for these additional variables yields an estimate of the treatment effect (0.199) that is similar to models with fewer controls and similar to the trial estimates. This provides further support for our conditional independence assumption – omitted factors driving treatment decisions ($v_{ih}^{\Delta}$) appear to be uncorrelated with omitted factors determining survival ($v_{ih}^{0}$).

The fact that we can replicate trial estimates with simple age-sex-race controls is context specific – reperfusion with thrombolytics is a non-invasive treatment whose benefits are not strongly related to mortality risk, unlike surgical interventions which are less likely to be successful in frail populations. Thus, after conditioning on age-sex-race, the factors that drive treatment in our setting are not strongly related to the factors that determine survival. In our application, the main advantage of the rich control variables available in the CCP is to allow more accurate estimation of variation in treatment propensity across patients and hospitals.[11] So the key takeaway is that the CCP controls clearly matter, but the basic patterns noted in the paper are robust to relying only on age-race-gender as control variables (additional robustness tests using age-sex-race as the only controls are provided in Appendix III).

*Comparative Advantage and Allocative Inefficiency Vary by Patient Characteristics*: while not an identifying assumption, one might think that comparative-advantage and allocative efficiency vary by patient characteristics—that they are larger or smaller for certain types of patients. This can be tested by letting the coefficients on the patient characteristics vary by the risk-adjusted hospital treatment rate (Θ). To do this, we estimate separate propensity models by three terciles of theta and correlate their

[11] For example, estimates of the propensity using only age-sex-race controls predict only 20% of the variation that is predicted by the propensity using the full CCP controls. Thus, if we re-estimate Table 2 controlling for age-sex-race (but using the original full-CCP propensity estimates) we get very similar results, but if we re-estimate Table 2 using age-sex-race controls in both the propensity and the survival equation we get similar signs and significance, but different magnitude estimates (see online Appendix III for these results)

predictions. If hospitals with high or low treatment rates weight patient characteristics differently, we would find weak correlations across the terciles. Empirically, the correlations across the indices is greater than 0.99 for terciles 1 and 2 and terciles 2 and 3. It is 0.98 for the correlation between tercile 1 and 3.

*Incorporating Costs*: We have assumed that hospitals maximize survival and don't think about costs. We could assume that hospitals aren't just maximizing survival, but rather, maximizing Net Benefit=Survival - λCost, where λ represents the value of life, measured as survival per $1000 of spending. λ captures the trade- off being made by the patient and physician between improved survival and increased costs. In our data, receiving reperfusion increases 1-year spending by $2.3k. Following the literature, we assume that every AMI patient who lives to one year, lives on average, to 5 years. A typical value of λ used in cost-effectiveness studies would place the societal value of a life year at around $100,000 (Cutler, 2004), although hospitals may use λ=0 if patients do not pay directly for the cost of treatment (e.g., because of insurance), or if providers have strong incentives to ignore costs (and social-welfare) and maximize survival (e.g., perhaps because of fee-for-service payments).

Using these numbers, an extremely low value of life would be to value each life-year at $20k, which would generate λ=0.01. For a value of life-year of $100k, we'd get λ=0.002. At the other extreme, one could assume that hospitals value a life-year at $300,000 (which is highly cost-ineffective); this would generate λ=0.00067. These values represent the increased probability of 1-year survival at which costs become relevant. Given that the reperfusion increases 1-year survival by 4 percentage points (0.04), and increases spending by $2.3k, costs become irrelevant, or at-best third-order important.[12]

### *III.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

[12] What is relevant for our analysis is whether more aggressive hospitals are less costly on a per unit basis (this would happen if they do more reperfusion which increases costs but this therapy has offsetting effects on 1-year hospital and physician spending). The difference in 1-year spending from reperfusion across the most and least intensive terciles of reperfusion is $164. There is no reasonable value of λ at which a $150-$200 difference in spending becomes salient to offset our emphasis on survival.

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 in the presence of comparative advantage, they say nothing 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 ($\theta_h, \alpha_h^0, \tau_h$). This approach will require us to make assumptions about the scale parameter ($\sigma_v$), which is the variance of the unobservable gain from treatment. Our earlier results on the presence of allocative-inefficiency were invariant to this scaling, but it will turn out to be central for the recovery of comparative-advantage. These estimates are then used to identify variation across hospitals in comparative advantage. We will rely on the linear-approximation for $g(I) = \lambda_0 + \lambda_1 I$ instead of the non-parametric control to simplify things (the linear assumption was justified by Figure IV and in Table II where we showed very similar results with this restriction compared to the fully non-parametric approach).

To motivate our approach to recovering hospital-level comparative advantages, recall that the treatment propensity was estimated using a random effect logit to estimate Equation 2:

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

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

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

Equation 7c is a logit model with a hospital-level random intercept ($\alpha_h^0$) and a hospital-level random coefficient on reperfusion ($\lambda_1\, \theta_h + \tau_h$). In estimating Tables II-IV, we used a 2-step approach that first estimated $\theta_h$ 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 ($\theta_h$) and the hospital-level intercept ($\alpha_h^0$) 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 ($\lambda_0, \lambda_1$) were estimated along with the variance and covariance of the hospital-level random coefficients by maximum likelihood.[13] Finally,

[13] 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

knowledge of $(\lambda_1\,\theta_h + \tau_h)$, $\lambda_1$ and $\theta_h$ allows one to restate all of the estimates in terms of $\tau_h$ rather than $(\lambda_1\,\theta_h + \tau_h)$.

We report the results from joint estimation of equation 2 and 7c in Table VII. We were reassured that the more complicated joint estimation procedure replicates the results and magnitudes from simpler models. For example, the coefficient on reperfusion from Table VII is 0.27 compared to 0.31 in the simpler logit model in Table IV (both coefficients are in log-odds). The benefit of reperfusion increases with the index with similar magnitudes in both models—0.29 in Table IV vs. 0.28 from Table VII. 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 $\tau_h$, 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 II, there is a negative correlation (-0.341) between $\tau_h$ and the reperfusion intercept $\theta_h$, 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 IV-V), 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 VII 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 $(\alpha_h^\Delta)$ and the treatment threshold $(\tau_h)$. Recall that $\theta_h = \left(\alpha_h^\Delta - \tau_h\right)/\sigma_v$, where $\sigma_v$ is the standard deviation of the patient-level idiosyncratic gains $(v_{ih}^\Delta)$, 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 $\sigma_v$ and the joint distribution of $(\theta_h, \tau_h)$. 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 $\sigma_v$ to calculate the standard deviation of $\alpha_h^\Delta$ and its correlation with $\tau_h$. These are presented in Appendix IV. Interestingly, these estimates bound the standard deviation of $\alpha_h^\Delta$ to be above 0.3. Thus, our estimates imply that the variation across hospitals in

---

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.

comparative advantage is at least as large as the variation across hospitals in the treatment threshold (SD=.327) and possibly much larger. For $\sigma_v$ < 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 $\tau_h$ and $\alpha_h^{\Delta}$, and that the two are strongly positively correlated (between 0.4 and 1), meaning that hospitals with low comparative advantage tend to have low thresholds and overuse reperfusion. Such a positive correlation (which was also noted above in Table VII) would arise if all hospitals *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 IIb). 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 IV adding interactions of treatment with a number of hospital-characteristics such as ownership, teaching status and size (Table VIII). 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 VIII implies that differences across hospitals with these characteristics in reperfusion rates reflect differences in comparative-advantage 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 IIb.[14] 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 $\alpha_h^{\Delta}$ 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 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 $\tau_h$ 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 $\tau_h$ as arising from an inaccurate belief about $\alpha_h^{\Delta}$, rather than assuming that hospitals know $\alpha_h^{\Delta}$ and consciously set $\tau_h \neq 0$ to achieve other objectives. A negative $\tau_h$ 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 $\sigma_{\omega}^2$ (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 $\sigma_{\alpha}^2$ is the variance of $\alpha^{\Delta}$ across hospitals), then the optimal prediction of $\alpha^{\Delta}$ 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.

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

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 VII estimated 6 reduced-form moments (the variances and covariances of $\theta$, $\tau$, and $\alpha_0$) which are a function of the 6 unknown structural parameters in this framework (the variances and covariance of $\alpha^{\Delta}$ and $\alpha_0$, the reliability of the signal r, the weight placed on the signal w, and the scale parameter from the logit $\sigma_v$).[15] 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 moments exactly because they are one-to-one function of the reduced form parameters. These equations are derived and listed in Appendix V. 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 IX. 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 hospitals receive about their comparative advantage is estimated to have very low reliability (r=.065), but hospitals 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 IX), 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 $\theta$, $\tau$, and $\alpha_0$ that are significantly different from those estimated in Table VII. More specifically, if w=r then the hospital's prediction ($\theta$) is optimal and, therefore, should be uncorrelated with the prediction error ($\tau$). The fact that we estimated a significant negative correlation of -0.34 between $\theta$ and $\tau$ in Table VII 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

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

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 IX 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 IX implies reduced-form variances and correlations of $\theta$, $\tau$, and $\alpha_0$ 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 IX) 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. WELFARE IMPLICATIONS

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 IIb, a patient in a hospital which uses a $\tau$ that is different than zero experience a welfare loss, where:

$$\text{(8a)} \qquad Welfare\ Loss_{ih} = \frac{1}{2}.\overbrace{(\tau_h)}^{height}\overbrace{(\tau_h)\frac{dPr(Treatment_{ih})}{\mathrm{d}\tau}}^{base}$$

The welfare loss due to allocative inefficiency is the average reduction in (logodds) survival per patient.[16] Noting that $E(\tau_h{}^2) = Var(\tau_h) + [E(\tau_h)]^2$, and letting $\frac{dPr(Treatment)}{\mathrm{d}\tau}$ represent the average effect of a change in $\tau_h$ across all patients, the average welfare loss across all hospitals is given by:

$$\text{(8b)} \qquad E(Welfare\ Loss_{ih}) = \frac{1}{2}.Var(\tau_h)\frac{dPr(Treatment)}{\mathrm{d}\tau} + \frac{1}{2}.[E(\tau_h)]^2\frac{dPr(Treatment)}{\mathrm{d}\tau}$$

[16] The welfare loss is measured in the same units as $\tau$ (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.

The welfare loss can be broken into two parts – the first due to variance in $\tau_h$ and the second due to systematic bias in $\tau_h$, across all hospitals. We estimate each in turn.

For the first term, we can return to the hierarchical-logit model in Table VII, where we estimated $SD(\tau_h)$=0.327.[17] 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 $\tau_h$, divided it by our estimate of the scale factor ($\sigma_v$) of 0.44 to turn it into how much change that would create in the hospitals risk-adjusted treatment rate $\theta_h$. Adding this change in $\theta_h$ to each patient's propensity index yielded an estimate of the average effect across all patients of $dPr(Treatment)/d\tau$ =-0.26. So the welfare loss from variation in allocative inefficiency is (1/2)*(0.327^2)*(-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 (0.265 in log-odds for the average treated patient from Table VII), who comprise 19 percent of the patient population for a total average benefit of 0.05. This means that we could increase the effectiveness of treatment by about 28% (.014/.05) if we removed the allocative inefficiency across hospitals.

The second term of the welfare loss equation says that there is additional welfare loss if there is systematic overuse across all hospitals, i.e. if the mean of τ is not equal to zero. This part of the welfare calculation is more speculative. A good guess about systematic overuse across all hospitals comes from the average treatment effect among patients with very low propensity, since the treatment effect asymptotes at $\tau_h$ as the propensity goes to zero in the limit. In Figure IV, patients with the lowest propensity being plotted (with about a 5% chance of treatment) have an average treatment effect of roughly -0.25 (a bit higher for conservative hospitals, a bit lower for aggressive hospitals), suggesting that $E(\tau_h) < -0.25$. Therefore, the additional welfare loss from systematic overuse should be at least (1/2)*(-0.25^2)*(-0.26)=-0.008, which would raise the overall loss from .014 (28% ) to .022 (44%). This calculation suggests that systematic overuse adds substantially to the welfare loss from the overall variation.

Put differently, these estimates mean that a policy which provided better information about treatment effect heterogeneity across hospitals could improve patient welfare by reducing the prediction error $\tau_h$, where $\tau_h = \alpha_h^{\Delta} - \theta_h$. Our more structural estimates from Table IX suggest that it is very difficult for hospitals to predict their own comparative advantage – they receive very unreliable signals, and then overweight these noisy signals. But if all hospitals had perfect information about $\alpha_h^{\Delta}$, and acted on it, we would eliminate allocative inefficiency entirely and improve the average value of the treatment

[17] The simpler model in Table 4 yielded a very similar estimate of 0.31.

by 44% (based on the calculation above, assuming perfect information eliminates both the bias and variation in $\tau_h$).

Our estimates also allow us to consider other counterfactuals. One alternative counterfactual would be to encourage more complete standardization in treatment rates across hospitals, along the lines often suggested by policy makers concerned about unwarranted variation across hospitals in treatment rates. In our model, this is equivalent to getting hospitals to ignore the signal (S) about their own comparative advantage, and instead all use a common $\theta$ as an estimate of their own comparative advantage. Policy-makers often have the sense that $\theta$ should be set to the average true comparative-advantage. The main benefit from this type of standardization is it can potentially eliminate the welfare loss from systematic bias in $\tau_h$. As we calculated above, eliminating this bias improves welfare by at least 16%. But there is also a cost from this type of standardization as it ignores true differences across hospitals in comparative advantage, and through this channel, increases allocative inefficiency. If hospitals were perfectly predicting the variation in $\alpha_h^{\Delta}$ (so that $\tau_h = 0$), then standardization would lead to greater prediction errors resulting in $Var(\tau_h) = Var(\alpha_h^{\Delta})$. Using the estimate of the standard deviation of $\alpha_h^{\Delta}$ from Table IX (0.317), we estimate that the welfare loss from standardization (relative to the first best of perfect prediction of $\alpha_h^{\Delta}$) is (1/2)*(0.317^2)*(-0.26)=-0.013. Note that this is similar to the welfare loss from variation in $\tau_h$ under current practice (-.014) because hospitals are doing such a poor job of predicting their own comparative advantage. Thus, in this case, there is no cost (relative to current practice) from reducing variation in treatment rates across hospitals and some benefit if standardization were able to remove the overall bias toward over-treatment.

A third, perhaps more realistic, counterfactual would be to encourage more standardization in treatment rates across patients. One possibility would be just prohibiting hospitals from treating patients below a certain cutoff and giving hospitals discretion in treating patients above the cutoff. This is similar to having evidence-based guidelines, and discouraging hospitals from treating outside of guidelines. Note that if there were no systematic bias toward over-treatment (i.e., $E(\tau_h) = 0$) then there would be no benefit from this strategy. If there were no bias then even low propensity patients would, on average, have positive treatment effects on the treated (although there would be negative treatment effects in some hospitals). Thus, this strategy is a second-best solution to the bias problem – rather than eliminating the bias (which benefits all patients), we eliminate treatment for low-propensity patients who are most negatively affected by the bias toward over-treatment. To calculate how much would be gained from this strategy, we compared the existing average benefit from treatment (.050) to a counterfactual in which we set treatment to 0 for all patients for whom the $\lambda_0 + \lambda_1 * X_i\beta < 0$, using estimates of the treatment effect on the treated for these patients (including the hospital random effects) derived from

the model in Table VII. Just over 18% of patients currently getting the treatment are predicted to have negative treatment effects, and not treating these patients increases the average benefit across all patients by 14% to .057. The gains from this second-best approach (.007) are nearly as large as the gains we estimated from eliminating bias (.008), suggesting once again that systematic overuse is a central component of welfare losses.

## VI. Conclusion

Using a Roy model of treatment to motivate our empirical framework, we found significant evidence of allocative inefficiency across hospitals. 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 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.

Our work suggests three 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 (2018), which finds evidence that patients select on gains in terms of choosing hospitals.

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.

Finally, and related to the point about heterogeneous benefits, our works speaks to using more economic structure at the time of estimating simpler reduced-form estimates of the marginal value of health. Regression-discontinuity designs such as those involving birthweight cuts are recovering the average marginal effect of a treatment across all hospitals and all patients. These approaches often focus on removing a demand-side confounder like patient illness. The Roy-model approach notes that there is substantial heterogeneity in the treatment effect across hospitals and across patients, with the marginal effect of more treatment depending on which patient receives the treatment, and which hospital delivers it. These are supply-side explanations, and ignoring such heterogeneity, across patients and hospitals, creates problems for the interpretation of simpler approaches.

But RCTs and other approaches can also be deployed to test other implications of our model, and we believe that a greater use of these methods to test approaches like ours is central. These methods could be deployed to test selection on gains—perhaps by evaluating whether the average treatment effect is smaller than treatment on the treated. One setting to validate our paper would be one where it became slightly harder to access treatment in a hospital. We would predict that the extra costs would reduce treatment for the marginal patient, and that this marginal patient will be different in different hospitals: in more intensive hospitals (as measured by risk-adjusted treatment rates), the marginal patient will be less appropriate in the global distribution of patient characteristics. Another test would be to give hospitals more information about their comparative-advantage and see if this information is used by hospitals to change their treatment decisions. Our model would predict that if hospitals respond to their information then (1) some less appropriate patients will no longer be treated (2) overall outcomes will improve and that this will be driven by a lack of overuse in less appropriate patients, and (3) these effects will be concentrated in smaller hospitals. These insights flow from a Roy model of treatment allocation, but not from theory-agnostic approaches that measure average treatment effects across all patients and hospitals.

**HARVARD UNIVERSITY AND THE NBER**
**DARTMOUTH COLLEGE AND THE NBER**

## REFERENCES


Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106, no. 12 (2016): 3730-64.

Baicker, Katherine, Amitabh Chandra, and Jonathan S. Skinner. "Saving Money or Just Saving Lives? Improving the Productivity of US Health Care Spending." *Annual Review of Economics* 4, no. 1 (2012): 33-56.

Barnato, Amber E., F. Lee Lucas, Douglas Staiger, David E. Wennberg, and Amitabh Chandra. "Hospital-Level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes." *Medical Care* 43, no. 4 (2005): 308.

Burstein, Harold J., Christina Lacchetti, Holly Anderson, Thomas A. Buchholz, Nancy E. Davidson, Karen A. Gelmon, Sharon H. Giordano et al. "Adjuvant Endocrine Therapy for Women With Hormone Receptor-Positive Breast Cancer: ASCO Clinical Practice Guideline Focused Update." *J Clin Oncol* 37, no. 5 (2018): 423-438.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. "Productivity Dispersion in Medicine and Manufacturing." *American Economic Review* 106, no. 5 (2016): 99-103.

—— "Health Care exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review* 106, no. 8 (2016): 2110-44.

Chandra, Amitabh, and Douglas O. Staiger. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy* 115, no. 1 (2007): 103-140.

—— *Identifying Provider Prejudice in Healthcare*. No. w16382. National Bureau of Economic Research, 2010.

Chandra, Amitabh, and Jonathan Skinner. "Technology Growth and Expenditure Growth in Health Care." *Journal of Economic Literature* 50, no. 3 (2012): 645-80.

Currie, Janet, and W. Bentley MacLeod. "Diagnosing Expertise: Human Capital, Decision Making, and Performance Among Physicians." *Journal of Labor Economics* 35, no. 1 (2017): 1-43.

Doyle, Joseph J. "Returns to Local-Area Health Care Spending: Evidence from Health Shocks to Patients Far from Home." *American Economic Journal: Applied Economics* 3, no. 3 (2011): 221-43.

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

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131, no. 4 (2016): 1681-1726.

Fibrinolytic Therapy Trialists' Collaborative Group. "Indications for Fibrinolytic Therapy in Suspected Acute Myocardial Infarction: Collaborative Overview of Early Mortality and Major Morbidity Results from All Randomised Trials of More Than 1000 Patients." *The Lancet* 343, no. 8893 (1994): 311-322.

Fisher, Elliott S., David E. Wennberg, Thrse A. Stukel, Daniel J. Gottlieb, F. Lee Lucas, and Etoile L. Pinder. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content,

Quality, and Accessibility of Care." *Annals of Internal Medicine* 138, no. 4 (2003a): 273-287.

—— "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138, no. 4 (2003b): 288-298.

Garber, Alan M., and Jonathan Skinner. "Is American Health Care Uniquely Inefficient?" *Journal of Economic Perspectives* 22, no. 4 (2008): 27-50.

Gawande, Atul. "The Cost Conundrum." *The New Yorker* 1 (2009): 36-44.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *The Review of Economics and Statistics* 88, no. 3 (2006): 389-432.

Hull, Peter. "Estimating Hospital Quality with Quasi-Experimental Data." *Available at SSRN 3118358* (2018).

Marciniak, Thomas A., Edward F. Ellerbeck, Martha J. Radford, Timothy F. Kresowik, Jay A. Gold, Harlan M. Krumholz, Catarina I. Kiefe, Richard M. Allman, Robert A. Vogel, and Stephen F. Jencks. "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, no. 17 (1998): 1351-1357.

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?: Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272, no. 11 (1994): 859-866.

O'Connor, Gerald T., Hebe B. Quinton, Neal D. Traven, Lawrence D. Ramunno, T. Andrew Dodds, Thomas A. Marciniak, and John E. Wennberg. "Geographic Variation in the Treatment of Acute Myocardial Infarction: The Cooperative Cardiovascular Project." *Journal of the American Medical Association* 281, no. 7 (1999): 627-633.

O'Gara, Patrick T., Frederick G. Kushner, Deborah D. Ascheim, Donald E. Casey, Mina K. Chung, James A. De Lemos, Steven M. Ettinger et al. "2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines." *Journal of the American College of Cardiology* 61, no. 4 (2013): e78-e140.

Phelps, Charles E. "Information Diffusion and Best Practice Adoption." In *Handbook of Health Economics*, vol. 1, pp. 223-264. Elsevier, 2000.

Skinner, Jonathan. "Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*, vol. 2, pp. 45-93. Elsevier, 2011.

Skinner, Jonathan, Douglas Staiger. "Technology Diffusion and Productivity Growth in Health Care," Review of Economics and Statistics 97, no.5 (2015): 951-964.

Suri, Tavneet. "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79, no. 1 (2011): 159-209.

Syverson, Chad. "What Determines Productivity?" *Journal of Economic Literature* 49, no. 2 (2011): 326-65.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner, and Amitabh Chandra. "Hospital Quality and Intensity of Spending: Is There an Association?" *Health Affairs* 28, no. 4 (2009): w566-w572.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2002.

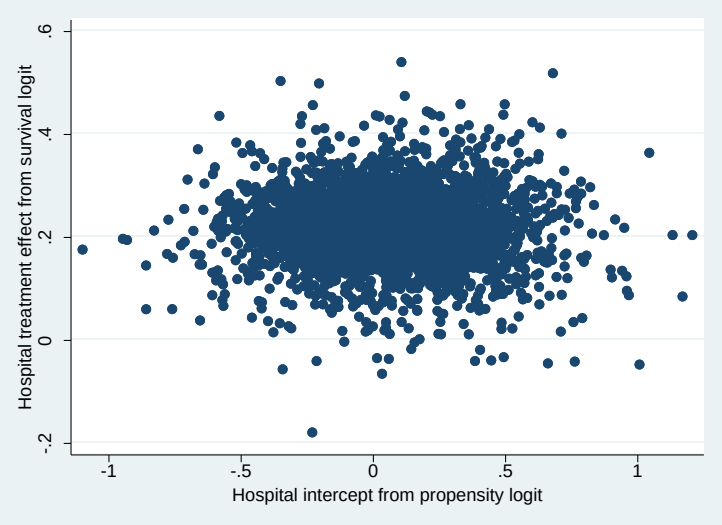


FIGURE I

Motivating Facts from the CCP Data

PANEL A

Survival benefit associated with reperfusion versus risk-adjusted rate of reperfusion at the hospital level (correlation = -0.02, Number of hospitals =4690)

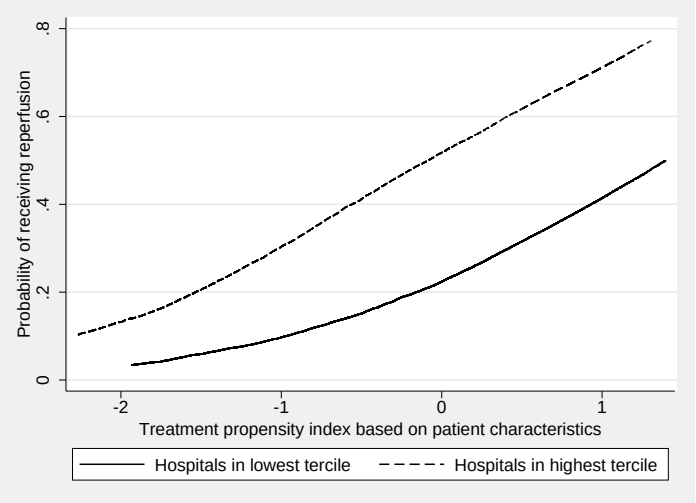


FIGURE I

Motivating Facts from the CCP Data

PANEL B

Probability of Receiving Reperfusion According to Patient Characteristics and Hospital Treatment Intensity

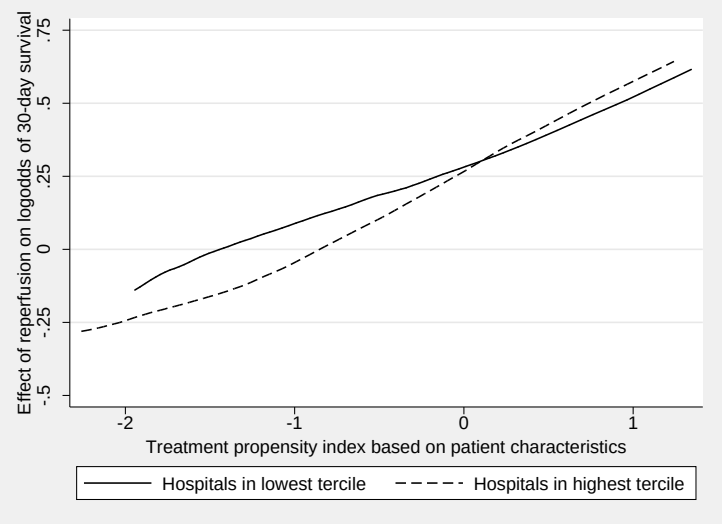


FIGURE I

Motivating Facts from the CCP Data

PANEL C

Survival Benefit Associated with Reperfusion According to Patient Characteristics and Hospital Treatment Intensity

In Panel A, the risk-adjusted rate of reperfusion at the hospital level is an empirical Bayes estimate of the hospital-level intercept from a random-effect logit model of whether a patient received reperfusion within 12 hours regressed on a rich set of covariates derived from patient charts (see Appendix I for a full list) and a hospital-level random intercept. The survival benefit associated with reperfusion is an empirical Bayes estimate of the hospital-level coefficient on reperfusion from a random-coefficient logit model of whether a patient survived 30 days after their heart attack regressed on whether the patient received reperfusion within 12 hours, controlling for the full set of patient covariates, and allowing for a hospital-level random intercept and (possibly correlated) random coefficient on reperfusion.

In Panel B, we non-parametrically plot the probability of treatment as a function of the treatment propensity index from the reperfusion logit based on patient covariates (without the hospital intercept), estimating separate lines for the probability of treatment in hospitals in the highest (most aggressive) and lowest (most conservative) terciles of risk-adjusted treatment rates (the hospital-level reperfusion intercept).

In Panel C, we non-parametrically plot the effect of reperfusion on survival as a function of the treatment propensity index based on patient characteristics (without the hospital intercept), separately for hospitals in the highest (most aggressive) and lowest (most conservative) terciles of risk-adjusted treatment rates. The survival benefit at each point in the distribution of the propensity index was estimated flexibly using a local-logistic regression of 30-day survival on reperfusion that controlled for the detailed risk adjusters available in the CCP. The lines in Panels B and C are based on local regressions with a triangular kernel that included 30% of the sample on either side.

All estimates are based on data from the CCP for 138,957 AMI patients treated in 4690 hospitals.

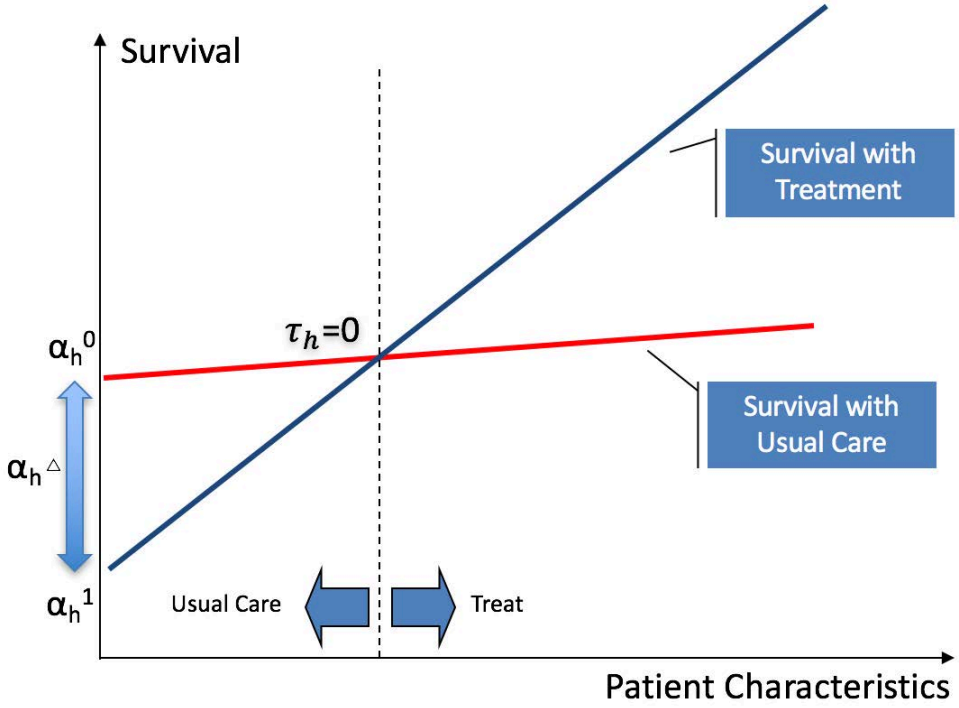


FIGURE IIa
A Roy Model of Treatment at the Hospital Level

The two lines denote patient survival if a hospital treats a given patient with usual care (intercept is $\alpha_h^0$) or with reperfusion treatment (intercept is $\alpha_h^1$) 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 $\alpha_h^0$ and $\alpha_h^1$ 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 ($\tau_h = 0$).

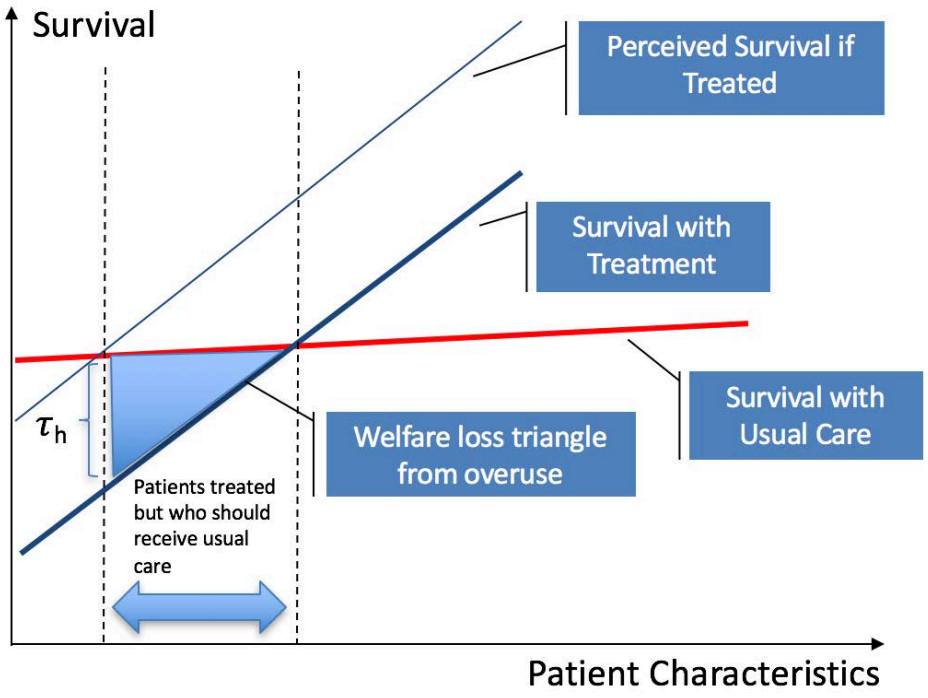


FIGURE IIb
A Roy Model of Treatment at the Hospital Level with Allocative Inefficiency

The figure illustrates the presence of allocative inefficiency. Here, misperceptions 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 ($\tau_h$). The shaded triangle represents the welfare loss.

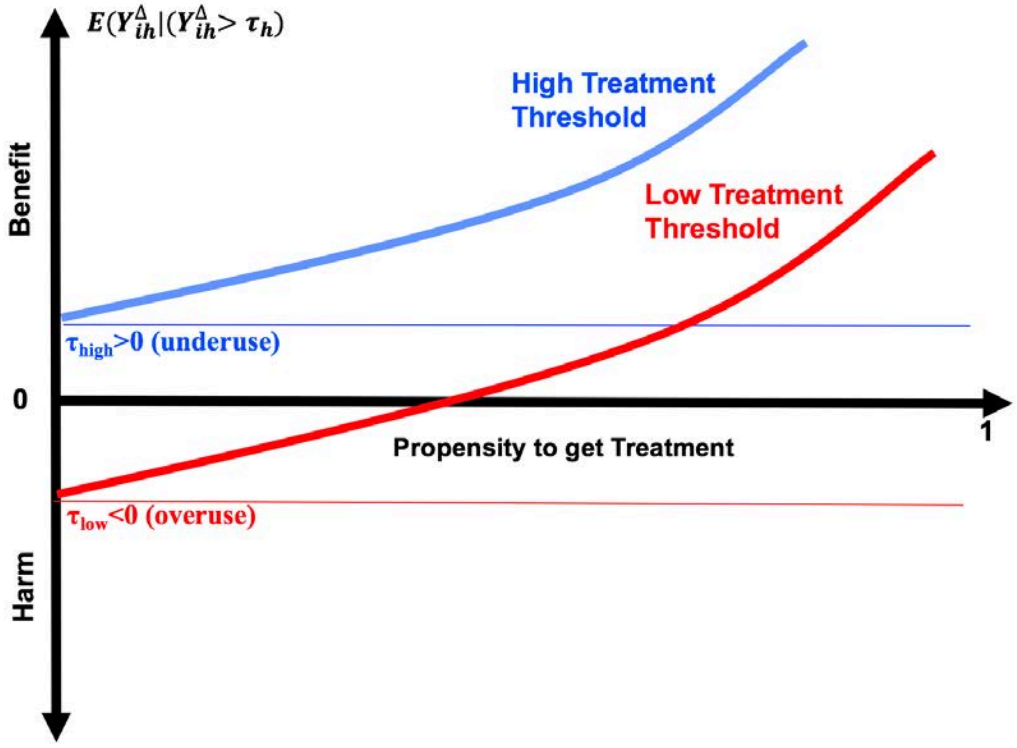


FIGURE III
Distinguishing Underuse and Overuse Using the Propensity to Receive Treatment

The figure illustrates the relationship between the expected benefit from treatment, $E(Y_{ih}^{\Delta} | Y_{ih}^{\Delta} > \tau_h)$, 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).

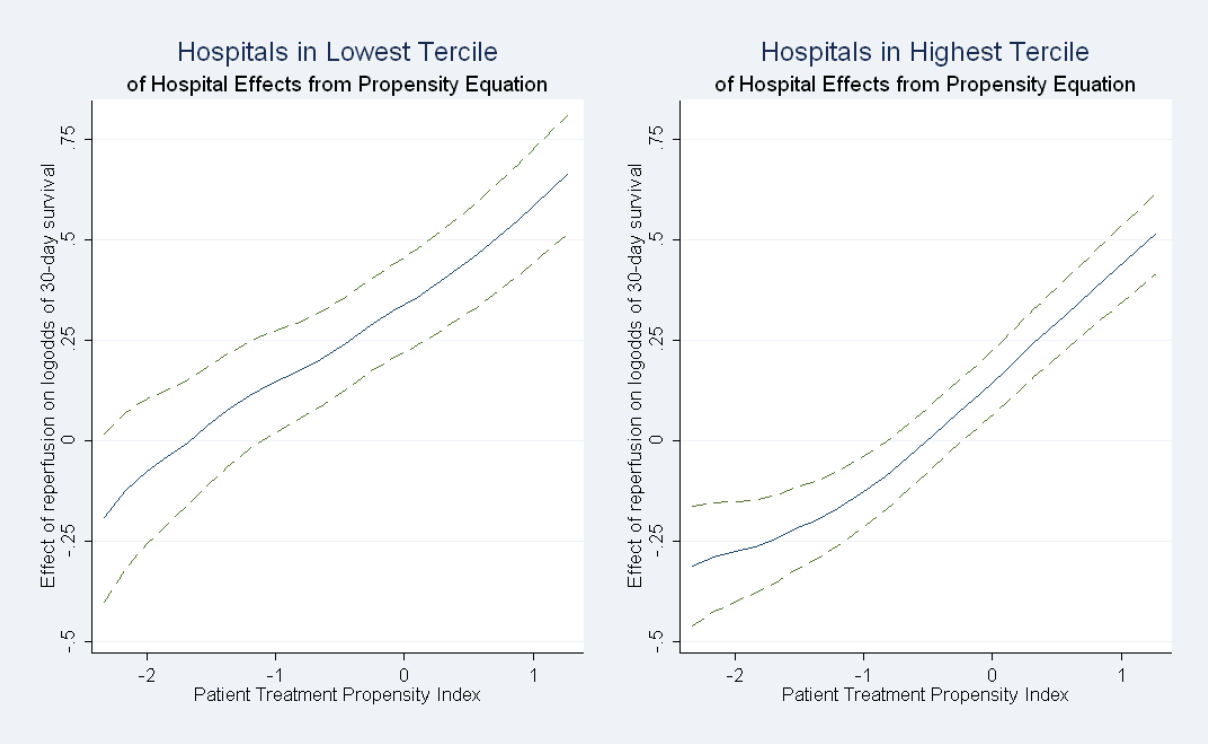


FIGURE IV
Survival Benefit from Reperfusion by Patient's Treatment Propensity,
Low-Treatment-Rate (Left) and High-Treatment-Rate (Right) Hospitals.

The figures plot the estimated survival benefit (and 95% confidence intervals) from reperfusion against a patient's treatment propensity index for hospitals 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 II. 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.

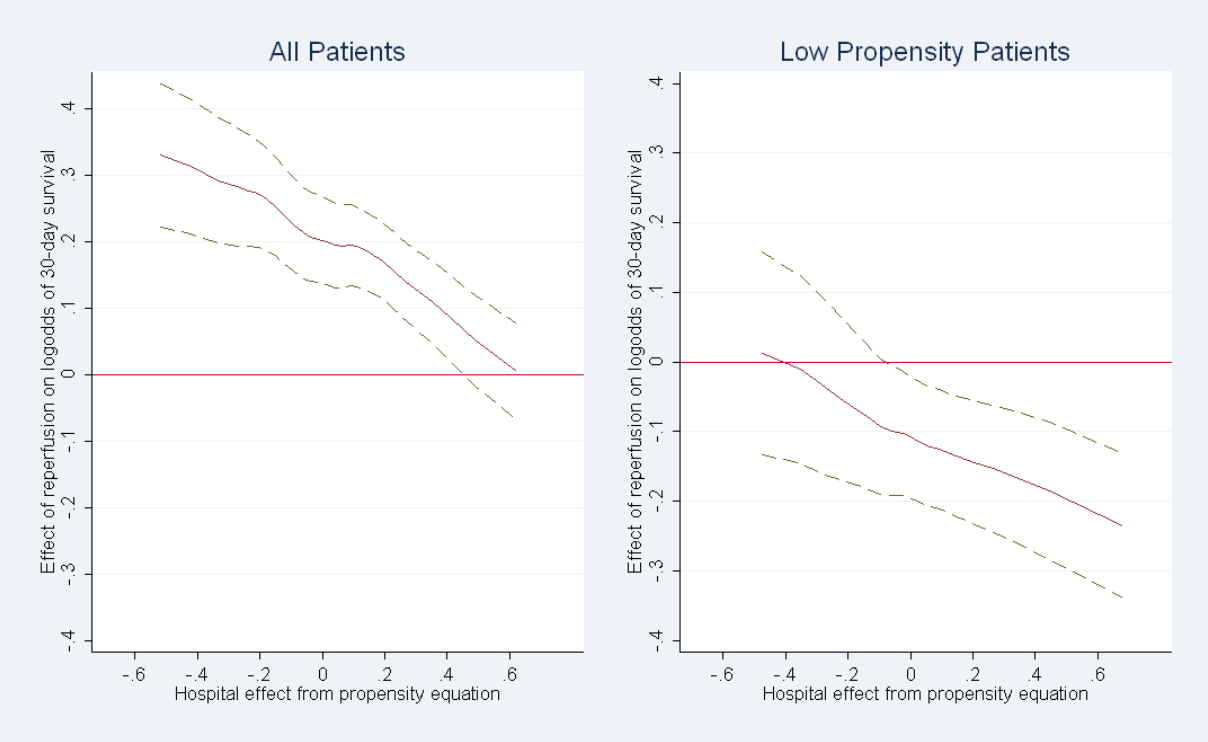


FIGURE V
Survival Benefit from Reperfusion by Risk-Adjusted Hospital Treatment Rate,
All Patients (Left) and Low-Propensity Patients (Right)

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 II. 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.

## TABLE I: 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 (in years) | 76.7 | 73.5 | 77.4 |
| Female (percent) | 49.5 | 43.2 | 50.9 |
| Black (percent) | 5.9 | 4.0 | 6.4 |
| Previous diagnoses: | | | |
| Congestive Heart Failure | 22% | 7% | 25% |
| Hypertension | 62% | 56% | 63% |
| Diabetes | 30% | 23% | 32% |
| Dementia | 6% | 2% | 7% |
| Percent treated at Non-Profit Hospital | 19.6 | 20.7 | 19.3 |
| Percent treated at For-Profit Hospital | 10.6 | 11.0 | 10.5 |
| Percent treated at Government Hospital | 13.2 | 12.7 | 13.4 |
| Percent treated at Major Teaching | 20.2 | 20.5 | 18.7 |
| Percent treated at Minor Teaching | 14.3 | 14.1 | 14.2 |
| Number of Beds at Treating Hospital | 275.3 | 274.6 | 275.5 |
| Number of observations | 138,957 | 25,876 | 113,081 |

Note: Sample is from the Cooperative Cardiovascular Project (CCP), which included data derived from patient charts for all Medicare beneficiaries with acute myocardial infarction admitted during selected months during 1994 and 1995. There are 4,690 hospitals in the sample. Full-list of variables and additional detail on the sample and data collection is in Appendix I.

**TABLE II: Effect of Reperfusion on 30-day Survival, OLS and Logit Estimates**

| | OLS (1) | OLS (2) | OLS (3) | OLS (4) | Logit (5) | Logit (6) | Logit (7) |
|---|---|---|---|---|---|---|---|
| Reperfusion | 0.039 | 0.043 | 0.044 | | 0.328 | 0.344 | |
| | (0.003) | (0.003) | (0.003) | | (0.027) | (0.027) | |
| Reperfusion*Propensity index | 0.040 | 0.042 | | | 0.291 | | |
| | (0.002) | (0.002) | | | (0.018) | | |
| Reperfusion* Hospital Rate (θ) | | -0.031 | -0.037 | -0.037 | -0.211 | -0.251 | -0.252 |
| | | (0.009) | (0.009) | (0.009) | (0.076) | (0.076) | (0.077) |
| Hospital Fixed-Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Control for Propensity Index | None | Linear | Mills Ratio | Non-Parametric | Linear | Mills Ratio | Non-Parametric |

Note: Dependent variable is 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 II. 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 5 include linear controls for propensity-index. Columns 3 and 6 use a mills-ratio but retain OLS/Logit for the survival equation for comparability. Columns 4 and 7 include 100 percentiles of propensity-index interacted with the receipt of Reperfusion. Sample-size in every regression is 138,957.

**TABLE III: 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 | 0.218<br>(0.031) | non-parametric |
| Reperfusion * Propensity index | 0.356<br>(0.021) | non-parametric |
| Reperfusion * Hospital Treatment Rate (θ) | -0.271<br>(0.087) | -0.325<br>(0.088) |
| Control for Propensity Index | Linear | Non-Parametric |
| **Panel B: 365 Day Survival** | | |
| Reperfusion | 0.393<br>(0.023) | non-parametric |
| Reperfusion * Propensity index | 0.176<br>(0.017) | non-parametric |
| Reperfusion * Hospital Treatment Rate (θ) | -0.147<br>(0.066) | -0.192<br>(0.067) |
| 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 II. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. Column 1 reports equation 7a and Column 2 reports equation 7b. All models include all CCP risk-adjusters. Sample-size in every regression is 138,957.

**TABLE IV: Effect of Reperfusion on 30-day Survival, Mixed-Logit Estimates**

| | (1) | (2) |
|---|---|---|
| Reperfusion | 0.297 | 0.314 |
| | (0.022) | (0.024) |
| Reperfusion*Propensity index | 0.289 | 0.292 |
| | (0.017) | (0.017) |
| Std dev of hospital intercept ($\alpha^0$) | 0.188 | 0.198 |
| | (0.015) | (0.017) |
| Hospital Level Random-Intercept ($\alpha^0$) | Yes | Yes |
| Hospital Level Random Coefficient on Reperfusion ($\tau$) | No | Yes |
| Std dev of hospital coefficient on reperfusion (identifies $\tau$; hospital level thresholds) | | 0.313 |
| | | (0.056) |
| corr(hospital level intercept, coefficient on reperfusion) (identifies corr ($\alpha^0$, $\tau$)) | | -0.331 |
| | | (0.154) |
| Number of Hospitals | 4,690 | 4,690 |

Note: Coefficients are log-odds. Propensity Equation is Pr(Reperfusion)=F(Xb+ Hospital Effect) and is estimated using a logit model; see Appendix II. Propensity index refers to the logit index (XB+Hospital Effect). It is demeaned to the average value of patients receiving reperfusion. Table reports estimating Equation 6a. All models include all CCP risk-adjusters. Sample-size in every regression is 138,957.

## TABLE V: Testing for violations of the Single-Index Assumption, Logit Estimates

| | Hospital Characteristic | | |
|---|---|---|---|
| | Hospital Treatment Rate | Ln (Volume) | Major Teaching Hospital |
| Reperfusion | 0.330<br>(0.027) | 0.178<br>(0.118) | 0.296<br>(0.027) |
| Reperfusion * Index | 0.295<br>(0.021) | 0.289<br>(0.088) | 0.268<br>(0.020) |
| Reperfusion * Hospital Treatment Rate | -0.216<br>(0.077) | | |
| Reperfusion * Hospital Treatment Rate * Index | -0.019<br>(0.056) | | |
| Reperfusion * ln (Volume) | | 0.031<br>(0.029) | |
| Reperfusion * ln (Volume) * Index | | -0.001<br>(0.022) | |
| Reperfusion * Major Teaching Hospital | | | 0.026<br>(0.058) |
| Reperfusion * Major Teaching Hospital* Index | | | 0.086<br>(0.044) |
| Control for Propensity Index | Linear | 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 II. Index refers to the logit index (XB+Hospital Effect) from this equation. It is demeaned to the average value of patients receiving reperfusion. Hospital treatment rate refers to the risk-adjusted treatment rate. All models include all CCP risk-adjusters. Sample-size in every regression is 138,957.

**TABLE VI: Sensitivity of the Effect of Reperfusion on 30-day Survival to Controls, Logit Estimates**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Reperfusion | 0.424 | 0.192 | 0.207 | 0.199 |
| | (0.020) | (0.021) | (0.025) | (0.026) |
| Controls: | | | | |
| Age-sex-race | No | Yes | Yes | Yes |
| Full CCP controls | No | No | Yes | Yes |
| Zipcode Characteristics | No | No | No | Yes |
| AHA/ACC Criteria | No | No | No | Yes |

Note: Coefficients are log-odds. Sample-size in every regression is 138,957. Age-sex-race were full interactions between five-year age categories, race categories, and gender categories. Zipcode characteristics included ln(average income), % with high school diploma, and % with college degree. Ideal for Cath and reperfusions reflect American Hospital Association and American College of Cardiology (AHA/ACC) criteria for reperfusion that were created for each patient by expert reviewers using the CCP data but were not in the information that admitting physicians saw. Ideal patients for a treatment are those for whom the treatment would almost always be indicated, and less-than-ideal candidates are patients for whom the therapy would be controversial.

**TABLE VII: 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 ($\alpha_0$; 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 ($\alpha^0$) | -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 Section III.C for details.

**TABLE VIII: 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 IV: 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 ($\alpha_0$) | 0.198 | 0.204 | 0.200 | 0.200 |
| | (0.0167) | (0.017) | (0.016) | (0.016) |
| Std. Dev ($\alpha^{\Delta}$) | 0.317 | 0.407 | 0.337 | 0.336 |
| | (0.058) | (0.059) | (0.057) | (0.057) |
| Corr($\alpha^{\Delta}$, $\alpha_0$) | -0.390 | -0.438 | -0.457 | -0.438 |
| | (0.145) | (0.130) | (0.148) | (0.154) |
| $\sigma_v$ | 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 $\theta$, $\tau$, and $\alpha^0$) as a function of the unknown structural parameters in this framework (the variance and covariance of $\alpha^{\Delta}$ and $\alpha^0$, the reliability of the signal r, the weight placed on the signal w, and the scale parameter from the logit $\sigma_v$). 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.